THE EUROPEAN CONFERENCE ON MACHINE LEARNING & PRINCIPLES AND PRACTICE OF KNOWLEDGE DISCOVERY IN DATABASES

SEPTEMBER 18 - 22, 2017 Skopje, Macedonia









<u>добредојдовте</u>

Dear Colleagues,

Welcome to Macedonia, welcome to Skopje, and welcome to ECML PKDD 2017!

This is the 10th edition of ECML PKDD as a single conference. While ECML and PKDD have been organized jointly since 2001, they only officially merged in 2008. Following the growth of the field and the community, the conference has diversified and grown over the last decade in terms of content, form and attendance. ECML PKDD attracted nearly 600 participants this year, keeping stable the record number of participants from last year.

We are proud to present a rich scientific program, including high-profile keynotes and many technical presentations in different tracks (research, journal, applied data science, nectar and demo), fora (EU projects, PhD), workshops, tutorials and discovery challenges. We are also happy to offer an extensive social program, including a reception, conference banquet and catered poster sessions. We hope that this will provide ample opportunities for exciting exchanges of ideas, pleasurable networking and (simply) fun.

Many people have put in countless hours of work to make this event happen: To them we express our heartfelt thanks: This includes the organization team listed at the end of the brochure, i.e., the program chairs of the different tracks and fora, workshops and tutorials, and discovery challenges, as well as the awards committee, production and public relations chairs, local organizers, sponsorship chairs and proceedings chairs. In addition, we'd like to thank the program committees of the different conference tracks, the organizers of the workshops and their respective committees, the Cankarjev Dom congress agency, and the student volunteers. Furthermore, many thanks to our sponsors for their generous financial support. Finally, thanks to all authors that submitted their work for presentation at ECML PKDD 2017.

Last, but certainly not least, we would like to thank you for coming to the conference and helping us make it a memorable event. Enjoy the conference and your stay in Skopje!

Sašo Džeroski and Michelangelo Ceci, general chairs On behalf of the ECML PKDD 2017 organizing team



Aleksandar Palace Main Conference Venue

<u>Address</u>

Bul. 8th of September no.15, 1000 Skopje, Macedonia [Bul. 8-mi Septemvri br.15, 1000 Skopje, Makedonija]

42.010818, 21.405563

Macedonian National Theatre

<u>Address</u>

11th of March st., 1000 Skopje, Macedonia [ul. 11ti Mart, 1000 Skopje, Makedonija]

41.998533, 21.432317

Macedonian Opera & Ballet

<u>Address</u>

Dimitar Vlahov walk, 1000 Skopje, Macedonia [Kej Dimitar Vlahov, 1000 Skopje, Makedonija]

41.997486, 21.436974

Registration Desk

Sunday, September 17 from 17:00 to 20:00

Monday, September 18 - Friday, September 22 from 08:00 to 17:40

Workshops & Tutorials

Monday, September 18 & Friday, September 22 from 09:00 to 17:40 Invited Talks

Tuesday, September 19 - Thursday, September 21 from 09:00 to 10:00

Best Paper Award Presentations

Tuesday, September 19 - Thursday, September 21 from 10:00 to 10:30 Conference & Journal Track

Tuesday, September 19 - Thursday, September 21 from 09:00 to 17:40 Applied Data Science Track

Tuesday, September 19 - Thursday, September 21 from 09:00 to 17:40

Nectar Track

Tuesday, September 19 & Wednesday, September 20 from 14:00 to 15:40

Discovery Challenges

Tuesday, September 19 & Wednesday, September 20 from 16:00 to 17:40

Community Meeting

Wednesday, September 22 from 17:00 to 18:00

Conference Banquet

Wednesday, September 22 from 20:00 to 23:00

Opening Ceremony

Monday, September 18 from 19:00 to 19:30

Invited Talk

Monday, September 18 from 19:30 to 20:30

Welcome Reception

Monday, September 18 from 20:30 to 23:00

Invited Talks

Tuesday, September 19 & Thursday, September 21 from 19:00 to 20:00

Demo Spotlights

Tuesday, September 19 & Thursday, September 21 from 20:00 to 20:20

Demo & Poster Sessions

Tuesday, September 19 & Thursday, September 21 from 20:00 to 23:00

SCHEDULE AT A GLANCE

<u>Monday</u>

<u>tuesday</u>

WEDNESDAY



THURSDAY



FRIDAY



Invited Talk Cordelia Schmid

> DEMO & POSTER SESSION 2

<u>COMBINED TUTORIALS</u> <u>WITH WORKSOPS</u>

MONDAY

09:00 - 17:40	IAL 2017 Interactive Adaptive Learning
09:00 - 17:40	IoT Large Scale Learning from Data Streams

FRIDAY

09:00 - 19:00 AutoML Automatic Selection, Configuration, and Composition of Machine Learning Algorithms

WORKSHOPS

MONDAY

09:00 - 17:40	MLSA 2017	Machine Learning and Data Mining for Sports Analytics
09:00 - 17:40	PAP 2017	1st International Workshop of Personal Analytics and Privacy
09:00 - 17:40	DyNO 2017+ TD-LSG 2017	Joint Workshop on Large-Scale Evolv- ing Networks & Graphs
09:00 - 12:40	SoGood 2017	Data Science for Social Good
09:00 - 12:40	SURL 2017	Scaling-Up Reinforcement Learning
14:00 - 17:40	MIDAS 2017	2nd Workshop on MIning DAta for financial applicationS
FRIDAY		
09:00 - 17:40	NFMCP 2017	The 6th International Workshop on New Frontiers in Mining Complex Patterns
09:00 - 17:40	LIDTA 2017	Learning with Imbalanced Domains: Theory and Applications
09:00 - 17:40	DLPM 2017	Deep Learning for Precision Medicine
09:00 - 12:40	DMNLP 2017	4th Workshop on Interactions between Data Mining and Natural Language Processing
09:00 - 12:40	DMSC 2017	Data Mining with Secure Computation
14:00 - 19:00	KnowMe	1st International Workshop on Knowledge Discovery from Mobility and Transportation Systems
14:00 - 17:40	DARE 2017	5th International Workshop on Data Analytics for Renewable Energy Integration

TUTORIALS

MONDAY

FRIDAY

09:00 - 12:40	Deep Learning for Computer Vision Applications: Robotics and Driving
14:00 - 17:40	Machine learning with fossil data: Analyzing Environmental and Climate Change

GENERAL INFORMATION

EACH PARTICIPANT WILL RECEIVE A NAME TAG AT REGISTRATION. PLEASE BRING IT WITH YOU AT ALL CONFERENCE EVENTS.

Additional tickets for accompanying persons to attend social events can be purchased at the registration desk.

COFFEE AND LUNCH BREAKS

Food and drinks served in the Breaks Area are included in the registration fee.

<u>Wi-Fi</u>

Wi-Fi will be available at the Conference venue during the whole conference. Please connect to the wireless networks ECMLPKDD2017 or ECMLPKDD2017-5 and use the password "SK2017ECMLPKDD".

CONFERENCE APP

We created an event for the conference on EventBase with all details about the program. You can browse it by downloading the EventBase app from the app store and searching for "ECML PKDD 2017".

<u>OR CODES</u>

The QR codes throughout the brochure, contain links to the camera ready versions of the accepted papers.

MOBILE PHONES

Please make sure to switch off your mobile phones during the sessions.

Note that roaming charges may apply in case of making phone calls, sending SMS and making data transfer.

EMERGENCY CONTACTS

Emergencies related to the conference should be directed to:

- +389 77 527 604 (Ivica Dimitrovski, Local Organization Chair),
- +389 78 749 804 (Dragi Kocev, Production & PR Chair), or
- +389 78 749 839 (Registration Desk).
- These three numbers are not toll free.

The Emergency Call Center in Macedonia uses the number 112.

TRANSPORT

For the three events (on Monday, Tuesday and Thursday) held outside the Main Conference Venue, there will be <u>organized bus transport</u>.

The buses will depart at 18.00 from the parking lot in front of Main Conference Venue.

The buses will start their trip back to Main Conference Venue at 22.15, the last one departing at 22.45. The buses on their trip back will make a stop at Hotel Karpos and proceed to the Main Conference Venue.

CURRENCY

The official currency of Macedonia is Denar (MKD). It has a stable exchange rate with the Euro: 1 EUR = 61.5 MKD. International credit cards are accepted for payments in most hotels, restaurants and shops. ATM machines are available throughout the city.

INSURANCE

The organizers of ECML PKDD 2017 cannot be held liable for any injury, loss or damage occurring during the conference.

ELECTRICITY

Macedonia uses a 230 volt 50 Hz system with European Type F plugs.

<u>INSTRUCTIONS FOR</u> PRESENTERS AND CHAIRS

FOR ALL GENERAL ENQUIRERS REGARDING PROJECTOR/COMPUTER/INTERNET ISSUES **PLEASE CONTACT THE REGISTRATION DESK**.

INSTRUCTIONS FOR SPEAKERS

- » Each of the conference rooms will have a projector and a laptop. The laptop will have Powerpoint and .pdf viewing software pre-installed.
- » You should be in the conference room <u>10 minutes</u> before your session and report to the Session Chair. In addition, you will be asked to copy your presentation to the laptop and check if everything works fine. All of the presentations will be deleted after the end of the session.
- » A volunteer will be available in every main session room to provide assistance.
- » The time allocated to each speaker, including time to set up and time for questions, is :
 - Conference, Journal, Applied Data Science and Nectar Tracks: 20 minutes.
 - Demo track elevator pitches: 4 minutes.
 - Workshops, Forums and Discovery Challenges: per workshop/forum/challenge instructions.

INSTRUCTIONS FOR SESSION CHAIRS

- » Please make sure to show up 10 minutes before the session starts to check the equipment works.
- » Please stick to the schedule. If a speaker fails to show up, please announce a short break.
- » Please moderate questions.
- » Please do not start sessions or talks early as attendees may need some time to move between rooms.

INSTRUCTIONS FOR POSTER PRESENTERS

- » The two poster sessions are scheduled:
- Tuesday, September 19, from 20:20 to 22:30 (for papers presented orally on Tuesday and Wednesday before the lunch break i.e.12:40)
- Thursday, September 21, from 20:20 to 22:30
- (for papers presented orally on Wednesday after the lunch break i.e. 12:40)
- » Both poster sessions will take place in the Macedonian Opera and Ballet and will be catered events.
- » The poster size should be A0 (841 x 1189 mm) in portrait orientation.
- » Presenters should set up their on the day of the poster session, before the last keynote lecture for the day. The presenters should remove their posters after the end of the poster session.
- » Volunteers will be available to assist you.

INSTRUCTIONS FOR SOFTWARE DEMOS

- » Demo sessions will be held together with the poster sessions in the Macedonian Opera and Ballet.
- » The demos are divided into two sessions. The elevator pitches are scheduled immediately after the two afternoon invited talks. The two demo sessions are scheduled:
 - Tuesday, from 20:20 to 22:30
 - Thursday, from 20:20 to 22:30
- » Demonstration desks will be set up next to the poster area.
- » Presenters should show up in advance to set up their demos and remove them at the end of the session.
- » Each demo will be provided with a desk, two chairs, a monitor and a (4 or 5 way) power socket. Internet connection will be available.





KEYNOTE SPEAKERS



Frank Hutter

University of Freiburg, Germany

Monday, September 18 2017 @ 19:30 Macedonian National Theatre

TOWARDS END-TO-END LEARNING & OPTIMIZATION

Deep learning has recently helped AI systems to achieve human-level performance in several domains, including speech recognition, object classification, and playing several types of games. The major benefit of deep learning is that it enables end-to-end learning of representations of the data on several levels of abstraction. However, the overall network architecture and the learning algorithms' sensitive hyperparameters still need to be set manually by human experts. In this talk, I will discuss extensions of Bayesian optimization for handling this problem effectively, thereby paving the way to fully automated end-to-end learning. I will focus on speeding up Bayesian optimization by reasoning over data subsets and initial learning curves, sometimes resulting in 100-fold speedups in finding good hyperparameter settings. I will also show competition-winning practical systems for automated machine learning (AutoML) and briefly show related applications to the end-to-end optimization of algorithms for solving hard combinatorial problems.

Chair: Sašo Džeroski

About the speaker :

Frank Hutter is an Emmy Noether Research Group Lead (eq. Asst. Prof.) at the Computer Science Department of the University of Freiburg (Germany). He received his PhD from the University of British Columbia (2009). Frank's main research interests span artificial intelligence, machine learning, combinatorial optimization, and automated algorithm design. He received a doctoral dissertation award from the Canadian Artificial Intelligence Association and, with his coauthors, several best paper awards (including from JAIR and IJCAI) and prizes in international competitions on machine learning, SAT solving, and AI planning. In 2016 he received an ERC Starting Grant for a project on automating deep learning based on Bayesian optimization, Bayesian neural networks, and deep reinforcement learning.

Sponsored by





Google DeepMind, UK

Tuesday, September 19 2017 @ 09:00 CONGRESS HALL 1, Aleksandar Palace



FRONTIERS IN RECURRENT NEURAL NETWORK RESEARCH

In the last few years, recurrent neural networks (RNNs) have become the Swiss army knife of sequence processing for machine learning. Problems involving long and complex data streams, such as speech recognition, machine translation and reinforcement learning from raw video are now routinely tackled with RNNs. However, significant limitations still exist for such systems, such as their ability to retain large amounts of information in memory, and the challenges of gradient-based training on very long sequences. My talk will review some of the new architectures and training strategies currently being developed to extend the frontiers of this exciting field.

Chair: Jaakko Hollmén

About the speaker :

Alex Graves completed a BSc in Theoretical Physics at the University of Edinburgh, Part III Maths at the University of Cambridge, a PhD in artificial intelligence at IDSIA with Jürgen Schmidhuber, followed by postdocs at the Technical University of Munich and with Geoff Hinton at the University of Toronto. He is now a research scientist at DeepMind. His contributions include the Connectionist Temporal Classification algorithm for sequence labelling (now widely used for commercial speech and handwriting recognition), stochastic gradient variational inference, and the Neural Turing Machine / Differentiable Neural Computer architectures.



John Quackenbush

Dana-Farber Cancer Institute, USA Harvard TH Chan School of Public Health, USA

Tuesday, September 19 2017 @ 19:00 Macedonian Opera & Ballet

USING NETWORKS TO LINK GENOTYPE TO PHENOTYPE

We know that genotype influences phenotype, but aside from a few highly penetrant Mendelian disorders, the link between genotype and phenotype is not well understood. We have used gene expression and genetic data to explore gene regulatory networks, to study phenotypic state transitions, and to analyze the connections between genotype and phenotype. I will describe how networks and their structure provide unique insight into how small effect variants influence phenotype.

Chair: Celine Vens

About the speaker :

John Quackenbush received his PhD in theoretical particle physics from UCLA in 1990. Following a postdoctoral fellowship in experimental high-energy physics, he received an NIH research award to work on the Human Genome Project and helped map chromosome 11 and sequence chromosomes 21 and 4. After four years of working in genomics and computational biology at the Salk Institute and then at Stanford University, John joined The Institute for Genomic Research (TIGR), pioneering microarray expression technologies and analytical methods. In 2005, he joined the Dana-Farber Cancer Institute and the Harvard T.H. Chan School of Public Health, where he uses computational and systems biology methods to explore the complexities of human disease, including cancer. In 2011, he cofounded Genospace, a precision-medicine software company acquired by Hospital Corporation of America in 2017. John's many awards include recognition as a White House Champion of Change for making genomic data useful and widely accessible.

Inderjit Dhillon

University of Texas at Austin, USA

Wednesday, September 20 2017 @ 09:00 CONGRESS HALL 1, Aleksandar Palace



MULTI-TARGET PREDICTION VIA LOW-RANK EMBEDDINGS

Linear prediction methods, such as linear regression and classification, form the bread-and-butter of modern machine learning. The classical scenario is the presence of data with multiple features and a single target variable. However, there are many recent scenarios where there are multiple target variables. For example, recommender systems, predicting bid words for a web page (where each bid word acts as a target variable), or predicting diseases linked to a gene. In many of these scenarios, the target variables might themselves be associated with features. In these scenarios, bilinear and nonlinear prediction via low-rank embeddings have been shown to be extremely powerful. The low-rank embeddings serve a dual purpose: (i) they enable tractable computation even in the face of millions of data points as well as target variables, and (ii) they exploit correlations among the target variables, even when there are many missing observations. We illustrate our methodology on various modern machine learning problems: recommender systems, multi-label learning and inductive matrix completion, and present results on some standard benchmarks as well as an application that involves prediction of gene-disease associations.

Chair: Michelangelo Ceci

About the speaker :

Inderjit Dhillon is the Gottesman Family Centennial Professor of Computer Science and Mathematics at UT Austin, where he is also the Director of the ICES Center for Big Data Analytics. Currently he is on leave from UT Austin and works as Amazon Fellow at A9/Amazon, where he is developing and deploying state-of-the-art machine learning methods for Amazon search. His main research interests are in big data, machine learning, network analysis, linear algebra and optimization. He received his B.Tech. degree from IIT Bombay, and Ph.D. from UC Berkeley. Inderjit has received several awards, including the ICES Distinguished Research Award, the SIAM Outstanding Paper Prize, the Moncrief Grand Challenge Award, the SIAM Linear Algebra Prize, the University Research Excellence Award, and the NSF Career Award. He has published over 175 journal and conference papers, and has served on the Editorial Board of the Journal of Machine Learning Research, the IEEE Transactions of Pattern Analysis and Machine Intelligence, Foundations and Trends in Machine Learning and the SIAM Journal for Matrix Analysis and Applications. Inderjit is an ACM Fellow, an IEEE Fellow, a SIAM Fellow and an AAAS Fellow.



Pierre-Philippe Mathieu

ESA/ESRIN, EO Science, Applications and New Technologies, Italy

Thursday, September 21 2017 @ 09:00 CONGRESS HALL 1, Aleksandar Palace

ENABLING A SMARTER PLANET WITH EARTH OBSERVATION

Nowadays, teams of researchers around the world can easily access a wide range of open data across disciplines and remotely process them on the Cloud, combining them with their own data to generate knowledge, develop information products for societal applications, and tackle complex integrative complex problems that could not be addressed a few years ago. Such rapid exchange of digital data is fostering a new world of data-intensive research, characterized by openness, transparency, and scrutiny and traceability of results, access to large volume of complex data, availability of community open tools, unprecedented level of computing power, and new collaboration among researchers and new actors such as citizen scientists. The EO scientific community is now facing the challenge of responding to this new paradigm in science 2.0 in order to make the most of the large volume of complex and diverse data delivered by the new generation of EO missions, and in particular the Sentinels. In this context, ESA is supporting a variety of activities in partnership with research communities to ease the transition and make the most of the data. These include the generation of new open tools and exploitation platforms, exploring new ways to disseminate data, building new partnership with citizen scientists, and training the new generation of data scientists. The talk will give a brief overview of some of ESA activities aiming to facilitate the exploitation of large amounts of data from EO missions in a collaborative, cross-disciplinary, and open way, for uses ranging from science to applications and education.

Chair: Ljupčo Todorovski

About the speaker :

Pierre-Philippe Mathieu is an Earth Observation Data Scientist at the European Space Agency in ESRIN (Frascati, Italy). He has spent 20+ years working in the field of environmental and ocean modelling, weather risk management and remote sensing. He has a degree in mechanical engineering and an M.Sc from the University of Liege (Belgium), a Ph.D. in oceanography from the University of Louvain (Belgium), and a Management degree from the University of Reading Business School (UK).





Cordelia Schmid

INRIA, France

Thursday, September 21 2017 @ 19:00 Macedonian Opera & Ballet

AUTOMATIC UNDERSTANDING OF THE VISUAL WORLD

One of the central problems of artificial intelligence is machine perception, i.e., the ability to understand the visual world based on input from sensors, such as cameras. Computer vision is the area which analyzes visual input. In this talk, I will present recent progress in visual understanding. It is for the most part due to the design of robust visual representations and learned models capturing the variability of the visual world based on state-of-the-art machine learning techniques, including convolutional neural networks. Progress has resulted in technology for a variety of applications. I will present in particular results for human action recognition.

Chair: Ivica Dimitrovski

About the speaker :

Cordelia Schmid holds an M.S. degree in Computer Science from the University of Karlsruhe and a Doctorate, also in Computer Science, from the Institut National Polytechnique de Grenoble (INPG). Her doctoral thesis received the best thesis award from INPG in 1996. Dr. Schmid was a postdoctoral research assistant in the Robotics Research Group of Oxford University in 1996--1997. Since 1997 she has held a permanent research position at INRIA Grenoble Rhone-Alpes, where she is a research director and directs an INRIA team. Dr. Schmid has been an Associate Editor for IEEE PAMI (2001--2005) and for IJCV (2004--2012), editor-in-chief for IJCV (2013---), a program chair of IEEE CVPR 2005 and ECCV 2012 as well as a general chair of IEEE CVPR 2015 and ECCV 2020. In 2006, 2014 and 2016, she was awarded the Longuet-Higgins prize for fundamental contributions in computer vision that have withstood the test of time. She is a fellow of IEEE. She was awarded an ERC advanced grant in 2013, the Humboldt research award in 2015 and the Inria & French Academy of Science Grand Prix in 2016. She was elected to the German National Academy of Sciences, Leopoldina, in 2017.



PROGRAM

PROGRAM AT A GLANCE



MACEDONIAN NATIONAL THEATRE

19:00 - 19:30	OPENING CEREMONY
19:30 - 20:30	Invited Talk FRANK HUTTER
	TOWARDS END-TO-END LEARNING & OPTIMIZATION
20:30 - 23:00	WEI COME RECEPTION
20.00 20.00	

EU PROJECTS FORUM

CONGRESS HALL 4

09:00 - 17:40

EU Projects Forum

Chairs :

Petra Kralj Novak, Jožef Stefan Institute, Slovenia Nada Lavrač, Jožef Stefan Institute, Slovenia

The EU Projects Forum at ECML PKDD 2017 is a novel initiative that encourages the dissemination of EU projects and their results to the targeted scientific audience of the conference participants. It aims at becoming an annual venue for the EU funded projects to present their vision and work to the conference audience, and an opportunity for the ECML PKDD audience to learn about the European scientific success stories in their research field. The inaugural EU Projects Forum (scheduled as a full-day satellite event of ECML PKDD) encompasses a talk by the EU project officer Salvatore Spinello on the main features and evaluation process of FET Open, a tutorial on how to write successful project proposals by Richard Wheeler, as well as the presentations of several ERC grants and EU projects.

Collection of abstracts: http://ecmlpkdd2017.ijs.si/EUPFabstracts.pdf

Invited talk FET OPEN: MAIN FEATURES AND EVALUATION PROCESS

Salvatore Spinello

Research Programme Officer at REA, Belgium

FET Open is one of the most attractive research programme under Horizon 2020. FET-Open supports the early-stage, highrisk research around new ideas towards radically new future technologies. It explores an open range of new and disruptive technological possibilities in all areas of Science & Technology, inspired by cutting edge science, unconventional collaborations and pioneering new ways to create the optimum conditions for serendipity to occur.

In these first 3 years of Horizon 2020, a total of 2.648 proposals were submitted to the FET-Open programme and covered a wide range of disciplines: from Physics to Life Sciences, from Information Sciences and Engineering to Chemistry. Most proposals show indeed high degree of interdisciplinarity.

During my presentation I will focus on Research and Innovation-Actions (RIA) and the so-called "gatekeepers" that every excellent proposal should address. I will then present the evaluation process that allows the selection of the best proposals, resulting in a continuously growing portfolio of high quality interdisciplinary projects. I will conclude providing some statistics in terms of country and organization participations, scientific fields covered and interdisciplinarity.

Invited talk HINTS ON HOW TO WRITE A SUCCESSFUL PROJECT PROPOSAL

Richard Wheeler

Edinburgh Scientific, UK

EU Funding has never been more important, nor harder to get. In this very practical talk, Richard Wheeler will provide an insider's view to securing European Union funding, including tips and tricks on good proposal writing, what really happens in EU review meetings, why most proposals fail, good project management methods, and more. Attendees will have the opportunity to ask questions and discuss their ideas in an informal workshop environment.

EU Projects Forum Program at a glance

09:00 - 09:10		<mark>Opening</mark> Nada Lavrač, Petra Kralj Novak
09:10 - 10:00	INVITED TALK	FET Open: Main Features and Evaluation Process Salvatore Spinello
10:00 - 10:30	ERC GRANT	FORSIED: Formalizing Subjective Interestingness in Exploratory Data Mining Tijl De Bie, Jefrey Lijffijt
10:30 - 11:00		Coffee & Tea Break
11:00 - 11:20	H2020 PROJECT	SoBigData Research Infrastructure Fosca Giannotti
11:20 - 11:40	H2020 PROJECT	REnnovates: Towards Net-Zero Energy Communities and Beyond Hussain Kazmi
11:40 - 12:00	H2020 PROJECT	CF-Web: ClowdFlows Data and Text Analytics Marketplace on the Web Martin Žnidaršič
12:00 - 12:20	H2020 PROJECT	AURORA: Advanced User-Centric Efficiency Metrics for Air Traffic Performance Analytics Brian Mac Namee
12:20 - 12:40	FP7 FET OPEN	MAESTRA: Learning from Massive, Incompletely Annotated, and Structured Data Sašo Džeroski
12:40 - 14:00		Lunch Break
14:00 - 14:30	ERC GRANT	CAUSALPATH: Next Generation Causal Analysis Inspired By the Induction of Biological Pathways from Cytometry Data Vincenzo Lagani, Ioannis Tsamardinos
14:30 - 15:40	INVITED TALK	Hints on How to Write a Successful Project Proposals Part 1 Richard Wheeler
15:40 - 16:00		Coffee & Tea Break
16:00 - 16:30	INVITED TALK	Hints on How to Write a Successful Project Proposals Part 2 Richard Wheeler
17:00 - 17:30		Late Breaking Project Presentations
17:30 - 17:40		Closing Remarks

Combined Tutorial with Workshop

CONGRESS HALL 1

09:00 - 17:40

IAL 2017 Interactive Adaptive Learning

Organizers: Georg Krempl, University Magdeburg, Germany Vincent Lemaire, Orange Labs, France Robi Polikar, Rowan University, USA Bernhard Sick, University of Kassel, Germany Daniel Kottke, University of Kassel, Germany Adrian Calma, University of Kassel, Germany

Webpage :

http://www.uni-kassel.de/go/ial2017

This workshop on interactive adaptive learning aims at discussing techniques and approaches for optimizing the whole learning process, including the interaction with human supervisors, processing systems, and includes adaptive, active, semi-supervised, and transfer learning techniques, and combinations thereof in interactive and adaptive machine learning systems. Our objective is to bridge the communities researching and developing these techniques and systems in machine learning and data mining. Therefore we welcome contributions that present a novel problem setting, propose a novel approach, or report experience with the practical deployment of such a system and raise unsolved questions to the research community.

Tutorial

Introduction to Stream Mining

This part starts with the classic stream mining paradigm. In its context, we discuss the challenges posed by non-stationarity and limitations in processing, storage, and supervision capacities. We briefly summarize related techniques, e.g. for incremental processing, forgetting, and change detection. This part concludes by an overview on further challenges that are investigated in the state-of-the-art research.

Active Learning

In this part of the tutorial, we focus on techniques for optimizing the interaction of a machine learning system with an oracle such as a human supervisor. We review active machine learning techniques, with focus on adaptive active learning for evolving and streaming data. We discuss recent advances and conclude with an overview on open research questions in adaptive active machine learning.

Semi-Supervised and Transfer Learning

This part of the tutorial addresses the problem of learning with incomplete or delayed supervision. We focus on the problem of learning with verification latency, and review techniques from change mining, semi-supervised and (unsupervised) transfer learning in non-stationary environments. We conclude with an overview on open challenges. Evaluation, Applications and Emerging Trends

This last part of the tutorial takes an integrative view on the previous parts, with focus on industrial applications and open challenges of adaptive interactive mining systems as a whole. We briefly discuss the related issues of evaluation and deployment, applications, reported challenges and solutions, and highlight potential directions for future research.

Invited talk ENSEMBLE LEARNING FROM DATA STREAMS WITH ACTIVE AND SEMI-SUPERVISED APPROACHES

Bartosz Krawczyk Virginia Commonwealth University, USA

IAL 2017 Program at a glance

09:00 - 09:10	Welcome
09:10 - 09:50	Introduction to Stream Mining Georg Krempl
09:50 - 10:40	Active Learning Daniel Kottke
10:40 - 11:00	Coffee & Tea Break
11:00 - 11:40	Semi-Supervised and Transfer Learning Georg Krempl
11:40 - 12:30	Evaluation, Applications and Emerging Trends Vincent Lemaire
12:30 - 12:40	Spotlights on Poster Session
	Probabilistic Expert Knowledge Elicitation of Feature Relevances in Sparse Linear Regression
	Pedram Daee, Tomi Peltola, Marta Soare, Samuel Kaski
	Users Behavioural Inference with Markovian Decision Process and Active Learning
	Firas Jarboui, Vincent Rocchisani, Wilfried Kirchenmann
	Multi-Arm Active Transfer Learning for Telugu Sentiment Analysis Subba Reddy Oota, Vijayasaradhi Indurthi, Mounika Reddy Marreddy, Sandeep Sricharan Mukku, Radhika Mamidi
12:40 - 14:00	Lunch Break + Poster Session
14:00 - 14:20	Probabilistic Active Learning with Structure-Sensitive Kernels Dominik Lang, Daniel Kottke, Georg Krempl, Bernhard Sick
14:20 - 14:40	Transfer Learning for Time Series Anomaly Detection Vincent Vercruyssen, Wannes Meert, Jesse Davis
14:40 - 15:40	Ensemble Learning from Data Streams with Active & Semi-Supervised Approaches Bartosz Krawczyk
15:40 - 16:00	Coffee & Tea Break + Poster Session
16:00 - 16:20	Simulation of Annotators for Active Learning: Uncertain Oracles Adrian Calma and Bernhard Sick
16:20 - 16:40	Interactive Anonymization for Privacy-Aware Machine Learning Bernd Malle, Peter Kieseberg, Andreas Holzinger
16:40 - 17:40	Panel Discussion George Kachergis, Bartosz Krawczyk, Myra Spiliopoulou, Jerzy Stefanowski

Combined Tutorial with Workshop

CONGRESS HALL 2

09:00 - 17:40

IoT Large Scale Learning from Data Streams

Organizers : Moamar Sayed-Mouchaweh, High Engineering School of Mines, Douai, France Albert Bifet, Telecom-ParisTech, France Hamid Bouchachia, University of Bournemouth, UK João Gama, University of Porto, Portugal Rita Ribeiro, University of Porto, Portugal

Webpage

https://abifet.wixsite.com/iotstreaming2017

The volume of data is rapidly increasing due to the development of the technology of information and communication. This data comes mostly in the form of streams. Learning from this ever-growing amount of data requires flexible learning models that self-adapt over time. In addition, these models must take into account many constraints: (pseudo) real-time processing, high-velocity, and dynamic multi-form change such as concept drift and novelty. This workshop welcomes novel research about learning from data streams in evolving environments. It will provide the researchers and participants with a forum for exchanging ideas, presenting recent advances and discussing challenges related to data streams processing. It solicits original work, already completed or in progress. Position papers are also considered. This workshop is combined with a tutorial treating the same topic and will be presented in the same day.

Tutorial

Gianmarco De Francisci Morales, Albert Bifet, Latifur Khan, Moamar Sayed-Mouchaweh, Joao Gama, Wei Fan

The challenge of deriving insights from the Internet of Things (IoT) has been recognized as one of the most exciting and key opportunities for both academia and industry. Advanced analysis of big data streams from sensors and devices is bound to become a key area of data mining research as the number of applications requiring such processing increases. Dealing with the evolution over time of such data streams, i.e., with concepts that drift or change completely, is one of the core issues in IoT stream mining. This tutorial is a gentle introduction to mining IoT big data streams. The first part introduces data stream learners for classification, regression, clustering, and frequent pattern mining. The second part deals with scalability issues inherent in IoT applications, and discusses how to mine data streams on distributed engines such as Spark, Flink, Storm, and Samza

Invited talk LEARNING FROM NON-STATIONARY DISTRIBUTIONS

Geoff Webb Monash University, Australia

IoT Large Scale Learning from Data Streams Program at a glance

09:00	Welcome
09:00 - 10:40	Tutorial 1: IoT Fundamentals and Stream Mining Algorithms Gianmarco De Francisci Morales, Albert Bifet, Latifur Khan, Moamar Sayed-Mouchaweh, Joao Gama, Wei Fan
10:40 - 11:00	Coffee & Tea Break
11:00 - 12:40	Tutorial: 2. IoT Distributed Big Data Stream Mining and Applications Gianmarco De Francisci Morales, Albert Bifet, Latifur Khan, Moamar Sayed-Mouchaweh, Joao Gama, Wei Fan
12:40 - 14:00	Lunch Break
	SESSION 1 - Chair : Albert Bifet
14:00 - 14:45	Learning from Non-Stationary Distributions Geoff Webb
15:00 - 15:15	A Sliding Window Filter for Time Series Streams Gordon Lesti and Stephan Spiegel
15:20 - 15:35	Evolutive Deep Models for Online Learning on Data Streams with No Storage
	Andrey Besedin, Pierre Blanchart, Michel Crucianu, Marin Ferecatu
15:40 - 16:00	Coffee & Tea Break
	SESSION 2 - Chair : Moamar Sayed-Mouchaweh
16:00 - 16:15	Hybrid Self Adaptive Learning Scheme for Simple and Multiple Drift-Like Fault Diagnosis in Wind Turbine Pitch Sensors Houari Toubakh, Moamar Sayed-Mouchaweh
16:20 - 16:35	Aggregation Algorithm Vs. Average For Time Series Prediction Waqas Jamil, Yuri Kaliniskan, Abdelhamid Bouchachia
16:40 - 16:55	Self-Adaptive Ensemble Classifier for Handling Complex Concept Drift Imen Khamassi, Moamar Sayed-Mouchaweh
17:00 - 17:15	Summary Extraction on Data Streams in Embedded Systems Sebastian Buschjäger, Katharina Morik

CONGRESS HALL 3

09:30 - 17:40

MLSA 2017 Machine Learning and Data Mining for Sports Analytics

Organizers : Jesse Davis, KU Leuven, Belgium Mehdi Kaytoue, INSA de Lyon, France Albrecht Zimmermann, Université Caen, France

Webpage https://dtai.cs.kuleuven.be/events/MLSA17/

Sports Analytics has been a steadily growing and rapidly evolving area over the last decade, both in US professional sports leagues and in European football leagues. The majority of techniques used in the field so far are statistical. However, there has been growing interest in the Machine Learning and Data Mining community about this topic as this setting is interesting, challenging and offers new sources of data. The workshop concerns all aspects of applying machine learning and data mining techniques for sports problems such as match strategy, tactics, and analysis; player acquisition, player valuation, and team spending; injury prediction and prevention; match outcome and league table prediction; and tournament design and scheduling among others.

Invited talk DECISION-MAKING, ANTICIPATION AND GAZE CONTROL IN SPORTS

Matt Dicks University of Portsmouth, UK

Invited talk SCISPORTS: ENABLING DATA-DRIVEN DECISION MAKING IN PROFESSIONAL FOOTBALL

Jan Van Haaren KU Leuven, Belgium

MLSA 2017 Program at a glance

09:30 - 09:40	Opening Remarks
09:40 - 10:40	Decision-Making, Anticipation and Gaze Control in Sports Matt Dicks
10:40 - 11:00	Coffee & Tea Break
11:00 - 11:25	Predicting the Potential of Professional Soccer Players Ruben Vroonen, Tom Decroos, Jan Van Haaren, Jesse Davis
11:25 - 11:50	STARSS: A Spatio-Temporal Action Rating System for Soccer Tom Decroos, Jan Van Haaren, Vladimir Dzyuba, Jesse Davis
11:50 - 12:15	Who Is Going to Get Hurt? Predicting Injuries in Professional Soccer
	Alessio Rossi, Luca Pappalardo, Paolo Cintia, Javier Fernandez, F. Marcello Iaia, Daniel Medina
12:15 - 12:40	Enabling Training Personalization By Predicting the Session Rate of Perceived Exertion Gilles Vandewiele, Youri Geurkink, Maarten Lievens, Femke Ongenae, Filip De Turck, Jan Boone
12:40 - 14:10	Lunch Break
14:10 - 15:10	SciSports: Enabling Data-Driven Decision Making in Professional Football
	Jan Van Haaren
15:10 - 15:35	An Artificial Neural Network-Based Prediction Model for Underdog Teams in NBA Matches Paolo Giuliodori
15:35 - 16:00	Coffee & Tea Break
16:00 - 16:25	Dynamic Winner Prediction in Twenty20 Cricket: Based on Relative Team Strengths Sasank Viswanadha, Kaustubh Sivalenka, Madan Gopal Jhawar, Vikram Pudi
16:25 - 16:50	Linking Event Mentions From Cricket Match Reports to Commentaries Manish Gupta
16:50 - 17:15	Honest Mirror: Quantitative Assessment of Player Performance in an ODI Cricket Match Madan Gopal Jhanwar, Vikram Pudi
17:15 - 17:40	Wrap-Up

BANQUET HALL

08:55 - 12:45

SURL 2017 Scaling-Up Reinforcement Learning

Organizers : Felipe Leno da Silva, University of São Paulo, Brazil Ruben Glatt, University of São Paulo, Brazil

Webpage http://surl.tirl.info/

Reinforcement Learning (RL) has achieved many successes over the years in training autonomous agents to perform simple tasks. However, one of the major remaining challenges in RL is scaling it to high-dimensional, real-world applications. Although many works have already focused on strategies to scale-up RL techniques and to find solutions for more complex problems with reasonable successes, many issues still exist. This workshop encourages to discuss diverse approaches to accelerate and generalize RL, such as the use of approximations, abstractions, hierarchical approaches, and Transfer Learning. Scaling-up RL methods has major implications on the research and practice of complex learning problems and will eventually lead to successful implementations in real-world applications.

This workshop intends to bridge the gap between conventional and scalable RL approaches. We aim to bring together researchers working on different approaches to scale-up RL with the goal to solve more complex or larger scale problems. We intend to make this an exciting event for researchers worldwide, not only for the presentation of top quality papers, but also to spark the discussion of opportunities and challenges for future research directions.

Invited talk SCALING UP RL WITH OFFLINE TASK HIERARCHIES

Devin Schwab

Carnegie Mellon University, USA

Invited talk SCALING UP POLICY SEARCH METHODS FOR ROBOTICS

Herke van Hoof McGill University, Canada

SURL 2017 Program at a glance

08:55 - 09:00	Opening
	SESSION 1
09:00 - 09:20	Case-Based Policy Inference for Transfer in Reinforcement Learning R. Glatt, F. L. Silva, A. H. R. Costa
09:20 - 09:40	Stabilising Experience Replay for Deep Multi-Agent Reinforcement Learning J. Foerster, N. Nardelli, G. Farquhar, T. Afouras, P. Torr, P. Kohli, S. Whiteson
09:40 - 10:00	Constrained Bayesian Reinforcement Learning Via Approximate Linear Programming J. Lee, Y. Jang, P. Poupart, K. Kim
10:00 - 10:40	Scaling Up RL with Offline Task Hierarchies Devin Schwab
10:40 - 11:00	Coffee & Tea Break
	SESSION 2
11:00 - 11:20	Automatic Object-Oriented Curriculum Generation for Reinforcement Learning F. L. Silva and A. H. R. Costa
11:20 - 11:40	Count-Based Exploration in Feature Space for Reinforcement Learning J. Martin, S. Narayanan S., T. Everitt , M. Hutter
11:40 - 12:00	Learning Multimodal Transition Dynamics for Model-Based Reinforcement Learning T. M. Moerland, J. Broekens, C. M. Jonker
12:00 - 12:40	Scaling Up Policy Search Methods for Robotics Herke Van Hoof
12:40 - 12:45	Closing / Community Meeting / Integration

09:00 - 12:45

SoGood 2017 Data Science for Social Good

Organizers : Ricard Gavaldà, UPC BarcelonaTech, Spain Irena Koprinska, University of Sidney, Australia Stefan Kramer, JGU Mainz, Germany

Webpage

https://sites.google.com/site/ecmlpkddsogood2017/

This workshops aims to attract papers presenting applications of Data Science to Social Good, or else that take into account social aspects of Data Science methods and techniques. Application domains should be as varied as possible. The novelty of the application and its social impact will be major selection criteria.

Invited talk USING SOCIAL MEDIA TO UNDERSTAND COLLECTIVE AND PERSONAL EVENTS: CHALLENGES AND APPLICATIONS

Alexandra Olteanu

IBM T.J. Watson Research Center, USA

SoGood 2017 Program at a glance	09:00 - 09:10	Welcome
		SESSION 1
	09:10 - 09:30	Some Like It Hoax: Automated Fake News Detection in Social Networks Eugenio Tacchini, Gabriele Ballarin, Marco Della Vedova, Stefano Moret, Luca De Alfaro
	09:30 - 10:30	Using Social Media to Understand Collective and Personal Events: Challenges and Applications
		Alexandra Olteanu
	10:30 - 10:40	Discussion
	11:40 - 12:00	Coffee & Tea Break
		SESSION 2
	11:00 - 11:20	Learning from Administrative Health Registries Jonathan Rebane, Isak Karlsson, Lars Asker, Henrik Boström, Panagiotis Papapetrou
	11:20 - 11:40	Online Topic Modeling: Keeping Track of News Topics for Social Good
		Zahra Ahmadi, Sophie Burkhardt, Stefan Kramer
	11:40 - 12:00	A Bayesian Reputation Framework for Citizen Science
		Joan Garriga, Jaume Piera, Frederic Bartumeus
	12:00 - 12:40	Panel & Discussion

MIDAS 2017 MIning DAta for financial applicationS

Organizers : Ilaria Bordino, UniCredit, R&D Department, Italy Ilaria Bordino, IMT School for Advanced Studies Lucca, Italy Fabio Fumarola, UniCredit, R&D Department, Italy Francesco Gullo, UniCredit, R&D Department, Italy Tiziano Squartini, IMT School for Advanced Studies Lucca, Italy

Webpage

http://networks.imtlucca.it/conferences/midas2017

Like the famous King Midas, popularly remembered in Greek mythology for his ability to turn everything he touched with his hand into gold, we believe that the wealth of data generated by modern technologies, with widespread presence of computers, users and media connected by Internet, is a goldmine for tackling a variety of problems in the financial domain.

The MIDAS workshop is aimed at discussing challenges, potentialities, and applications of leveraging data-mining tasks to tackle problems in the financial domain. The workshop provides a premier forum for sharing findings, knowledge, insights, experience and lessons learned from mining data generated in various application domains.

Invited talk

EVOLVING DATA, EVOLVING MODELS IN ECONOMY AND FINANCE Joao Gama

University of Porto, Portugal

MIDAS 2017 Program at a glance	14:00 - 14:10	Opening
		SESSION 1
	14:10 - 15:10	Evolving Data, Evolving Models in Economy and Finance Joao Gama
	15:10 - 15:35	Influence Analysis in Business Social Media Flora Amato, Vincenzo Moscato, Antonio Picariello, Giovanni Ponti, Giancarlo Sperli
	15:40 - 16:00	Coffee & Tea Break
		SESSION 2
	16:00 - 16:25	Event Recognition Strategies Applied in the Mercurio Project Davide Azzalini, Fabio Azzalini, Davide Greco, Miriana Mazuran, Letizia Tanca
	16:25 - 16:50	BoostEMM - Transparent Boosting Using Exceptional Model Mining Simon Van Der Zon, Oren Zeev Ben Mordehai, Tom Vrijdag, Werner Van Ipenburg, Jan Veldsink, Wouter Duivesteijn, Mykola Pechenizkiy
	16:50 - 17:15	Profiling Smart Contracts Interactions Tensor Decomposition and Graph Mining
		Jérémy Charlier, Sofiane Lagraa, Radu State, Jérôme François
	17:15 - 17:40	Machine Learning for Multi-Step Ahead Forecasting of Volatility Proxies
		Jacopo De Stefani, Olivier Caelen, Dalila Hattab and Gianluca Bontempi
	17:40	Closing

09:00 - 17:40

PAP 2017 Personal Analytics & Privacy

Organizers : Serge Abiteboul, Inria, ENS Paris, France Riccardo Guidotti, KDDLab, ISTI-CNR Pisa, Italy Anna Monreale, University of Pisa, Italy Dino Pedreschi, University of Pisa, Italy

Webpage

http://kdd.isti.cnr.it/pap2017

Personal data analytics and individual privacy protection are the key elements to leverage nowadays services to a new type of systems. The availability of personal analytics tools able to extract hidden knowledge from individual data while protecting the privacy right can help the society to move from organization-centric systems to user-centric systems, where the user is the owner of her personal data and is able to manage, understand, exploit, control and share her own data and the knowledge deliverable from them in a completely safe way.

The purpose of PAP is to encourage principled research that will lead to the advancement of personal data analytics, personal services development, privacy, data protection and privacy risk assessment. The workshop will seek top-quality submissions addressing important issues related to personal analytics, personal data mining and privacy in the context where real individual data (spatio-temporal data, call details records, tweets, mobility data, transactional data, social networking data, etc.) are used for developing a data-driven service, for realizing a social study aimed at understanding nowadays society, and for publication purposes.

Invited talk THE RISE OF DECENTRALIZED PERSONAL DATA MARKETS Bruno Lepri

Fondazione Bruno Kessler, Italy

Invited talk PERSONAL KNOWLEDGE MANAGEMENT SYSTEMS

Serge Abiteboul Inria and ENS, France
PAP 2017 Program at a glance

09:00 - 09:15	Welcome & Workshop Overview
09:15 - 10:00	The Rise of Decentralized Personal Data Markets Bruno Lepri
10:00 - 10:30	Your Privacy, My Privacy? On Leakage Risk Assessment in Online Social Networks Ruggero G. Pensa, Livio Bioglio
10:30 - 11:00	Coffee & Tea Break
11:00 - 11:30	Assessing Privacy Risk in Retail Data Roberto Pellungrini, Francesca Pratesi, Luca Pappalardo
11:30 - 12:00	Differential Privacy and Neural Networks Giuseppe Manco, Giuseppe Pirrò
12:00 - 12:30	Co-Clustering for Differentially Private Synthetic Data Generation Tarek Benkhelif, Françoise Fessant, Fabrice Clérot, Guillaume Raschia
12:30 - 14:00	Lunch Break
14:00 - 14:40	Personal Knowledge Management Systems Serge Abiteboul
14:40 - 15:10	From Self-Data to Self-Preferences: Towards Preference Elicitation in Personal Information Management Systems Tristan Allard, Tassadit Bouadi, Joris Duguépéroux, Virginie Sans
15:10 - 15:40	Evaluating the Impact of Friends in Predicting User'S Availability in Online Social Networks Andrea De Salve, Paolo Mori, Laura Ricci
15:40 - 16:00	Coffee & Tea Break
16:00 - 16:30	Movement Behaviour Recognition for Water Activities Mirco Nanni, Roberto Trasarti
16:30 - 17:00	Guess the Movie - Linking Facebook Pages to IMDb Movies Paolo Fornacciari, Barbara Guidi, Monica Mordonini, Jacopo Orlandini, Laura Sani, Michele Tomaiuolo
17:00 - 17:30	Research on Online Digital Cultures - Community Extraction and Analysis By Markov and K-Means Clustering Tobias Blanke, Giles Greenway, Mark Coté, Jennifer Pybus
17:30 - 17·40	Conclusive Remarks

37

Joint Workshop

09:00 - 17:00

DyNO 2017 + TD-LSG 2107 Large-Scale Evolving Networks & Graphs

Organizers : Giulio Rossetti, ISTI-CNR Pisa, Italy Rémy Cazabet, LIP6, UPMC/CNRS, Sorbonne Universités, France Letizia Milli, University of Pisa, Italy Sabeur Aridhi, University of Lorraine, France José Fernandes de Macedo, Universidade Federale do Ceara, Brazil Engelbert Mephu Nguifo, LIMOS, Blaise Pascal University, France Karine Zeitouni, DAVID, Université de Versailles Saint-Quentin, France

Webpage

http://kdd.isti.cnr.it/dyno/ & http://tdlsg-ecmlpkdd17.isima.fr/

The aim of this joint workshop on Large-scale Evolving Networks and Graphs is to bring together active scholars and practitioners of dynamic graphs, complex and temporal networks. In this workshop, we aim to discuss the problem of mining large-scale time-dependent graphs, since there are many real world applications deal with a large volumes of spatio-temporal data (e.g. moving objects' trajectories). Managing and analyzing large-scale time-dependent graphs is very challenging since this requires sophisticated methods and techniques for creating, storing, accessing and processing such graphs in a distributed environment, because centralized approaches do not scale in a Big Data scenario. Contributions will clearly point out answers to one of these challenges focusing on large-scale graphs.

In the last years, we also witnessed a shift from static network analysis to dynamic ones, i.e., the study of networks whose structure changes over time. As time goes by, all the perturbations which occur in the network topology due to the rise and fall of nodes and edges have repercussions on the network phenomena we are used to observing. Nowadays, one of the most fascinating challenges is to analyze the structural dynamics of real world networks and how they impact on the processes which occur on them, i.e. the spreading of social influence and diffusion of innovations. Results in this field will enable a better understanding of important aspects of human behaviors as well as to a more detailed characterization of the complex interconnected society we inhabit. The purpose of this workshop is also to encourage research that will lead to the advancement of the social science in time-evolving networks.

Invited talk DEALING WITH BETWEENNESS IN EVOLVING GRAPHS AND IMPOSED SYSTEM WORKLOAD IMBALANCE

Nicolas Kourtellis Telefonica I+D, Spain

Invited talk LEARNING FROM HIDDEN TIME-DEPENDENT GRAPHS

Jan Ramon **INRIA**. France

Large-Scale Evolving Networks & Graphs Program at a glance

09:00	Welcome
	MORNING SESSION 1 - SOCIAL NETWORKS
09:10 - 09:45	Dynamic Community Detection : State of the Art and First Empirical Comparisons Cazabet Rémy
09:45 - 10:10	Churn Prediction Using Dynamic RFM-Augmented Node2vec Sandra Mitrovic, Bart Baesens, Wilfried Lemahieu, Jochen De Weerdt
10:10 - 10:25	Influence Maximization-Based Event Organization on Social Networks Cheng-Te Li
10:30 - 11:00	Coffee & Tea Break
	SESSION 2 - EVOLVING (SOCIAL) GRAPHS
11:00 - 11:45	Dealing with Betweenness in Evolving Graphs and Imposed System Workload Imbalance Nicolas Kourtellis
11:45 - 12:10	Multi-Scale Community Detection in Temporal Networks Using Spectral Graph Wavelets Zhana Kuncheva, Giovanni Montana
12:10 - 12:35	Finding Simple Temporal Cycles in an Interaction Network Rohit Kumar, Toon Calders
12:40 - 14:00	Lunch Break
	AFTERNOON SESSION 1 -TRANSPORTATION NETWORKS
14:00 - 14:45	Learning from Hidden Time-Dependent Graphs Jan Ramon
14:45 - 15:00	Discovering Connections in Heterogeneous Transportation Datasets with Link++ Ali Masri, Karine Zeitouni, Zoubida Kedad
15:00 - 15:15	Finding the Nearest Service Provider on Time-Dependent Road Networks Lívia Almada Cruz, Francesco Lettich, Leopoldo Soares Júnior, Regis Pires Magalhães, José Antônio, Fernandes De Macedo
15:15 - 15:30	Design and Implementation Issues of a Time-Dependent Shortest Path Algorithm for Multimodal Transportation Network
	Abdelfattah Idri, Mariyem Oukarfi, Azedine Boulmakoul and Karine Zeitouni
15:40 - 16:00	Coffee & Tea Break
	AFTERNOON SESSION 2 - PERFORMANCE ISSUES OF LARGE-SCALE EVOLVING GRAPHS
16:00 - 16:25	Synthetic Graph Generation from Finely-Tuned Temporal Constraints Karim Alami, Radu Ciucanu and Engelbert Mephu Nguifo
16:25 - 16:40	A Distributed Framework for Large-Scale Time-Dependent Graph Analysis Wissem Inoubli, Livia Almada, Ticiana Linhares, Gustavo Coutinho, Lucas Peres, Regis Pires Magalhães, Jose Antonio F. De Macedo, Sabeur Aridhi, Engelbert Mephu Nguifo
16:40 - 17:00	Panel Discussion

TUTORIAL

BANQUET HALL

14:00 - 17:40

Core Decomposition of Networks: Concepts, Algorithms and Applications

Organizers : Fragkiskos D. Malliaros, UC San Diego, USA Apostolos N. Papadopoulos, Aristotle University of Thessaloniki, Greece Michalis Vazirgiannis, École Polytechnique, France

Webpage

http://fragkiskos.me/projects/core-tutorial/

Graph mining is an important research area with a plethora of practical applications. Core decomposition of networks is a fundamental operation strongly related to more complex mining tasks such as community detection, dense subgraph discovery, identification of influential nodes, network visualization, text mining, just to name a few. In this tutorial, we will present in detail the concept and properties of core decomposition in graphs, the associated algorithms for its efficient computation and important cross-disciplinary applications that benefit from it.

- 14:00 15:40 PART 1 : Definitions, Fundamental Concepts and Basic Algorithms
- 15:40 16:00 Coffee & Tea Break
- 16:00 17:40 PART 2 : Algorithms (continued), Applications, Open Problems

TOWARDS END-TO-END LEARNING & OPTIMIZATION Frank Hutter

University of Freiburg, Germany

Deep learning has recently helped AI systems to achieve human-level performance in several domains, including speech recognition, object classification, and playing several types of games. The major benefit of deep learning is that it enables end-to-end learning of representations of the data on several levels of abstraction. However, the overall network architecture and the learning algorithms' sensitive hyperparameters still need to be set manually by human experts. In this talk, I will discuss extensions of Bayesian optimization for handling this problem effectively, thereby paving the way to fully automated end-to-end learning. I will focus on speeding up Bayesian optimization by reasoning over data subsets and initial learning curves, sometimes resulting in 100-fold speedups in finding good hyperparameter settings. I will also show competition-winning practical systems for automated machine learning (AutoML) and briefly show related applications to the end-to-end optimization of algorithms for solving hard combinatorial problems.

Chair: Sašo Džeroski

Sponsored by



TUESDAY

PROGRAM AT A GLANCE



MACEDONIAN OPERA & BALLET



INVITED TALK

CONGRESS HALL 1

09:00 - 10:00

Frontiers in Recurrent Neural Network Research **Alex Graves**

Google DeepMind, UK

In the last few years, recurrent neural networks (RNNs) have become the Swiss army knife of sequence processing for machine learning. Problems involving long and complex data streams, such as speech recognition, machine translation and reinforcement learning from raw video are now routinely tackled with RNNs. However, significant limitations still exist for such systems, such as their ability to retain large amounts of information in memory, and the challenges of gradient-based training on very long sequences. My talk will review some of the new architectures and training strategies currently being developed to extend the frontiers of this exciting field.

Chair: Jaakko Hollmén

INVITED TALK

MACEDONIAN OPERA & BALLET

19:00 - 20:00

Using Networks to Link Genotype to Phenotype John Quackenbush

Dana-Farber Cancer Institute, USA

Harvard TH Chan School of Public Health, USA

We know that genotype influences phenotype, but aside from a few highly penetrant Mendelian disorders, the link between genotype and phenotype is not well understood. We have used gene expression and genetic data to explore gene regulatory networks, to study phenotypic state transitions, and to analyze the connections between genotype and phenotype. I will describe how networks and their structure provide unique insight into how small effect variants influence phenotype.

Chair: Celine Veins

45

Arbitrated Ensemble for Time Series Forecasting Vitor Cerqueira, Luis Torgo, Fábio Pinto & Carlos Soares

This paper proposes an ensemble method for time series forecasting tasks. Combining different forecasting models is a common approach to tackle these tasks. State-of-the-art methods track the loss of the available models and adapt their weights accordingly. Metalearning strategies such as stacking are also used in these tasks. We propose a metalearning approach for adaptively combining forecasting models that specializes them across the time series. Our assumption is that different forecasting models have different areas of expertise and a varying relative performance. Moreover, many time series show recurring structures due to factors such as seasonality. Therefore, the ability of a method to deal with changes in relative performance of models as well as recurrent changes in the data distribution can be very useful in dynamic environments. Our approach is based on an ensemble of heterogeneous forecasters, arbitrated by a metalearning model. This strategy is designed to cope with the different dynamics of time series and quickly adapt the ensemble to regime changes. We validate our proposal using time series from several real world domains. Empirical results show the competitiveness of the method in comparison to state-of-the-art approaches for combining forecasters.

Chair: Jaakko Hollmén

BEST STUDENT ML PAPER AWARD

Sponsored by





TUESDAY

SESSIONS AT A GLANCE

Neural Networks and Deep Learning 1

CONGRESS HALL 1 11:00-12:40

Chair: Jesse Read	11:00-12:40

11.00 11.00	FCNNs: Fourier Convolutional Neural Networks
11:00 - 11:20	Harry Pratt, Bryan Williams, Frans Coenen, Yalin Zheng
11.00 11.40	Multimodal Classification for Analysing Social Media - NO SHOW
11.20 - 11.40	Chi Thang Duong, Remi Lebret, Karl Aberer
11.40 10.00	Sequence Generation with Target Attention
11.40 - 12.00	Yingce Xia, Fei Tian, Tao Qin, Nenghai Yu, Tie-Yan Liu
10.00 10.00	Weightless Neural Networks for Open Set Recognition
12.00 - 12.20	Douglas O. Cardoso, João Gama, Felipe M. G. França
10.00 10.40	A Network Architecture for Multi-Multi Instance Learning
12.20 - 12.40	Alessandro Tibo, Paolo Frasconi, Manfred Jaeger

Time Series and Streams 1

CONGRESS HALL 2 11:00-12:40

Chair: Joao Gama

11:00 - 11:20	Cost-Sensitive Perceptron Decision Trees for Imbalanced Drifting Data Streams Bartosz Krawczyk, Przemyslaw Skryjomski
	Learning TSK Fuzzy Rules from Data Streams
11:20 - 11:40	Ammar Shaker, Waleri Heldt, Eyke Huellermeier
11.40 - 19.00	Non-Parametric Online AUC Maximization
11:40 - 12:00	Balazs Szorenyi, Snir Cohen, Shie Mannor
12:00 - 12:20	On-Line Dynamic Time Warping for Streaming Time Series
	Izaskun Oregi, Aritz Perez, Javier Del Ser, Jose Lozano
12:20 - 12:40	PowerCast: Mining and Forecasting Power Grid Sequences
	Hyun Ah Song, Bryan Hooi, Marko Jereminov, Amritanshu Pandey, Larry Pileggi, Christos Faloutsos

Ensembles and Meta Learning

Chair: Albrecht Zimmermann II:UU-I2:40

CONGRESS HALL 4

CONGRESS HALL 1

11:00 - 11:20	Adaptive Random Forests for Evolving Data Stream Classification
	Heitor M. Gomes, Albert Bifet, Jesse Read, Jean Paul Barddal, Fabrício Enembreck, Bernhard Pfharinger, Geoff Holmes, Talel Abdessalem
11.00 11.40	Classification of High-Dimensional Evolving Data Streams Via a Resource-Efficient Online Ensemble
11.20 - 11.40	Tingting Zhai, Yang Gao, Hao Wang, Longbing Cao
11.40 - 19.00	Dynamic Ensemble Selection with Probabilistic Classifier Chains
11:40 - 12:00	Anil Narassiguin, Haytham Elghazel, Alexandre Aussem
12:00 - 12:20	Ensemble-Compression: A New Method for Parallel Training of Deep Neural Networks
	Shizhao Sun, Wei Chen, Jiang Bian, Xiaoguang Liu, Tie-Yan Liu
12:20 - 12:40	Fast and Accurate Density Estimation with Extremely Randomized Cutset Networks
	Nicola Di Mauro, Antonio Vergari, Teresa M.A. Basile, Floriana Esposito

Applied Data Science 1

Chair: Yasemin Altun 11:00-12:40

11:00 - 11:20	Probabilistic Inference of Twitter Users' Age Based on What They Follow
	Benjamin Chamberlain, Marc Deisenroth, Clive Humby
11:20 - 11:40	Unsupervised Signature Extraction from Forensic Logs
	Stefan Thaler, Vlado Menkovski, Milan Petkovic
11:40 - 12:00	Event Detection and Summarization Using Phrase Networks: PhraseNet
	Sara Melvin, Wenchao Yu, Peng Ju, Sean Young, Wei Wang
12:00 - 12:20	Stance Classification of Tweets Using Skip Char NGrams
	Yaakov HaCohen-Kerner, Ziv Ido, Ronen Ya'Akobov
12:20 - 12:40	Optimal Client Recommendation for Market Makers in Illiquid Financial Products
	Dieter Hendricks, Stephen Roberts

Probabilistic Models and Methods 1

Chair: Giuseppe M	14:00-15:40	I
14:00 - 14:20	Bayesian Heatmaps: Probabilistic Classification with Multiple Unreliable Information Sources	
	Edwin Simpson, Steven Reece, Stephen Roberts	
	Bayesian Inference for Least Squares Temporal Difference Regularization	
14:20 - 14:40	Nikolaos Tziortziotis, Christos Dimitrakakis	
14:40 - 15:00	Discovery of Causal Models That Contain Latent Variables Through Bayesian Scoring of Independence Constraints	
	Fattaneh Jabbari, Gregory Cooper, Joseph Ramsey, Peter Spirtes	
15:00 15:00	Efficient Parameter Learning of Bayesian Network Classifiers	
15:00 - 15:20	Nayyar A. Zaidi, Geoffrey I. Webb, Mark J. Carman, François Petitjean, Wray Buntine, Mike Hynes, Hans De Sterck	
15:20 - 15:40	MixedTrails: Bayesian Hypothesis Comparison on Heterogeneous Sequential Data	
	Martin Becker, Florian Lemmerich, Philipp Singer, Markus Strohmaier, Andreas Hotho	

47

Networks and Graphs 1

CONGRESS HALL 2 14:00-15:40

CONGRESS HALL 3

CONGRESS HALL 4

14:00-15:40

14:00 - 14:20	Attributed Graph Clustering with Unimodal Normalized Cut
14.00 - 14.20	Wei Ye, Linfei Zhou, Xin Sun, Claudia Plant, Christian Böhm
14.20 - 14.40	Ensemble-Based Community Detection in Multilayer Networks
14.20 - 14.40	Andrea Tagarelli, Alessia Amelio, Francesco Gullo
14:40 15:00	K-Clique-Graphs for Dense Subgraph Discovery
14.40 - 15.00	Giannis Nikolentzos, Polykarpos Meladianos, Yannis Stavrakas, Michalis Vazirgiannis
15:00 - 15:00	Local Community Detection in Multilayer Networks
15.00 - 15.20	Roberto Interdonato, Andrea Tagarelli, Dino Ienco, Arnaud Sallaberry, Pascal Poncelet
15.00 15.40	Local Lanczos Spectral Approximation for Membership Identification
15.20 - 15.40	Pan Shi, Kun He, David Bindel, John Hopcroft

Pattern and Sequence Mining

Chair: Stefan Kramer

14:00 - 14:20	BeatLex: Summarizing and Forecasting Time Series with Patterns
	Bryan Hooi, Shenghua Liu, Asim Smailagic, Christos Faloutsos
14:00 14:40	Behavioral Constraint Template-Based Sequence Classification
14.20 - 14.40	Johannes De Smedt, Galina Deeva, Jochen De Weerdt
14.40 - 15.00	Efficient Sequence Regression By Learning Linear Models in All-Subsequence Space
14.40 - 15.00	Severin Gsponer, Barry Smyth, Georgiana Ifrim
15:00 - 15:20	Flexible Constrained Sampling with Guarantees for Pattern Mining
	Vladimir Dzyuba, Matthijs Van Leeuwen, Luc De Raedt
15:20 - 15:40	Subjectively Interesting Connecting Trees
	Florian Adriaens, Jefrey Lijffijt, Tijl De Bie

Applied Data Science 2

14:00-15:40 Chair: Yasemin Altun Automatic Detection and Recognition of Individuals in Patterned Species 14:00 - 14:20 Gullal Singh Cheema, Saket Anand Sequential Keystroke Behavioral Biometrics for User Identification Via Multi-View Deep Learning 14:20 - 14:40 Lichao Sun, Yuqi Wang, Bokai Cao, Philip Yu, Witawas Srisa-An, Alex Leow Modeling the Temporal Nature of Human Behavior for Demographics Prediction 14:40 - 15:00 Bjarke Felbo, Pål Sundsøy, Alex 'Sandy' Pentland, Sune Lehmann, Yves-Alexandre De Montjoye RSSI Based Supervised Learning for Uncooperative Direction-Finding 15:00 - 15:20 Tathagata Mukherjee, Michael Duckett, Piyush Kumar, Jared Paquet, Daniel Rodriguez, Mallory Haulcomb, Kevin George, Eduardo Pasiliao Urban Water Flow and Water Level Prediction Based on Deep Learning 15:20 - 15:40 Haytham Assem, Salem Ghariba, Gabor Makrai, Paul Johnston, Laurence Gill, Francesco Pilla

Nectar 1

CONGRESS HALL 1

CONGRESS HALL 2

16:00-17:40

Chairs: Donato Malerba & Jerzy Stefanowski

14:00 - 14:20	Comparing Hypotheses on Sequential Behavior: A Bayesian Approach and Its Applications
	Florian Lemmerich, Philipp Singer, Martin Becker, Lisette Espin-Noboa, Dimitar Dimitrov, Denis Helic, Andreas Hotho, Markus Strohmaier
14:00 14:40	Music Generation Using Bayesian Networks
14.20 - 14.40	Tetsuro Kitahara
14:40 15:00	Phenotype Inference from Text and Genomic Data
14:40 - 15:00	Maria Brbic, Matija Piskorec, Vedrana Vidulin, Anita Krisko, Tomislav Smuc, Fran Supek
15:00 - 15:20	Data-Driven Approaches for Smart Parking
	Fabian Bock, Sergio Di Martino, Monika Sester
15:20 - 15:40	Process-Based Modeling and Design of Dynamical Systems
	Jovan Tanevski, Nikola Simidijevski, Ljupco Todorovski, Sašo Džeroski

Probabilistic Models and Methods 2

Chair: Jaakko Hollmén

16:00 - 16:20	Gaussian Conditional Random Fields Extended for Directed Graphs
	Tijana Vujicic, Jesse Glass, Fang Zhou, Zoran Obradovic
16:20 - 16:40	Multi-View Generative Adversarial Networks
	Mickael Chen, Ludovic Denoyer
16:40 - 17:00	PAC-Bayesian Analysis for a Two-Step Hierarchical Mutliview Learning Approach
	Anil Goyal, Emilie Morvant, Pascal Germain, Massih-Reza Amini
17:00 - 17:20	Robust Multi-View Topic Modeling By Incorporating Detecting Anomalies
	Guoxi Zhang, Tomoharu Iwata, Hisashi Kashima
17:20 - 17:40	Semi-Supervised Bayesian Deep Multi-Modal Emotion Recognition - NO SHOW
	Changde Du, Changying Du, Jinpeng Li, Wei-long Zheng, Baoliang Lv, Huiguang He

Reinforcement Learning

16:00-17:40 Chair: Kurt Driessens Generalized Exploration in Policy Search 16:00 - 16:20 Herke Van Hoof, Daniel Tanneberg, Jan Peters Generalized Inverse Reinforcement Learning on Linearly Solvable MDP 16:20 - 16:40 Masahiro Kohjima, Tatsushi Matsubayashi, Hiroshi Sawada Max K-Armed Bandit: On the ExtremeHunter Algorithm and Beyond 16:40 - 17:00 Mastane Achab, Stephan Clémençon, Aurélien Garivier, Anne Sabourin, Claire Vernade Offline Reinforcement Learning With Task Hierarchies 17:00 - 17:20 Devin Schwab, Soumya Ray Variational Thompson Sampling for Relational Recurrent Bandits 17:20 - 17:40 Sylvain Lamprier, Thibault Gisselbrecht, Patrick Gallinari

Regression

CONGRESS HALL 3 16:00-17:40

Chair: Nikola Simidjievski

16:00 - 16:20	Adaptive Skip-Train Structured Regression for Temporal Networks Martin Pavlovski, Fang Zhou, Ivan Stojkovic, Ljupco Kocarev, Zoran Obradovic
16:20 - 16:40	ALADIN: A New Approach for DrugTarget Interaction Prediction
	Krisztian Buza, Ladislav Peska
16:40 - 17:00	Co-Regularised Support Vector Regression
	Katrin Ullrich, Michael Kamp, Thomas Gärtner, Martin Vogt, Stefan Wrobel
17:00 - 17:20	Online Regression with Controlled Label Noise Rate
	Edward Moroshko, Koby Crammer
17:20 - 17:40	Robust Regression Using Biased Objectives
	Matthew J. Holland, Kazushi Ikeda

Privacy and Security

CONGRESS HALL 4

16:00-17:40

LIMIT: TODIAS SCHETTER	Chair:	Tobias	Scheffer
------------------------	--------	--------	----------

16:00 - 16:20	Differentially Private Nearest Neighbor Classification
	Mehmet Emre Gursoy, Ali Inan, Mehmet Ercan Nergiz, Yucel Saygin
16:20 - 16:40	Malware Detection By Analysing Encrypted Network Traffic with Neural Networks
	Paul Prasse, Lukas Machlika, Tomas Pevny, Jiri Havelka, Tobias Scheffer
16:40 - 17:00	PEM: Practical Differentially Private System for Large-Scale Cross-Institutional Data Mining
	Li Yi, Yitao Duan, Wei Xu
17:00 - 17:20	Preserving Differential Privacy in Convolutional Deep Belief Networks
	NhatHai Phan, Xintao Wu, Dejing Dou
17:20 - 17:40	The Best Privacy Defense Is a Good Privacy Offense: Obfuscating a Search Engine User'S Profile
	Jörg Wicker, Stefan Kramer

Demo 1	MACEDONIAN Opera & Ballet
Chairs: Jesse Read & Marinka Žitnik	20:00

WHODID: Web-Based Interface for Human-Assisted Factory Operations in Fault Detection
Gouy-Pailler Cedric, Pierre Blanchart
TrAnET: Tracking and Analyzing the Evolution of Topics in Information Networks
Livio Bioglio, Ruggero Pensa, Valentina Rho
Framework for Exploring and Understanding Multivariate Correlations
Louis Kirsch, Niklas Riekenbrauck, Daniel Thevessen, Marcus Pappik, Axel Stebner, Julius Kunze, Alexander
Meissner, Arvind Kumar Shekar, Emmanuel Müller
Tetrahedron: Barycentric Measure Visualizer
Dariusz Brzezinski, Jerzy Stefanowski, Izabela Szczęch, Robert Susmaga
Delve: A Data Set Retrieval and Document Analysis System
Uchenna Akujuobi, Xiangliang Zhang

Chair: Dino lenco

TISELAC Challenge : Time Series Land Cover Classification

Nowadays, modern earth observation programs produce huge volumes of satellite images time series (SITS) that can be useful to monitor geographical areas through time. How to efficiently analyze such kind of information is still an open question in the remote sensing field. In the context of land cover classification, exploiting time series of satellite images, instead that one single image, can be fruitful to distinguish among classes based on the fact they have different temporal profiles.

The objective of this challenge is to bring closer the Machine Learning and Remote Sensing communities to work on such kind of data. The Machine Learning community has the opportunity to validate and test their approaches on real world data in an application context that is getting more and more attention due to the increasing availability of SITS data while, this challenge offers to the Remote Sensing experts a way to discover and evaluate new data mining and machine learning methods to deal with SITS data.

The challenge involves a multi-class single label classification problem where the examples to classify are pixels described by the time series of satellite images and the prediction is related to the land cover of associated to each pixel.

Organizers : Dino Ienco, UMR TETIS - IRSTEA, Montpellier, France Raffaele Gaetano, UMR TETIS - CIRAD, Montpellier, France

https://sites.google.com/site/dinoienco/tiselc

16:00 - 16:15	Introduction
	Dino lenco, Raffaele Gaetano
16:15 - 16:35	End-to-End Learning of Deep Spatio-Temporal Representations for Hyperspectral Image Time Series Classification - 1st Place
	Nicola Di Mauro, Antonio Vergari, Teresa M.A. Basile, Fabrizio G. Ventola and Floriana Esposito
16:35 - 16:55	Time Series Land Cover Classification Challenge: Solution Report - 2nd Place
	Sergey Ryabukhin
16:55 - 17:15	TiSeLaC: A Distance Metric Learning Approach
	Bac Nguyen Cong, Christina Papagiannopoulou
17:15 - 17:35	The IRISA/LETG Ensemble Classifier for the TiSeLaC Challenge
	Romain Tavenard. Simon Malinowski. Nicolas Courty

ECML PKDD 2017 SKOPJE - MACEDONIA

SESSIONS WITH ABSTRACTS

Neural Networks & Deep Learning 1

CONGRESS HALL 1

11:00-11:20 FCNNs: Fourier Convolutional Neural Networks



Harry Pratt, Bryan Williams, Frans Coenen, Yalin Zheng

The Fourier domain is used in computer vision and machine learning as image analysis tasks in the Fourier domain are analogous to spatial domain methods but are achieved using different operations. Convolutional Neural Networks (CNNs) use machine learning to achieve state-of-the-art results with respect to many computer vision tasks. One of the main limiting aspects of CNNs is the computational cost of updating a large number of convolution parameters. Further, in the spatial domain, larger images take exponentially longer than smaller image to train on CNNs due to the operations involved in convolution methods. Consequently, CNNs are often not a viable solution for large image computer vision tasks. In this paper a Fourier Convolution Neural Network (FCNN) is proposed whereby training is conducted entirely within the Fourier domain. The advantage offered is that there is a significant speed up in training time without loss of effectiveness. Using the proposed approach larger images can therefore be processed within viable computation time. The FCNN is fully described and evaluated. The evaluation was conducted using the benchmark Cifar10 and MNIST datasets, and a bespoke fundus retina image dataset. The results demonstrate that convolution in the Fourier domain gives a significant speed up without adversely affecting accuracy. For simplicity the proposed FCNN concept is presented in the context of a basic CNN architecture, however, the FCNN concept has the potential to improve the speed of any neural network system involving convolution.

11:20-11:40 Multimodal Classification for Analysing Social Media

Chi Thang Duong, Remi Lebret, Karl Aberer

This paper was accepted for presentation. However, it was not presented at the conference and is thus not published in the conference proceedings.

11:40-12:00 Sequence Generation with Target Attention



NO

SHOW

Yingce Xia, Fei Tian, Tao Qin, Nenghai Yu, Tie-Yan Liu

Source-target attention mechanism (briefly, source attention) has become one of the key components in a wide range of sequence generation tasks, such as neural machine translation, image caption, and open-domain dialogue generation. In these tasks, the attention mechanism, typically in control of information flow from the encoder to the decoder, enables to generate every component in the target sequence relying on different source components. While source attention mechanism has attracted many research interests, few of them turn eyes to if the generation of target sequence can additionally benefit from attending back to itself, which however is intuitively motivated by the nature of attention. To investigate the question, in this paper, we propose a new target-target attention mechanism (briefly, target attention). Along the progress of generating target sequence, target attention mechanism takes into account the relationship between the component to generate and its preceding context within the target sequence, such that it can better keep the coherent consistency and improve the readability of the generated sequence. Furthermore, it complements the information from source attention so as to further enhance semantic adequacy. After designing an effective approach to incorporate target attention in encoder-decoder framework, we conduct extensive experiments on both neural machine translation and image caption. Experimental results clearly demonstrate the effectiveness of our design of integrating both source and target attention for sequence generation tasks.

12:00 - 12:20 Weightless Neural Networks for Open Set Recognition

Douglas O. Cardoso, João Gama, Felipe M. G. França

Open set recognition is a classification-like task. It is accomplished not only by the identification of observations which belong to targeted classes (i.e., the classes among those represented in the training sample which should be later recognized) but also by the rejection of inputs from other classes in the problem domain. The need for proper handling of elements of classes beyond those of interest is frequently ignored, even in works found in the literature. This leads to the improper development of learning systems, which may obtain misleading results when evaluated in their test beds, consequently failing to keep the performance level while facing some real challenge. The adaptation of a classifier for open set recognition is not always possible: the probabilistic premises most of them are built upon are not valid in a open-set setting. Still, this paper details how this was realized for WiSARD a weightless artificial neural network model. Such achievement was based on an elaborate distancelike computation this model provides and the definition of rejection thresholds during training. The proposed methodology was tested through a collection of experiments, with distinct backgrounds and goals. The results obtained confirm the usefulness of this tool for open set recognition.

12:20 - 12:40 A Network Architecture for Multi-multi Instance Learning

Alessandro Tibo, Paolo Frasconi, Manfred Jaeger

We study an extension of the multi-instance learning problem where examples are organized as nested bags of instances (e.g., a document could be represented as a bag of sentences, which in turn are bags of words). This framework can be useful in various scenarios, such as graph classification, image classification and translation-invariant pooling in convolutional neural network. In order to learn multi-multi instance data, we introduce a special neural network layer, called bag-layer, whose units aggregate sets of inputs of arbitrary size. We prove that the associated class of functions contains all Boolean functions over sets of sets of instances. We present empirical results on semi-synthetic data showing that such class of functions can be actually learned from data. We also present experiments on citation graphs datasets where our model obtains competitive results.

Time Series & Streams 1

Cost-Sensitive Perceptron Decision Trees for Imbalanced Drifting 11:00 - 11:20 **Data Streams**

Bartosz Krawczyk, Przemyslaw Skryjomski

Mining streaming and drifting data is among the most popular contemporary applications of machine learning methods. Due to the potentially unbounded number of instances arriving rapidly, evolving concepts and limitations imposed on utilized computational resources, there is a need to develop efficient and adaptive algorithms that can handle such problems. These learning difficulties can be further augmented by appearance of skewed distributions during the stream progress. Class imbalance in non-stationary scenarios is highly challenging, as not only imbalance ratio may change over time, but also the class relationships. In this paper we propose an efficient and fast cost-sensitive decision tree learning scheme for handling online class imbalance. In each leaf of the tree we train a perceptron with output adaptation to compensate for skewed class distributions, while McDiarmid's bound is used for controlling the splitting attribute selection. The cost matrix automatically adapts itself to the current imbalance ratio in the stream, allowing for a smooth compensation of evolving class relationships. Furthermore, we analyze the characteristics of minority class instances and incorporate this information during the training process. It allows our classifier to focus on most difficult instances, while a sliding window keeps track of the changes in class structures. Experimental analysis carried out on a number of binary and multi-class imbalanced data streams indicate the usefulness of the proposed approach.

11:20 - 11:40 Learning TSK Fuzzy Rules from Data Streams

Ammar Shaker, Waleri Heldt, Eyke Huellermeier

Learning from data streams has received increasing attention in recent years, not only in the machine learning community but also in other research fields, such as computational intelligence and fuzzy systems. In particular, several rule-based methods for the incremental induction of regression models have been proposed. In this paper, we develop a method that combines the strengths of two existing approaches rooted in different learning paradigms. Our method induces a set of fuzzy rules, which, compared to conventional rules with Boolean antecedents, has the advantage of producing smooth regression functions. To do so, it makes use of an induction technique inspired by AMRules, a very efficient and effective learning algorithm that can be seen as the state of the art in machine learning. We conduct a comprehensive experimental study showing that a combination of the expressiveness of fuzzy rules with the algorithmic concepts of AMRules yields a learning system with superb performance.







CONGRESS HALL 2



TUESDAY **SEPTEMBER 19**

11:40 - 12:00 Non-Parametric Online AUC Maximization

Balazs Szorenyi, Snir Cohen, Shie Mannor

We consider the problems of online and one-pass maximization of the area under the ROC curve (AUC). AUC maximization is hard even in the offline setting and thus solutions often make some compromises. Existing results for the online problem typically optimize for some proxy defined via surrogate losses instead of maximizing the real AUC. This approach is confirmed by results showing that the optimum of these proxies, over the set of all (measurable) functions, maximize the AUC. The problem is that --- in order to meet the strong requirements for per round run time complexity---online methods typically work with restricted hypothesis classes and this, as we show, corrupts the above compatibility and causes the methods to converge to suboptimal solutions even in some simple stochastic cases. To remedy this, we propose a different approach and show that it leads to asymptotic optimality. Our theoretical claims and considerations are tested by experiments on real datasets, which provide empirical justification to them.

12:00 - 12:20 **On-Line Dynamic Time Warping for Streaming Time Series**

Izaskun Oregi, Aritz Perez, Javier Del Ser, Jose Lozano

Dynamic Time Warping is a well-known measure of dissimilarity between time series. Due to its flexibility to deal with non-linear distortions along the time axis, this measure has been widely utilized in machine learning models for this particular kind of data. Nowadays, the proliferation of streaming data sources has ignited the interest and attention of the scientific community around on-line learning models. In this work, we naturally adapt Dynamic Time Warping to the on-line learning setting. Specifically, we propose a novel on-line measure of dissimilarity for streaming time series which combines a warp constraint and a weighted memory mechanism to simplify the time series alignment and adapt to non-stationary data intervals along time. Computer simulations are analyzed and discussed so as to shed light on the performance and complexity of the proposed measure.

12:20 - 12:40 **PowerCast: Mining and Forecasting Power Grid Sequences**

Hyun Ah Song, Bryan Hooi, Marko Jereminov, Amritanshu Pandey, Larry Pileggi, Christos Faloutsos

What will be the power consumption of our institution at 8am for the upcoming days? What will happen to the power consumption of a small factory, if it wants to double (or half) its production? Technologies associated with the smart electrical grid are needed. Central to this process are algorithms that accurately model electrical load behavior, and forecast future electric power demand. However, existing power load models fail to accurately represent electrical load behavior in the grid. In this paper, we propose PowerCast, a novel domain-aware approach for forecasting the electrical power demand, by carefully incorporating domain knowledge. Our contributions are as follows: 1. Infusion of domain expert knowledge: We represent the time sequences using an equivalent circuit model, the "BIG" model, which allows for an intuitive interpretation of the power load, as the BIG model is derived from physics-based first principles. 2. Forecasting of the power load: Our PowerCast uses the BIG model, and provides (a) accurate prediction in multi-step-ahead forecasting, and (b) extrapolations, under what-if scenarios, such as variation in the demand (say, due to increase in the count of people on campus, or a decision to half the production in our factory etc.) 3. Anomaly detection: PowerCast can spot and, even explain, anomalies in the given time sequences. The experimental results based on two real world datasets of up to three weeks duration, demonstrate that PowerCast is able to forecast several steps ahead, with 59% error reduction, compared to the competitors. Moreover, it is fast, and scales linearly with the duration of the sequences.

Ensembles & Meta Learning

11:00 - 11:20 Adaptive Random Forests for Evolving Data Stream Classification

Heitor M. Gomes, Albert Bifet, Jesse Read, Jean Paul Barddal, Fabrício Enembreck, Bernhard Pfharinger, Geoff Holmes, Talel Abdessalem

Random forests is currently one of the most used machine learning algorithms in the non-streaming (batch) setting. This preference is attributable to its high learning performance and low demands with respect to input preparation and hyper-parameter tuning. However, in the challenging context of evolving data streams, there is no random forests algorithm that can be considered state-of-the-art in comparison to bagging and boosting based algorithms. In this work, we present the adaptive random forest (ARF) algorithm for classification of evolving data streams. In contrast to previous attempts of replicating random forests for data stream learning, ARF includes an effective resampling method and adaptive operators that can cope with different types of concept drifts without complex optimizations for different data sets. We present experiments with a parallel implementation of ARF which has no degradation in terms of classification performance in comparison to a serial implementation, since trees and adaptive operators are independent from one another. Finally, we compare ARF with state-of-the-art algorithms in a traditional testthen-train evaluation and a novel delayed labelling evaluation, and show that ARF is accurate and uses a feasible amount of resources.











11:20-11:40 Classification of High-dimensional Evolving Data Streams via a Resource-efficient Online Ensemble

Tingting Zhai, Yang Gao, Hao Wang, Longbing Cao

A novel online ensemble strategy, ensemble BPegasos(EBPegasos), is proposed to solve the problems simultaneously caused by concept drifting and the curse of dimensionality in classifying high-dimensional evolving data streams, which has not been addressed in the literature. First, EBPegasos uses BPegasos, an online kernelized SVM-based algorithm, as the component classifier to address the scalability and sparsity of high-dimensional data. Second, EBPegasos takes full advantage of the characteristics of BPegasos to cope with various types of concept drifts. Specifically, EBPegasos constructs diverse component classifiers by controlling the budget size of BPegasos; it also equips each component with a drift detector to monitor and evaluate its performance, and modifies the ensemble structure only when large performance degradation occurs. Such conditional structural modification strategy makes EBPegasos strike a good balance between exploiting and forgetting old knowledge. Lastly, we first prove experimentally that EBPegasos is more effective and resource-efficient than the tree ensembles on high-dimensional data. Then comprehensive experiments on synthetic and real-life datasets also show that EBPegasos can cope with various types of concept drifts significantly better than the state-of-the-art ensemble frameworks when all ensembles use BPegasos as the base learner.

11:40-12:00 **Dynamic Ensemble Selection with Probabilistic Classifier Chains**

Anil Narassiguin, Haytham Elghazel, Alexandre Aussem

Dynamic ensemble selection (DES) is the problem of finding, given an input x, a subset of models among the ensemble that achieves the best possible prediction accuracy. Recent studies have reformulated the DES problem as a multi-label classification problem and promising performance gains have been reported. However, their approaches may converge to an incorrect, and hence suboptimal, solution as they don't optimize the true - but non standard - loss function directly. In this paper, we show that the label dependencies have to be captured explicitly and propose a DES method based on Probabilistic Classifier Chains. Experimental results on 20 benchmark data sets show the effectiveness of the proposed method against competitive alternatives, including the aforementioned multi-label approaches. Keywords: Dynamic ensemble selection, Multi-label learning, Probabilistic Classifier Chains

12:00-12:20 Ensemble-Compression: A New Method for Parallel Training of Deep Neural Networks

Shizhao Sun, Wei Chen, Jiang Bian, Xiaoguang Liu, Tie-Yan Liu

Parallelization framework has become a necessity to speed up the training of deep neural networks (DNN) recently. Such framework typically employs the Model Average approach, denoted as MA-DNN, in which parallel workers conduct respective training based on their own local data while the parameters of local models are periodically communicated and averaged to obtain a global model which serves as the new start of local models. However, since DNN is a highly non-convex model, averaging parameters cannot ensure that such global model can perform better than those local models. To tackle this problem, we introduce a new parallel training framework called Ensemble-Compression, denoted as EC-DNN. In this framework, we propose to aggregate the local models by ensemble, i.e., averaging the outputs of local models instead of the parameters. As most of prevalent loss functions are convex to the output of DNN, the performance of ensemble-based global model is guaranteed to be at least as good as the average performance of local models. However, a big challenge lies in the explosion of model size since each round of ensemble can give rise to multiple times size increment. Thus, we carry out model compression after each ensemble, specialized by a distillation based method in this paper, to reduce the size of the global model to be the same as the local ones. Our experimental results demonstrate the prominent advantage of EC-DNN over MA-DNN in terms of both accuracy and speedup.

12:20-12:40 Fast and Accurate Density Estimation with Extremely Randomized Cutset Networks

Nicola Di Mauro, Antonio Vergari, Teresa M.A. Basile, Floriana Esposito

Cutset Networks (CNets) are recently introduced density estimators leveraging context-specific independences to provide exact inference in polynomial time. Learning a CNet is done first by building a weighted probabilistic OR tree and then estimating tractable distributions as its leaves. In practice, selecting an optimal OR split node requires cubic time in the number of the data features, and even approximate heuristics still scale in quadratic time. We introduce Extremely Randomized Cutset Networks (XCNets), CNets whose OR tree is learned by performing random conditioning. This simple yet surprisingly effective approach reduces the complexity of OR node selection to constant time. While the likelihood of an XCNet is slightly worse than an optimally learned CNet, ensembles of XCNets outperform state-of-the-art density estimators on a series of standard benchmark datasets, yet employing only a fraction of the time needed to learn the competitors.







Applied Data Science 1

11:00 - 11:20 Probabilistic Inference of Twitter Users' Age Based on What They Follow

Benjamin Chamberlain, Marc Deisenroth, Clive Humby

Twitter provides an open and rich source of data for studying human behaviour at scale and is widely used in social and network sciences. However, a major criticism of Twitter data is that demographic information is largely absent. Enhancing Twitter data with user ages would advance our ability to study social network structures, information flows and the spread of contagions. Approaches toward age detection of Twitter users typically focus on specific properties of tweets, e.g., linguistic features, which are language dependent. In this paper, we devise a languageindependent methodology for determining the age of Twitter users from data that is native to the Twitter ecosystem. The key idea is to use a Bayesian framework to generalise ground-truth age information from a few Twitter users to the entire network based on what/whom they follow. Our approach scales to inferring the age of 700 million Twitter accounts with high accuracy.

Unsupervised Signature Extraction from Forensic Logs 11:20 - 11:40

Stefan Thaler, Vlado Menkovski, Milan Petkovic

Signature extraction is a key part of forensic log analysis. It involves recognizing patterns in log lines such that log lines that originated from the same line of code are grouped together. A log signature consists of immutable parts and mutable parts. The immutable parts define the signature, and the mutable parts are typically variable parameter values. In practice, the number of log lines and signatures can be quite large, and the task of detecting and aligning immutable parts of the logs to extract the signatures becomes a significant challenge. We propose a novel method based on a neural language model that outperforms the current state-of-the-art on signature extraction. We use an RNN auto-encoder to create an embedding of the log lines. Log lines embedded in such a way can be clustered to extract the signatures in an unsupervised manner.

11:40 - 12:00 **Event Detection and Summarization Using Phrase Networks: PhraseNet**

Sara Melvin, Wenchao Yu, Peng Ju, Sean Young, Wei Wang

Identifying events in real-time data streams such as Twitter is crucial for many occupations to make timely, actionable decisions. It is however extremely challenging because of the subtle difference between "events" and trending topics, the definitive rarity of these events, and the complexity of modern Internet's text data. Existing approaches often utilize topic modeling technique and keywords frequency to detect events on Twitter, which have three main limitations: 1) supervised and semi-supervised methods run the risk of missing important, breaking news events; 2) existing topic/event detection models are base on words, while the correlations among phrases are ignored; 3) many previous methods identify trending topics as events. To address these limitations, we propose the model, PhraseNet, an algorithm to detect and summarize events from tweets. To begin, all topics are defined as a clustering of high-frequency phrases extracted from text. All trending topics are then identified based on temporal spikes of the phrase cluster frequencies. PhraseNet thus filters out high-confidence events from other trending topics using number of peaks and variance of peak intensity. We evaluate PhraseNet on a three month duration of Twitter data and show the both the efficiency and the effectiveness of our approach.

12:00 - 12:20 Stance Classification of Tweets Using Skip Char NGrams

Yaakov HaCohen-Kerner, Ziv Ido, Ronen Ya'akobov

In this research, we focus on automatic supervised stance classification of tweets. Given test datasets of tweets from five various topics, we try to classify the stance of the tweet authors as either in FAVOR of the target, AGAINST it, or NONE. We apply eight variants of seven supervised machine learning methods and three filtering methods using the WEKA platform. The macro-average results obtained by our algorithm are significantly better than the state-of-art results re-ported by the best macro-average results achieved in the SemEval 2016 Task 6-A for all the five released datasets. In contrast to the competitors of the SemEval 2016 Task 6-A, who did not use any char skip ngrams but rather used thousands of ngrams and hundreds of word embedding features, our algorithm uses a few tens of features mainly character-based features where most of them are skip char ngram features.



Optimal Client recommendation for Market Makers in Illiquid Financial Products

Dieter Hendricks, Stephen Roberts

The process of liquidity provision in financial markets can result in prolonged exposure to illiquid instruments for market makers. In this case, where a proprietary position is not desired, pro-actively targeting the right client who is likely to be interested can be an effective means to offset this position, rather than relying on commensurate interest arising through natural demand. In this paper, we consider the inference of a client profile for the purpose of corporate bond recommendation, based on typical recorded information available to the market maker. Given a historical record of corporate bond transactions and bond meta-data, we use a topic-modelling analogy to develop a probabilistic technique for compiling a curated list of client recommendations for a particular bond that needs to be traded, ranked by probability of interest. We show that a model based on Latent Dirichlet Allocation offers promising performance to deliver relevant recommendations for sales traders.

Probabilistic Models & Methods 1

Bayesian Heatmaps: Probabilistic Classification with Multiple Unreliable 14:00 - 14:20 Information Sources

Edwin Simpson, Steven Reece, Stephen Roberts

Unstructured data from diverse sources, such as social media and aerial imagery, can provide valuable up-to-date information for intelligent situation assessment. Mining these different information sources could bring major benefits to applications such as situation awareness in disaster zones and mapping the spread of diseases. Such applications depend on classifying the situation across a region of interest, which can be depicted as a spatial "heatmap". Annotating unstructured data using crowdsourcing or automated classifiers produces individual classifications at sparse locations that typically contain many errors. In this paper, we propose a novel Bayesian approach that models the relevance, error rates and bias of each information source, enabling us to learn a spatial Gaussian Process classifier by aggregating data from multiple sources with varying reliability and relevance. Our method does not require gold-labelled data and can make predictions at any location in an area of interest given only sparse observations. We show empirically that our approach can handle noisy and biased data sources, and that simultaneously inferring reliability and transferring information between neighbouring reports leads to more accurate predictions. We demonstrate our method on two real-world problems from disaster response, showing how our approach reduces the amount of crowdsourced data required and can be used to generate valuable heatmap visualisations from SMS messages and satellite images.

12:20 - 12:40

Bayesian Inference for Least Squares Temporal Difference Regularization

Nikolaos Tziortziotis, Christos Dimitrakakis

This paper proposes a fully Bayesian approach for Least-Squares Temporal Differences (LSTD), resulting in fully probabilistic inference of value functions that avoids the overfitting commonly experienced with classical LSTD when the number of features is larger than the number of samples. Sparse Bayesian learning provides an elegant solution through the introduction of a prior over value function parameters. This gives us the advantages of probabilistic predictions, a sparse model, and good generalisation capabilities, as irrelevant parameters are marginalised out. The algorithm efficiently approximates the posterior distribution through variational inference. We demonstrate the ability of the algorithm in avoiding overfitting experimentally.

14:40 - 15:00

Discovery of Causal Models That Contain Latent Variables through Bayesian Scoring of Independence Constraints

Fattaneh Jabbari, Gregory Cooper, Joseph Ramsey, Peter Spirtes

Discovering causal structure from observational data in the presence of latent variables remains an active research area. Constraint-based causal discovery algorithms are relatively efficient at discovering such causal models from data using independence tests. Typically, however, they derive and output only one such model. In contrast, Bayesian methods can generate and probabilistically score multiple models, outputting the most probable one; however, they are often computationally infeasible to apply when modeling latent variables. We introduce a hybrid method that derives a Bayesian probability that the set of independence tests associated with a given causal model are jointly correct. Using this constraintbased scoring method, we are able to score multiple causal models, which possibly contain latent variables, and output the most probable one. The structure-discovery performance of the proposed method is compared to an existing constraint-based method (RFCI) using data generated from several previously published Bayesian networks. The structural Hamming distances of the output models improved when using the proposed method compared to RFCI, especially for small sample sizes.





CONGRESS HALL 1



TUESDAY **SEPTEMBER 19**

15:00 - 15:20 Efficient Parameter Learning of Bayesian Network Classifiers

Nayyar A. Zaidi, Geoffrey I. Webb, Mark J. Carman, François Petitjean, Wray Buntine, Mike Hynes, Hans De Sterck

Recent advances have demonstrated substantial benefits from learning with both generative and discriminative parameters. On the one hand, generative approaches address the estimation of the parameters of the joint distribution-P(y, x), which for most network types is very computationally efficient (a notable exception to this are Markov networks) and on the other hand, discriminative approaches address the estimation of the parameters of the posterior distribution-and, are more effective for classification, since they fit P(y|x) directly. However, discriminative approaches are less computationally efficient as the normalization factor in the conditional log-likelihood precludes the derivation of closed-form estimation of parameters. This paper introduces a new discriminative parameter learning method for Bayesian network classifiers that combines in an elegant fashion parameters learned using both generative and discriminative methods. The proposed method is discriminative in nature, but uses estimates of generative probabilities to speed-up the optimization process. A second contribution is to propose a simple framework to characterize the parameter learning task for Bayesian network classifiers. We conduct an extensive set of experiments on 72 standard datasets and demonstrate that our proposed discriminative parameterization provides an efficient alternative to other state-of-the-art parameterizations.

MixedTrails: Bayesian Hypothesis Comparison on Heterogeneous 15:20 - 15:40 Sequential Data

Martin Becker, Florian Lemmerich, Philipp Singer, Markus Strohmaier, Andreas Hotho

Sequential traces of user data are frequently observed online and offline, e.g., as sequences of visited websites or as sequences of locations captured by GPS. However, understanding factors explaining the production of sequence data is a challenging task, especially since the data generation is often not homogeneous. For example, navigation behavior might change in different phases of browsing a website or movement behavior may vary between groups of users. In this work, we tackle this task and propose MixedTrails, a Bayesian approach for comparing the plausibility of hypotheses regarding the generative processes of heterogeneous sequence data. Each hypothesis is derived from existing literature, theory, or intuition and represents a belief about transition probabilities between a set of states that can vary between groups of observed transitions. For example, when trying to understand human movement in a city and given some data, a hypothesis assuming tourists to be more likely to move towards points of interests than locals can be shown to be more plausible than a hypothesis assuming the opposite. Our approach incorporates such hypotheses as Bayesian priors in a generative mixed transition Markov chain model, and compares their plausibility utilizing Bayes factors. We discuss analytical and approximate inference methods for calculating the marginal likelihoods for Bayes factors, give guidance on interpreting the results, and illustrate our approach with several experiments on synthetic and empirical data from Wikipedia and Flickr. Thus, this work enables a novel kind of analysis for studying sequential data in many application areas.

Networks & Graphs 1

14:00 - 14:20 Attributed Graph Clustering with Unimodal Normalized Cut

Wei Ye, Linfei Zhou, Xin Sun, Claudia Plant, Christian Böhm

Graph vertices are often associated with attributes. For example, in addition to their connection relations, people in friendship networks have personal attributes, such as interests, age, and residence. Such graphs (networks) are called attributed graphs. The detection of clusters in attributed graphs is of great practical relevance, e.g., targeting ads. Attributes and edges often provide complementary information. The effective use of both types of information promises meaningful results. In this work, we propose a method called UNCut (for Unimodal Normalized Cut) to detect cohesive clusters in attributed graphs. A cohesive cluster is a subgraph that has densely connected edges and has as many homogeneous (unimodal) attributes as possible. We adopt the normalized cut to assess the density of edges in a graph cluster. To evaluate the unimodality of attributes, we propose a measure called unimodality compactness which exploits Hartigans' dip test. Our method UNCut integrates the normalized cut and unimodality compactness in one framework such that the detected clusters have low normalized cut and unimodality compactness values. Extensive experiments on various synthetic and real-world data verify the effectiveness and efficiency of our method UNCut compared with state-of-the-art approaches.



CONGRESS HALL 2





14:20-14:40 Ensemble-Based Community Detection in Multilayer Networks

Andrea Tagarelli, Alessia Amelio, Francesco Gullo

The problem of community detection in a multilayer network can effectively be addressed by aggregating the community structures separately generated for each network layer, in order to infer a consensus solution for the input network. To this purpose, clustering ensemble methods developed in the data clustering field are naturally of great support. Bringing these methods into a community detection framework would in principle represent a powerful and versatile approach to reach more stable and reliable community structures. Surprisingly, research on consensus community detection is still in its infancy. In this paper, we propose a novel modularity-driven ensemble-based approach to multilayer community detection. A key aspect is that it finds consensus community structures that not only capture prototypical community analysis. Empirical evidence obtained on seven real-world multilayer networks sheds light on the effectiveness and efficiency of our proposed modularity-driven ensemble-based approach, which has shown to outperform state-of-the-art multilayer methods in terms of modularity, silhouette of community memberships, and redundancy assessment criteria, and also in terms of execution times.

14:40 - 15:00 k-clique-graphs for Dense Subgraph Discovery

Giannis Nikolentzos, Polykarpos Meladianos, Yannis Stavrakas, Michalis Vazirgiannis

Finding dense subgraphs in a graph is a fundamental graph mining task, with applications in several fields. Algorithms for identifying dense subgraphs are used in biology, in finance, in spam detection, etc. Standard formulations of this problem such as the problem of finding the maximum clique of a graph are hard to solve. However, some tractable formulations of the problem have also been proposed, focusing mainly on optimizing some density function, such as the degree density and the triangle density. However, maximization of degree density usually leads to large subgraphs with small density, while maximization of triangle density does not necessarily lead to subgraphs that are close to being cliques. In this paper, we introduce the k-clique-graph densest subgraph problem, $k \le 3$, a novel formulation for the discovery of dense subgraphs. Given an input graph, its k-clique-graph is a new graph created from the input graph where each vertex of the new graph corresponds to a k-clique of the input graph and two vertices are connected with an edge if they share a common (k - 1)-clique. We define a simple density function, the k-clique-graph density, which gives compact and at the same time dense subgraphs, and we project its resulting subgraphs back to the input graph. In this paper we focus on the triangle-graph densest subgraph problem obtained for k = 3. To optimize the proposed function, we provide an exact algorithm. Furthermore, we present an efficient greedy approximation algorithm that scales well to larger graphs. We evaluate the proposed algorithms on real datasets and compare them with other algorithms in terms of the size and the density of the extracted subgraphs. The results verify the ability of the proposed algorithms in finding high-quality subgraphs in terms of size and density. Finally, we apply the proposed method to the important problem of keyword extraction from textual documents.

15:00 - 15:20 Local Community Detection in Multilayer Networks

Roberto Interdonato, Andrea Tagarelli, Dino Ienco, Arnaud Sallaberry, Pascal Poncelet

The problem of local community detection in graphs refers to the identification of a community that is specific to a query node and relies on limited information about the network structure. Existing approaches for this problem are defined to work in dynamic network scenarios, however they are not designed to deal with complex real-world networks, in which multiple types of connectivity might be considered. In this work, we fill this gap in the literature by introducing the first framework for local community detection in multilayer networks (ML-LCD). We formalize the ML-LCD optimization problem and provide three definitions of the associated objective function, which correspond to different ways to incorporate within-layer and across-layer topological features. We also exploit our framework to generate multilayer global community structures. We conduct an extensive experimentation using seven real-world multilayer networks, which also includes comparison with state-of-the-art methods for single-layer local community detection and for multilayer global community detection. Results show the significance of our proposed methods in discovering local communities over multiple layers, and also highlight their ability in producing global community structures that are better in modularity than those produced by native global community detection approaches.

15:20-15:40 Local Lanczos Spectral Approximation for Membership Identification

Pan Shi, Kun He, David Bindel, John Hopcroft

We propose a novel approach called the Local Lanczos Spectral Approximation (LLSA) for identifying all latent members of a local community from very few seed members. To reduce the computation complexity, we first apply a fast heat kernel diffusing to sample a comparatively small subgraph covering almost all possible community members around the seeds. Then starting from a normalized indicator vector of the seeds and by a few steps of Lanczos iteration on the sampled subgraph, a local eigenvector is gained for approximating the eigenvector of the transition matrix with the largest eigenvalue. Elements of this local eigenvector is a relaxed indicator for the affiliation probability of the corresponding nodes to the target community. We conduct extensive experiments on real-world datasets in various domains as well as synthetic datasets. Results show that the proposed method outperforms state-of-the-art local community detection algorithms. To the best of our knowledge, this is the first work to adapt the Lanczos method for local community detection, which is natural and potentially effective. Also, we did the first attempt of using heat kernel as a sampling method instead of detecting communities directly, which is proved empirically to be very efficient and effective.







Pattern & Sequence Mining

60

14:00-14:20 BeatLex: Summarizing and Forecasting Time Series with Patterns

Bryan Hooi, Shenghua Liu, Asim Smailagic, Christos Faloutsos

Given time-series data such as electrocardiogram (ECG) readings, or motion capture data, how can we succinctly summarize the data in a way that robustly identifies patterns that appear repeatedly? How can we then use such a summary to identify anomalies such as abnormal heartbeats, and also forecast future values of the time series? Our main idea is a vocabulary-based approach, which automatically learns a set of common patterns, or 'beat patterns,' which are used as building blocks to describe the time series in an intuitive and interpretable way. Our summarization algorithm, BeatLex (Beat Lexicons for Summarization) is: 1) fast and online, requiring linear time in the data size and bounded memory; 2) effective, outperforming competing algorithms in labelling accuracy by 5.3 times, and forecasting accuracy by 1.8 times; 3) principled and parameter-free, as it is based on the Minimum Description Length principle of summarizing the data by compressing it using as few bits as possible, and automatically tunes all its parameters; 4) general: it applies to any domain of time series data, and can make use of multidimensional (i.e. coevolving) time series.

14:20-14:40 Behavioral Constraint Template-Based Sequence Classification

Johannes De Smedt, Galina Deeva, Jochen De Weerdt

In this paper we present the interesting Behavioral Constraint Miner (iBCM), a new approach towards classifying sequences. The prevalence of sequential data, i.e., a collection of ordered items such as text, website navigation patterns, traffic management, and so on, has incited a surge in research interest towards sequence classification. Existing approaches mainly focus on retrieving sequences of itemsets and checking their presence in labeled data streams to obtain a classifier. The proposed iBCM approach, rather than focusing on plain sequences, is template-based and draws its inspiration from behavioral patterns used for software verification. These patterns have a broad range of characteristics and go beyond the typical sequence mining representation, allowing for a more precise and concise way of capturing sequential information in a database. Furthermore, it is possible to also mine for negative information, i.e., sequences that do not occur. The technique is benchmarked against other state-of-the-art approaches and exhibits a strong potential towards sequence classification.

14:40-15:00 Efficient Sequence Regression By Learning Linear Models in All-Subsequence Space

Severin Gsponer, Barry Smyth, Georgiana Ifrim

We present a new approach for learning a sequence regression function, i.e., a mapping from sequential observations to a numeric score. Our learning algorithm employs coordinate gradient descent with Gauss-Southwell optimization in the feature space of all subsequences. We give a tight upper bound for the coordinate wise gradients of squared error loss which enables efficient Gauss-Southwell selection. The proposed bound is built by separating the positive and the negative gradients of the loss function and exploits the structure of the feature space. Extensive experiments on simulated as well as real-world sequence regression benchmarks show that the bound is effective and our proposed learning algorithm is efficient and accurate. The resulting linear regression model provides the user with a list of the most predictive features selected during the learning stage, adding to the interpretability of the method.

15:00 - 15:20 Flexible Constrained Sampling with Guarantees for Pattern Mining

Vladimir Dzyuba, Matthijs van Leeuwen, Luc De Raedt

Pattern sampling has been proposed as a potential solution to the infamous pattern explosion. Instead of enumerating all patterns that satisfy the constraints, individual patterns are sampled proportional to a given quality measure. Several sampling algorithms have been proposed, but each of them has its limitations when it comes to (1) flexibility in terms of quality measures and constraints that can be used, and/or (2) guarantees with respect to sampling accuracy. We therefore present Flexics, the first flexible pattern sampler that supports a broad class of quality measures and constraints, while providing strong guarantees regarding sampling accuracy. To achieve this, we leverage the perspective on pattern mining as a constraint satisfaction problem and build upon the latest advances in sampling solutions in SAT as well as existing pattern mining algorithms. Furthermore, the proposed algorithm is applicable to a variety of pattern languages, which allows us to introduce and tackle the novel task of sampling sets of patterns. We introduce and empirically evaluate two variants of Flexics: (1) a generic variant that addresses the well-known itemset sampling task and the novel pattern set sampling task as well as a wide range of expressive constraints within these tasks, and (2) a specialized variant that exploits existing frequent itemset techniques to achieve substantial speed-ups. Experiments show that Flexics is both accurate and efficient, making it a useful tool for pattern-based data exploration.









15:20-15:40 Subjectively Interesting Connecting Trees

Florian Adriaens, Jefrey Lijffijt, Tijl De Bie

Consider a large network, and a user-provided set of query nodes between which the user wishes to explore relations. For example, a researcher may want to connect research papers in a citation network, an analyst may wish to connect organized crime suspects in a communication network, or an internet user may want to organize their bookmarks given their location in the world wide web. A natural way to show how query nodes are related is in the form of a tree in the network that connects them. However, in sufficiently dense networks, most such trees will be large or somehow trivial (e.g. involving high degree nodes) and thus not insightful. In this paper, we define and investigate the new problem of mining subjectively interesting trees connecting a set of query nodes in a network, i.e., trees that are highly surprising to the specific user at hand. Using information theoretic principles, we formalize the notion of interestingness of such trees mathematically, taking in account any prior beliefs the user's prior belief state is however not necessarily computationally tractable. Yet, we show how a highly generic class of prior beliefs, namely about individual node degrees in combination with the density of particular sub-networks, can be dealt with in a tractable manner. Such types of beliefs can be used to model knowledge of a partial or total order of the network nodes, e.g. where the nodes represent events in time (such as papers in a citation network). An empirical validation of our methods on a large real network evaluates the different heuristics and validates the interestingness of the given trees.

Applied Data Science 2

14:00-14:20 Automatic Detection and Recognition of Individuals in Patterned Species

Gullal Singh Cheema, Saket Anand

Visual animal biometrics is rapidly gaining popularity as it enables a non-invasive and cost-effective approach for wildlife monitoring applications. Widespread usage of camera traps has led to large volumes of collected images, making manual processing of visual content hard to manage. In this work, we develop a framework for automatic detection and recognition of individuals in different patterned species like tigers, zebras and jaguars. Most existing systems primarily rely on manual input for localizing the animal, which does not scale well to large datasets. In order to automate the detection process while retaining robustness to blur, partial occlusion, illumination and pose variations, we use the recently proposed Faster-RCNN object detection framework to efficiently detect animals in images. We further extract features from AlexNet of the animal's flank and train a logistic regression (or Linear SVM) classifier to recognize the individuals. We primarily test and evaluate our framework on a camera trap tiger image dataset that contains images that vary in overall image quality, animal pose, scale and lighting. We also evaluate our recognition system on zebra and jaguar images to show generalization to other patterned species. Our framework gives perfect detection results in camera trapped tiger images and a similar or better individual recognition performance when compared with state-of-the-art recognition techniques.

14:20-14:40 Sequential Keystroke Behavioral Biometrics for User Identification via Multi-view Deep Learning

Lichao Sun, Yuqi Wang, Bokai Cao, Philip Yu, Witawas Srisa-An, Alex Leow

With the rapid growth in smartphone usage, more organizations begin to focus on providing better services for mobile users. User identification can help these organizations to identify their customers and then cater services that have been customized for them. Currently, the use of cookies is the most common form to identify users. However, cookies are not easily transportable (e.g., when a user uses a different login account, cookies do not follow the user). This limitation motivates the need to use behavior biometric for user identification. In this paper, we propose DEEPSERVICE, a new technique that can identify mobile users based on user's keystroke information captured by a special keyboard or web browser. Our evaluation results indicate that DEEPSERVICE is highly accurate in identifying mobile users (over 93% accuracy). The technique is also efficient and only takes less than 1 ms to perform identification.



CONGRESS HALL 4

TUESDAY

TUESDAY SEPTEMBER 19

14:40-15:00 Modeling the Temporal Nature of Human Behavior for Demographics Prediction

Bjarke Felbo, Pål Sundsøy, Alex 'Sandy' Pentland, Sune Lehmann, Yves-Alexandre de Montjoye

Mobile phone metadata is increasingly used for humanitarian purposes in developing countries as traditional data is scarce. Basic demographic information is however often absent from mobile phone datasets, limiting the operational impact of the datasets. For these reasons, there has been a growing interest in predicting demographic information from mobile phone metadata. Previous work focused on creating increasingly advanced features to be modeled with standard machine learning algorithms. We here instead model the raw mobile phone metadata directly using deep learning, exploiting the temporal nature of the patterns in the data. From high-level assumptions we design a data representation and convolutional network architecture for modeling patterns within a week. We then examine three strategies for aggregating patterns across weeks and show that our method reaches state-of-the-art accuracy on both age and gender prediction using only the temporal modality in mobile metadata. We finally validate our method on low activity users and evaluate the modeling assumptions.

15:00-15:20 RSSI Based Supervised Learning for Uncooperative Direction-Finding

Tathagata Mukherjee, Michael Duckett, Piyush Kumar, Jared Paquet, Daniel Rodriguez, Mallory Haulcomb, Kevin George, Eduardo Pasiliao

This paper studies supervised learning algorithms for the problem of uncooperative direction finding of a radio emitter using the received signal strength indicator (RSSI) from a rotating and uncharacterized antenna. Radio Direction Finding (RDF) is the task of finding the direction of a radio frequency emitter from which the received signal was transmitted, using a single receiver. We study the accuracy of radio direction finding for the 2.4 GHz WiFi band, and restrict ourselves to applying supervised learning algorithms for RSSI information analysis. We designed and built a hardware prototype for data acquisition using off-the-shelf hardware. During the course of our experiments, we collected more than three million RSSI values. We show that we can reliably predict the bearing of the transmitter with an error bounded by 11 degrees, in both indoor and outdoor environments. We do not explicitly model the multi-path, that inevitably arises in such situations and hence one of the major challenges that we faced in this work is that of automatically compensating for the multi-path and hence the associated noise in the acquired data.

15:20-15:40 Urban Water Flow and Water Level Prediction Based on Deep Learning

Haytham Assem, Salem Ghariba, Gabor Makrai, Paul Johnston, Laurence Gill, Francesco Pilla

The future planning, management and prediction of water demand and usage should be preceded by long-term variation analysis for related parameters in order to enhance the process of developing new scenarios whether for surface-water or ground-water resources. This paper aims to provide an appropriate methodology for long-term prediction for the water flow and water level parameters of the Shannon river in Ireland over a 30-year period from 1983 - 2013 through a framework that is composed of three phases: city wide scale analytics, data fusion, and domain knowledge data analytics phase which is the main focus of the paper that employs a machine learning model based on deep convolutional neural networks (DeepCNNs). We test our proposed deep learning model on three different water stations across the Shannon river and show it outperforms four well-known time-series forecasting models. We finally show how the proposed model simulate the predicted water flow and water level from 2013 - 2080. Our proposed solution can be very useful for the water authorities for better planning the future allocation of water resources among competing users such as agriculture, demotic and power stations. In addition, it can be used for capturing abnormalities by setting and comparing thresholds to the predicted water flow and water level.





Comparing Hypotheses on Sequential Behavior: A Bayesian Approach 14:00 - 14:20 and its Applications

Florian Lemmerich, Philipp Singer, Martin Becker, Lisette Espin-Noboa, Dimitar Dimitrov, Denis Helic, Andreas Hotho, Markus Strohmaier

Sequential data can be found in many settings, e.g., as sequences of visited websites or as location sequences of travelers. To improve the understanding of the underlying mechanisms that generate such sequences, the HypTrails approach provides for a novel data analysis method. Based on first-order Markov chain models and Bayesian hypothesis testing, it allows for comparing a set of hypotheses, i.e., beliefs about transitions between states, with respect to their plausibility considering observed data. HypTrails has been successfully employed to study phenomena in the online and the offline world. In this talk, we want to give an introduction to HypTrails and showcase selected real-world applications on urban mobility and reading behavior on Wikipedia.

14:20 - 14:40 **Music Generation Using Bayesian Networks**

Tetsuro Kitahara

Music generation has recently become popular as an application of machine learning. To generate polyphonic music, one must consider both simultaneity (the vertical consistency) and sequentiality (the horizontal consistency). Bayesian networks are suitable to model both simultaneity and sequentiality simultaneously. Here, we present music generation models based on Bayesian networks applied to chord voicing, four-part harmonization, and real-time chord prediction.

14:40 - 15:00 Phenotype Inference from Text and Genomic Data

Maria Brbic, Matija Piskorec, Vedrana Vidulin, Anita Krisko, Tomislav Smuc, Fran Supek

We describe ProTraits, a machine learning pipeline that systematically annotates microbes with phenotypes using a large amount of textual data from scientific literature and other online resources, as well as genome sequencing data. Moreover, by relying on a multi-view non-negative matrix factorization approach, ProTraits pipeline is also able to discover novel phenotypic concepts from unstructured text. We present the main components of the developed pipeline and outline challenges for the application to other fields.

15:00 - 15:20 **Data-Driven Approaches for Smart Parking**

Fabian Bock, Sergio Di Martino, Monika Sester

Finding a parking space is a key problem in urban scenarios, often due to the lack of actual parking availability information for drivers. Modern vehicles, able to identify free parking spaces using standard on-board sensors, have been proven to be effective probes to measure parking availability. Nevertheless, spatio-temporal datasets resulting from probe vehicles pose significant challenges to the machine learning and data mining communities, due to volume, noise, and heterogeneous spatio-temporal coverage. In this paper we summarize some of the approaches we proposed to extract new knowledge from this data, with the final goal to reduce the parking search time. First, we present a spatio-temporal analysis of the suitability of taxi movements for parking crowd-sensing. Second, we describe machine learning approaches to automatically generate maps of parking spots and to predict parking availability. Finally, we discuss some open issues for the ML/KDD community.

Process-Based Modeling and Design of Dynamical Systems 15:20 - 15:40

Jovan Tanevski, Nikola Simidjievski, Ljupco Todorovski, Sašo Džeroski

Process-based modeling is an approach to constructing explanatory models of dynamical systems from knowledge and data. The knowledge encodes information about potential processes that explain the relationships between the observed system entities. The resulting processbased models provide both an explanatory overview of the system components and closed-form equations that allow for simulating the system behavior. In this paper, we present three recent improvements of the process-based approach: (i) improving predictive performance of processbased models using ensembles, (ii) extending the scope of process-based models towards handling uncertainty and (iii) addressing the task of automated process-based design.







Probabilistic Models & Methods 2

Gaussian Conditional Random Fields Extended for Directed Graphs 16:00 - 16:20

Tijana Vujicic, Jesse Glass, Fang Zhou, Zoran Obradovic

For many real-world applications, structured regression is commonly used for predicting output variables that have some internal structure. Gaussian conditional random fields (GCRF) are a widely used type of structured regression model that incorporates the outputs of unstructured predictors and the correlation between objects in order to achieve higher accuracy. However, applications of this model are limited to objects that are symmetrically correlated, while interaction between objects is asymmetric in many cases. In this work we propose a new model, called Directed Gaussian conditional random fields (DirGCRF), which extends GCRF to allow modeling asymmetric relationships (e.g. friendship, influence, love, solidarity, etc.). The DirGCRF models the response variable as a function of both the outputs of unstructured predictors and the asymmetric structure. The effectiveness of the proposed model is characterized on six types of synthetic datasets and four real-world applications where DirGCRF was consistently more accurate than the standard GCRF model and baseline unstructured models.

16:20 - 16:40 **Multi-View Generative Adversarial Networks**

Mickael Chen, Ludovic Denoyer

Learning over multi-view data is a challenging problem with strong practical applications. Most related studies focus on the classification point of view and assume that all the views are available at any time. We consider an extension of this framework in two directions. First, based on the BiGAN model, the Multi-view BiGAN (MV-BiGAN) is able to perform density estimation from multi-view inputs. Second, it can deal with missing views and is able to update its prediction when additional views are provided. We illustrate these properties on a set of experiments over different datasets.

PAC-Bayesian Analysis for A two-step Hierarchical Mutliview Learning 16:40 - 17:00 Approach

Anil Goyal, Emilie Morvant, Pascal Germain, Massih-Reza Amini

We study a two-level multiview learning with more than two views under the PAC-Bayesian framework. This approach, sometimes referred as late fusion, consists in learning sequentially multiple view-specific classifiers at the first level, and then combining these view-specific classifiers at the second level. Our main theoretical result is a generalization bound on the risk of the majority vote which exhibits a term of diversity in the predictions of the view-specific classifiers. From this result it comes out that controlling the trade-off between diversity and accuracy is a key element for multiview learning, which complements other results in multiview learning. Finally, we experiment our principle on multiview datasets extracted from the Reuters RCV1/RCV2 collection.

17:00 - 17:20 **Robust Multi-View Topic Modeling by Incorporating Detecting Anomalies**

Guoxi Zhang, Tomoharu Iwata, Hisashi Kashima

Multi-view text data consist of texts from different sources. For in- stance, multilingual Wikipedia corpora contain articles in different languages which are created by different group of users. Because multi-view text data are often created in distributed fashion, information from different sources may not be consistent. Such inconsistency introduce noise to analysis of such kind of data. In this paper, we propose a probabilistic topic model for multi-view data, which is robust against noise. The proposed model can also be used for detecting anoma-lies. In our experiments on Wikipedia data sets, the proposed model is more ro- bust than existing multi-view topic models in terms of held-out perplexity.

Semi-Supervised Bayesian Deep Multi-modal Emotion Recognition 17:20 - 17:40

Changde Du, Changying Du, Jinpeng Li, Wei-long Zheng, Baoliang Lv, Huiguang He

This paper was accepted for presentation. However, it was not presented at the conference and is thus not published in the conference proceedinas.



CONGRESS HALL 1





CONGRESS HALL 2

16:00 - 16:20 Generalized Exploration in Policy Search

Herke van Hoof, Daniel Tanneberg, Jan Peters

To learn control policies in unknown environments, learning agents need to explore by trying actions deemed suboptimal. In prior work, such exploration is performed by either perturbing the actions at each time-step independently, or by perturbing policy parameters over an entire episode. Since both of these strategies have certain advantages, a more balanced trade-off could be beneficial. We introduce a unifying view on step-based and episode-based exploration that allows for such balanced trade-offs. This trade-off strategy can be used with various reinforcement learning algorithms. In this paper, we study this generalized exploration strategy in a policy gradient method and in relative entropy policy search. We evaluate the exploration strategy on four dynamical systems and compare the results to the established step-based and episode-based exploration strategies. Our results show that a more balanced trade-off can yield faster learning and better final policies, and illustrate some of the effects that cause these performance differences.

16:20-16:40 Max K-Armed bandit: On the ExtremeHunter algorithm and beyond

Mastane Achab, Stephan Clémençon, Aurélien Garivier, Anne Sabourin, Claire Vernade

This paper is devoted to the study of the max K-armed bandit problem, which consists in sequentially allocating resources in order to detect extreme values. Our contribution is twofold. We first significantly refine the analysis of the ExtremeHunter algorithm carried out in Carpentier and Valko (2014), and next propose an alternative approach, showing that, remarkably, Extreme Bandits can be reduced to a classical version of the bandit problem to a certain extent. Beyond the formal analysis, these two approaches are compared through numerical experiments.

16:40 - 17:00 Generalized Inverse Reinforcement Learning on Linearly Solvable MDP

Masahiro Kohjima, Tatsushi Matsubayashi, Hiroshi Sawada

In this paper, we consider a generalized variant of inverse reinforcement learning (IRL) that estimates both a cost (negative reward) function and a transition probability from observed optimal behavior. In theoretical studies of standard IRL, which estimates only the cost function, it is well known that IRL involves a non-identifiable problem, i.e., the cost function cannot be determined uniquely. This problem has been solved by using a new class of Markov decision process (MDP) called a linearly solvable MDP (LMDP). In this paper, we investigate whether a non-identifiable problem occurs in the generalized variant of IRL (gIRL) using the framework of LMDP and construct a new gIRL method. The contributions of this study are summarized as follows: (i) We point out that gIRL with LMDP suffers from a non-identifiable problem. (ii) We propose a Bayesian method to escape the non-identifiable problem. (iii) We validate the proposed method by performing an experiment on synthetic data and real car probe data.

17:00 - 17:20 Offline Reinforcement Learning With Task Hierarchies

Devin Schwab, Soumya Ray

In this work, we build upon the observation that offline reinforcement learning (RL) is synergistic with task hierarchies that decompose large Markov decision processes (MDPs). Task hierarchies can allow more efficient sample collection from large MDPs, while offline algorithms can learn better policies than the so-called "recursively optimal" or even hierarchically optimal policies learned by standard hierarchical RL algorithms. To enable this synergy, we study sample collection strategies for offline RL that are consistent with a provided task hierarchy while still providing good exploration of the state-action space. We show that naïve extensions of uniform random sampling do not work well in this case and design a strategy that has provably good convergence properties. We also augment the initial set of samples using additional information from the task hierarchy, such as state abstraction. We use the augmented set of samples to learn a policy offline. Given a capable offline RL algorithm, this policy is then guaranteed to have a value greater than or equal to the value of the hierarchically optimal policy. We evaluate our approach on several domains and show that samples generated using a task hierarchy with a suitable strategy allow significantly more sample-efficient convergence to policies with value greater than or equal to hierarchical RL approach.



UESDAY

WEDNESDAY

TUESDAY **SEPTEMBER 19**

17:20 - 17:40 Variational Thompson Sampling for Relational Recurrent Bandits

Sylvain Lamprier, Thibault Gisselbrecht, Patrick Gallinari

In this paper, we introduce a novel non-stationary bandit setting, called relational recurrent bandit, where reward expectations at successive time steps are interdependent. The aim is to discover temporal and structural dependencies between arms in order to maximize the cumulative collected reward. Two algorithms are proposed: while a first one directly models temporal dependencies between arms, a second one assumes the existence of hidden states of the system to explain rewards of arms. For both approaches, we develop a Variational Thompson Sampling method, which approximates distributions on every hidden variable via variational inference, and uses the estimated distributions to sample reward expectations at each iteration of the process. Experiments conducted on both synthetic and real data demonstrate the effectiveness of our approaches.

Rearession

16:00 - 16:20 Adaptive Skip-Train Structured Regression for Temporal Networks

Martin Pavlovski, Fang Zhou, Ivan Stojkovic, Ljupco Kocarev, Zoran Obradovic

A broad range of high impact applications involve learning a predictive model in a temporal network environment. In weather forecasting, predicting effectiveness of treatments, outcomes in healthcare and in many other domains, networks are often large, while intervals between consecutive time moments are brief. Therefore, models are required to forecast in a more scalable and efficient way, without compromising accuracy. The Gaussian Conditional Random Field (GCRF) is a widely used graphical model for performing structured regression on networks. However, GCRF is not applicable to large networks and it cannot capture different network substructures (communities) since it considers the entire network while learning. In this study, we present a novel model, Adaptive Skip-Train Structured Ensemble (AST-SE), which is a samplingbased structured regression ensemble for prediction on top of temporal networks. AST-SE takes advantage of the scheme of ensemble methods to allow multiple GCRFs to learn from several subnetworks. The proposed model is able to automatically skip the entire training or some phases of the training process. The prediction accuracy and efficiency of AST-SE were assessed and compared against alternatives on synthetic temporal networks and the H3N2 Virus Influenza network. The obtained results provide evidence that (1) AST-SE is ~140 times faster than GCRF as it skips retraining quite frequently; (2) It still captures the original network structure more accurately than GCRF while operating solely on partial views of the network; (3) It outperforms both unweighted and weighted GCRF ensembles which also operate on subnetworks but require retraining at each timestep.

ALADIN: A New Approach for Drug--Target Interaction Prediction 16:20 - 16:40

Krisztian Buza, Ladislav Peska

Due to its pharmaceutical applications, one of the most prominent machine learning challenges in bioinformatics is the prediction of drugtarget interactions. State-of-the-art approaches are based on various techniques, such as matrix factorization, restricted Boltzmann machines, network-based inference and bipartite local models (BLM). In this paper, we extend BLM by the incorporation of a hubness-aware regression technique coupled with an enhanced representation of drugs and targets in a multi-modal similarity space. Additionally, we propose to build a projection-based ensemble. Our Advanced Local Drug-Target Interaction Prediction technique (ALADIN) is evaluated on publicly available realworld drug-target interaction datasets. The results show that our approach statistically significantly outperforms BLM-NII, a recent version of BLM, as well as NetLapRLS and WNN-GIP.

16:40 - 17:00 **Co-Regularised Support Vector Regression**

Katrin Ullrich, Michael Kamp, Thomas Gärtner, Martin Vogt, Stefan Wrobel

We consider a semi-supervised learning scenario for regression, where only few labelled examples, many unlabelled instances and different data representations (multiple views) are available. For this setting, we extend support vector regression with a co-regularisation term and obtain co-regularised support vector regression (CoSVR). In addition to labelled data, co-regularisation includes information from unlabelled examples by ensuring that models trained on different views make similar predictions. Ligand affinity prediction is an important real-world problem that fits into this scenario. The characterisation of the strength of protein-ligand bonds is a crucial step in the process of drug discovery and design. We introduce variants of the base CoSVR algorithm and discuss their theoretical and computational properties. For the CoSVR function class we provide a theoretical bound on the Rademacher complexity. Finally, we demonstrate the usefulness of CoSVR for the affinity prediction task and evaluate its performance empirically on different protein-ligand datasets. We show that CoSVR outperforms co-regularised least squares regression as well as existing state-of-the-art approaches for affinity prediction.











17:00 - 17:20 Online Regression with Controlled Label Noise Rate

Edward Moroshko, Koby Crammer

Many online regression (and adaptive filtering) algorithms are linear, use additive update and designed for the noise-free setting. We consider the practical setting where the algorithm's feedback is noisy, rather than a clean label. We propose a new family of algorithms which modifies the learning rate based on the noise-variance of the feedback (labels), by shrinking both inputs and feedbacks, based on the amount of noise per input instance. We consider both settings, where the noise is either given or estimated. Empirical study with both synthetic and real-world speech data shows that our algorithms improve the overall performance of the regressor, even when there is no additional explicit information (i.e. amount of noise). We also consider a more general setting where an algorithm can sample more than single (noisy) label, yet there is a total (or average) budget for the feedback. We propose a few strategies how to effectively spend the given budget, which are based on noise-variance estimation and our shrinkage rule. We show empirically that our approach outperforms other naive approaches.

17:20 - 17:40 Robust Regression using Biased Objectives

Matthew J. Holland, Kazushi Ikeda

For the regression task in a non-parametric setting, designing the objective function to be minimized by the learner is a critical task. In this paper we propose a principled method for constructing and minimizing robust losses, which are resilient to errant observations even under small samples. Existing proposals typically utilize very strong estimates of the true risk, but in doing so require a priori information that is not available in practice. As we abandon direct approximation of the risk, this lets us enjoy substantial gains in stability at a tolerable price in terms of bias, all while circumventing the computational issues of existing procedures. We analyze existence and convergence conditions, provide practical computational routines, and also show empirically that the proposed method realizes superior robustness over wide data classes with no prior knowledge assumptions.

Privacy & Security

16:00 - 16:20

Differentially Private Nearest Neighbor Classification

Mehmet Emre Gursoy, Ali Inan, Mehmet Ercan Nergiz, Yucel Saygin

Instance-based learning, and the k-nearest neighbors algorithm (k-NN) in particular, provide simple yet effective classification algorithms for data mining. Classifiers are often executed on sensitive information such as medical or personal data. Differential privacy has recently emerged as the accepted standard for privacy protection in sensitive data. However, straightforward applications of differential privacy to k-NN classification yield rather inaccurate results. Motivated by this, we develop algorithms to increase the accuracy of private instance-based classification. We first describe the radius neighbors classifier (r-N) and show that its accuracy under differential privacy can be greatly improved by a non-trivial sensitivity analysis. Then, for k-NN classification, we build algorithms that convert k-NN classifiers to r-N classifiers. We experimentally evaluate the accuracy of both classifiers using various datasets. Experiments show that our proposed classifiers significantly outperform baseline private classifiers (i.e., straightforward applications of differential privacy) and executing the classifiers on a dataset published using differential privacy. In addition, the accuracy of our proposed k-NN classifiers are at least comparable to, and in many cases better than, the other differentially private machine learning techniques.

16:20-16:40 Malware Detection By Analysing Encrypted Network Traffic with Neural Networks

Paul Prasse, Lukas Machlika, Tomas Pevny, Jiri Havelka, Tobias Scheffer

We study the problem of detecting malware on client computers based on the analysis of HTTPS traffic. Here, malware has to be detected based on the host address, timestamps, and data volume information of the computer's network traffic. We develop a scalable protocol that allows us to collect network flows of known malicious and benign applications as training data and derive a malware-detection method based on a neural embedding of domain names and a long short-term memory network that processes network flows. We study the method's ability to detect new malware in a large-scale empirical study.



67



CONGRESS HALL 4



TUESDAY SEPTEMBER 19

16:40-17:00 PEM: Practical Differentially Private System for Large-Scale

Cross-Institutional Data Mining Li Yi, Yitao Duan, Wei Xu

Privacy has become a serious concern in data mining. Achieving adequate privacy is especially challenging when the scale of the problem is large. Fundamentally, designing a practical privacy-preserving data mining system involves trade-offs among several factors such as the privacy guarantee, the accuracy or utility of the mining result, the computation efficiency and the generality of the approach. In this paper, we present PEM, a practical system that tries to strike the right balance among these factors. We use a combination of noise-based and noise-free techniques to achieve provable differential privacy at a low computational overhead while obtaining more accurate result than previous approaches. PEM provides an efficient private gradient descent that can be the basis for many practical data mining and machine learning algorithms, like logistic regression, k-means, and Apriori. Evaluate them on three real- world open datasets in a cloud computing environment. The results show that PEM achieves good accuracy, high scalability, low computation cost while maintaining differential privacy.

17:00-17:20 Preserving Differential Privacy in Convolutional Deep Belief Networks

NhatHai Phan, Xintao Wu, Dejing Dou

The remarkable development of deep learning in medicine and healthcare domain presents obvious privacy issues, when deep neural networks are built on users' personal and highly sensitive data, e.g., clinical records, user profiles, biomedical images, etc. However, only a few scientific studies on preserving privacy in deep learning have been conducted. In this paper, we focus on developing a private convolutional deep belief network (pCDBN), which essentially is a convolutional deep belief network (CDBN) under differential privacy. Our main idea of enforcing ε -differential privacy is to leverage the functional mechanism to perturb the energy-based objective functions of traditional CDBNs, rather than their results. One key contribution of this work is that we propose the use of Chebyshev expansion to derive the approximate polynomial representation of objective functions. Our theoretical analysis shows that we can further derive the sensitivity and error bounds of the approximate polynomial representation. As a result, preserving differential privacy in CDBNs is feasible. We applied our model in a health social network, i.e., YesiWell data, and in a handwriting digit dataset, i.e., MNIST data, for human behavior prediction, human behavior classification, and handwriting digit recognition tasks. Theoretical analysis and rigorous experimental evaluations show that the pCDBN is highly effective. It significantly outperforms existing solutions.

17:20-17:40 The Best Privacy Defense Is a Good Privacy Offense: Obfuscating a Search Engine User's Profile

Jörg Wicker, Stefan Kramer

User privacy on the Internet is an important and unsolved problem. So far,no sufficient and comprehensive solution has been proposed that helps a user to protect his or her privacy while using the Internet. Data are collected and assembled by numerous service providers. Solutions so far focused on the side of the service providers to store encrypted or transformed data that can be still used for analysis. This has a major flaw, as it relies on the service providers to do this. The user has no chance of actively protecting his or her privacy. In this work, we suggest a new approach, empowering the user to take advantage of the same tool the other side has, namely data mining to produce data which obfuscates the user's profile. We apply this approach to search engine queries and use feedback of the search engines in terms of personalized advertisements in an algorithm similar to reinforcement learning to generate new queries potentially confusing the search engine. We evaluated the approach using a real-world data set. While evaluation is hard, we achieve results that indicate that it is possible to influence the user's profile that the search engine generates. This shows that it is feasible to defend a user's privacy from a new and more practical perspective.





WHODID: Web-Based interface for Human-assisted Factory Operations in Fault Detection

Gouy-Pailler Cedric, Pierre Blanchart

We present WHODID: a turnkey intuitive web-based interface for fault detection, identification and diagnosis in production units. Fault detection and identification is an extremely useful feature and is becoming a necessity in modern production units. Moreover, the large deployment of sensors within the stations of a production line has enabled the close monitoring of products being manufactured. In this context, there is a high demand for computer intelligence able to detect and isolate faults inside production lines, and to additionally provide a diagnosis for maintenance on the identified faulty production device, with the purpose of preventing subsequent faults caused by the diagnosed faulty device behavior. We thus introduce a system which has fault detection, isolation, and identification features, for retrospective and on-the-fly monitoring and maintenance of complex dynamical production processes. It provides real-time answers to the questions: "is there a fault?," "where did it happen?," "for what reason?". The method is based on a posteriori analysis of decision sequences in XGBoost tree models, using recurrent neural networks sequential models of tree paths. The particularity of the presented system is that it is robust to missing or faulty sensor measurements, it does not require any modeling of the underlying, possibly exogenous manufacturing process, and provides fault diagnosis along with confidence level in plain English formulations. The latter can be used as maintenance directions by a human operator in charge of production monitoring and control.

TrAnET: Tracking and Analyzing the Evolution of Topics in Information Networks

Livio Bioglio, Ruggero Pensa, Valentina Rho

This paper presents a system for tracking and analyzing the evolution and transformation of topics in an information network. The system consists of four main modules for pre-processing, adaptive topic modeling, network creation and temporal network analysis. The core module is built upon an adaptive topic modeling algorithm adopting a sliding time window technique that enables the discovery of groundbreaking ideas as those topics that evolve rapidly in the network.

Framework for Exploring and Understanding Multivariate Correlations

Louis Kirsch, Niklas Riekenbrauck, Daniel Thevessen, Marcus Pappik, Axel Stebner, Julius Kunze, Alexander Meissner, Arvind Kumar Shekar, Emmanuel Müller

Feature selection is an essential step to identify relevant and non-redundant features for target class prediction. In this context, the number of feature combinations grows exponentially with the dimension of the feature space. This hinders the user's understanding of the feature-target relevance and feature-feature redundancy. We propose an interactive Framework for Exploring and Understanding Multivariate Correlations (FEXUM), which embeds these correlations using a force-directed graph. In contrast to existing work, our framework allows the user to explore the correlated feature space and guides in understanding multivariate correlations through interactive visualizations.

Tetrahedron: Barycentric Measure Visualizer

Dariusz Brzezinski, Jerzy Stefanowski, Izabela Szczęch, Robert Susmaga

Each machine learning task comes equipped with its own set of performance measures. For example, there is a plethora of classification measures that assess predictive performance, a myriad of clustering indices, and equally many rule interestingness measures. Choosing the right measure requires careful thought, as it can influence model selection and thus the performance of the final machine learning system. However, analyzing and understanding measure properties is a difficult task. Here, we present Tetrahedron, a web-based visualization tool that aids the analysis of complete ranges of performance measures based on a two-by-two contingency matrix. The tool operates in a barycentric coordinate system using a 3D tetrahedron, which can be rotated, zoomed, cut, parameterized, and animated. The application is capable of visualizing predefined measures (86 currently), as well as helping prototype new measures by visualizing user-defined formulas.

Delve: A Data Set Retrieval and Document Analysis System

Uchenna Akujuobi, Xiangliang Zhang

Academic search engines (e.g., Google scholar or Microsoft academic) provide a medium for retrieving various information on scholarly documents. However, most of these popular scholarly search engines overlook the area of data set retrieval, which should provide information on relevant data sets used for academic research. Due to the increasing volume of publications, it has become a challenging task to locate suitable data sets on a particular research area for benchmarking or evaluations. We propose Delve, a web-based system for data set retrieval and document analysis. This system is different from other scholarly search engines as it provides a medium for both data set retrieval and real time visual exploration and analysis of data sets and documents.

WEDNESDAY

PROGRAM AT A GLANCE



NVITED TALK

CONGRESS HALL 1

09:00 - 10:00

Multi-Target Prediction via Low-Rank Embeddings Inderjit Dhillon

University of Texas at Austin, USA

Linear prediction methods, such as linear regression and classification, form the bread-and-butter of modern machine learning. The classical scenario is the presence of data with multiple features and a single target variable. However, there are many recent scenarios where there are multiple target variables. For example, recommender systems, predicting bid words for a web page (where each bid word acts as a target variable), or predicting diseases linked to a gene. In many of these scenarios, the target variables might themselves be associated with features. In these scenarios, bilinear and nonlinear prediction via low-rank embeddings have been shown to be extremely powerful. The low-rank embeddings serve a dual purpose: (i) they enable tractable computation even in the face of millions of data points as well as target variables, and (ii) they exploit correlations among the target variables, even when there are many missing observations. We illustrate our methodology on various modern machine learning problems: recommender systems, multi-label learning and inductive matrix completion, and present results on some standard benchmarks as well as an application that involves prediction of gene-disease associations.

Chair: Michelangelo Ceci
Random k-Labelsets: An Ensemble Method for Multilabel Classification

Grigorios Tsoumakas & Ioannis Vlahavas

This paper proposes an ensemble method for multilabel classification. The RAndom k-labELsets (RAKEL) algorithm constructs each member of the ensemble by considering a small random subset of labels and learning a single-label classifier for the prediction of each element in the powerset of this subset. In this way, the proposed algorithm aims to take into account label correlations using single-label classifiers that are applied on subtasks with manageable number of labels and adequate number of examples per label. Experimental results on common multilabel domains involving protein, document and scene classification show that better performance can be achieved compared to popular multilabel classification approaches.

Chair: Michelangelo Ceci

Sponsored by



SESSIONS AT A GLANCE

Learning & Optimization 1

Chair: Georgiana Ifrim

CONGRESS HALL 1

11:00-12:40

Cost-Sensitive Label Embedding for Multi-Label Classification

11:00 - 11:20	
	Kuan-Hao Huang, Hsuan-Tien Lin
11:20 - 11:40 11:40 - 12:00	Crossprop: Learning Representations By Stochastic Meta-Gradient Descent in Neural Networks
	Vivek Veeriah, Shangtong Zhang, Richard Sutton
	Group Online Adaptive Learning
	Alon Zweig, Gal Chechik
12:00 - 12:20	Speeding Up Hyper-Parameter Optimization By Extrapolation of Learning Curves Using Previous Builds
	Akshay Chandrashekaran, lan Lane
12:20 - 12:40	Thompson Sampling for Optimizing Stochastic Local Search
	Tong Yu, Branislav Kveton, Ole Mengshoel

Unsupervised & Semi-supervised Learning 1

Chair: Gjorgji Madjarov

11:00 - 11:00	Constraint-Based Clustering Selection
11.00 - 11.20	Toon Van Craenendonck, Hendrik Blockeel
11.00 11.40	k ² -Means for Fast and Accurate Large Scale Clustering
11.20 - 11.40	Eirikur Agustsson, Radu Timofte, Luc Van Gool
11.40 10.00	Learning Constraints in Spreadsheets and Tabular Data
11.40 - 12.00	Samuel Kolb, Sergey Paramonov, Tias Guns, Luc De Raedt
10.00 10.00	On Temporal-Constrained Sub-Trajectory Cluster Analysis
12.00 - 12.20	Nikos Pelekis, Panagiotis Tampakis, Marios Vodas, Christos Doulkeridis, Yannis Theodoridis
10.00 10.40	Pivot-Based Distributed k-Nearest Neighbor Mining
12.20 - 12:40	Caitlin Kuhlman, Yizhou Yan, Lei Cao, Elke Rundensteiner

.....

CONGRESS HALL 2 11:00-12:40

Feature Selection & Extraction

Chair: Dragi Kocev

11:00 - 11:20	Deep Discrete Hashing with Self-Supervised Labels
	Jingkuan Song, Tao He, Hangbo Fan, Lianli Gao
11:20 - 11:40	Including Multi-Feature Interactions and Redundancy for Feature Ranking in Mixed Datasets
	Arvind Kumar Shekar, Tom Bocklisch, Patricia Iglesias Sanchez, Christoph-Nikolas Straehle, Emmanuel Müller
11.40 - 19.00	Non-Redundant Spectral Dimensionality Reduction
11.40 - 12.00	Yochai Blau, Tomer Michaeli
12:00 - 12:20	Rethinking Unsupervised Feature Selection: From Pseudo Labels to Pseudo Must-Links
	Xiaokai Wei, Sihong Xie, Bokai Cao, Philip Yu
12:20 - 12:40	SetExpan: Corpus-Based Set Expansion Via Context Feature Selection and Rank Ensemble
	Jiaming Shen, Zeqiu Wu, Dongming Lei, Jingbo Shang, Xiang Ren, Jiawei Han

Applied Data Science 3

CONGRESS HALL 4

CONGRESS HALL 1

11:00-12:40

Chair: Taneli Mielikäinen

11:00 - 11:20	Koopman Spectral Kernels for Comparing Complex Dynamics: Application to Multiagent Sport Plays Keisuke Fujii, Yuki Inaba, Yoshinobu Kawahara		
11:20 - 11:40	Boosting Based Multiple Kernel Learning and Transfer Regression for Electricity Load Forecasting Di Wu, Boyu Wang, Doina Precup, Benoit Boulet		
11:40 - 12:00	DC-Prophet: Predicting Catastrophic Machine Failures in DataCenters You-Luen Lee, Da-Cheng Juan, Xuan-An Tseng, Yu-Ting Chen, Shih-Chieh Chang		
12:00 - 12:20	Quantifying Heterogeneous Causal Treatment Effects in World Bank Development Finance Projects Jianing Zhao, Daniel Runfola, Peter Kemper		
12:20 - 12:40	Taking It for a Test Drive: A Hybrid Spatio-Temporal Model for Wildlife Poaching Prediction Evaluated Through a Controlled Field Test		
	Shahrzad Gholami, Benjamin Ford, Fei Fang, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Mustapha Nsubaga, Joshua Mabonga		

Neural Networks & Deep Learning 2

Chair: Marko Robnik-Šikonja

 14:00-15:40

 14:00 - 14:20

 CON-S2V: A Generic Framework for Incorporating Extra-Sentential Context Into Sen2Vec

 Tanay Kumar Saha, Shafia, laty Mehammad Al Hasan

lanay Kumar Saha, Shafiq Joty, Mohammad Al Hasan	
Deep Over-Sampling Framework for Classifying Imbalanced Data	
Shin Ando, Chun Yuan Huang	
Joint User Modeling Across Aligned Heterogeneous Sites Using Neural Networks	
Xuezhi Cao, Yong Yu	
Learning Deep Kernels in the Space of Dot Product Polynomials	
Michele Donini, Fabio Aiolli	
Wikipedia Vandal Early Detection: from User Behavior to User Embedding	
Shuhan Yuan, Panpan Zheng, Xintao Wu, Yang Xiang	

Kernel Methods 1

Chair: Andrea Passerini

CONGRESS HALL 2 14:00-15:40

14:00 - 14:20	Fair Kernel Learning
	Adrian Perez-Suay, Valero Laparra, Gonzalo Mateo-García, Jordi Muñoz-Marí, Luis Gómez-Chova, Gustau Camps-Valls
14:20 - 14:40	GaKCo: a Fast Gapped K-Mer String Kernel Using Counting
	Ritambhara Singh, Arshdeep Sekhon, Andrew Norton, Jack Lanchantin, Kamran Kowsari, Beilun Wang, Yanjun Qi
14:40 - 15:00	Graph Enhanced Memory Networks for Sentiment Analysis
	Zhao Xu, Romain Vial, Kristian Kersting
15:00 - 15:20	Kernel Sequential Monte Carlo
	Ingmar Schuster, Heiko Strathmann, Brooks Paige, Dino Sejdinovic
15:20 - 15:40	Learning Lukasiewicz Logic Fragments By Quadratic Programming
	Francesco Giannini, Michelangelo Diligenti, Marco Gori, Marco Maggini

Nectar 2

CONGRESS HALL 3

CONGRESS HALL 4

14:00-15:40

Chairs: Donato Malerba & Jerzy Stefanowski

14:00 - 14:20	Recent Advances in Kernel-Based Graph Classification
	Nils Kriege, Christopher Morris
14:20 - 14:40	Activity-Driven Influence Maximization in Social Networks
	Rohit Kumar, Muhammad Saleem, Toon Calders, Torben Pedersen, Xike Xie
14:40 - 15:00	QuickScorer: Efficient Traversal of Large Ensembles of Decision Trees
	Claudio Lucchese, Franco Maria Nardini, Salvatore Orlando, Raffaele Perego, Nicola Tonellotto, Rossano Venturini
15:00 - 15:20	An AI Planning System for Data Cleaning
	Roberto Boselli, Mirko Cesarini, Fabio Mercorio, Mario Mezzanzanica
15:20 - 15:40	Image Representation, Annotation and Retrieval with Predictive Clustering Trees
	Ivica Dimitrovski, Dragi Kocev, Suzana Loskovska, Sašo Džeroski

Applied Data Science 4

Chair: Taneli Mielik	äinen 14:00-15:4
14:00 - 14:20	Analyzing Granger Causality in Climate Data with Time Series Classification Methods
	Christina Papagiannopoulou, Stijn Decubber, Diego Miralles, Matthias Demuzere, Niko Verhoest, Willem Waegeman
14:20 - 14:40	CREST - Risk Prediction for Clostridium Difficile Infection Using Multimodal Data Mining
	Cansu Sen, Thomas Hartvigsen, Elke Rundensteiner, Kajal Claypool

SPONSOR PRESENTATION 14:40 - 15:40

Networks & Graphs 2

Chair: Nikolaj Tatti	10:00-17:00

16:00 - 16:20	Regularizing Neural Knowledge Graph Embeddings Via Equivalence and Inversion Axioms
	Pasquale Minervini, Luca Costabello, Emir Munoz, Vit Novacek, Pierre-Yves Vandenbussche
	The Network-Untangling Problem: From Interactions to Activity Timelines
16:20 - 16:40	Polina Rozenstein, Nikolaj Tatti, Aristides Gionis
16:40 - 17:00	TransT: Type-Based Multiple Embedding Representations for Knowledge Graph Completion
	Shiheng Ma, Jianhui Ding, Weijia Jia, Kun Wang, Minyi Guo

Kernel Methods 2 CONGRESS H/		
Chair: Celine Vens 16:00-17		
16:00 - 16:20	Bayesian Nonlinear Support Vector Machines for Big Data	
	Florian Wenzel, Theo Galy-Fajou, Matthäus Deutsch, Marius Kloft	
16:20 - 16:40	Entropic Trace Estimation for Log Determinants	
	Jack Fitzsimons, Diego Granziol, Kurt Cutajar, Michael Osborne, Maurizio Filippone, Stephen Roberts	
16:40 - 17:00	Nystrom Sketching	
	Daniel Perry, Braxton Osting, Ross Whitaker	

Learning & Optimization 2	CONGRESS HALL 3
Chair: Ljupčo Todorovski	16:00-17:00

16:00 - 16:20	A Constrained L1 Minimization Approach for Estimating Multiple Sparse Gaussian Or Nonparanormal Graphical Models
	Beilun Wang, Ritambhara Singh, Yanjun Qi
16:20 - 16:40	Distributed Stochastic Optimization of the Regularized Risk Via Saddle-Point Problem
	Shin Matsushima, Hyokun Yun, Xinhua Zhang, S.V.N. Vishwanathan
16:40 - 17:00	Sparse Probit Linear Mixed Model
	Stephan Mandt, Florian Wenzel, Shinichi Nakajima, John Cunningham, Christoph Lippert, Marius Kloft

Subgroup Disco	very Congress Hall 4
Chair: Jilles Vreeke	en 16:00-17:00
40.00 40.00	Explaining Deviating Subsets Through Explanation Networks
16:00 - 16:20	Antti Ukkonen, Vladimir Dzyuba, Matthijs Van Leeuwen
16:00 - 16:40	Flash Points: Discovering Exceptional Pairwise Behaviors in Vote Or Rating Data
16:20 - 16:40	Adnene Belfodil, Sylvie Cazalens, Philippe Lamarre, Marc Plantevit
16:40 - 17:00	Identifying Consistent Statements About Numerical Data with Dispersion-Corrected Subgroup Discovery
	Mario Boley, Bryan R. Goldsmith, Luca M. Ghiringhelli, Jilles Vreeken

77

Discovery Challenge 2

Chair: Roberto Corizzo

Multi-Plant Photovoltaic Energy Forecasting Challenge

The urgent need to reduce pollution emission has made renewable energy a strategic European Union (EU) and international sector. This has resulted in an increasing presence of renewable energy sources and thus, significant distributed power generation. The main challenges faced by this new energy market are grid integration, load balancing and energy trading.

In order to face these challenges, it is of paramount importance to monitor the production and consumption of energy, both at the local and global level, to store historical data and to design new, reliable prediction tools.

In this challenge, we focus our attention on photovoltaic (PV) power plants, due to their wide distribution in Europe. During the last years, the forecast of PV energy production has received significant attention since photovoltaics are becoming a major source of renewable energy for the world.

Forecast may apply to a single renewable power generation system, or refer to an aggregation of large numbers of systems spread over an extended geographic area.

Organizers : Roberto Corizzo, University of Bari, Italy

http://193.204.187.201:8080/pv_challenge_website/

14.00 - 14.15	Introduction
14.00 - 14.15	Roberto Corizzo
14:15 - 14:35	Multi-Plant Photovoltaic Energy Forecasting with Regression Tree Ensembles and Hourly Average Forecasts - 1st Place
	Kathrin Bujna, Martin Wistuba
14.25 14.55	Multi-Plant Photovoltaic Energy Forecasting - 2nd Place
14.30 - 14:55	Clément Gautrais, Yann Dauxais, Maël Guilleme

SESSIONS WITH ABSTRACTS

Learning & Optimization 1

CONGRESS HALL 1

11:00 - 11:20 **Cost-Sensitive Label Embedding for Multi-Label Classification**



Label embedding (LE) is an important family of multi-label classification algorithms that digest the label information jointly for better performance. Different real-world applications evaluate performance by different cost functions of interest. Current LE algorithms often aim to optimize one specific cost function, but they can suffer from bad performance with respect to other cost functions. In this paper, we resolve the performance issue by proposing a novel cost-sensitive LE algorithm that takes the cost function of interest into account. The proposed algorithm, costsensitive label embedding with multidimensional scaling (CLEMS), approximates the cost information with the distances of the embedded vectors by using the classic multidimensional scaling approach for manifold learning. CLEMS is able to deal with both symmetric and asymmetric cost functions, and effectively makes cost-sensitive decisions by nearest-neighbor decoding within the embedded vectors. We derive theoretical results that justify how CLEMS achieves the desired cost-sensitivity. Furthermore, extensive experimental results demonstrate that CLEMS is significantly better than a wide spectrum of existing LE algorithms and state-of-the-art cost-sensitive algorithms across different cost functions.

Crossprop: Learning Representations by Stochastic Meta-gradient 11:20 - 11:40 **Descent in Neural Networks** Vivek Veeriah, Shangtong Zhang, Richard Sutton



Representations are fundamental to artificial intelligence. The performance of a learning system depends on the type of representation used for representing the data. Typically, these representations are hand-engineered using domain knowledge. More recently, the trend is to learn these representations through stochastic gradient descent in multi-layer neural networks, usually called backprop. Learning the representations directly from the incoming data stream reduces the human labour involved in designing a learning system. More importantly, this allows in scaling of a learning system for difficult tasks. In this paper, we introduce a new incremental learning algorithm called crossprop, which learns incoming weights of hidden units based on the meta-gradient descent approach, that was previously introduced by Sutton (1992) and Schraudolph (1999) for learning step-sizes. The final update equation introduces an additional memory parameter for each of these weights and generalizes the backprop update equation. From our experiments, we show that crossprop learns and reuses its feature representation while tackling new and unseen tasks whereas backprop relearns a new feature representation.

11:40 - 12:00 Group Online Adaptive Learning

Alon Zweig, Gal Chechik

Sharing information among multiple learning agents can accelerate learning. It could be particularly useful if learners operate in continuously changing environments, because a learner could benefit from previous experience of another learner to adapt to their new environment. Such group-adaptive learning has numerous applications, from predicting financial time-series, through content recommendation systems, to visual understanding for adaptive autonomous agents. Here we address the problem in the context of online adaptive learning. We formally define the learning settings of Group Online Adaptive Learning and derive an algorithm named Shared Online Adaptive Learning (SOAL) to address it. SOAL avoids explicitly modeling changes or their dynamics, and instead shares information continuously. The key idea is that learners share a common small pool of experts, which they can use in a weighted adaptive way. We define group adaptive regret and prove that SOAL maintains known bounds on the adaptive regret obtained for single adaptive learners. Furthermore, it quickly adapts when learning tasks are related to each other. We demonstrate the benefits of the approach for two domains: vision and text. First, in the visual domain, we study a visual navigation task where a robot learns to navigate based on outdoor video scenes. We show how navigation can improve when knowledge from other robots in related scenes is available. Second, in the text domain, we create a new dataset for the task of assigning submitted papers to relevant editors. This is, inherently, an adaptive learning task due to the dynamic nature of research fields evolving in time. We show how learning to assign editors improves when knowledge from other editors is available. Together, these results demonstrate the benefits for sharing information across learners in concurrently changing environments.

WEDNESDAY **SEPTEMBER 20**

Speeding Up Hyper-parameter Optimization by Extrapolation of 12:00 - 12:20 Learning Curves using Previous Builds

Akshay Chandrashekaran, Ian Lane

Recent work has shown that the usage of extrapolation of learning curves to determine when to terminate a training build has been shown to be effective in reducing the number of epochs of training required for finding a good performing hyper-parameter configuration. However, the current technique uses the information only from the current build to make the prediction. We propose the usage of a simple regression based extrapolation model that uses the trajectories from previous builds to make predictions of new builds. This can be used to terminate poorly performing builds and thus, speed up hyper-parameter search with performance comparable to non-augmented hyper-parameter optimization techniques. We compare the predictions made by our model against that of the existing extrapolation technique in different tasks. We incorporate our approach into a pre-existing termination criterion. We incorporate this termination criterion into an existing hyper-parameter optimization toolkit. We analyze the performance of our approach and contrast it against a baseline in terms of quality of prediction in three different tasks. We show that our approach yields builds with performance comparable to the non-augmented version with fewer epochs, and outperforms an existing parametric extrapolation technique for two out of three tasks in terms of number of required epochs.

12:20 - 12:40 Thompson Sampling for Optimizing Stochastic Local Search

Tong Yu, Branislav Kveton, Ole Mengshoel

Stochastic local search (SLS), like many other stochastic optimization algorithms, has several parameters that need to be optimized in order for the algorithm to find high quality solutions within a short amount of time. In this paper, we formulate a stochastic local search bandit (SLSB), which is a novel learning variant of SLS based on multi-armed bandits. SLSB optimizes SLS over a sequence of stochastic optimization problems and achieves high average cumulative reward. In SLSB, we study how SLS can be optimized via low degree polynomials in its noise and restart parameters. To determine the coefficients of the polynomials, we present polynomial Thompson Sampling (PolyTS). We derive a regret bound for PolyTS and validate its performance on synthetic problems of varying difficulty as well as on feature selection problems. Compared to bandits with no assumptions of the reward function and other parameter optimization approaches, our PolyTS assuming polynomial structure can provide substantially better parameter optimization for SLS. In addition, due to its simple model update, PolyTS has low computational cost compared to other SLS parameter optimization methods.

Unsupervised & Semi-supervised Learning 1

11:00 - 11:20 **Constraint-Based Clustering Selection**

Toon Van Craenendonck, Hendrik Blockeel

Clustering requires the user to define a distance metric, select a clustering algorithm, and set the hyperparameters of that algorithm. Getting these right, so that a clustering is obtained that meets the users subjective criteria, can be difficult and tedious. Semi-supervised clustering methods make this easier by letting the user provide must-link or cannot-link constraints. These are then used to automatically tune the similarity measure and/or the optimization criterion. In this paper, we investigate a complementary way of using the constraints: they are used to select an unsupervised clustering method and tune its hyperparameters. It turns out that this very simple approach outperforms all existing semi-supervised methods. This implies that choosing the right algorithm and hyperparameter values is more important than modifying an individual algorithm to take constraints into account. In addition, the proposed approach allows for active constraint selection in a more effective manner than other methods.

11:20 - 11:40 k²-means for Fast and Accurate Large Scale Clustering

Eirikur Agustsson, Radu Timofte, Luc Van Gool

We propose k²-means, a new clustering method which efficiently copes with large numbers of clusters and achieves low energy solutions. k²-means builds upon the standard k-means (Lloyd's algorithm) and combines a new strategy to accelerate the convergence with a new low time complexity divisive initialization. The accelerated convergence is achieved through only looking at k n nearest clusters and using triangle inequality bounds in the assignment step while the divisive initialization employs an optimal 2-clustering along a direction. The worst-case time complexity per iteration of our k^2-means is O(nk nd+k²d), where d is the dimension of the n data points and k is the number of clusters and usually n >> k >> k_n. Compared to k-means' O(nkd) complexity, our k²-means complexity is significantly lower, at the expense of slightly increasing the memory complexity by O(nk_n+k²). In our extensive experiments k²-means is an order of magnitude faster than standard methods in computing accurate clusterings on several standard datasets and settings with hundreds of clusters and high dimensional data. Moreover, the proposed divisive initialization generally leads to clustering energies comparable to those achieved with the standard k-means++ initialization, while being significantly faster.











11:40 - 12:00 Learning Constraints in spreadsheets and tabular data

Samuel Kolb, Sergey Paramonov, Tias Guns, Luc De Raedt

Spreadsheets, comma separated value files and other tabular data representations are in wide use today. However, writing, maintaining and identifying good formulas for tabular data and spreadsheets can be time-consuming and error-prone. We investigate the automatic learning of constraints (formulas and relations) in raw tabular data in an unsupervised way. We represent common spreadsheet formulas and relations through predicates and expressions whose arguments must satisfy the inherent properties of the constraint. The challenge is to automatically infer the set of constraints present in the data, without labeled examples or user feedback. We propose a two-stage generate and test method where the first stage uses constraint solving techniques to efficiently reduce the number of candidates, based on the predicate signatures. Our approach takes inspiration from inductive logic programming, constraint learning and constraint satisfaction. We show that we are able to accurately discover constraints in spreadsheets from various sources.

12:00 - 12:20 **On Temporal-Constrained Sub-Trajectory Cluster Analysis**

Nikos Pelekis, Panagiotis Tampakis, Marios Vodas, Christos Doulkeridis, Yannis Theodoridis

Cluster analysis over Moving Object Databases (MODs) is a challenging research topic that has attracted the attention of the mobility data mining community. In this paper, we study the temporal-constrained sub-trajectory cluster analysis problem, where the aim is to discover clusters of sub-trajectories given an ad-hoc, user-specified temporal constraint within the dataset's lifetime. The problem is challenging because: (a) the time window is not known in advance, instead it is specified at query time, and (b) the MOD is continuously updated with new trajectories. Existing solutions first filter the trajectory database according to the temporal constraint, and then apply a clustering algorithm from scratch on the filtered data. However, this approach is extremely inefficient, when considering explorative data analysis where multiple clustering tasks need to be performed over different temporal subsets of the database, while the database is updated with new trajectories. To address this problem, we propose an incremental and scalable solution to the problem, which is built upon a novel indexing structure, called Representative Trajectory Tree (ReTraTree). ReTraTree acts as an effective spatio-temporal partitioning technique; partitions in ReTraTree correspond to groupings of sub-trajectories, which are incrementally maintained and assigned to representative (sub-)trajectories. Due to the proposed organization of sub-trajectories, the problem under study can be efficiently solved as simply as executing a guery operator on ReTraTree, while insertion of new trajectories is supported. Our extensive experimental study performed on real and synthetic datasets shows that our approach outperforms a state-of-the-art in-DBMS solution supported by PostgreSQL by orders of magnitude.

12:20 - 12:40 **Pivot-Based Distributed k-Nearest Neighbor Mining**

Caitlin Kuhlman, Yizhou Yan, Lei Cao, Elke Rundensteiner

k-nearest-neighbor (kNN) search is a fundamental data mining task critical to many data analytics methods. Yet no effective techniques to date scale kNN search to large datasets. In this work we present PkNN, an exact distributed method that by leveraging modern distributed architectures for the first time scales kNN search to billion point datasets. The key to the PkNN strategy is a multi-round kNN search that exploits pivot-based data partitioning at each stage. This includes an outlier-driven partition adjustment mechanism that effectively minimizes data duplication and achieves a balanced workload across the compute cluster. Aggressive data-driven bounds along with a tiered support assignment strategy ensure correctness while limiting computation costs. Our experimental study on multi-dimensional real-world data demonstrates that PkNN achieves significant speedup over the state-of-the-art and scales effectively in data cardinality and dimension.

Feature Selection & Extraction

11:00 - 11:20 **Deep Discrete Hashing with Self-Supervised Labels**

Jingkuan Song, Tao He, Hangbo Fan, lianli Gao

Hashing methods have been widely used for applications of large-scale image retrieval and classification. Non-deep hashing methods using handcrafted features have been significantly outperformed by deep hashing methods due to their better feature representation and end-to-end learning framework. However, the most striking successes in deep hashing have mostly involved discriminative models, which require labels. In this paper, we propose a novel unsupervised deep hashing method, named Deep Discrete Hashing (DDH), for large-scale image retrieval and classification. In the proposed framework, we address two main problems: 1) how to directly learn discrete binary codes? 2) how to equip the binary representation with the ability of accurate image retrieval and classification in an unsupervised way? We resolve these problems by introducing an intermediate variable and a loss function steering the learning process, which is based on the neighborhood structure in the original space. Experiments on real datasets show that our method can significantly outperform other unsupervised methods to achieve the stateof-the-art performance for image retrieval and object recognition.





WEDNESDAY SEPTEMBER 20

11:20-11:40 Including Multi-feature interactions and Redundancy for Feature Ranking in Mixed Datasets

Arvind Kumar Shekar, Tom Bocklisch, Patricia Iglesias Sanchez, Christoph-Nikolas Straehle, Emmanuel Müller



Feature ranking is beneficial to gain knowledge and to identify the relevant features from a high-dimensional dataset. However, in several datasets, few features by itself might have small correlation with the target classes, but by combining these features with some other features, they can be strongly correlated with the target. This means that multiple features exhibit interactions among themselves. It is necessary to rank the features based on these interactions for better analysis and classifier performance. However, evaluating these interactions on large datasets is computationally challenging. Furthermore, datasets often have features with redundant information. Using such redundant features hinders both efficiency and generalization capability of the classifier. The major challenge is to efficiently rank the features based on relevance and redundance on mixed datasets. In this work, we propose a filter-based framework based on Relevance and Redundancy (RaR), RaR computes a single score that quantifies the feature relevance by considering interactions between features and redundancy. The top ranked features of RaR are characterized by maximum relevance and non-redundance. The evaluation on synthetic and real world datasets demonstrates that our approach outperforms several state-of-the-art feature selection techniques.

11:40-12:00 Non-Redundant Spectral Dimensionality Reduction

Yochai Blau, Tomer Michaeli



12:00-12:20 Rethinking Unsupervised Feature Selection: From Pseudo Labels to Pseudo Must-Links

Xiaokai Wei, Sihong Xie, Bokai Cao, Philip Yu



High-dimensional data are prevalent in various machine learning applications. Feature selection is a useful technique for alleviating the curse of dimensionality. Unsupervised feature selection problem tends to be more challenging than its supervised counterpart due to the lack of class labels. State-of-the-art approaches usually use the concept of pseudo labels to select discriminative features by their regression coefficients and the pseudo-labels derived from clustering is usually inaccurate. In this paper, we propose a new perspective for unsupervised feature selection by Discriminatively Exploiting Similarity (DES). Through forming similar and dissimilar data pairs, implicit discriminative information can be exploited. The similar/dissimilar relationship of data pairs can be used as guidance for feature selection. Based on this idea, we propose hypothesis testing based and classification based methods as instantiations of the DES framework. We evaluate the proposed approaches extensively using six real-world datasets. Experimental results demonstrate that our approaches achieve significantly outperforms the state-of-the-art unsupervised methods. More surprisingly, our unsupervised method even achieves performance comparable to a supervised feature selection method.

12:20-12:40 SetExpan: Corpus-Based Set Expansion via Context Feature Selection and Rank Ensemble



Jiaming Shen, Zeqiu Wu, Dongming Lei, Jingbo Shang, Xiang Ren, Jiawei Han

Corpus-based set expansion (i.e., finding the "complete" set of entities belonging to the same semantic class, based on a given corpus and a tiny set of seeds) is a critical task in knowledge discovery. It may facilitate numerous downstream applications, such as information extraction, taxonomy induction, question answering, and web search. To discover new entities in an expanded set, previous approaches either make one-time entity ranking based on distributional similarity, or resort to iterative pattern-based bootstrapping. The core challenge for these methods is how to deal with noisy context features derived from free-text corpora, which may lead to entity intrusion and semantic drifting. In this study, we propose a novel framework, SetExpan, which tackles this problem, with two techniques: (1) a context feature selection method that selects clean context features for calculating entity-entity distributional similarity, and (2) a ranking-based unsupervised ensemble method for expanding entity set based on denoised context features. Experiments on three datasets show that SetExpan is robust and outperforms previous state-of-the-art methods in terms of mean average precision.

11:00 - 11:20

Koopman Spectral kernels for comparing complex dynamics: Application to multiagent sport plays Keisuke Fujii, Yuki Inaba, Yoshinobu Kawahara



Understanding the complex dynamics in the real-world such as in multi-agent behaviors is a challenge in numerous engineering and scientific fields. Spectral analysis using Koopman operators has been attracting attention as a way of obtaining a global modal description of a nonlinear dynamical system, without requiring explicit prior knowledge. However, when applying this to the comparison or classification of complex dynamics, it is necessary to incorporate the Koopman spectra of the dynamics into an appropriate metric. One way of implementing this is to design a kernel that reflects the dynamics via the spectra. In this paper, we introduced Koopman spectral kernels to compare the complex dynamics by generalizing the Binetâ€"Cauchy kernel to nonlinear dynamical systems without specifying an underlying model. We applied this to strategic multiagent sport plays wherein the dynamics can be classified, e.g., by the success or failure of the shot. We mapped the latent dynamic characteristics of multiple attacker–defender distances to the feature space using our kernels and then evaluated the scorability of the play by using the features in different classification models.

Boosting Based Multiple Kernel Learning and Transfer Regression for 11:20 - 11:40 **Electricity Load Forecasting**



Accurate electricity load forecasting is of crucial importance for power system operation and smart grid energy management. Different factors, such as weather conditions, lagged values, and day types may affect electricity load consumption. We propose to use multiple kernel learning (MKL) for electricity load forecasting, as it provides more flexibilities than traditional kernel methods. Computation time is an important issue for short-term load forecasting, especially for energy scheduling demand. However, conventional MKL methods usually lead to complicated optimization problems. Another practical aspect of this application is that there may be very few data available to train a reliable forecasting model for a new building, while at the same time we may have prior knowledge learned from other buildings. In this paper, we propose a boosting based framework for MKL regression to deal with the aforementioned issues for short-term load forecasting. In particular, we first adopt boosting to learn an ensemble of multiple kernel regressors, and then extend this framework to the context of transfer learning. Experimental results on residential data sets show the effectiveness of the proposed algorithms.

11:40 - 12:00 **DC-Prophet: Predicting Catastrophic Machine Failures in DataCenters**



You-Luen Lee, Da-Cheng Juan, Xuan-An Tseng, Yu-Ting Chen, Shih-Chieh Chang

When will a server fail catastrophically in an industrial datacenter? Is it possible to forecast these failures so preventive actions can be taken to increase the reliability of a datacenter? To answer these questions, we have studied what are probably the largest, publicly available datacenter traces, containing more than 104 million events from 12,500 machines. Among these samples, we observe and categorize three types of machine failures, all of which are catastrophic and may lead to information loss, or even worse, reliability degradation of a datacenter. We further propose a two-stage framework "DC-Prophet" based on One-Class Support Vector Machine and Random Forest. DC-Prophet extracts surprising patterns and accurately predicts the next failure of a machine. Experimental results show that DC-Prophet achieves an AUC of 0.93 in predicting the next machine failure, and a F3-score of 0.88 (out of 1). On average, DC-Prophet outperforms other classical machine learning methods by 39.45% in F3-score.

Quantifying Heterogeneous Causal Treatment Effects in World Bank 12:00 - 12:20 **Development Finance Projects**



The World Bank provides billions of dollars in development finance to countries across the world every year. As many projects are related to the environment, we want to understand the World Bank projects impact to forest cover. However, the global extent of these projects results in substantial heterogeneity in impacts due to geographic, cultural, and other factors. Recent research by Athey and Imbens has illustrated the potential for hybrid machine learning and causal inferential techniques which may be able to capture such heterogeneity. We apply their approach using a geolocated dataset of World Bank projects, and augment this data with satellite-retrieved characteristics of their geo- graphic context (including temperature, precipitation, slope, distance to urban areas, and many others). We use this information in conjunction with causal tree (CT) and causal forest (CF) approaches to contrast 'control' and 'treatment' geographic locations to estimate the impact of World Bank projects on vegetative cover.

WEDNESDAY

WEDNESDAY SEPTEMBER 20

12:20 - 12:40

Taking It for a Test Drive: A Hybrid Spatio-temporal Model for Wildlife Poaching Prediction Evaluated through a Controlled Field Test

Shahrzad Gholami, Benjamin Ford, Fei Fang, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Mustapha Nsubaga, Joshua Mabonga

Worldwide, conservation agencies employ rangers to protect conservation areas from poachers. However, agencies lack the manpower to have rangers effectively patrol these vast areas frequently. While past work has modeled poachers' behavior so as to aid rangers in planning future patrols, those models' predictions were not validated by extensive field tests. In this paper, we present a hybrid spatio-temporal model that predicts poaching threat levels and results from a five-month field test of our model in Uganda's Queen Elizabeth Protected Area (QEPA). To our knowledge, this is the first time that a predictive model has been evaluated through such an extensive field test in this domain. We present two major contributions. First, our hybrid model consists of two components: (i) an ensemble model which can work with the limited data common to this domain and (ii) a spatio-temporal model to boost the ensemble's predictions when sufficient data are available. When evaluated on real-world historical data from QEPA, our hybrid model achieves significantly better performance than previous approaches with either temporally-aware dynamic Bayesian networks or an ensemble of spatially-aware models. Second, in collaboration with the Wildlife Conservation Society and Uganda Wildlife Authority, we present results from a five-month controlled experiment, where rangers patrolled over 450 sq km across QEPA. We demonstrate that our model successfully predicted (1) where snaring activity would occur and (2) where it would not occur; in areas where we predicted a high rate of snaring activity, rangers found more snares than in areas of lower predicted activity. These findings demonstrate that (1) our model's predictions are selective, (2) our model's superior laboratory performance extends to the real world, and (3) these predictive models can aid rangers in focusing their efforts to prevent wildlife poaching and save animals.

Neural Networks & Deep Learning 2

14:00-14:20 CON-S2V: A Generic Framework for Incorporating Extra-Sentential Context Into Sen2Vec

Tanay Kumar Saha, Shafiq Joty, Mohammad Al Hasan

We present a novel approach to learning distributed representation of sentences from unlabeled data by modeling content and context of a sentence. The content model learns sentence representation by predicting its words. The context model comprises a neighbor prediction component and a regularizer to model distributional and proximity hypotheses, respectively. We propose an online algorithm to train the model components jointly. For the first time, we evaluate the models in a setup, where contextual information is available to infer the sentence vectors. The experimental results on tasks involving classifying, clustering, and ranking sentences show that our model outperforms best existing models by a wide margin across multiple datasets.

14:20-14:40 Deep Over-Sampling Framework for Classifying Imbalanced Data

Shin Ando, Chun Yuan Huang

Class imbalance is a challenging issue in practical classification problems for deep learning models as well as traditional models. Traditionally successful countermeasures such as synthetic over-sampling have had limited success with complex, structured data handled by deep learning models. In this paper, we propose Deep Over-sampling (DOS), a framework for extending the synthetic over-sampling method to the deep feature space acquired by a convolutional neural network (CNN). Its key feature is an explicit, supervised representation learning, for which the training data presents each raw input sample with a synthetic embedding target in the deep feature space, which is sampled from the linear subspace of in-class neighbors. We implement an iterative process of training the CNN and updating the targets, which induces smaller in-class variance among the embeddings, to increase the discriminative power of the deep representation. We present an empirical study using public benchmarks, which shows that the DOS framework not only counteracts class imbalance better than the existing method, but also improves the performance of the CNN in the standard, balanced settings.

84





Joint User Modeling Across Aligned Heterogeneous Sites using

Neural Networks Xuezhi Cao, Yong Yu

14:40 - 15:00

The quality of user modeling is crucial for personalized recommender systems. Traditional within-site recommender systems aim at modeling user preferences using only actions within target site, thus suffer from cold-start problem. To alleviate such problem, researchers propose cross-domain models to leverage user actions from other domains within same site. Joint user modeling is later proposed to further integrate user actions from aligned sites for data enrichment. However, there are still limitations in existing works regarding the modeling of heterogeneous actions, the requirement of full alignment and the design of preferences coupling. To tackle these, we propose JUN: a joint user modeling framework using neural network. We take advantage of neural network's capability of capturing different data types and its ability for mining high-level non-linear correlations. Specifically, in additional to site-specific preferences models, we further introduce an auxiliary neural network to transfer knowledge between sites by fine-tuning the user embeddings using alignment information. We adopt JUN for item-based and text-based site to demonstrate its performance. Experimental results indicate that JUN outperforms both within-site and cross-site models. Specifically, JUN achieves relative improvement of 2.96% and 2.37% for item-based and text-based sites (5.77% and 13.54% for cold-start scenarios). Besides performance gain, JUN also achieves great generality and significantly extends the use scenarios.

15:00 - 15:20 Learning Deep Kernels in the Space of Dot Product Polynomials

Michele Donini, Fabio Aiolli

Recent literature has shown the merits of having deep representations in the context of neural networks. An emerging challenge in kernel learning is the definition of similar deep representations. In this paper, we propose a general methodology to define a hierarchy of base kernels with increasing expressiveness and combine them via multiple kernel learning (MKL) with the aim to generate overall deeper kernels. As a leading example, this methodology is applied to learning the kernel in the space of Dot-Product Polynomials (DPPs), that is a positive combination of homogeneous polynomial kernels (HPKs). We show theoretical properties about the expressiveness of HPKs that make their combination empirically very effective. This can also be seen as learning the coefficients of the Maclaurin expansion of any definite positive dot product kernel thus making our proposed method generally applicable. We empirically show the merits of our approach comparing the effectiveness of the kernel generated by our method against baseline kernels (including homogeneous and non homogeneous polynomials, RBF, etc...) and against another hierarchical approach on several benchmark datasets.

15:20-15:40 Wikipedia Vandal Early Detection: from User Behavior to User Embedding

Shuhan Yuan, Panpan Zheng, Xintao Wu, Yang Xiang

Wikipedia is the largest online encyclopedia that allows anyone to edit articles. In this paper, we propose the use of deep learning to detect vandals based on their edit history. In particular, we develop a multi-source long-short term memory network (M-LSTM) to model user behaviors by using a variety of user edit aspects as inputs, including the history of edit reversion information, edit page titles and categories. With M-LSTM, we can encode each user into a low dimensional real vector, called user embedding. Meanwhile, as a sequential model, M-LSTM updates the user embedding each time after the user commits a new edit. Thus, we can predict whether a user is benign or vandal dynamically based on the up-to-date user embedding. Furthermore, those user embeddings are crucial to discover collaborative vandals.

Kernel Methods 1

14:00 - 14:20 Fair Kernel Learning

Adrian Perez-Suay, Valero Laparra, Gonzalo Mateo-García, Jordi Muñoz-Marí, Luis Gómez-Chova, Gustau Camps-Valls

New social and economic activities massively exploit big data and machine learning algorithms to do inference on people's lives. Applications include automatic curricula evaluation, wage determination, and risk assessment for credits and loans. Recently, many governments and institutions have raised concerns about the lack of fairness, equity and ethics in machine learning to treat these problems. It has been shown that not including sensitive features that bias fairness, such as gender or race, is not enough to mitigate the discrimination when other related features are included. Instead, including fairness in the objective function has been shown to be more efficient. We present novel fair regression and dimensionality reduction methods built on a previously proposed fair classification framework. Both methods rely on using the Hilbert Schmidt independence criterion as the fairness term. Unlike previous approaches, this allows us to simplify the problem and to use multiple sensitive variables simultaneously. Replacing the linear formulation by kernel functions allows the methods to deal with nonlinear problems. This simplifies the evaluation of the solutions for different trade-off values between the predictive error and fairness terms. We illustrate the usefulness of the proposed methods in toy examples, and evaluate their performance on real world datasets to predict income using gender and/or race discrimination as sensitive variables, and contraceptive method prediction under demographic and socio-economic sensitive descriptors.







WEDNESDAY **SEPTEMBER 20**

14:20 - 14:40 GaKCo: A Fast Gapped k-mer String Kernel using Counting

Ritambhara Singh, Arshdeep Sekhon, Andrew Norton, Jack Lanchantin, Kamran Kowsari, Beilun Wang, Yanjun Qi

String Kernel (SK) techniques, especially those using gapped k-mers as features (gk), have obtained great success in classifying sequences like DNA, protein, and text. However, the state-of-the-art gk-SK runs extremely slow when we increase the dictionary size or allow more number of mismatches. This is because current gk-SK uses a trie-based algorithm to calculate co-occurrence of mismatched substrings resulting in a time cost proportional to dictionary size and mismatches. We propose a fast algorithm for calculating Gapped k-mer Kernel using counting (GaKCo). GaKCo uses associative arrays to calculate the co-occurrence of substrings using cumulative counting. This algorithm is fast, scalable to larger dictionary and mismatches, and naturally parallelizable. We provide a rigorous asymptotic analysis that compares GaKCo with the state-of-the-art gk-SK. Theoretically, the time cost of GaKCo is independent of the dictionary term that slows down the trie-based approach. Experimentally, we observe that GaKCo achieves the same accuracy as the state-of-the-art and outperforms its speed by factors of 2, 100, and 4, on classifying sequences of DNA (5 datasets), protein (12 datasets), and character-based English text (2 datasets).

Graph Enhanced Memory Networks for Sentiment Analysis 14:40 - 15:00

Zhao Xu, Romain Vial, Kristian Kersting

Memory networks model information and knowledge as memories that can be manipulated for prediction and reasoning about questions of interest. In many cases, there exists complicated relational structure in the data, by which the memories can be linked together into graphs to propagate information. Typical examples include tree structure of a sentence and knowledge graph in a dialogue system. In this paper, we present a novel graph enhanced memory network GEMN to integrate relational information between memories for prediction and reasoning. Our approach introduces graph attentions to model the relations, and couples them with content-based attentions via an additional neural network layer. It thus can better identify and manipulate the memories related to a given question, and provides more accurate prediction about the final response. We demonstrate the effectiveness of the proposed approach with aspect based sentiment classification. The empirical analysis on real data shows the advantages of incorporating relational dependencies into the memory networks.

15:00 - 15:20 Kernel Sequential Monte Carlo

Ingmar Schuster, Heiko Strathmann, Brooks Paige, Dino Sejdinovic

We propose kernel sequential Monte Carlo (KSMC), a framework for sampling from static target densities. KSMC is a family of sequential Monte Carlo (SMC) algorithms that are based on building emulator models of the current particle system in a reproducing kernel Hilbert space. We here focus on modelling nonlinear covariance structure and gradients of the target. The emulator's geometry is adaptively updated and subsequently used to inform local proposals. Unlike in adaptive Markov chain Monte Carlo (MCMC), continuous adaptation does not compromise convergence of the sampler. KSMC combines the strengths of SMC and kernel methods: superior performance for multimodal targets and the ability to estimate model evidence as compared to MCMC, and the emulator's ability to represent targets that exhibit high degrees of nonlinearity. As KSMC does not require access to target gradients, it is particularly applicable on targets whose gradients are unknown or prohibitively expensive. We describe necessary tuning details and demonstrate the benefits of the the proposed methodology on a series of challenging synthetic and real-world examples.

15:20 - 15:40 Learning Lukasiewicz Logic Fragments By Quadratic Programming

Francesco Giannini, Michelangelo Diligenti, Marco Gori, Marco Maggini

In this paper we provide a framework to embed logical constraints into the classical learning scheme of kernel machines, that gives rise to a learning algorithm based on a quadratic programming problem. In particular, we show that, once the constraints are expressed using a specific fragment from the Lukasiewicz logic, the learning objective turns out to be convex. We formulate the primal and dual forms of a general multitask learning problem, where the functions to be determined are predicates (of any arity) defined on the feature space. The learning set contains both supervised examples for each predicate and unsupervised examples exploited to enforce the constraints. We give some properties of the solutions constructed by the framework along with promising experimental results.





15:20 - 15:40

Recent Advances in Kernel-Based Graph Classification 14:00 - 14:20

Nils Kriege, Christopher Morris

We review our recent progress in the development of graph kernels. We discuss the hash graph kernel framework, which makes the computation of kernels for graphs with vertices and edges annotated with real-valued information feasible for large data sets. Moreover, we summarize our general investigation of the benefits of explicit graph feature maps in comparison to using the kernel trick. Our experimental studies on real-world data sets suggest that explicit feature maps often provide sufficient classification accuracy while being computed more efficiently. Finally, we describe how to construct valid kernels from optimal assignments to obtain new expressive graph kernels. These make use of the kernel trick to establish one-to-one correspondences. We conclude by a discussion of our results and their implication for the future development of graph kernels.

14.20 - 14.40 Activity-Driven Influence Maximization in Social Networks

Rohit Kumar, Muhammad Saleem, Toon Calders, Torben Pedersen, Xike Xie

Interaction networks consist of a static graph with a timestamped list of edges over which interaction took place. Examples of interaction networks are social networks whose users interact with each other through messages or location-based social networks where people interact by checking in to locations. Previous work on finding influential nodes in such networks mainly concentrate on the static structure imposed by the interactions or are based on fixed models for which parameters are learned using the interactions. In two recent works, however, we proposed an alternative activity data-driven approach based on the identification of influence propagation patterns. In the first work, we identify so-called information-channels to model potential pathways for information spread, while the second work exploits how users in a location-based social network check-in to locations in order to identify influential locations. To make our algorithms scalable, approximate versions based on sketching techniques from the data streams domain have been developed. Experiments show that in this way it is possible to efficiently find good seed sets for influence propagation in social networks.

14:40 - 15:00 QuickScorer: Efficient Traversal of Large Ensembles of Decision Trees

Claudio Lucchese, Franco Maria Nardini, Salvatore Orlando, Raffaele Perego, Nicola Tonellotto, Rossano Venturini

Machine-learnt models based on additive ensembles of binary regression trees are currently deemed the best solution to address complex classification, regression, and ranking tasks. Evaluating these models is a computationally demanding task as it needs to traverse thousands of trees with hundreds of nodes each. The cost of traversing such large forests of trees significantly impacts their application to big and stream input data, when the time budget available for each prediction is limited to guarantee a given processing throughput. Document ranking in Web search is a typical example of this challenging scenario, where the exploitation of tree-based models to score query-document pairs, and finally rank lists of documents for each incoming query, is the state-of-art method for ranking (a.k.a. Learning-to-Rank). This paper presents QuickScorer, a novel algorithm for the traversal of huge decision trees ensembles that, thanks to a cache- and CPU-aware design, provides a ~9x speedup over best competitors.

An Al Planning System for Data Cleaning 15:00 - 15:20

Roberto Boselli, Mirko Cesarini, Fabio Mercorio, Mario Mezzanzanica

Data Cleaning represents a crucial and error prone activity in KDD that might have unpredictable effects on data analytics, affecting the believability of the whole KDD process. In this paper we describe how a bridge between AI Planning and Data Quality communities has been made, by expressing both the data quality and cleaning tasks in terms of AI planning. We also report a real-life application of our approach

> Image Representation, Annotation and Retrieval with Predictive **Clustering Trees**

Ivica Dimitrovski, Dragi Kocev, Suzana Loskovska, Sašo Džeroski

In this work, we summarize our work on using the predictive clustering framework for image analysis. More specifically, we used predictive clustering trees to generate image representations, that can then be used to perform image retrieval and/or image annotation. We evaluated the proposed method for performing image retrieval on general purpose images, and annotation of general purpose images, medical images and diatom images.



Applied Data Science 4

14:00-14:20 Analyzing Granger Causality in Climate Data with Time Series Classification Methods



Christina Papagiannopoulou, Stijn Decubber, Diego Miralles, Matthias Demuzere, Niko Verhoest, Willem Waegeman

Attribution studies in climate science aim for scientifically ascertaining the influence of climatic variations on natural or anthropogenic factors. Many of those studies adopt the concept of Granger causality to infer statistical cause-effect relationships, while utilizing traditional autoregressive models. In this article, we investigate the potential of state-of-the-art time series classification techniques to enhance causal inference in climate science. We conduct a comparative experimental study of different types of algorithms on a large test suite that comprises a unique collection of datasets from the area of climate-vegetation dynamics. The results indicate that specialized time series classification methods are able to improve existing inference procedures. Substantial differences are observed among the methods that were tested.

14:20-14:40 CREST - Risk Prediction for Clostridium Difficile Infection Using Multimodal Data Mining



Cansu Sen, Thomas Hartvigsen, Elke Rundensteiner, Kajal Claypool

Clostridium difficile infection (CDI) is a common hospital acquired infection with a \$1B annual price tag that resulted in ~30,000 deaths in 2011. Studies have shown that early detection of CDI significantly improves the prognosis for the individual patient and reduces the overall mortality rates and associated medical costs. In this paper, we present CREST: CDI Risk Estimation, a data-driven framework for early and continuous detection of CDI in hospitalized patients. CREST uses a three-pronged approach for high accuracy risk prediction. First, CREST builds a rich set of highly predictive features from Electronic Health Records. These features include clinical and non-clinical phenotypes, key biomarkers from the patient's laboratory tests, synopsis features processed from time series vital signs, and medical history mined from clinical notes. Given the inherent multimodality of clinical data, CREST bins these features into three sets: time-invariant, time-variant, and temporal synopsis features. CREST then learns classifiers for each set of features, evaluating their relative effectiveness. Lastly, CREST employs a second-order meta learning process to ensemble these classifiers for optimized estimation of the risk scores. We evaluate the CREST framework using publicly available critical care data collected for over 12 years from Beth Israel Deaconess Medical Center, Boston. Our results demonstrate that CREST predicts the probability of a patient acquiring CDI with an AUC of 0.76 five days prior to diagnosis. This value increases to 0.80 and even 0.82 for prediction two days and one day prior to diagnosis, respectively.

Networks & Graphs 2

16:00-16:20 Regularizing Neural Knowledge Graph Embeddings via Equivalence and Inversion Axioms



Pasquale Minervini, Luca Costabello, Emir Munoz, Vit Novacek, Pierre-Yves Vandenbussche

Learning embeddings of entities and relations using neural architectures is an effective method of performing statistical learning on largescale relational data, such as knowledge graphs. In this paper, we consider the problem of regularizing the training of neural knowledge graph embeddings by leveraging external background knowledge. We propose a principled and scalable method for leveraging equivalence and inversion axioms during the learning process, by imposing a set of model-dependent soft constraints on the predicate embeddings. The method has several advantages: i) the number of introduced constraints does not depend on the number of entities in the knowledge base; ii) regularities in the embedding space effectively reflect available background knowledge; iii) it yields more accurate results in link prediction tasks over non-regularized methods; and iv) it can be adapted to a variety of models, without affecting their scalability properties. We demonstrate the effectiveness of the proposed method on several large knowledge graphs. Our evaluation shows that it consistently improves the predictive accuracy of several neural knowledge graph embedding models (for instance,the MRR of TransE on WordNet increases by 11%) without compromising their scalability properties.

16:20-16:40 The Network-untangling Problem: From Interactions to Activity Timelines



Polina Rozenstein, Nikolaj Tatti, Aristides Gionis

In this paper we study a problem of determining when entities are active based on their interactions with each other. More formally, we consider a set of entities V and a sequence of time-stamped edges E among the entities. Each edge (u,v,t) in E denotes an interaction between entities u and v that takes place at time t. We view this input as a temporal network. We then assume a simple activity model in which each entity is active during a short time interval. An interaction (u,v,t) can be explained if at least one of u or v are active at time t. Our goal is to reconstruct the activity intervals, for all entities in the network, so as to explain the observed interactions. This problem, which we refer to as the network-untangling problem, can be applied to discover timelines of events from complex interactions among entities. We provide two formulations for the networkuntangling problem: (i) minimizing the total interval length over all entities, and (ii) minimizing the maximum interval length. We show that the sum problem is NP-hard, while, surprisingly, the max problem can be solved optimally in linear time, using a mapping to 2-SAT. For the sum problem we provide efficient and effective algorithms based on realistic assumptions. Furthermore, we complement our study with an extensive evaluation on synthetic and real-world datasets, which demonstrates the validity of our concepts and the good performance of our algorithms.

16:40-17:00 TransT: Type-Based Multiple Embedding Representations for Knowledge Graph Completion

Shiheng Ma, Jianhui Ding, Weijia Jia, Kun Wang, Minyi Guo

Knowledge graph completion with representation learning predicts new entity-relation triples from the existing knowledge graphs by embedding entities and relations into a vector space. Most existing methods focus on the structured information of triples and maximize the likelihood of them. However, they neglect semantic information contained in most knowledge graphs and the prior knowledge indicated by the semantic information. To overcome this drawback, we propose an approach that integrates the structured information and entity types which describe the categories of entities. Our approach constructs relation types from entity types and utilizes type-based semantic similarity of the related entities and relations to capture prior distributions of entities and relations. With the type-based prior distributions, our approach generates multiple embedding representations of each entity in different contexts and estimates the posterior probability of entity and relation prediction. Extensive experiments show that our approach outperforms previous semantics-based methods.

Kernel Methods 2

CONGRESS HALL 2

16:00 - 16:20 Bayesian Nonlinear Support Vector Machines for Big Data

Florian Wenzel, Theo Galy-Fajou, Matthäus Deutsch, Marius Kloft

We propose a fast inference method for Bayesian nonlinear support vector machines that leverages stochastic variational inference and inducing points. Our experiments show that the proposed method is faster than competing Bayesian approaches and scales easily to millions of data points. It provides additional features over frequentist competitors such as accurate predictive uncertainty estimates and automatic hyperparameter search.

WEDNESDAY **SEPTEMBER 20**

16:20 - 16:40 Entropic Trace Estimation for Log Determinants

Jack Fitzsimons, Diego Granziol, Kurt Cutajar, Michael Osborne, Maurizio Filippone, Stephen Roberts

The scalable calculation of matrix determinants has been a bottleneck to the widespread application of many machine learning methods such as determinantal point processes, Gaussian processes, generalised Markov random fields, graph models and many others. In this work, we estimate log determinants under the framework of maximum entropy, given information in the form of moment constraints from stochastic trace estimation. The estimates demonstrate a significant improvement on state-of-the-art alternative methods, as shown on a wide variety of matrices from the SparseSuite Matrix Collection. By taking the example of a general Markov random field, we also demonstrate how this approach can significantly accelerate inference in large-scale learning methods involving the log determinant.

16:40 - 17:00 Nystrom Sketching

Daniel Perry, Braxton Osting, Ross Whitaker

Despite prolific success, kernel methods become difficult to use in many large-scale unsupervised problems because of the evaluation and storage of the full Gram matrix. Here we overcome this difficulty by proposing a novel approach: compute the optimal small, out-of-sample Nystrom sketch which allows for fast approximation of the Gram matrix via the Nystrom method. We demonstrate and compare several methods for computing the optimal Nystrom sketch and show how this approach outperforms previous state-of-the-art Nystrom subset-based methods of similar size.

Learning & Optimization 2

A Constrained I1 minimization approach for estimating multiple Sparse 16:00 - 16:20 **Gaussian or Nonparanormal Graphical Models** Beilun Wang, Ritambhara Singh, Yanjun Qi

Identifying context-specific entity networks from aggregated data is an important task, arising often in bioinformatics and neuroimaging applications. Computationally, this task can be formulated as jointly estimating multiple different, but related, sparse undirected graphical models (UGM) from aggregated samples across several contexts. Previous joint-UGM studies have mostly focused on sparse Gaussian graphical models (sGGMs) and can't identify context-specific edge patterns directly. We, therefore, propose a novel approach, SIMULE (detecting Shared and Individual parts of MULtiple graphs Explicitly) to learn multi-UGM via a constrained I1 minimization. SIMULE automatically infers both specific edge patterns that are unique to each context and shared interactions preserved among all the contexts. Through the I1 constrained formulation, this problem is cast as multiple independent subtasks of linear programming that can be solved efficiently in parallel. In addition to Gaussian data, SIMULE can also handle multivariate Nonparanormal data that greatly relaxes the normality assumption that many real-world applications do not follow. We provide a novel theoretical proof showing that SIMULE achieves a consistent result at the rate O(log(Kp)/ntot). On multiple synthetic datasets and two biomedical datasets, SIMULE shows significant improvement over state-of-the-art multi-sGGM and single-UGM baselines (SIMULE implementation and the used datasets @https://github.com/QData/SIMULE).

Distributed Stochastic Optimization of the Regularized Risk via Saddle-16:20 - 16:40 **Point Problem**

Shin Matsushima, Hyokun Yun, Xinhua Zhang, S.V.N. Vishwanathan

Many machine learning algorithms minimize a regularized risk, and stochastic optimization is widely used for this task. When working with massive data, it is desirable to perform stochastic optimization in parallel. Unfortunately, many existing stochastic optimization algorithms cannot be parallelized efficiently. In this paper we show that one can rewrite the regularized risk minimization problem as an equivalent saddle-point problem, and propose an efficient distributed stochastic optimization (DSO) algorithm. We prove the algorithm's rate of convergence; remarkably, our analysis shows that the algorithm scales almost linearly with the number of processors. We also verify with empirical evaluations that the proposed algorithm is competitive with other parallel, general purpose stochastic and batch optimization algorithms for regularized risk minimization.









16:40 - 17:00 Sparse Probit Linear Mixed Model

Stephan Mandt, Florian Wenzel, Shinichi Nakajima, John Cunningham, Christoph Lippert, Marius Kloft

Linear mixed models (LMMs) are important tools in statistical genetics. When used for feature selection, they allow to find a sparse set of genetic traits that best predict a continuous phenotype of interest, while simultaneously correcting for various confounding factors such as age, ethnicity and population structure. Formulated as models for linear regression, LMMs have been restricted to continuous phenotypes. We introduce the sparse probit linear mixed model (Probit-LMM), where we generalize the LMM modeling paradigm to binary phenotypes. As a technical challenge, the model no longer possesses a closed-form likelihood function. In this paper, we present a scalable approximate inference algorithm that lets us fit the model to high-dimensional data sets. We show on three real-world examples from different domains that in the setup of binary labels, our algorithm leads to better prediction accuracies and also selects features which show less correlation with the confounding factors.

Subgroup Discovery

CONGRESS HALL 4

16:00 - 16:20 Explaining Deviating Subsets Through Explanation Networks

Antti Ukkonen, Vladimir Dzyuba, Matthijs Van Leeuwen

We propose a novel approach to finding explanations of deviating subsets, often called subgroups. Existing approaches for subgroup discovery rely on various quality measures that nonetheless often fail to find subgroup sets that are diverse, of high quality, and most importantly, provide good explanations of the deviations that occur in the data. To tackle this issue we introduce explanation networks, which provide a holistic view on all candidate subgroups and how they relate to each other, offering elegant ways to select high-quality yet diverse subgroup sets. Explanation networks are constructed by representing subgroups by nodes and having weighted edges represent the extent to which one subgroup explains another. Explanatory strength is defined by extending ideas from database causality, in which interventions are used to quantify the effect of one query on another. Given an explanatory network, existing network analysis techniques can be used for subgroup discovery. In particular, we study the use of PageRank for pattern ranking and seed selection (from influence maximization) for pattern set selection. Experiments on synthetic and real data show that the proposed approach finds subgroup sets that are more likely to capture the generative processes of the data than other methods.

Flash Points: Discovering exceptional pairwise behaviors in Vote or

16:20 - 16:40

Rating Data Adnene Belfodil, Sylvie Cazalens, Philippe Lamarre, Marc Plantevit

We address the problem of discovering contexts that lead well-distinguished collections of individuals to change their pairwise agreement w.r.t. their usual one. For instance, in the European parliament, while in overall, a strong disagreement is witnessed between deputies of the far-right French party Front National and deputies of the left party Front de Gauche, a strong agreement is observed between these deputies in votes related to the thematic: External relations with the union. We devise the method DSC (Discovering Similarities Changes) which relies on exceptional model mining to uncover three-set patterns that identify contexts and two collections of individuals where an unex- pected strengthening or weakening of pairwise agreement is observed. To efficiently explore the search space, we define some closure operators and pruning techniques using upper bounds on the quality measure. In addition of handling usual attributes (e.g. numerical, nominal), we propose a novel pattern domain which involves hierarchical multi-tag attributes that are present in many datasets. A thorough empirical study on two real-world datasets (i.e., European parliament votes and collaborative movie reviews) demonstrates the efficiency and the effectiveness of our approach as well as the interest and the actionability of the patterns.

16:40-17:00 Identifying Consistent Statements About Numerical Data with Dispersion-Corrected Subgroup Discovery

Mario Boley, Bryan R. Goldsmith, Luca M. Ghiringhelli, Jilles Vreeken

Existing algorithms for subgroup discovery with numerical targets do not optimize the error or target variable dispersion of the groups they find. This often leads to unreliable or inconsistent statements about the data, rendering practical applications, especially in scientific domains, futile. Therefore, we here extend the optimistic estimator framework for optimal subgroup discovery to a new class of objective functions: we show how tight estimators can be computed efficiently for all functions that are determined by subgroup size (non-decreasing dependence), the subgroup median value, and a dispersion measure around the median (non-increasing dependence). In the important special case when dispersion is measured using the mean absolute deviation from the median, this novel approach yields a linear time algorithm. Empirical evaluation on a wide range of datasets shows that, when used within branch-and-bound search, this approach is highly efficient and indeed discovers subgroups with much smaller errors.





THURSDAY

PROGRAM AT A GLANCE



MACEDONIAN OPERA & BALLET



INVITED TALK

CONGRESS HALL 1

09:00 - 10:00

Enabling a Smarter Planet with Earth Observation

Pierre-Philippe Mathieu

ESA/ESRIN, EO Science, Applications and New Technologies, Italy

Nowadays, teams of researchers around the world can easily access a wide range of open data across disciplines and remotely process them on the Cloud, combining them with their own data to generate knowledge, develop information products for societal applications, and tackle complex integrative complex problems that could not be addressed a few years ago. Such rapid exchange of digital data is fostering a new world of data-intensive research, characterized by openness, transparency, and scrutiny and traceability of results, access to large volume of complex data, availability of community open tools, unprecedented level of computing power, and new collaboration among researchers and new actors such as citizen scientists. The EO scientific community is now facing the challenge of responding to this new paradigm in science 2.0 in order to make the most of the large volume of complex and diverse data delivered by the new generation of EO missions, and in particular the Sentinels. In this context, ESA is supporting a variety of activities in partnership with research communities to ease the transition and make the most of the data. These include the generation of new open tools and exploitation platforms, exploring new ways to disseminate data, building new partnership with citizen scientists, and training the new generation of data scientists. The talk will give a brief overview of some of ESA activities aiming to facilitate the exploitation of large amounts of data from EO missions in a collaborative, cross-disciplinary, and open way, for uses ranging from science to applications and education.

Chair: Ljupčo Todorovski

Sponsored by



MACEDONIAN OPERA & BALLET

NVITED TALK

19:00 - 20:00

Automatic Understanding of the Visual World Cordelia Schmid

INRIA, France

One of the central problems of artificial intelligence is machine perception, i.e., the ability to understand the visual world based on input from sensors, such as cameras. Computer vision is the area which analyzes visual input. In this talk, I will present recent progress in visual understanding. It is for the most part due to the design of robust visual representations and learned models capturing the variability of the visual world based on state-of-the-art machine learning techniques, including convolutional neural networks. Progress has resulted in technology for a variety of applications. I will present in particular results for human action recognition.

Chair: Ivica Dimitrovski

CONGRESS HALL 1

Learning and Scaling Directed Networks via **Graph Embedding**

Mikhail Drobyshevskiy, Anton Korshunov & Denis Turdakov

Reliable evaluation of network mining tools implies significance and scalability testing. This is usually achieved by picking several graphs of various size from different domains. However, graph properties and thus evaluation results could be dramatically different from one domain to another. Hence the necessity of aggregating results over a multitude of graphs within each domain. The paper introduces an approach to automatically learn features of a directed graph from any domain and generate similar graphs while scaling input graph size with a real-valued factor. Generating multiple graphs with similar size allows significance testing, while scaling graph size makes scalability evaluation possible. The proposed method relies on embedding an input graph into low-dimensional space, thus encoding graph features in a set of node vectors. Edge weights and node communities could be imitated as well in optional steps. We demonstrate that embedding-based approach ensures variability of synthetic graphs while keeping degree and subgraphs distributions close to the original graphs. Therefore, the method could make significance and scalability testing of network algorithms more reliable without the need to collect additional data. We also show that embeddingbased approach preserves various features in generated graphs which can't be achieved by other generators imitating a given graph.

Chair: Ljupčo Todorovski

Sponsored by

ita Mining and



SESSIONS AT A GLANCE

Networks & Graphs 3

CONGRESS HALL 1

11:00-12:40

Chair: Toon Calders Graph-Based Predictable Feature Analysis 11:00 - 11:20 Björn Weghenkel, Asja Fischer, Laurenz Wiskott Lagrangian Relaxations for Multiple Network Alignment 11:20 - 11:40 Eric Malmi, Sanjay Chawla, Aristides Gionis Measuring and Moderating Opinion Polarization in Social Networks 11:40 - 12:00 Antonis Matakos, Evimaria Terzi, Panayiotis Tsaparas Micro-Review Synthesis for Multi-Entity Summarization 12:00 - 12:20 Thanh-Son Nguyen, Hady W. Lauw, Panayiotis Tsaparas Survival Factorization for Topical Cascades on Diffusion Networks 12:20 - 12:40 Nicola Barbieri, Giuseppe Manco, Ettore Ritacco

Unsupervised & Semi-supervised Learning 2

CONGRESS HALL 2

11:00-12:40

Chair: Ulf Brefeld

11.00 - 11.20	An Exponential Family Framework For Learning to Predict Unseen Classes
11.00 11.20	Vinay Verma, Wenlin Wang, Piyush Rai
11.00 11.40	An Expressive Similarity Measure for Relational Clustering Using Neighbourhood Trees
11:20 - 11:40	Sebastijan Dumančić, Hendrik Blockeel
11.40 - 19.00	DeepCluster: A General Clustering Framework Based on Deep Learning
11.40 - 12.00	Kai Tian, Shuigeng Zhou, Jihong Guan
10.00 - 10.00	Local PurTree Subspace Spectral Clustering for Customer Transaction Data - NO SHOW
12.00 - 12.20	Xiaojun Chen, JianZhe Zhang, Wenya Sun, Joshua Huang, Qingyao Wu
10.00 10.40	Multi-View Spectral Clustering on Conflicting Views
12.20 - 12:40	Xiao He, Limin Li, Damian Roqueiro, Karsten Borgwardt

Anomaly Detection

Chair: Matthijs van Leeuwen II:UU-12:40

11.00 11.00	Concentration Free Outlier Detection
11:00 - 11:20	Fabrizio Angiulli
11.00 - 11.40	Efficient Top Rank Optimization with Gradient Boosting for Supervised Anomaly Detection
11.20 - 11.40	Jordan Frery, Marc Sebban, Amaury Habrard, Olivier Caelen, Liyun Guelton
11.40 10.00	Robust, Deep and Inductive Anomaly Detection
11.40 - 12.00	Raghavendra Chalapathy, Aditya Krishna Menon, Sanjay Chawla
10.00 - 10.00	Sentiment Informed Cyberbullying Detection in Social Media
12.00 - 12.20	Harsh Dani, Jundong Li, Huan Liu
10.00 10.40	ZooRank: Ranking Suspicious Activities in Time-Evolving Tensors
12.20 - 12.40	Hemank Lamba, Bryan Hooi, Kijung Shin, Christos Faloutsos, Juergen Pfeffer

Applied Data Science 5 Chair: Kamalika Das

11:00 - 11:20	A Novel Framework for Online Sales Burst Prediction
	Rui Chen, Jiajun Liu
11:20 - 11:40	MRNet-Product2Vec: A Multi-Task Recurrent Neural Network for Product Embeddings
	Arijit Biswas, Mukul Bhutani, Subhajit Sanyal
11:40 - 12:00	Disjoint-Support Factors and Seasonality Estimation in E-Commerce
	Abhay Jha
12:00 - 12:20	Generalising Random Forest Parameter Optimisation to Include Stability and Cost
	Chak Hin Bryan Liu, Benjamin Chamberlain, Duncan Little, Angelo Cardoso
12:20 - 12:40	Session-Based Fraud Detection in Online E-Commerce Transactions Using Recurrent Neural Networks
	Shuhao Wang, Cancheng Liu, Xiang Gao, Hongtao Qu, Wei Xu

Recommendati	ON CONGRESS HALL I
Chair: Myra Spilic	opoulu 14:00-15:40
14:00 - 14:20	A Regularization Method with Inference of Trust and Distrust in Recommender Systems
	Dimitrios Rafailidis, Fabio Crestani
14.20 - 14.40	A Unified Contextual Bandit Framework for Long- and Short-Term Recommendations
14.20 - 14.40	Maryam Tavakol, Ulf Brefeld
14:40 - 15:00	Perceiving the Next Choice with Comprehensive Transaction Embeddings for Online Recommendation
14.40 - 15.00	Shoujin Wang, Liang Hu, Longbing Cao, Xiaoshui Huang
15.00 - 15.20	Social Regularized Von Mises-Fisher Mixture Model for Item Recommendation
10.00 - 10.20	Aghiles Salah, Mohamed Nadif
15:20 - 15:40	Tour Recommendation for Groups

Aris Anagnostopoulos, Reem Atassi, Luca Becchetti, Adriano Fazzone, Fabrizio Silvestri

00100000 11411 4

Transfer & Multi-Task Learning 1

Chair: Gianvito Pio I4:UU-15:4U

CO	NG	RE	SS	5 H	AL	L	2
		1	4:	00	-15	:4	0

CONGRESS HALL 3

CONGRESS HALL 4

14:00-15:40

14.00 - 14.20	Lifelong Machine Learning with Gaussian Processes
14.00 14.20	Christopher Clingerman, Eric Eaton
14.20 - 14.40	Personalized Tag Recommendation for Images Using Deep Transfer Learning
14.20 - 14.40	Hanh T.H. Nguyen, Martin Wistuba, Lars Schmidt-Thieme
14:40 - 15:00	Theoretical Analysis of Domain Adaptation with Optimal Transport
14.40 - 15.00	levgen Redko, Amaury Habrard, Marc Sebban
15.00 - 15.20	TSP: Learning Task-Specific Pivots for Unsupervised Domain Adaptation
13.00 - 15.20	Xia Cui, Frans Coenen, Danushka Bollegala
15:00 15:40	Varying-Coefficient Models for Geospatial Transfer Learning
15.20 - 15:40	Matthias Bussas, Christoph Sawade, Nicolas Kühn, Tobias Scheffer, Niels Landwehr

Time Series & Streams 2

Chair: Panče Panov

14:00 - 14:20	A Multiscale Bezier-Representation for Time Series That Supports Elastic Matching
	Frank Hoeppner, Tobias Sobek
14:20 - 14:40	Cost Sensitive Time-Series Classification
	Shoumik Roychoudhury, Mohamed Ghalwash, Zoran Obradovic
14:40 - 15:00	Efficient Temporal Kernels Between Feature Sets for Time Series Classification
	Romain Tavenard, Simon Malinowski, Laetitia Chapel, Adeline Bailly, Heider Sanchez, Benjamin Bustos
15:00 - 15:20	Forecasting and Granger Modelling with Non-Linear Dynamical Dependencies
	Magda Gregorova, Alexandros Kalousis, Stephan Marchand-Maillet
15:20 - 15:40	UAPD: Predicting Urban Anomalies from Spatial-Temporal Data
	Xian Wu, Yuxiao Dong, Chao Huang, Jian Xu, Dong Wang, Nitesh Chawla

Applied Data Science 6

Applieu Dala	
Chair: Kamalika	Das 14:00-15:40
14:00 - 14:20	Predicting Self-Reported Customer Satisfaction of Interactions with a Corporate Call Center
	Joseph Bockhorst, Shi Yu, Luisa Polania Cabrera, Glenn Fung Moo
14:20 - 14:40	Have It Both Ways - from A/B Testing to A&B Testing with Exceptional Model Mining
	Wouter Duivesteijn, Tara Farzami, Thijs Putman, Evertjan Peer, Hilde J.P. Weerts, Jasper Adegeest, Gerson Foks, Mykola Pechenizkiy
14:40 - 15:00	Structural Semantic Models for Automatic Analysis of Urban Areas
14:40 - 15:00	Gianni Barlacchi, Alberto Rossi, Bruno Lepri, Alessandro Moschitti
15:00 15:00	Using Machine Learning for Labour Market Intelligence
15:00 - 15:20	Roberto Boselli, Mirko Cesarini, Fabio Mercorio, Mario Mezzanzanica
15:20 - 15:40	SINAS: Suspect Investigation Using Offenders' Activity Space
	Mohammad Tayebi, Uwe Glässer, Patricia Brantingham, Hamed Yaghoubi Shahir

Probabilistic Models & Methods 3

Chair: Indrė Žliobaitė

CONGRESS HALL 2

CONGRESS HALL 3

16:00-17:40

16:00 - 16:20	Knowledge Elicitation Via Sequential Probabilistic Inference for High-Dimensional Prediction
	Pedram Daee, Tomi Peltola, Marta Soare, Samuel Kaski
16:20 - 16:40	Labeled DBN Learning with Community Structure Knowledge
	Etienne Auclair, Nathalie Peyrard, Régis Sabbadin
16:40 - 17:00	Online Sparse Collapsed Hybrid Variational-Gibbs Algorithm for Hierarchical Dirichlet Process Topic Models
	Sophie Burkhardt, Stefan Kramer
17.00 - 17.00	Partial Device Fingerprints
17.00 - 17.20	Michael Ciere, Carlos Ganan, Michel Van Eeten
17:20 - 17:40	Vine Copulas for Mixed Data : Multi-View Clustering for Mixed Data Beyond Meta-Gaussian Dependencies
	Lavanya Sita Tekumalla, Vaibhav Rajan, Chiranjib Bhattacharyya

Matrix & Tensor Factorization

Chair: Katharina Morik

16:00 - 16:20	C-SALT: Mining Class-Specific ALTerations in Boolean Matrix Factorization Sibylle Hess, Katharina Morik
16:20 - 16:40	Comparative Study of Inference Methods for Bayesian Nonnegative Matrix Factorisation
	Thomas Brouwer, Jes Frellsen, Pietro Lio
16:40 - 17:00	Content-Based Social Recommendation with Poisson Matrix Factorization
	Eliezer De Souza Da Silva, Helge Langseth, Heri Ramampiaro
17:00 - 17:20	Feature Extraction for Incomplete Data Via Low-Rank Tucker Decomposition
	Qiquan Shi, Yiu-Ming Cheung, Qibin Zhao
17:20 - 17:40	Structurally Regularized Non-Negative Tensor Factorization for Spatio-Temporal Pattern Discoveries
	Koh Takeuchi, Yoshinobu Kawahara, Tomoharu Iwata

Transfer & Multi-Task Learning 2

Chair: Pavel Brazd	il 16:00-17:40	
16:00 - 16:20	A Novel Rating Pattern Transfer Model for Improving Non-Overlapping Cross-Domain Collaborative Filtering	
	Yizhou Zang, Xiaohua Hu	
16:20 - 16:40	Distributed Multi-Task Learning for Sensor Network	
	Jiyi Li, Tomohiro Arai, Yukino Baba, Hisashi Kashima, Shotaro Miwa	
16:40 - 17:00	Learning Task Structure Via Sparsity Grouped Multitask Learning	
	Meghana Kshirsagar, Eunho Yang, Aurelie Lozano	
17:00 - 17:20	Ranking Based Multitask Learning of Scoring Functions	
	Ivan Stojkovic, Mohamed Ghalwash, Zoran Obradovic	

Computer Vision

Chair: Ivica Dimitrovski

CONGRESS HALL 4 16:00-17:40

16.00 - 16.20	Alternative Semantic Representations for Zero-Shot Human Action Recognition
10.00 - 10.20	Qian Wang, Ke Chen
16.20 - 16.40	Early Active Learning with Pairwise Constraint for Person Re-Identification
10.20 - 10.40	Wenhe Liu, Xiaojun Chang, Ling Chen, Yi Yang, Alexander Hauptmann
16.40 17.00	Guiding InfoGAN with Semi-Supervision
16:40 - 17:00	Adrian Spurr, Emre Aksan, Otmar Hilliges
17.00 17.00	Scatteract: Automated Extraction of Data from Scatter Plots
17:00 - 17:20	Mathieu Cliche, David Rosenberg, Connie Yee, Dhruv Madeka
10.00 10.40	Unsupervised Diverse Colorization Via Generative Adversarial Networks
17:20 - 17:40	Yun Cao, Zhiming Zhou, Weinan Zhang, Yong Yu

Demo 2

MACEDONIAN **OPERA & BALLET**

20:00

Chair: Jesse Read & Marinka Žitnik

Monitoring Physical Activity and Mental Stress Using Wrist-Worn Device and a Smartphone Božidara Cvetković, Martin Gjoreski, Jure Šorn, Pavel Maslov, Mitja Luštrek Lit@EVE: Explainable Recommendation Based on Wikipedia Concept Vectors Muhammad Atif Qureshi, Derek Greene TF Boosted Trees: A Scalable TensorFlow Based Framework for Gradient Boosting Natalia Ponomareva, Soroush Radpour, Gilbert Hendry, Salem Haykal, Thomas Colthurst, Petr Mitrichev, Alexander Grushetsky ASK-the-Expert: Active Learning Based Knowledge Discovery Using the Expert Kamalika Das, Ilya Avrekh, Bryan Matthews, Manali Sharma, Nikunj Oza TrajViz: A Tool for Visualizing Patterns and Anomalies in Trajectory Yifeng Gao, Qingzhe Li, Xiaosheng Li, Jessica Lin, Huzefa Rangwala, Ranjeev Mittu

SESSIONS WITH ABSTRACTS

Networks & Graphs 3

11:00-11:20 Graph-Based Predictable Feature Analysis

Björn Weghenkel, Asja Fischer, Laurenz Wiskott

We propose graph-based predictable feature analysis (GPFA), a new method for unsupervised learning of predictable features from highdimensional time series, where high predictability is understood very generically as low variance in the distribution of the next data point given the previous ones. We show how this measure of predictability can be understood in terms of graph embedding as well as how it relates to the information-theoretic measure of predictive information in special cases. We confirm the effectiveness of GPFA on different datasets, comparing it to three existing algorithms with similar objectives—namely slow feature analysis, forecastable component analysis, and predictable feature analysis—to which GPFA shows very competitive results.

11:20-11:40 Lagrangian Relaxations for Multiple Network Alignment

Eric Malmi, Sanjay Chawla, Aristides Gionis

We propose a principled approach for the problem of aligning multiple partially overlapping networks. The objective is to map multiple graphs into a single graph while preserving vertex and edge similarities. The problem is inspired by the task of integrating partial views of a family tree (genealogical network) into one unified network, but it also has applications, for example, in social and biological networks. Our approach, called Flan, introduces the idea of generalizing the facility location problem by adding a non-linear term to capture edge similarities and to infer the underlying entity network. The problem is solved using an alternating optimization procedure with a Lagrangian relaxation. Flan has the advantage of being able to leverage prior information on the number of entities, so that when this information is available, Flan is shown to work robustly without the need to use any ground truth data for fine-tuning method parameters. Additionally, we present three multiple-network extensions to an existing state-of-the-art pairwise alignment method called Natalie. Extensive experiments on synthetic, as well as real-world datasets on social networks and genealogical networks, attest to the effectiveness of the proposed approaches which clearly outperform a popular multiple network alignment method called IsoRankN.

11:40-12:00 Measuring and Moderating Opinion Polarization in Social Networks

Antonis Matakos, Evimaria Terzi, Panayiotis Tsaparas

The polarization of society over controversial social issues has been the subject of study in social sciences for decades (Isenberg in J Personal Soc Psychol 50(6):1141–1151, 1986, Sunstein in J Polit Philos 10(2):175–195, 2002). The widespread usage of online social networks and social media, and the tendency of people to connect and interact with like-minded individuals has only intensified the phenomenon of polarization (Bakshy et al. in Science 348(6239):1130–1132, 2015). In this paper, we consider the problem of measuring and reducing polarization of opinions in a social network. Using a standard opinion formation model (Friedkin and Johnsen in J Math Soc 15(3–4):193–206, 1990), we define the polarization index, which, given a network and the opinions of the individuals in the network, it quantifies the polarization observed in the network. Our measure captures the tendency of opinions to concentrate in network communities, creating echo-chambers. Given this numeric measure of polarization, we then consider the problem of reducing polarization in the network by convincing individuals (e.g., through education, exposure to diverse viewpoints, or incentives) to adopt a more neutral stand towards controversial issues. We formally define the ModerateInternal and ModerateExpressed problems, and we prove that both our problems are NP-hard. By exploiting the linear- algebraic characteristics of the opinion formation model we design polynomial-time algorithms for both problems. Our experiments with real-world datasets demonstrate the validity of our metric, and the efficiency and the effectiveness of our algorithms in practice.



THURSDAY SEPTEMBER 21

12:00-12:20 Micro-Review Synthesis for Multi-Entity Summarization

Thanh-Son Nguyen, Hady W. Lauw, Panayiotis Tsaparas

Location-based social networks (LBSNs), exemplified by Foursquare, are fast gaining popularity. One important feature of LBSNs is microreview. Upon check-in at a particular venue, a user may leave a short review (up to 200 characters long), also known as a tip. These tips are an important source of information for others to know more about various aspects of an entity (e.g., restaurant), such as food, waiting time, or service. However, a user is often interested not in one particular entity, but rather in several entities collectively, for instance within a neighborhood or a category. In this paper, we address the problem of summarizing the tips of multiple entities in a collection, by way of synthesizing new micro-reviews that pertain to the collection, rather than to the individual entities per se. We formulate this problem in terms of first finding a representation of the collection, by identifying a number of "aspects" that link common threads across two or more entities within the collection. We express these aspects as dense subgraphs in a graph of sentences derived from the multi-entity corpora. This leads to a formulation of maximal multi-entity quasi-cliques, as well as a heuristic algorithm to find K such quasi-cliques maximizing the coverage over the multi-entity corpora. To synthesize a summary tip for each aspect, we select a small number of sentences from the corresponding quasi-clique, balancing conciseness and representativeness in terms of a facility location problem. Our approach performs well on collections of Foursquare entities based on localities and categories, producing more representative and diverse summaries than the baselines.

12:20-12:40 Survival Factorization for Topical Cascades on Diffusion Networks

Nicola Barbieri, Giuseppe Manco, Ettore Ritacco

In this paper we propose a survival factorization framework that models information cascades by tying together social influence patterns, topical structure and temporal dynamics. This is achieved through the introduction of a latent space which encodes: (a) the relevance of a information cascade on a topic; (b) the topical authoritativeness and the susceptibility of each individual involved in the information cascade, and (c) temporal topical patterns. By exploiting the cumulative properties of the survival function and of the likelihood of the model on a given adoption log, which records the observed activation times of users and side-information for each cascade, we show that the inference phase is linear in the number of users and in the number of adoptions. The evaluation on both synthetic and real-world data shows the effectiveness of the model in detecting the interplay between topics and social influence patterns, which ultimately provides high accuracy in predicting users activation times.

Unsupervised & Semi-supervised Learning 2

11:00-11:20 An Exponential Family Framework For Learning to Predict Unseen Classes

Vinay Verma, Wenlin Wang, Piyush Rai

We present a simple generative framework for learning to predict previously unseen classes, based on estimating class-attribute-gated classconditional distributions. We model each class-conditional distribution as an exponential family distribution and the parameters of the distribution of each seen/unseen class are defined as functions of the respective observed class attributes. These functions can be learned using only the seen class data and can be used to predict the parameters of the class-conditional distribution of each unseen class. Unlike most existing methods for zero-shot learning that represent classes as fixed embeddings in some vector space, our generative model naturally represents each class as a probability distribution. It is simple to implement and also allows leveraging additional unlabeled data from unseen classes to improve the estimates of their class-conditional distributions using transductive/semi-supervised learning. Moreover, it extends seamlessly to few-shot learning by easily updating these distributions when provided with a small number of additional labelled examples from unseen classes. Through a comprehensive set of experiments on several benchmark data sets, we demonstrate the efficacy of our framework.







An Expressive Similarity Measure for Relational Clustering Using Neighbourhood Trees

Sebastijan Dumančić, Hendrik Blockeel

Clustering is an underspecified task: there are no universal criteria for what makes a good clustering. This is especially true for relational data, where similarity can be based on the features of individuals, the relationships between them, or a mix of both. Existing methods for relational clustering have strong and often implicit biases in this respect. In this paper, we introduce a novel dissimilarity measure for relational data. It is the first approach to incorporate a wide variety of types of similarity, including similarity of attributes, similarity of relational context, and proximity in a hypergraph. We experimentally evaluate the proposed dissimilarity measure on both clustering and classification tasks using data sets of very different types. Considering the quality of the obtained clustering, the experiments demonstrate that (a) using this dissimilarity in standard clustering methods consistently gives good results, whereas other measures work well only on data sets that match their bias; and (b) on most data sets, the novel dissimilarity outperforms even the best among the existing ones. On the classification tasks, the proposed method outperforms the competitors on the majority of data sets, often by a large margin. Moreover, we show that learning the appropriate bias in an unsupervised way is a very challenging task, and that the existing methods offer a marginal gain compared to the proposed similarity method, and can even hurt performance. Finally, we show that the asymptotic complexity of the proposed dissimilarity measure is similar to the existing state-of-the-art approaches. The results confirm that the proposed dissimilarity measure is indeed versatile enough to capture relevant information, regardless of whether that comes from the attributes of vertices, their proximity, or connectedness of vertices, even without parameter tuning.

11:40-12:00 DeepCluster: A General Clustering Framework Based on Deep Learning

Kai Tian, Shuigeng Zhou, Jihong Guan

In this paper, we propose a general framework DeepCluster to integrate traditional clustering methods into deep learning (DL) models and adopt Alternating Direction of Multiplier Method (ADMM) to optimize it. While most existing DL based clustering techniques have separate feature learning (via DL) and clustering (with traditional clustering methods), DeepCluster simultaneously learns feature representation and does cluster assignment under the same framework. Furthermore, it is a general and flexible framework that can employ different networks and clustering methods. We demonstrate the effectiveness of DeepCluster by integrating two popular clustering methods: K-means and Gaussian Mixture Model (GMM) into deep networks. The experimental results shown that our method can achieve state-of-the-art performance on learning representation for clustering analysis.

12:00-12:20 Local PurTree Subspace Spectral Clustering for Customer Transaction Data

Xiaojun Chen, JianZhe Zhang, Wenya Sun, Joshua Huang, Qingyao Wu

This paper was accepted for presentation. However, it was not presented at the conference and is thus not published in the conference proceedings.

12:20-12:40 Multi-View Spectral Clustering on Conflicting Views

Xiao He, Limin Li, Damian Roqueiro, Karsten Borgwardt

In a growing number of application domains, multiple feature representations or views are available to describe objects. Multi-view clustering tries to find similar groups of objects across these views. This task is complicated when the corresponding clusterings in each view show poor agreement (conflicting views). In such cases, traditional multi-view clustering methods will not benefit from using multi-view data. Here, we propose to overcome this problem by combining the ideas of multi-view spectral clustering with alternative clustering through kernel-based dimensionality reduction. Our method automatically determines feature transformations in each view that lead to an optimal clustering w.r.t to a new proposed objective function for conflicting views. In our experiments, our approach outperforms state-of-the-art multi-view clustering methods by more accurately detecting the ground truth clustering supported by all views.





NO

SHOW

11:20 - 11:40

Fabrizio Angiulli

Anomaly Detection

11:00 - 11:20

We present a novel notion of outlier, called Concentration Free Outlier Factor (CFOF), having the peculiarity to resist concentration phenomena that affect other scores when the dimensionality of the feature space increases. Indeed we formally prove that CFOF does not concentrate in intrinsically high-dimensional spaces. Moreover, CFOF is adaptive to different local density levels and it does not require the computation of exact neighbors in order to be reliably computed. We present a very efficient technique, named fast-CFOF, for detecting outliers in very large highdimensional datasets. The technique is efficiently parallelizable, and we provide a MIMD-SIMD implementation. Experimental results witness for scalability and effectiveness of the technique and highlight that CFOF exhibits state of the art detection performances.

Efficient Top Rank Optimization with Gradient Boosting for Supervised 11:20 - 11:40 **Anomaly Detection**

Jordan Frery, Marc Sebban, Amaury Habrard, Olivier Caelen, Liyun Guelton

In this paper we address the anomaly detection problem in a supervised setting where positive examples might be very sparse. We tackle this task with a learning to rank strategy by optimizing a differentiable smoothed surrogate of the so-called Average Precision (AP). Despite its nonconvexity, we show how to use it efficiently in a stochastic gradient boosting framework. We show that using AP is much better to optimize the top rank alerts than the state of the art measures. We demonstrate on anomaly detection tasks that the interest of our method is even reinforced in highly unbalanced scenarios.

11:40 - 12:00 **Robust, Deep and Inductive Anomaly Detection**

Concentration Free Outlier Detection

Raghavendra Chalapathy, Aditya Krishna Menon, Sanjay Chawla

PCA is a classical statistical technique whose simplicity and maturity has seen it find widespread use as an anomaly detection technique. However, it is limited in this regard by being sensitive to gross perturbations of the input, and by seeking a linear subspace that captures normal behaviour. The first issue has been dealt with by robust PCA, a variant of PCA that explicitly allows for some data points to be arbitrarily corrupted; however, this does not resolve the second issue, and indeed introduces the new issue that one can no longer inductively find anomalies on a test set. This paper addresses both issues in a single model, the robust autoencoder. This method learns a nonlinear subspace that captures the majority of data points, while allowing for some data to have arbitrary corruption. The model is simple to train and leverages recent advances in the optimisation of deep neural networks. Experiments on a range of real-world datasets highlight the model's effectiveness.

Sentiment Informed Cyberbullying Detection in Social Media 12:00 - 12:20

Harsh Dani, Jundong Li, Huan Liu

Cyberbullying is a phenomenon which negatively affects the individuals, the victims suffer from various mental issues, ranging from depression, loneliness, anxiety to low self-esteem. In parallel with the pervasive use of social media, cyberbullying is becoming more and more prevalent. Traditional mechanisms to fight against cyberbullying include the use of standards and guidelines, human moderators, and blacklists based on the profane words. However, these mechanisms fall short in social media and cannot scale well. Therefore, it is necessary to develop a principled learning framework to automatically detect cyberbullying behaviors. However, it is a challenging task due to short, noisy and unstructured content information and intentional obfuscation of the abusive words or phrases by social media users. Motivated by sociological and psychological findings on bullying behaviors and the correlation with emotions, we propose to leverage sentiment information to detect cyberbullying behaviors in social media by proposing a sentiment informed cyberbullying detection framework. Experimental results on two real-world, publicly available social media datasets show the superiority of the proposed framework. Further studies validate the effectiveness of leveraging sentiment information for cyberbullying detection.

ZooRank: Ranking Suspicious Activities in Time-Evolving Tensors 12:20 - 12:40

Hemank Lamba, Bryan Hooi, Kijung Shin, Christos Faloutsos, Juergen Pfeffer

Most user-based websites such as social networks(Twitter,Facebook) and e-commerce websites (Amazon) have been targets of group fraud (multiple users working together for malicious purposes). How can we better rank malicious entities in such cases of group-fraud? Most of the existing work in group anomaly detection detects lock-step behavior by detecting dense blocks in matrices, and recently, in tensors. However, there is no principled way of scoring the users based on their participation in these dense blocks. In addition, existing methods do not take into account temporal features while detecting dense blocks, which are crucial to uncover bot-like behaviors. In this paper (a) we propose a systematic way of handling temporal information; (b) we give a list of axioms that any individual suspiciousness metric satisfies; (c) we propose ZOORANK, an algorithm that finds and ranks suspicious entities (users, targeted products, days, etc.) effectively in real-world datasets. Experimental results on multiple real-world datasets show that ZOORANK detected and ranked the suspicious entities with a nearly accurate precision, while outperforming the baseline approach.









11:00-11:20 A Novel Framework for Online Sales Burst Prediction

Rui Chen, Jiajun Liu



With the rapid growth of e-commerce, a large number of on- line transactions are processed every day. In this paper, we take the initiative to conduct a systematic study of the challenging prediction problems of sales bursts. Here, we propose a novel model to detect bursts, find the bursty features, namely the start time of the burst, the peak value of the burst and the o -burst value, and predict the entire burst shape. Our model analyzes the features of similar sales bursts in the same category, and applies them to generate the prediction. We argue that the framework is capable of capturing the seasonal and categorical features of sales burst. Based on the real data from JD.com, we conduct extensive experiments and discover that the proposed model makes a relative MSE improvement of 71% and 30% over LSTM and ARMA.

11:20-11:40 MRNet-Product2Vec: A Multi-Task Recurrent Neural Network for Product Embeddings



Arijit Biswas, Mukul Bhutani, Subhajit Sanyal

E-commerce websites such as Amazon, Alibaba, Flipkart, and Walmart sell billions of products. Machine learning (ML) algorithms involving products are often used to improve the customer experience and increase revenue, e.g., product similarity, recommendation, and price estimation. The products are required to be represented as features before training an ML algorithm. In this paper, we propose an approach called MRNet-Product2Vec for creating generic embeddings of products within an e-commerce ecosystem. We learn a dense and low-dimensional embedding where a diverse set of signals related to a product are explicitly injected into its representation. We train a Discriminative Multi-task Bidirectional Recurrent Neural Network (RNN), where the input is a product title fed through a Bidirectional RNN and at the output, product labels corresponding to fifteen different tasks are predicted. The task set includes several intrinsic characteristics about a product such as price, weight, size, color, popularity, and material. We evaluate the proposed embedding quantitatively and qualitatively. We demonstrate that they are almost as good as sparse and extremely high-dimensional TF-IDF representation in spite of having less than 3% of the TF-IDF dimension. We also use a multimodal autoencoder for comparing products from different language-regions and show preliminary yet promising qualitative results.

11:40-12:00 Disjoint-Support Factors and Seasonality Estimation in E-Commerce



Successful inventory management in retail entails accurate demand forecasts for many weeks/months ahead. Forecasting models use seasonality: recurring pattern of sales every year, to make this forecast. In e-commerce setting, where the catalog of items is much larger than brick and mortar stores and hence includes a lot of items with short history, it is infeasible to compute seasonality for items individually. It is customary in these cases to use ideas from factor analysis and express seasonality by a few factors/basis vectors computed together for an entire assortment of related items. In this paper, we demonstrate the effectiveness of choosing vectors with disjoint support as basis for seasonality when dealing with a large number of short time-series. We give theoretical results on computation of disjoint support factors that extend the state of the art, and also discuss temporal regularization necessary to make it work on walmart e-commerce dataset. Our experiments demonstrate a marked improvement in forecast accuracy for items with short history.

12:00-12:20 Generalising Random Forest Parameter Optimisation to Include Stability and Cost

Chak Hin Bryan Liu, Benjamin Chamberlain, Duncan Little, Angelo Cardoso

Random forests are among the most popular classification and regression methods used in industrial applications. To be effective, the parameters of random forests must be carefully tuned. This is usually done by choosing values that minimize the prediction error on a held out dataset. We argue that error reduction is only one of several metrics that must be considered when optimizing random forest parameters for commercial applications. We propose a novel metric that captures the stability of random forest predictions, which we argue is key for scenarios that require successive predictions. We motivate the need for multi-criteria optimization by showing that in practical applications, simply choosing the parameters that lead to the lowest error can introduce unnecessary costs and produce predictions that are not stable across independent runs. To optimize this multi-criteria trade-off, we present a new framework that efficiently finds a principled balance between these three considerations using Bayesian optimisation. The pitfalls of optimising forest parameters purely for error reduction are demonstrated using two publicly available real world datasets. We show that our framework leads to parameter settings that are markedly different from the values discovered by error reduction metrics alone.

TUESDAY

THURSDAY SEPTEMBER 21

12:20-12:40 Session-Based Fraud Detection in Online E-Commerce Transactions Using Recurrent Neural Networks

Shuhao Wang, Cancheng Liu, Xiang Gao, Hongtao Qu, Wei Xu



Recommendation

CONGRESS HALL 1

14:00-14:20 A Regularization Method with Inference of Trust and Distrust in Recommender Systems

Dimitrios Rafailidis, Fabio Crestani



In this study we investigate the recommendation problem with trust and distrust relationships to overcome the sparsity of users' preferences, accounting for the fact that users trust the recommendations of their friends, and they do not accept the recommendations of their foes. In addition, not only users' preferences are sparse, but also users' social relationships. So, we first propose an inference step with multiple random walks to predict the implicit-missing trust relationships that users might have in recommender systems, while considering users' explicit trust and distrust relationships during the inference. We introduce a regularization method and design an objective function with a social regularization term to weigh the influence of friends' trust and foes' distrust degrees on users' preferences. We formulate the objective function of our regularization method as a minimization problem with respect to the users' and items' latent features and then we solve our recommendation problem via gradient descent. Our experiments confirm that our approach preserves relatively high recommendation accuracy in the presence of sparsity in both the users' preferences and social relationships, significantly outperforming several state-of-the-art methods.

14:20-14:40 A Unified Contextual Bandit Framework for Long- and Short-Term Recommendations

Maryam Tavakol, Ulf Brefeld

We present a unified contextual bandit framework for recommendation problems that is able to capture long- and short-term interests of users. The model is devised in dual space and the derivation is consequentially carried out using Fenchel-Legrende conjugates and thus lever- ages to a wide range of tasks and settings. We detail two instantiations for regression and classification scenarios and obtain well-known algorithms for these special cases. The resulting general and unified framework allows for quickly adapting contextual bandits to different applications at-hand. The empirical study demonstrates that the proposed long- and short-term framework outperforms both, short-term and long-term models on data. Moreover, a tweak of the combined model proves beneficial in cold start problems.

14:40-15:00 Perceiving the Next Choice with Comprehensive Transaction Embeddings for Online Recommendation Shouin Wang Ling Hu Longhing Cao, Xiaoshui Huang



Shoujin Wang, Liang Hu, Longbing Cao, Xiaoshui Huang

To predict customers next choice in the context of what he/she has bought in a session is interesting and critical in the transaction domain especially for online shopping. Precise prediction leads to high quality recommendations and thus high benefit. Such kind of recommendation is usually formalized as transaction-based recommender systems (TBRS). Existing TBRS either tend to recommend popular items while ignore infrequent and newly-released ones (e.g., pattern-based RS) or assume a rigid order between items within a transaction(e.g., Markov Chainbased RS) which does not satisfy real-world cases in most time. In this paper, we propose a neural network based comprehensive transaction embedding model (NTEM) which can effectively perceive the next choice in a transaction context. Specifically, we learn these comprehensive embeddings of both items and their features from relaxed ordered transactions. The relevance between items revealed by the transactions is encoded into such embeddings. With rich information embedded, such embeddings are powerful to predict the next choices given those already bought items. NTEM is a shallow wide-in-wide-out network, which is more efficient than deep networks considering large numbers of items and transactions. Experimental results on real-world datasets show that NTEM outperforms three typical TBRS models FPMC, PRME and GRU4Rec in terms of recommendation accuracy and novelty. Our implementation is available at https://github.com/shoujin88/NTEM-model

Social Regularized Von Mises-Fisher Mixture Model for Item 15:00 - 15:20 Recommendation

Aghiles Salah, Mohamed Nadif

Collaborative filtering (CF) is a widely used technique to guide the users of web applications towards items that might interest them. CF approaches are severely challenged by the characteristics of user-item preference matrices, which are often high dimensional and extremely sparse. Recently, several works have shown that incorporating information from social networks-such as friendship and trust relationships-into traditional CF alleviates the sparsity related issues and yields a better recommendation quality, in most cases. More interestingly, even with comparable performances, social-based CF is more beneficial than traditional CF; the former makes it possible to provide recommendations for cold start users. In this paper, we propose a novel model that leverages information from social networks to improve recommendations. While existing social CF models are based on popular modelling assumptions such as Gaussian or Multinomial, our model builds on the von Mises-Fisher assumption which turns out to be more adequate, than the aforementioned assumptions, for high dimensional sparse data. Setting the estimate of the model parameters under the maximum likelihood approach, we derive a scalable learning algorithm for analyzing data with our model. Empirical results on several real-world datasets provide strong support for the advantages of the proposed model.

15:20 - 15:40 **Tour Recommendation for Groups**

Aris Anagnostopoulos, Reem Atassi, Luca Becchetti, Adriano Fazzone, Fabrizio Silvestri

Consider a group of people who are visiting a major touristic city, such as NY, Paris, or Rome. It is reasonable to assume that each member of the group has his or her own interests or preferences about places to visit, which in general may differ from those of other members. Still, people almost always want to hang out together and so the following question naturally arises: What is the best tour that the group could perform together in the city? This problem underpins several challenges, ranging from understanding people's expected attitudes towards potential points of interest, to modeling and providing good and viable solutions. Formulating this problem is challenging because of multiple competing objectives. For example, making the entire group as happy as possible in general conflicts with the objective that no member becomes disappointed. In this paper, we address the algorithmic implications of the above problem, by providing various formulations that take into account the overall group as well as the individual satisfaction and the length of the tour. We then study the computational complexity of these formulations, we provide effective and efficient practical algorithms, and, finally, we evaluate them on datasets constructed from real city data.

Transfer & Multi-Task Learning 1

14:00 - 14:20 Lifelong Machine Learning with Gaussian Processes

Christopher Clingerman, Eric Eaton

Recent developments in lifelong machine learning have demonstrated that it is possible to learn multiple tasks consecutively, transferring knowledge between those tasks to accelerate learning and improve performance. However, these methods are limited to using linear parametric base learners, substantially restricting the predictive power of the resulting models. We present a lifelong learning algorithm that can support non-parametric models, focusing on Gaussian processes. To enable efficient online transfer between Gaussian process models, our approach assumes a factorized formulation of the covariance functions, and incrementally learns a shared sparse basis for the models' parameterizations. We show that this lifelong learning approach is highly computationally efficient, and outperforms existing methods on a variety of data sets.

Personalized Tag Recommendation for Images Using Deep Transfer 14:20 - 14:40 Learning

Hanh T.H. Nguyen, Martin Wistuba, Lars Schmidt-Thieme

Image tag recommendation in social media systems provides the users with personalized tag suggestions which facilitate the users' tagging task and enable automatic organization and many image retrieval tasks. Factorization models are a widely used approach for personalized tag recommendation and achieve good results. These methods rely on the user's tagging preferences only and ignore the contents of the image. However, it is obvious that especially the contents of the image, such as the objects appearing in the image, colors, shapes or other visual aspects, strongly influence the user's tagging decisions. We present a personalized content-aware image tag recommendation approach that combines both historical tagging information and image-based features in a factorization model. Employing transfer learning, we apply state of the art deep learning image classification and object detection techniques to extract powerful features from the images. Both, image information and tagging history, are fed to an adaptive factorization model to recommend tags. Empirically, we can demonstrate that the visual and object-based features can improve the performance up to 1.5 percent over the state of the art.





THURSDAY SEPTEMBER 21

14:40-15:00 Theoretical Analysis of Domain Adaptation with Optimal Transport

levgen Redko, Amaury Habrard, Marc Sebban

Domain adaptation (DA) is an important and emerging field of machine learning that tackles the problem occurring when the distributions of training (source domain) and test (target domain) data are similar but different. Current theoretical results show that the efficiency of DA algorithms depends on their capacity of minimizing the divergence between source and target probability distributions. In this paper, we provide a theoretical study on the advantages that concepts borrowed from optimal transportation theory can bring to DA. In particular, we show that the Wasserstein metric can be used as a divergence measure between distributions to obtain generalization guarantees for three different learning settings: (i) classic DA with unsupervised target data (ii) DA combining source and target labeled data, (iii) multiple source DA. Based on the obtained results, we provide some insights showing when this analysis can be tighter than other existing frameworks. We think that these results open the door to novel ideas and directions for DA.

15:00-15:20 TSP: Learning Task-Specific Pivots for Unsupervised Domain Adaptation

Xia Cui, Frans Coenen, Danushka Bollegala

Unsupervised Domain Adaptation (UDA) considers the problem of adapting a classifier trained using labelled training instances from a source domain to a different target domain, without having access to any labelled training instances from the target domain. Projection-based methods, where the source and target domain instances are first projected onto a common feature space on which a classifier can be trained and applied have produced state-of-the-art results for UDA. However, a critical pre-processing step required by these methods is the selection of a set of common features (aka. pivots), this is typically done using heuristic approaches,applied prior to performing domain adaptation. In contrast to the one of heuristics, we propose a method for learning Task-Specific Pivots (TSPs) in a systematic manner by considering both the labelled and unlabelled data available from both domains. We evaluate TSPs against pivots selected using alternatives in two cross-domain sentiment classification applications. Our experimental results show that the proposed TSPs significantly outperform previously proposed selection strategies in both tasks. Moreover, when applied in a cross-domain sentiment classification task, TSP captures many sentiment-bearing pivots.

15:20-15:40 Varying-Coefficient Models for Geospatial Transfer Learning

Matthias Bussas, Christoph Sawade, Nicolas Kühn, Tobias Scheffer, Niels Landwehr

We study prediction problems in which the conditional distribution of the output given the input varies as a function of task variables which, in our applications, represent space and time. In varying-coefficient models, the coefficients of this conditional are allowed to change smoothly in space and time; the strength of the correlations between neighboring points is determined by the data. This is achieved by placing a Gaussian process (GP) prior on the coefficients. Bayesian inference in varying-coefficient models is generally intractable. We show that with an isotropic GP prior, inference in varying-coefficient models resolves to standard inference for a GP that can be solved efficiently. MAP inference in this model resolves to multitask learning using task and instance kernels. We clarify the relationship between varying-coefficient models and the hierarchical Bayesian multitask model and show that inference for hierarchical Bayesian multitask models can be carried out efficiently using graph-Laplacian kernels. We explore the model empirically for the problems of predicting rent and real-estate prices, and predicting the ground motion during seismic events. We find that varying-coefficient models with GP priors excel at predicting rents and real-estate prices. The ground-motion model predicts seismic hazards in the State of California more accurately than the previous state of the art.

Time Series & Streams 2

14:00-14:20 A Multiscale Bezier-Representation for Time Series That Supports Elastic Matching

Frank Hoeppner, Tobias Sobek

Common time series similarity measures that operate on the full series (like Euclidean distance or Dynamic Time Warping DTW) do not correspond well to the visual similarity as perceived by a human. Based on the interval tree of scale, we propose a multiscale Bezier representation of time series, that supports the definition of elastic similarity measures that overcome this problem. With this representation the matching can be performed efficiently as similarity is measured segment-wise rather than element-wise (as with DTW). We effectively restrict the set of warping paths considered by DTW and the results do not only correspond better to the analysts intuition but improve the accuracy in the standard 1NN time series classification.






14:20 - 14:40 Cost Sensitive Time-Series Classification

14:40 - 15:00



This paper investigates the problem of highly imbalanced time-series classification using shapelets. The current state-of-the-art approach learns generalized shapelets along with weights of the classification hyperplane via a classical cost-insensitive loss function. Cost-insensitive loss functions tend to treat different misclassification errors equally and thus, models are usually biased towards examples of majority class. The rare class (which will be referred to as positive class) is usually the important class and a false negative is always costlier than a false positive. Traditional cost-insensitive loss functions fail to differentiate between these two types of misclassification errors. In this paper, the generalized shapelets learning framework is extended and a cost-sensitive learning model is proposed. Instead of incorporating the misclassification cost as a prior knowledge, as was done by other published methods, we formulate a constrained optimization problem to learn the unknown misclassification costs along with the shapelets and their weights. First, we demonstrated the effectiveness of the proposed method on two case studies, with the objective to detect true alarms from life threatening cardiac arrhythmia dataset from Physionets MIMIC II repository. The results show improved true alarm detection rates over the current state-of-the-art method. Next, we compared to the state-of-the-art learning shapelet method on 16 balanced dataset from UCR time-series repository. The results show evidence that the proposed method outperforms the state-of-the-art method. Finally, we performed extensive experiments across additional 18 imbalanced time-series datasets. The results provide evidence that the proposed method achieves comparable results with state-of-the-art sampling/non-sampling based approaches for highly imbalanced time-series datasets. However, our method is highly interpretable which is advantage over many other methods.

Efficient Temporal Kernels Between Feature Sets for Time Series Classification

Romain Tavenard, Simon Malinowski, Laetitia Chapel, Adeline Bailly, Heider Sanchez, Benjamin Bustos

In the time-series classification context, the majority of the most accurate core methods are based on the Bag-of-Words framework, in which sets of local features are first extracted from time series. A dictionary of words is then learned and each time series is finally represented by a histogram of word occurrences. This representation induces a loss of information due to the quantization of features into words as all the time series are represented using the same fixed dictionary. In order to overcome this issue, we introduce in this paper a kernel operating directly on sets of features. Then, we extend it to a time-compliant kernel that allows one to take into account the temporal information. We apply this kernel in the time series classification context. Proposed kernel has a quadratic complexity with the size of input feature sets, which is problematic when dealing with long time series. However, we show that kernel approximation techniques can be used to define a good trade-off between accuracy and complexity. We experimentally demonstrate that the proposed kernel can significantly improve the performance of time series classification algorithms based on Bag-of-Words.

15:00-15:20 Forecasting and Granger Modelling with Non-linear Dynamical Dependencies

Magda Gregorova, Alexandros Kalousis, Stephan Marchand-Maillet

Traditional linear methods for forecasting multivariate time series are not able to satisfactorily model the non-linear dependencies that may exist in non-Gaussian series. We build on the theory of learning vector-valued functions in the reproducing kernel Hilbert space and develop a method for learning prediction functions that accommodate such non-linearities. The method not only learns the predictive function but also the matrix-valued kernel underlying the function search space directly from the data. Our approach is based on learning multiple matrix-valued kernels, each of those composed of a set of input kernels and a set of output kernels learned in the cone of positive semi-definite matrices. In addition to superior predictive performance in the presence of strong non-linearities, our method also recovers the hidden dynamic relationships between the series and thus is a new alternative to existing graphical Granger techniques.

15:20-15:40 UAPD: Predicting Urban Anomalies from Spatial-Temporal Data

Xian Wu, Yuxiao Dong, Chao Huang, Jian Xu, Dong Wang, Nitesh Chawla

Urban anomalies such as noise complaints and potholes in streets negatively affect our everyday life and need to be addressed in a timely manner. While significant effort has been made in detecting whether there exist anomalies in current urban data, the prediction of future urban anomalies is much less well explored and understood. In this work, we formalize the anomaly prediction problem in urban environments, such that those can be addressed in a more timely and efficient manner. We develop the Urban Anomaly PreDiction (UAPD) framework, which addresses a number of challenges, including the dynamics of different categories of anomalies, as well as varied spatial-temporal distributions. Given the urban anomaly data to date, UAPD first detects its change point in the temporal dimension, then a tensor decomposition based model decouples the interrelations among the spatial, temporal, and categorical dimensions of the urban data to detect the potential anomalies. Subsequently, based on the latent tensor of each dimension, we apply the autoregression method to predict which category of anomalies will happen at each region in the future. We perform extensive experiments to evaluate the proposed UAPD framework in two urban environments, namely New York City and Pittsburgh. Experimental results demonstrate that UAPD outperforms alternative baselines across various settings, including different region and time-frame scales, as well as diverse categories of anomalies.



Applied Data Science 6

CONGRESS HALL 4

14:00-14:20 Predicting Self-Reported Customer Satisfaction of Interactions with a Corporate Call Center

Joseph Bockhorst, Shi Yu, Luisa Polania Cabrera, Glenn Fung Moo



Timely identification of dissatisfied customers would provide corporations and other customer serving enterprises the opportunity to take meaningful interventions. This work describes a fully operational system we have developed at a large US insurance company for predicting customer satisfaction following all incoming phone calls at our call center. To capture call relevant information, we integrate signals from multiple heterogeneous data sources including: speech-to-text transcriptions of calls, call metadata (duration, waiting time, etc.), customer profiles and insurance policy information. Because of its ordinal, subjective, and often highly-skewed nature, self-reported survey scores presents several modeling challenges. To deal with these issues we introduce a novel modeling workflow: First, a ranking model is trained on the customer call data fusion. Then, a convolutional fitting function is optimized to map the ranking scores to actual survey satisfaction scores. This approach produces more accurate predictions than standard regression and classification approaches that directly fit the survey scores with call data, and can be easily generalized to other customer satisfaction prediction problems. Source code and data are available at https://github.com/cyberyu/ ecml2017.

14:20-14:40 Have It Both Ways - from A/B Testing to A&B Testing with Exceptional Model Mining



Wouter Duivesteijn, Tara Farzami, Thijs Putman, Evertjan Peer, Hilde J.P. Weerts, Jasper Adegeest, Gerson Foks, Mykola Pechenizkiy

In traditional A/B testing, we have two variants of the same product, a pool of test subjects, and a measure of success. In a randomized experiment, each test subject is presented with one of the two variants, and the measure of success is aggregated per variant. The variant of the product associated with the most success is retained, while the other variant is discarded. This, however, presumes that the company producing the products only has enough capacity to maintain one of the two product variants. If more capacity is available, then advanced data science techniques can extract more profit for the company from the A/B testing results. Exceptional Model Mining is one such advanced data science technique, which specializes in identifying subgroups that behave differently from the overall population. Using the association model class for EMM, we can find subpopulations that prefer variant A where the general population prefers variant B, and vice versa. This data science technique is applied on data from StudyPortals, a global study choice platform that ran an A/B test on the design of aspects of their website.

14:40-15:00 Structural Semantic Models for Automatic Analysis of Urban Areas



The growing availability of data from cities (e.g., traffic flow, human mobility and geographical data) open new opportunities for predicting and thus optimizing human activities. For example, the automatic analysis of land use enables the possibility of better administrating a city in terms of resources and provided services. However, such analysis requires specific information, which is often not available for privacy concerns. In this paper, we propose a novel machine learning representation based on the available public information to classify the most predominant land use of an urban area, which is a very common task in urban computing. In particular, in addition to standard feature vectors, we encode geo-social data from Location-Based Social Networks (LBSNs) into a conceptual tree structure that we call Geo-Tree. Then, we use such representation in kernel machines, which can thus perform accurate classification exploiting hierarchical substructure of concepts as features. Our extensive comparative study on the areas of New York and its boroughs shows that Tree Kernels applied to Geo-Trees are very effective improving the state of the art up to 18% in Macro-F1.

15:00-15:20 Using Machine Learning for Labour Market Intelligence



Roberto Boselli, Mirko Cesarini, Fabio Mercorio, Mario Mezzanzanica

The rapid growth of Web usage for advertising job positions provides a great opportunity for real-time labour market monitoring. This is the aim of Labour Market Intelligence (LMI), a field that is becoming increasingly relevant to EU Labour Market policies design and evaluation. The analysis of Web job vacancies, indeed, represents a competitive advantage to labour market stakeholders with respect to classical survey-based analyses, as it allows for reducing the time-to-market of the analysis by moving towards a fact-based decision making model. In this paper, we present our approach for automatically classifying million Web job vacancies on a standard taxonomy of occupations. We show how this problem has been expressed in terms of text classification via machine learning. Then, we provide details about the classification pipelines we evaluated and implemented, along with the outcomes of the validation activities. Finally, we discuss how machine learning contributed to the LMI needs of the European Organisation that supported the project.

15:20 - 15:40 SINAS: Suspect Investigation Using Offenders' Activity Space

Mohammad Tayebi, Uwe Glässer, Patricia Brantingham, Hamed Yaghoubi Shahir

Suspect investigation as a critical function of policing determines the truth about how a crime occurred, as far as it can be found. Understanding of the environmental elements in the causes of a crime incidence inevitably improves the suspect investigation process. Crime pattern theory concludes that offenders, rather than venture into unknown territories, frequently commit opportunistic and serial violent crimes by taking advantage of opportunities they encounter in places they are most familiar with as part of their activity space. In this paper, we present a suspect investigation method, called SINAS, which learns the activity space of offenders using an extended version of the random walk method based on crime pattern theory, and then recommends the top-K potential suspects for a committed crime. Our experiments on a large real-world crime dataset show that SINAS outperforms the baseline suspect investigation methods we used for the experimental evaluation.

Probabilistic Models & Methods 3

Knowledge Elicitation via Seguential Probabilistic Inference for High-16:00 - 16:20 **Dimensional Prediction**

Pedram Daee, Tomi Peltola, Marta Soare, Samuel Kaski

Prediction in a small-sized sample with a large number of covariates, the "small n, large p" problem, is challenging. This setting is encountered in multiple applications, such as in precision medicine, where obtaining additional data can be extremely costly or even impossible, and extensive research effort has recently been dedicated to finding principled solutions for accurate prediction. However, a valuable source of additional information, domain experts, has not yet been efficiently exploited. We formulate knowledge elicitation generally as a probabilistic inference process, where expert knowledge is sequentially queried to improve predictions. In the specific case of sparse linear regression, where we assume the expert has knowledge about the relevance of the covariates, or of values of the regression coefficients, we propose an algorithm and computational approximation for fast and efficient interaction, which sequentially identifies the most informative features on which to query expert knowledge. Evaluations of the proposed method in experiments with simulated and real users show improved prediction accuracy already with a small effort from the expert.

16:20 - 16:40 Labeled DBN Learning with Community Structure Knowledge

Etienne Auclair, Nathalie Peyrard, Régis Sabbadin

Learning interactions between dynamical processes is a widespread but difficult problem in ecological or human sciences. Unlike in other domains (bioinformatics, for example), data is often scarce, but expert knowledge is available. We consider the case where knowledge is about a limited number of interactions that drive the processes dynamics, and on a community structure in the interaction network. We propose an original framework, based on Dynamic Bayesian Networks with labeled-edge structure and parsimonious parameterization, and a Stochastic Block Model prior, to integrate this knowledge. Then we propose a restoration-estimation algorithm, based on 0-1 Linear Programing, that improves network learning when these two types of expert knowledge are available. The approach is illustrated on a problem of ecological interaction network learning.

Online Sparse Collapsed Hybrid Variational-Gibbs Algorithm for 16:40 - 17:00 **Hierarchical Dirichlet Process Topic Models**

Sophie Burkhardt, Stefan Kramer

Topic models for text analysis are most commonly trained using either Gibbs sampling or variational Bayes. Recently, hybrid variational-Gibbs algorithms have been found to combine the best of both worlds. Variational algorithms are fast to converge and more efficient for inference on new documents. Gibbs sampling enables sparse updates since each token is only associated with one topic instead of a distribution over all topics. Additionally, Gibbs sampling is unbiased. Although Gibbs sampling takes longer to converge, it is guaranteed to arrive at the true posterior after infinitely many iterations. By combining the two methods it is possible to reduce the bias of variational methods while simultaneously speeding up variational updates. This idea has previously been applied to standard latent Dirichlet allocation (LDA). We propose a new sampling method that enables the application of the idea to the nonparametric version of LDA, hierarchical Dirichlet process topic models. Our fast sampling method leads to a significant speedup of variational updates as compared to other sampling methods. Experiments show that training of our topic model converges to a better loglikelihood than previously existing variational methods and converges faster than Gibbs sampling in the batch setting.



CONGRESS HALL 1



111

THURSDAY **SEPTEMBER 21**

17:00 - 17:20 Partial Device Fingerprints

Michael Ciere, Carlos Ganan, Michel van Eeten

In computing, remote devices may be identified by means of device fingerprinting, which works by collecting a myriad of client-side attributes such as the device's browser and operating system version, installed plugins, screen resolution, hardware artifacts, Wi-Fi settings, and anything else available to the server, and then merging these attributes into uniquely identifying fingerprints. This technique is used in practice to present personalized content to repeat website visitors, detect fraudulent users, and stop masquerading attacks on local networks. However, device fingerprints are seldom uniquely identifying. They are better viewed as partial device fingerprints, which do have some discriminatory power but not enough to uniquely identify users. How can we infer from partial fingerprints whether different observations belong to the same device? We present a mathematical formulation of this problem that enables probabilistic inference of the correspondence of observations. We set out to estimate a correspondence probability for every pair of observations that reflects the plausibility that they are made by the same user. By extending probabilistic data association techniques previously used in object tracking, traffic surveillance and citation matching, we develop a general-purpose probabilistic method for estimating correspondence probabilities with partial fingerprints. Our approach exploits the natural variation in fingerprints and allows for use of situation-specific knowledge through the specification of a generative probability model. Experiments with a real-world dataset show that our approach gives calibrated correspondence probabilities. Moreover, we demonstrate that improved results can be obtained by combining device fingerprints with behavioral models.

Vine Copulas for Mixed Data : Multi-View Clustering for Mixed Data 17:20 - 17:40 **Beyond Meta-Gaussian Dependencies**

Lavanya Sita Tekumalla, Vaibhav Rajan, Chiranjib Bhattacharyya

Copulas enable flexible parameterization of multivariate distributions in terms of constituent marginals and dependence families. Vine copulas, hierarchical collections of bivariate copulas, can model a wide variety of dependencies in multivariate data including asymmetric and tail dependencies which the more widely used Gaussian copulas, used in Meta-Gaussian distributions, cannot. However, current inference algorithms for vines cannot fit data with mixed-a combination of continuous, binary and ordinal-features that are common in many domains. We design a new inference algorithm to fit vines on mixed data thereby extending their use to several applications. We illustrate our algorithm by developing a dependency-seeking multi-view clustering model based on Dirichlet Process mixture of vines that generalizes previous models to arbitrary dependencies as well as to mixed marginals. Empirical results on synthetic and real datasets demonstrate the performance on clustering single-view and multi-view data with asymmetric and tail dependencies and with mixed marginals.

Matrix & Tensor Factorization

C-SALT: Mining Class-Specific ALTerations in Boolean Matrix Factorization

Sibylle Hess, Katharina Morik

Given labeled data represented by a binary matrix, we consider the task to derive a Boolean matrix factorization which identifies commonalities and specifications among the classes. While existing works focus on rank-one factorizations which are either specific or common to the classes, we derive class-specific alterations from common factorizations as well. Therewith, we broaden the applicability of our new method to datasets whose class-dependencies have a more complex structure. On the basis of synthetic and real-world datasets, we show on the one hand that our method is able to filter structure which corresponds to our model assumption, and on the other hand that our model assumption is justified in real-world application. Our method is parameter-free.

16:20 - 16:40

16:00 - 16:20

Comparative Study of Inference Methods for Bayesian Nonnegative Matrix Factorisation

Thomas Brouwer, Jes Frellsen, Pietro Lio

In this paper, we study the trade-offs of different inference approaches for Bayesian matrix factorisation methods, which are commonly used for predicting missing values, and for finding patterns in the data. In particular, we consider Bayesian nonnegative variants of matrix factorisation and tri-factorisation, and compare non-probabilistic inference, Gibbs sampling, variational Bayesian inference, and a maximum-a-posteriori approach. The variational approach is new for the Bayesian nonnegative models. We compare their convergence, and robustness to noise and sparsity of the data, on both synthetic and real-world datasets. Furthermore, we extend the models with the Bayesian automatic relevance determination prior, allowing the models to perform automatic model selection, and demonstrate its efficiency.









Content-Based Social Recommendation with Poisson Matrix 16:40 - 17:00

Factorization

Eliezer De Souza da Silva, Helge Langseth, Heri Ramampiaro

We introduce Poisson Matrix Factorization with Content and Social trust information (PoissonMF-CS), a latent variable probabilistic model for recommender systems with the objective of jointly modeling social trust, item content and user's preference using Poisson matrix factorization framework. This probabilistic model is equivalent to collectively factorizing a non-negative user--item interaction matrix and a non-negative item-content matrix. The user--item matrix consists of sparse implicit (or explicit) interactions counts between user and item, and the item--content matrix consists of words or tags counts per item. The model imposes additional constraints given by the social ties between users, and the homophily effect on social networks -- the tendency of people with similar preferences to be socially connected. Using this model we can account for and fine-tune the weight of content-based and social-based factors in the user preference. We develop approximate variational inference algorithm and perform experiments comparing PoissonMF-CS with competing models. The experimental evaluation indicates that PoissonMF-CS achieves superior predictive performance on held-out data for the top-M recommendations task. Also, we observe that PoissonMF-CS generates compact latent representations when compared with alternative models while maintaining superior predictive performance.

Feature Extraction for Incomplete Data via Low-Rank Tucker 17:00 - 17:20 Decomposition

Qiquan Shi, Yiu-ming Cheung, Qibin Zhao

Extracting features from incomplete tensors is a challenging task which is not well explored. Due to the data with missing entries, existing feature extraction methods are not applicable. Although tensor completion techniques can estimate the missing entries well, they focus on data recovery and do not consider the relationships among tensor samples for effective feature extraction. To solve this problem of feature extraction for incomplete data, we propose an unsupervised method, TDVM, which incorporates low-rank Tucker Decomposition with feature Variance Maximization in a unified framework. Based on Tucker decomposition, we impose nuclear norm regularization on the core tensors while minimizing reconstruction errors, and meanwhile maximize the variance of core tensors (i.e., extracted features). Here, the relationships among tensor samples are explored via variance maximization while estimating the missing entries. We thus can simultaneously obtain lower-dimensional core tensors and informative features directly from observed entries. The alternating direction method of multipliers approach is utilized to solve the optimization objective. We evaluate the features extracted from two real data with different missing entries for face recognition tasks. Experimental results illustrate the superior performance of our method with a significant improvement over the state-of-the-art methods.

Structurally Regularized Non-Negative Tensor Factorization for 17:20 - 17:40 **Spatio-temporal Pattern Discoveries**

Koh Takeuchi, Yoshinobu Kawahara, Tomoharu Iwata

Understanding spatio-temporal activities in a city is a typical problem of spatio-temporal data analysis. For this analysis, tensor factorization methods have been widely applied for extracting a few essential patterns into latent factors. Non-negative Tensor Factorization (NTF) is popular because of its capability of learning interpretable factors from non-negative data, simple computation procedures, and dealing with missing observation. However, since existing NTF methods are not fully aware of spatial and temporal dependencies, they often fall short of learning latent factors where a large portion of missing observation exist in data. In this paper, we present a novel NTF method for extracting smooth and flat latent factors by leveraging various kinds of spatial and temporal structures. Our method incorporates a unified structured regularizer into NTF that can represent various kinds of auxiliary information, such as an order of timestamps, a daily and weekly periodicity, distances between sensor locations, and areas of locations. For the estimation of the factors for our model, we present a simple and efficient optimization procedure based on the alternating direction method of multipliers. In missing value interpolation experiments of traffic flow data and bike-sharing system data, we demonstrate that our proposed method improved interpolation performances from existing NTF, especially when a large portion of missing values exists.

Transfer & Multi-Task Learning 2

CONGRESS HALL 3

A Novel Rating Pattern Transfer Model for Improving Non-Overlapping 16:00 - 16:20 **Cross-Domain Collaborative Filtering** Yizhou Zang, Xiaohua Hu

Cross-Domain Collaborative Filtering (CDCF) has attracted various research works in recent years. However, an important problem setting, i.e., "users and items in source and target domains are totally different", has not received much attention yet. We coin this problem as Non-Overlapping Cross-Domain Collaborative Filtering (NOCDCF). In order to solve this challenging CDCF task, we propose a novel 3-step rating pattern transfer model, i.e. low-rank knowledge transfer via factorization machines (LKT-FM). Our solution is able to mine high quality knowledge from large and sparse source matrices, and to integrate the knowledge without losing much information contained in the target matrix via exploiting Factorization Machine (FM). Extensive experiments on real world datasets show that the proposed LKT-FM model outperforms the state-of-the-art CDCF solutions.







THURSDAY **SEPTEMBER 21**

Distributed Multi-Task Learning for Sensor Network 16:20 - 16:40

Jiyi Li, Tomohiro Arai, Yukino Baba, Hisashi Kashima, Shotaro Miwa

A sensor in a sensor network is expected to be able to make prediction or decision utilizing the models learned from the data observed on this sensor. However, in the early stage of using a sensor, there may be not a lot of data available to train the model for this sensor. A solution is to leverage the observation data from other sensors which have similar conditions and models with the given sensor. We thus propose a novel distributed multi-task learning approach which incorporates neighborhood relations among sensors to learn multiple models simultaneously in which each sensor corresponds to one task. It may be not cheap for each sensor to transfer the observation data from other sensors; broadcasting the observation data of a sensor in the entire network is not satisfied for the reason of privacy protection; each sensor is expected to make real-time prediction independently from neighbor sensors. Therefore, this approach shares the model parameters as regularization terms in the objective function by assuming that neighbor sensors have similar model parameters. We conduct the experiments on two real datasets by predicting the temperature with the regression. They verify that our approach is effective, especially when the bias of an independent model which does not utilize the data from other sensors is high such as when there is not plenty of training data available.

16:40 - 17:00 Learning Task structure via sparsity grouped multitask learning

Meghana Kshirsagar, Eunho Yang, Aurelie Lozano

Sparse mapping has been a key methodology in many high-dimensional scientific problems. When multiple tasks share the set of relevant features, learning them jointly in a group drastically improves the quality of relevant feature selection. However, in practice this technique is used limitedly since such grouping information is usually hidden. In this paper, our goal is to recover the group structure on the sparsity patterns and leverage that information in the sparse learning. Toward this, we formulate a joint optimization problem in the task parameter and the group membership, by constructing an appropriate regularizer to encourage sparse learning as well as correct recovery of task groups. We further demonstrate that our proposed method recovers groups and the sparsity patterns in the task parameters accurately by extensive experiments.

17:00 - 17:20 Ranking Based Multitask Learning of Scoring Functions

Ivan Stojkovic, Mohamed Ghalwash, Zoran Obradovic

Scoring functions are an important tool for quantifying properties of interest in many domains; for example, in healthcare, a disease severity scores are used to diagnose the patient's condition and to decide its further treatment. Scoring functions might be obtained based on the domain knowledge or learned from data by using classification, regression or ranking techniques - depending on the type of supervised information. Although learning scoring functions from collected data is beneficial, it can be challenging when limited data are available. Therefore, learning multiple distinct, but related, scoring functions together can increase their quality as shared regularities may be easier to identify. We propose a multitask formulation for ranking-based learning of scoring functions, where the model is trained from pairwise comparisons. The approach uses mixed-norm regularization to impose structural regularities among the tasks. The proposed regularized objective function is convex; therefore, we developed an optimization approach based on alternating minimization and proximal gradient algorithms to solve the problem. The increased predictive accuracy of the presented approach, in comparison to several baselines, is demonstrated on synthetic data and two different realworld applications; predicting exam scores and predicting tolerance to infections score.

Computer Vision

16:00 - 16:20

Alternative Semantic Representations for Zero-Shot Human Action Recognition

Qian Wang, Ke Chen

A proper semantic representation for encoding side information is key to the success of zero-shot learning. In this paper, we explore two alternative semantic representations especially for zero-shot human action recognition: textual descriptions of human actions and deep features extracted from still images relevant to human actions. Such side information are accessible on Web with little cost, which paves a new way in gaining side information for large-scale zero-shot human action recognition. We investigate different encoding methods to generate semantic representations for human actions from such side information. Based on our zero-shot visual recognition method, we conducted experiments on UCF101 and HMDB51 to evaluate two proposed semantic representations . The results suggest that our proposed text- and image-based semantic representations outperform traditional attributes and word vectors considerably for zero-shot human action recognition. In particular, the image-based semantic representations yield the favourable performance even though the representation is extracted from a small number of images per class.







CONGRESS HALL 4



16:20-16:40 Early Active Learning with Pairwise Constraint for Person Re-Identification

Wenhe Liu, xiaojun Chang, Ling Chen, Yi Yang, Alexander Hauptmann

Research on person re-identification (re-id) has attached much attention in the machine learning field in recent years. With sufficient labeled training data, supervised re-id algorithm can obtain promising performance. However, producing labeled data for training supervised re-id models is an extremely challenging and time-consuming task because it requires every pair of images across no-overlapping camera views to be labeled. Moreover, in the early stage of experiments, when labor resources are limited, only a small number of data can be labeled. Thus, it is essential to design an effective algorithm to select the most representative samples. This is referred as early active learning or early stage experimental design problem. The pairwise relationship plays a vital role in the re-id problem, but most of the existing early active learning algorithms fail to consider this relationship. To overcome this limitation, we propose a novel and efficient early active learning algorithm with a pairwise constraint for person re-identification in this paper. By introducing the pairwise constraint, the closeness of similar representations of instances is enforced in active learning. This benefits the performance of active learning for re-id. Extensive experimental results on four benchmark datasets confirm the superiority of the proposed algorithm.

16:40 - 17:00 Guiding InfoGAN with Semi-Supervision

Adrian Spurr, Emre Aksan, Otmar Hilliges

In this paper we propose a new semi-supervised GAN architecture (ss-InfoGAN) for image synthesis that leverages information from few labels (as little as 0.22%, max. 10% of the dataset) to learn semantically meaningful and controllable data representations where latent variables correspond to label categories. The architecture builds on Information Maximizing Generative Adversarial Networks (InfoGAN) and is shown to learn both continuous and categorical codes and achieves higher quality of synthetic samples compared to fully unsupervised settings. Furthermore, we show that using small amounts of labeled data speeds-up training convergence. The architecture maintains the ability to disentangle latent variables for which no labels are available. Finally, we contribute an information-theoretic reasoning on how introducing semi-supervision increases mutual information between synthetic and real data.

17:00 - 17:20 Scatteract: Automated Extraction of data from scatter plots

Mathieu Cliche, David Rosenberg, Connie Yee, Dhruv Madeka

Charts are an excellent way to convey patterns and trends in data, but they do not facilitate further modeling of the data or close inspection of individual data points. We present a fully automated system for extracting the numerical values of data points from images of scatter plots. We use deep learning techniques to identify the key components of the chart, and optical character recognition together with robust regression to map from pixels to the coordinate system of the chart. We focus on scatter plots with linear scales, which already have several interesting challenges. Previous work has done fully automatic extraction for other types of charts, but to our knowledge this is the first approach that is fully automatic for scatter plots. Our method performs well, achieving successful data extraction on 89% of the plots in our test set.

17:20 - 17:40 Unsupervised Diverse Colorization via Generative Adversarial Networks

Yun Cao, Zhiming Zhou, Weinan Zhang, Yong Yu

Colorization of grayscale images is a hot topic in computer vision. Previous research mainly focuses on producing a color image to recover the original one in a supervised learning fashion. However, since many colors share the same gray value, an input grayscale image could be diversely colorized while maintaining its reality. In this paper, we design a novel solution for unsupervised diverse colorization. Specifically, we leverage conditional generative adversarial networks to model the distribution of real-world item colors, in which we develop a fully convolutional generator with multi-layer noise to enhance diversity, with multi-layer condition concatenation to maintain reality, and with stride 1 to keep spatial information. With such a novel network architecture, the model yields highly competitive performance on the open LSUN bedroom dataset. The Turing test on 80 humans further indicates our generated color schemes are highly convincible.

115





TUESDAY



Monitoring Physical Activity and Mental Stress Using Wrist-worn Device and a Smartphone

Božidara Cvetković, Martin Gjoreski, Jure Šorn, Pavel Maslov, Mitja Luštrek

The paper presents a smartphone application for monitoring physical activity and mental stress. The application utilizes sensor data from a wristband and/or a smartphone, which can be worn in various pockets or in a bag in any orientation. The presence and location of the devices are used as contexts for the selection of appropriate machine-learning models for activity recognition and the estimation of human energy expenditure. The stress-monitoring method uses two machine-learning models, the first one relying solely on physiological sensor data and the second one incorporating the output of the activity monitoring and other context information. The evaluation showed that we recognize a wide range of atomic activities with an accuracy of 87%, and that we outperform the state-of-the art consumer devices in the estimation of energy expenditure. In stress monitoring we achieved an accuracy of 92% in a real-life setting.

Lit@EVE: Explainable Recommendation Based on Wikipedia Concept Vectors

Muhammad Atif Qureshi, Derek Greene

We present an explainable recommendation system for novels and authors, called Lit@EVE, which is based on Wikipedia concept vectors. In this system, each novel or author is treated as a concept whose definition is extracted as a concept vector through the application of an explainable word embedding technique called EVE. Each dimension of the concept vector is labelled as either a Wikipedia article or a Wikipedia category name, making the vector representation readily interpretable. In order to recommend items, the Lit@EVE system uses these vectors to compute similarity scores between a target novel or author and all other candidate items. Finally, the system generates an ordered list of suggested items by showing the most informative features as human-readable labels, thereby making the recommendation explainable.

TF Boosted Trees: A Scalable TensorFlow based framework for Gradient boosting

Natalia Ponomareva, Soroush Radpour, Gilbert Hendry, Salem Haykal, Thomas Colthurst, Petr Mitrichev, Alexander Grushetsky

TF Boosted Trees (TFBT) is a new open-sourced framework for the distributed training of gradient boosted trees. It is based on TensorFlow, and its distinguishing features include a novel architecture, automatic loss differentiation, layer-by-layer boosting that results in smaller ensembles and faster prediction, principled multi-class handling, and a number of regularization techniques to prevent overfitting.

ASK-the-Expert: Active Learning based knowledge discovery using the expert

Kamalika Das, Ilya Avrekh, Bryan Matthews, Manali Sharma, Nikunj Oza

Often the manual review of large data sets, either for purposes of labeling unlabeled instances or for classifying meaningful results from uninteresting (but statistically significant) ones is extremely resource intensive, especially in terms of subject matter expert (SME) time. Use of active learning has been shown to reduce this review time significantly. However, since active learning is an iterative process of learning a classifier based on a small number of SME-provided labels at each iteration, the lack of an enabling tool can hinder the process of adoption of these technologies in real-life, in spite of their labor-saving potential. In this demo we present ASK-the-Expert, an interactive tool that allows SMEs to review instances from a data set and provide labels within a single framework. ASK-the-Expert is powered by an active learning algorithm for training a classifier in the backend. We demonstrate this system in the context of an aviation safety application, but the tool can be adopted to work as a simple review and labeling tool as well, without the use of active learning.

TrajViz: A Tool for Visualizing Patterns and Anomalies in Trajectory

Yifeng Gao, Qingzhe Li, Xiaosheng Li, Jessica Lin, Huzefa Rangwala, Ranjeev Mittu

Visualizing frequently occurring patterns and potentially unusual behaviors in trajectory can provide valuable insights into activities behind the data. In this paper, we introduce TrajViz, a motif (frequently repeated subsequences) based visualization software that detects patterns and anomalies by inducing ``grammars" from discretized spatial trajectories. We consider patterns as a set of sub-trajectories with unknown lengths that are spatially similar to each other. We demonstrate that TrajViz has the capacity to help users visualize anomalies and patterns effectively.







MONDAY
TUESDAY

117

FRDAY

PROGRAM AT A GLANCE



PhD Forum

BANQUET HALL

09:00 - 15:40

PhD Forum

Chairs : Tomislav Šmuc, Rudjer Boškovič Institute, Croatia Bernard Ženko, Jožef Stefan Institute, Slovenia

The PhD forum aims at attracting junior PhD students at the beginning of their scientific career to present and discuss their work in progress, and to receive useful comments, advice and ideas for their further work. The forum will also be attended by senior researchers with experience in supervising PhD students who will provide constructive feedback to the participants. This will be an excellent opportunity for developing person-to-person networks of PhD students, which can be extremely valuable for their future careers.

Sponsored by

IBM Research

Invited talk 50 WAYS TO TWEAK YOUR PAPER

Johannes Fürnkranz

TU Darmstadt, Germany

Very often, we see papers that have good research ideas be rejected because the quality of the write-up does not quite live up to the quality of the idea. Whether you like it or not, the presentation of your work can often make the difference that tilts a borderline paper one way or the other. In particular for conference-style reviewing, where reviewers have to make recommendations for multiple papers in a very narrow time frame, they are often influenced by the writing and the appearance of the paper (whether they like it or not). In this talk, targeted towards junior Ph. D. students, we will make a few suggestions how to improve the presentation of your work. Most of them are obvious, but we nevertheless often see them violated in practice.

Invited talk TIPS FOR A SUCCESSFUL PHD, AND HOW TO WIN AN AWARD WITH IT

Tias Guns

Vrije Universiteit Brussel (VUB), Belgium

The talk will dive into some of the every day challenges a PhD student faces: working with your promoter, writing a paper, conference visits, how to get your work more widely known and up to winning an award with your thesis. The talk will have anecdotes based on my or colleagues' experiences and indispensable references to PhD comics, XKCD and other highly valuable sources.

09:00	Opening Tomislav Šmuc, Bernard Ženko	PhD Forum
	SESSION 1	Program at a giance
09:00 - 09:35	50 Ways to Tweak Your Paper Johannes Fürnkranz	
09:35 - 09:45	Automatic Object-Oriented Curriculum Generation for Reinforceme Felipe Leno Da Silva, Anna Helena Real Costa	nt Learning
09:45 - 09:55	Scalable Generalized Dynamic Topic Models Florian Wenzel, Patrick Jähnichen, Marius Kloft, Stephan Mandt	
09:55 - 10:05	Heterogeneous Network Analysis for Semantic Data Mining Jan Kralj, Marko Robnik-Šikonja and Nada Lavrač	
10:05 - 10:15	Burg Matrix Divergence Based Multi-Metric Learning Yan Wang, Han-Xiong	
10:15 - 10:21	Automatically Correcting Semantic Errors in Programming Assignme Ben Trevett, Donald Reay, Nick Taylor	ents
10:21 - 10:27	Gene Network Reconstruction Via Transfer Learning Across Multiple Paolo Mignone, Gianvito Pio, Michelangelo Ceci	e Organisms
10:27 - 10:33	Temporal Pattern Mining from Evolving Networks Angelo Impedovo, Corrado Loglisci, Michelangelo Ceci	
10:33 - 10:40	Session Wrap-Up	
10:40 - 11:00	Coffee & Tea Break	
	SESSION 2	
11:00 - 11:35	Tips for a Successful PhD, and How to Win an Award with It Tias Guns	t
11:35 - 11:45	Integrating Knowledge Graph Information Into Neural Conversationa Boulos El Asmar	al Models
11:45 - 11:55	Adaptive Shallow Architectures for Streamed Data with Parallel Dee Craig Bower, Ashiq Anjum	ep Incremental Feature Learning
11:55 - 12:05	Differentiating Between Exogenous and Endogenous Information Pr Matija Piškorec, Tomislav Šmuc, Mile Šikić	ropagation in Social Networks
12:05 - 12:15	Construction and Exploration of Redescription Sets Matej Mihelčić, Sašo Džeroski, Nada Lavrač, Tomislav Šmuc	
12:15 - 12:21	Synergy of Natural Language Processing and Statistics to Explore F Data and Knowledge Tome Eftimov, Peter Korošec, Barbara Koroušić Seljak	Food- and Nutrition-Related
12:21 - 12:27	Feature Ranking in Structured Data Matej Petković, Sašo Džeroski	
12:27 - 12:40	Session and PhD Forum Wrap-Up	
12:40 - 14:00	Lunch Break	
14:00 - 15:40	Poster Session	

TUESDAY

WEDNESDAY

THURSDAY

FRIDAY

Combined Tutorial with Workshop

CONGRESS HALL 1

09:00 - 18:40

AutoML Automatic Selection, Configuration and Composition of Machine Learning Algorithms

Organizers : Pavel Brazdil, LIAAD Inesc Tec., Portugal Joaquin Vanschoren, Eindhoven University of Technology, The Netherlands Holger H. Hoos, Universiteit Leiden, The Netherlands Frank Hutter, University of Freiburg, Germany

Tutorial Webpage

https://sites.google.com/site/automl2017ecmlpkdd/tutorial/

Workshop Webpage

https://sites.google.com/site/automl2017ecmlpkdd/workshop/

Tutorial

This tutorial will introduce and discuss state of the art methods in meta-learning, algorithm selection, and algorithm configuration, which are increasingly relevant today. Researchers and practitioners from all areas of science and technology face a large choice of parameterized machine learning algorithms, with little guidance as to when and how to use which technique. Data mining challenges frequently remind us that algorithm selection and configuration play a crucial role in achieving cutting-edge performance, and are indispensable in industrial applications.

Meta-learning leverages knowledge of past applications of algorithms applications to learn how to select the best techniques for future applications, and offers effective techniques that are superior to humans both in terms of the quality of the end result and even more so in the time required to achieve it. Recent approaches include also (preferably very fast) partial probing runs on a given problem with the aim of determining the best strategy to be used from there onwards. This may include further probing or recommending an algorithm to be used to solve the given problem. A recent trend is to incorporate such techniques into software platforms. This synergy leads to new advances that recommend combinations of algorithms and hyperparameter settings simultaneously, and that speed up algorithm configuration by learning which parameter settings are likely most useful for dealing with the data at hand.

After motivating and introducing the concepts of algorithm selection and configuration, we elucidate how they arise in machine learning and data mining, but also in other domains, such as optimization. We demonstrate how meta-learning techniques can be effectively used in this context, exploiting information gleaned from past experiments as well as by probing the data at hand. Moreover, many current applications require the use of machine learning or data mining workflows that consist of several different processes or operations. Constructing such complex systems or workflows requires extensive expertise, as well as existing meta-data and software, and can be greatly facilitated by leveraging the methodologies presented at this tutorial.

Workshop

This workshop will provide a platform for discussing the recent developments in the area of algorithm selection and configuration, which arises in many diverse domains, such as machine learning, data mining, optimization and automated reasoning. Algorithm selection and configuration are increasingly relevant today. Researchers and practitioners from all areas of science and technology face a large choice of parameterized machine learning algorithms, with little guidance as to which techniques to use in a given application context. Moreover, data mining challenges frequently remind us that algorithm selection and configuration are crucial in order to achieve cutting-edge performance, and drive industrial applications.

Meta-learning leverages knowledge of past algorithm applications to select the best techniques for future applications, and offers effective techniques that are superior to humans both in terms of the end result and especially in the time required to achieve it. In this workshop, we will discuss different ways of exploiting meta-learning techniques to identify the potentially best algorithm(s) for a new task, based on meta-level information, prior experiments on both past datasets and the current one. Many contemporary problems require the use of workflows that consist of several processes or operations. Constructing such complex workflows requires extensive expertise, and could be greatly facilitated by leveraging planning, meta-learning and intelligent system design. This task is inherently interdisciplinary, as it builds on expertise in various areas of AI.

Invited talk

STOCHASTIC GRADIENT DESCENT: GOING AS FAST AS POSSIBLE BUT NOT FASTER

Michele Sebag

CNRS, France

AutoML Program at a glance

09:00	Opening Pavel Brazdil, Joaquin Vanschoren, Frank Hutter, Holger Hoos	
	TUTORIAL SESSION 1	
09:00 - 10:30	Introduction Blackbox Hyperparameter Optimization and Automated Machine Learning Algorithm Configuration Meta-Learning for Algorithm Selection Pavel Brazdil, Joaquin Vanschoren, Frank Hutter, Holger Hoos	MONDAY
10:40 - 11:00	Coffee & Tea Break	
	TUTORIAL SESSION 2	
11:00 - 12:40	Beyond Blackbox Optimization Automating Workflow Design AutoML Systems and Applications The Big Picture: Machine Learning for Automated Algorithm Design Pavel Brazdil, Joaquin Vanschoren, Frank Hutter, Holger Hoos	TUESDAY
12:40 - 14:00	Lunch Break	
	WORKSHOP SESSION 1	N
14:00 - 14:05	Welcome	ESI
14:05 - 14:45	Stochastic Gradient Descent: Going As Fast As Possible But Not Faster Michele Sebag	WEDN
14:45 - 15:00	Poster Spotlights 1	
15:00 - 15:40	Poster Session 1	~
15:40 - 16:00	Coffee & Tea Break	SDA
	WORKSHOP SESSION 2	l Si
16:00 - 16:25	Towards Automated Relational Data Wrangling G. Verbruggen, L. De Raedt	Ē
16:25 - 16:50	A Meta-Learning Approach to One-Step Active-Learning G. Contardo, L. Denoyer, T. Artieres	
16:50 - 17:05	Poster Spotlights 2	IDA
17:05 - 17:40	Poster Session 2	Æ
	WORKSHOP SESSION 3	
17:40 - 18:05	Poster Session 3	
18:05 - 18:35	Panel & Commentaries on Some Papers & Research Lines	
18:35 - 18:40	Closing	

CONGRESS HALL 2

09:00 - 19:00

DLPM 2017 Deep Learning for Precision Medicine

Organizers :

Cesare Furlanello, Fondazione Bruno Kessler, Italy Bertram Müller- Myhsok, University of Liverpool, UK Max Planck Institute of Psychiatry, Germany

Webpage

http://dlpm2017.fbk.eu/

Deep Learning is beginning to exert a disruptive impact for functional genomics, with applications of high industrial and ethical relevance in pharmacogenomics and toxicogenomics. Moreover, in less than one year, Deep Learning has emerged with solutions in diagnostic imaging and pathology that have reached best human expertise, as for the success in games. Examples in miRNA prediction already demonstrated the potential for deriving implicit features with high predictive accuracy, and novel methods for genomewide association studies and prediction of molecular traits following suite are appearing both as scientific initiatives as well as key technologies of AI startups. Following the success of DLPM2016, also colocated with ECML/PKDD in 2016, we thus wish to discuss about the best options for the adoption of deep learning models, both for improved accuracy as well as for better biomedical understanding. Questions such as end-to-end modeling from structure to functionality and biological impact as well as architectures for integration of genotype, expression and epigenetics would be of immediate interest for the workshop. Further topics of interest are described in the call.

This event is intended to be a one-day workshop within ECML-PKDD 2017 including lectures and discussions on this very timely, pressing and active subject. We aim to create a connection between machine learning experts and leaders in the Precision Medicine initiatives in Europe and the USA. In particular, the workshop aims to link experts from the FDA SEQC2 initiative on Precision Medicine, which will pave the way for defining optimal procedures for the development of actionable drugs that can target phenotype-selected patient groups. We wish to discuss also technical challenges such as working with very large cohorts (e.g. from 60K to 300K to 1M subjects in molecular psychiatry studies) that are now amenable for modeling with deep learning. Further, family cohorts will challenge machine learning and bioinformatics experts for new efficient solutions. In summary, both methodological aspects from deep learning, machine learning, information technology, statistics as well as actual applications, pitfalls and (medical) needs are to be featured.

Invited talk MULTIVARIATE PREDICTIVE MODELS IN PSYCHIATRY -PROMISES, PROBLEMS, PERSPECTIVES

Tim Hahn University of Münster, Germany

Invited talk TRAINING NEURAL NETWORKS TO CALL GENOTYPES, APPLICATIONS TO KIDNEY TRANSPLANTATION AND NEPHROLOGY

Fabien Campagne Cornell University, USA

Invited talk

SYSTEMS HEALTH - CHALLENGES AND OPPORTUNITIES

Kristel Van Steen Giga Liege Universite, Belgium

DLPM 2017 Program at a glance

09:00 - 09:10	DLPM2017 Opening Cesare Furlanello, Bertram Müller-Myhsok
	SESSION 1
09:10 - 10:00	Multivariate Predictive Models in Psychiatry - Promises, Problems, Perspectives Tim Hahn
10:00 - 10:25	Tree-LSTM and Cross-Corpus Training for Extracting Biomedical Relationships from Text Joël Legrand, Yannick Toussaint, Chedy Raïssi, Adrien Coulet
10:25 - 10:40	Prediction of Cancer of Unknown Primary Site (CUP) Using Deep Learning Methods Luyao Ren, Jingcheng Yang, Bin Li, Chen Suo, Ying Yu, Yuanting Zheng, Cesare Furlanello, Leming Shi
10:40 - 11:00	Coffee & Tea Break
11:00 - 11:40	Training Neural Networks to Call Genotypes, Applications to Kidney Transplantation and Nephrology Fabien Campagne
10:25 - 10:40	Augmenting Deep Learning with Random Forest for Stratification of Breast Cancer Recurrence Risk
	Noah Eyal-Altman, Mark Last, Eitan Rubin
10:25 - 10:40	Ph-CNN: a Keras Layer for Feature-Wise Structured Data Ylenia Giarratano, Claudio Agostinelli, Valerio Maggio, Diego Fioravanti, Marco Chierici, Calogero Zarbo, Giuseppe Jurman, Cesare Furlanello
12:40 - 14:00	Lunch Break
	SESSION 2
14:00 - 14:50	Systems Health – Challenges and Opportunities Kristel Van Steen
14:50 - 15:15	Deep Learning Based Integration of Clinical and Genomics Data Mooez Subhani
15:15 - 15:40	MinCall — MinION End2end Convolutional Deep Learning Basecaller Neven Miculinić, Marko Ratković, Mile Šikić
15:40 - 16:00	Coffee & Tea Break
16:00 - 16:25	Read Classification Using Semi-Supervised Deep Learning Tomislav Šebrek, Jan Tomljanović, Josip Krapac, Mile Šikić
16:25 - 18:30	Tutorial: Deep Learning for Precision Medicine with Keras and Tensorflow Valerio Maggio
18:30 - 19:00	Discussion and Summary – Perspectives – All Contributors and Audience Moderators: Kristel Van Steen, Fabien Campaigne, Tim Hahn, Cesare Furlanello, Bertram Müller-Myhsok
19:00	Closing

125

TUESDAY

WEDNESDAY

THURSDAY

FRIDAY

NFMCP 2017 New Frontiers in Mining Complex Patterns

Organizers: Annalisa Appice, University of Bari Aldo Moro, Italy Corrado Loglisci, University of Bari Aldo Moro, Italy Giuseppe Manco, ICAR-CNR, Italy Elio Masciari, ICAR-CNR, Italy Zbigniew W. Ras, University of North Carolina, USA

Webpage

http://www.di.uniba.it/~loglisci/NFmcp17/index.html

Modern automatic systems are able to collect huge volumes of data, often with a complex structure (e.g. multi-table data, XML data, web data, time series and sequences, graphs and trees). The massive and complex data pose new challenges for current research in Knowledge Discovery and Data Mining. They require new methods for storing, managing and analysing them by taking into account various complexity aspects: Complex structures (e.g. multi-relational, time series and sequences, networks, and trees) as input/output of the data mining process; Massive amounts of high dimensional data collections flooding as high-speed streams and requiring (near) real time processing and model adaptation to concept drifts; New application scenarios involving security issues, interaction with other entities and real-time response to events triggered by sensors.

The purpose of the workshop is to bring together researchers and practitioners of data mining and machine learning interested in analysis of complex data. We welcome submissions focusing on recent advances and latest developments in the analysis of complex and massive data sources such as blogs, event or log data, medical data, spatio-temporal data, social networks, mobility data, sensor data and streams. Submissions discussing and introducing new algorithmic foundations and representation formalisms in complex pattern discovery are also welcome. We encourage submissions from the areas of statistics, machine learning and big data analytics, which present techniques that take advantage of the informative richness of complex massive data for efficiently and effectively identifying new patterns. Finally, submissions describing preliminary and promising studies are also welcome.

Invited talk WHICH IS MORE INFLUENTIAL, "WHO" OR "WHEN" FOR A USER TO RATE IN ONLINE REVIEW SITE?

Hiroshi Motoda AFOSR/AOARD, USA and Osaka University, Japan

09:00 - 09:30	Opening Annalisa Appice, Corrado Loglisci, Giuseppe Manco, Elio Masciari, Elio Masciari		
	MORNING SESSION 1	NFMCP 20	17
09:30 - 10:40	Which Is More Influential, "Who" Or "When" for a User to Rate in Online Review Site? Hiroshi Motoda	Program at a gla	nce
10:40 - 11:00	Coffee & Tea Break		
	MORNING SESSION 2 - PROBABILISTIC MODELING		
11:00 - 11:20	Bayesian User Behavior Models Jan Reubold, Ahcène Boubekki, Thorsten Strufe, Ulf Brefeld		_
11:20 - 11:40	Generative Probabilistic Models for Positive-Unlabeled Learning Teresa M.A. Basile, Nicola Di Mauro, Floriana Esposito, Stefano Ferilli, Antonio Vergari		DAY
	MORNING SESSION 3 - CLASSIFICATION AND REGRESSION		INO
11:40 - 12:00	Complex Localization in the Multiple Instance Learning Context Dan-Ovidiu Graur, Razvan-Alexandru Maris, Rodica Potolea, Mihaela Dinsoreanu, Camelia	a Lemnaru	
12:00 - 12:20	Decomposition of the Output Space in Multi-Label Classification Using Feature Stevanche Nikoloski, Dragi Kocev, Sašo Džeroski	Ranking	7
12:20 - 12:40	Understanding Climate-Vegetation Interactions in Global Rainforests Through a GP Model-Tree Analysis		UESDA
	Anuradha Kodali, Marcin Szubert, Sangram Ganguly, Joshua Bongard, Kamalika Das		
12:40 - 14:00	Lunch Break		
	AFTERNOON SESSION 1 - BIOINFORMATICS		X
14:00 - 14:20	Identifying LncRNA-Disease Relationships Via Heterogeneous Clustering Emanuele Pio Barracchia, Gianvito Pio, Donato Malerba, Michelangelo Ceci		INESD
14:20 - 14:40	Phenotype Prediction with Semi-Supervised Learning Jurica Levatić, Maria Brbić, Tomaž Stepišnik Perdih, Dragi Kocev, Vedrana Vidulin, Tomisla Džeroski	av Šmuc, Fran Supek, Sašo	WED
14:40 - 15:00	Learning Association Rules for Pharmacogenomic Studies Giuseppe Agapito, Pietro Hiram Guzzi, Mario Cannataro		AY
15:00 - 15:20	Mining a Sub-Matrix of Maximal Sum Vincent Branders, Pierre Schaus, Pierre Dupont		IURSD
15:20 - 15:40	Community-Based Semantic Subgroup Discovery Blaž Škrlj, Anže Vavpetič, Jan Kralj, Nada Lavrač		Ē
15:40 - 16:00	Coffee & Tea Break		
	AFTERNOON SESSION 2 - WEB AND MULTIMEDIA		A
16:00 - 16:20	Segment-Removal Based Stuttered Speech Remediation Pierre Arbajian, Ayman Hajja, Zbigniew Ras, Alicja Wieczorkowska		FRID
16:20 - 16:40	Discovering of Alternative Marketplaces on the Web for Mobile App Security M Massimo Guarascio, Ettore Ritacco, Francesco Sergio Pisani, Daniele Biondo, Rocco Ma	<mark>onitoring</mark> ammoliti, Alessandra Toma	
	AFTERNOON SESSION 3 - TIME SERIES		
16:40 - 17:00	A Scaled-Correlation Based Approach for Generating and Analyzing Functiona Signals	l Networks from EEG	
	Samuel Dolean, Attila Geiszt, Raul C. Mureşan, Mihaela Dînşoreanu, Rodica Potolea, Ioar	na Țincaș	
17:00 - 17:20	Is Unsupervised Ensemble Learning Useful for Aggregated Or Clustered Load Forecasting?		
	Peter Laurinec and Maria Lucka		
17:20 - 17:40	Closing	ECML PKDD 2017 SKOPJE - MACEDONIA	127

MEZZANINE HALL 2

09:00 - 17:40

LIDTA 2017 Learning with Imbalanced Domains:Theory & Applications

Organizers : Luís Torgo, University of Porto; LIAAD - INESC Tec, Portugal Bartosz Krawczyk, Virginia Commonwealth University, USA Paula Branco, University of Porto; LIAAD - INESC Tec, Portugal Nuno Moniz, University of Porto; LIAAD - INESC Tec, Portugal

Webpage

http://lidta.dcc.fc.up.pt/

Many real-world data-mining applications involve obtaining and evaluating predictive models using data sets with strongly imbalanced distributions of the target variable. Frequently, the least-common values are associated with events that are highly relevant for end users. This problem has been thoroughly studied in the last decade with a specific focus on classification tasks. However, the research community has started to address this problem within other contexts. It is now recognized that imbalanced domains are a broader and important problem posing relevant challenges for both supervised and unsupervised learning tasks, in an increasing number of real world applications. This workshop invites inter-disciplinary contributions to tackle the problems that many real-world domains face today. With the growing attention that this problem has collected, it is crucial to promote its development and to tackle its theoretical and application challenges.

Invited talk MARKING THE 15-YEAR ANNIVERSARY OF SMOTE: ORIGIN, PROGRESS AND OPPORTUNITIES

Nitesh Chawla University of Notre Dame, USA

128 ECML PKDD 2017 SKOPJE - MACEDONIA

LIDTA 2017 Program at a glance

- 09:00 09:10 **Opening Remarks**
- 09:10 10:15 Marking the 15-Year Anniversary of SMOTE: Origin, Progress and Opportunities Nitesh Chawla

10:15 - 10:40 IDEA IN A NUTSHELL (POSTERS)

Tunable Plug-In Rules with Reduced Posterior Certainty Loss in Imbalanced Datasets E. Krasanakis, E. Spyromitros-Xioufis, S. Papadopoulos, Y. Kompatsiaris

Evaluation of Ensemble Methods in Imbalanced Regression Tasks

N. Moniz, P. Branco, L. Torgo

Controlling Imbalanced Error in Deep Learning with the Log Bilinear Loss

Y. Resheff, A. Mandelbom, D. Weinshall

Unsupervised Classification of Speaker Profiles As a Point Anomaly Detection Task C. Fayet, A. Delhay, D. Lolive, P.-F. Marteau

Dealing with the Task of Imbalanced, Multidimensional Data Classification Using Ensembles of Exposers

P. Ksieniewicz, M. Wozniak

10:40 - 11:00 Coffee & Tea Break

11:00 - 12:40 PAPER PRESENTATIONS 1

Predicting Defective Engines Using Convolutional Neural Networks on Temporal Vibration Signals N. Guennemann, J. Pfeffer

Influence of Minority Class Instance Types on SMOTE Imbalanced Data Oversampling P. Skryjomski, B. Krawczyk

A Network Perspective on Stratification of Multi-Label Data

P. Szymański, T. Kajdanowicz

Stacked-MLkNN: A Stacking Based Improvement to Multi-Label K-Nearest Neighbours A. Pakrashi, B. Mac Namee

Sampling a Longer Life: Binary Versus One-Class Classification Revisited S. Sharma, C. Bellinger, O. Zaiane, N. Japkowicz

12:40 - 14:00 Lunch Break

14:00 - 15:00 PAPER PRESENTATIONS 2

Improving Resampling-Based Ensemble in Churn Prediction Z. Bing, S. Vanden Broucke, B. Baesens, S. Maldonado

SMOGN: a Pre-Processing Approach for Imbalanced Regression P. Branco, L. Torgo, R. P. Ribeiro

Effect of Data Imbalance on Unsupervised Domain Adaptation of Part-of-Speech Tagging and Pivot Selection Strategies X. Cui, F. Coenen, D. Bollegala

15:00 - 15:40 Poster Session

- 15:40 16:00 Coffee & Tea Break
- 16:00 17:30 Discussion Table: What'S Next?
- 17:30 17:40 Closing Remarks

129

HURSDAY

FUESDAY

09:00 - 12:40

DMSC 2017 Data Mining with Secure Computation

Organizers : Mykola Pechenizkiy, Eindhoven University of Technology, The Netherlands Stefan Kramer, University of Mainz, Germany Niek J. Bouman, Eindhoven University of Technology, The Netherlands

Webpage

http://www.soda-project.eu/dmsc-workshop

Security and privacy aspects of data analytics become of central importance in many application areas. New legislation also pushes companies and research communities to address challenges of privacy-preserving data analytics. In our data mining community, questions about data privacy and security have been predominantly approached from the perspective of k-anonymity and differential privacy.

The aim of this workshop is to draw attention to secure multiparty computation (MPC), a subfield of cryptology, as the key foundation for building privacy-preserving data mining (DM) and machine learning (ML). In this approach, sensitive data is typically secret-shared over multiple players, such that those players can jointly perform DM / ML computations, but individual players (or collusions) cannot learn anything about the data, beyond the result of the computations.

Invited talk INTRODUCTION TO SECURE MULTIPARTY COMPUTATION

Berry Schoenmakers

Eindhoven University of Technology, The Netherlands

Invited talk SECURE MULTIPARTY COMPUTATION FOR PRIVACY-PRESERVING STATISTICS

Kurt Nielsen

Partisia, Denmark

DMSC 2017 Program at a glance

09:00 - 09:05	Opening
09:05 - 09:45	Introduction to Secure Multiparty Computation Berry Schoenmakers
09:45 - 10:25	Secure Data Mining with the Sharemind Platform, and a Use Case in Privacy-Preserving Tax Fraud Detection Sander Siim
10:25 - 10:40	Privacy-Preserving Patent Search: Additively Homomorphic Encryption Techniques for Private Text Mining Over Public Datasets Jennifer Katherine Fernick, Sebastian Verschoor
10:40 - 11:00	Coffee & Tea Break
11:00 - 11:40	Secure Multiparty Computation for Privacy-Preserving Statistics Kurt Nielsen
11:40 - 12:10	Processing Encrypted Data Using Homomorphic Encryption Anthony Barnett, Charlotte Bonte, Carl Bootland, Joppe W. Bos, Wouter Castryck, Anamaria Costache, Louis Goubin, Ilia Iliashenko, Tancrède Lepoint, Michele Minelli, Pascal Paillier, Nigel P. Smart, Frederik Vercauteren, Srinivas Vivek, Adrian Waller
12:10 - 12:25	Decentralised and Privacy-Aware Learning of Traversal Time Models Thanh Le Van, Aurélien Bellet, Jan Ramon
12:25 - 12:40	Closing Panel Discussion: Challenges and Trade-Offs in Secure Computation in a Data Mining Context

09:00 - 12:40

DMNLP 2017 Interactions between Data Mining and Natural Language Processing

Organizers :	Peggy Cellierm, INSA Rennes, France
	Thierry Charnois, Université de Paris, France
	Andreas Hotho, University of Kassel, Germany
	Marie-Francine Moens, KU Leuven, Belgium
	Stan Matwin, Dalhousie University, Canada
	Yannick Toussaint, INRIA, France

Webpage

http://dmnlp.loria.fr/

The workshop will favor the use of symbolic methods. Indeed, statistical and machine learning methods (CRF, SVM, Naive Bayes) holds a predominant position in NLP researches and "may have been too successful (...) as there is no longer much room for anything else". They have proved their effectiveness for some tasks but one major drawback is that they do not provide human readable models. By contrast, symbolic machine learning methods are known to provide more human-readable model that could be an end in itself (e.g., for stylistics) or improve, by combination, further methods including numerical ones. Research in Data Mining has progressed significantly in the last decades, through the development of advanced algorithms and techniques to extract knowledge from data in different forms. In particular, for two decades Pattern Mining has been one of the most active field in Knowledge Discovery.

Recently, a new field has emerged taking benefit of both domains: Data Mining and NLP. The objective of DMNLP is thus to provide a forum to discuss how Data Mining can be interesting for NLP tasks, providing symbolic knowledge, but also how NLP can enhance data mining approaches by providing richer and/or more complex information to mine and by integrating linguistic knowledge directly in the mining process. The workshop aims at bringing together researchers from both communities in order to stimulate discussions about the cross-fertilization of those two research fields. The idea of this workshop is to discuss future directions and new challenges emerging from this cross-fertilization of Data Mining and NLP and in the same time to initiate collaborations between researchers of both communities.

DMNLP 2017 Program at a glance

09:00 - 09:05 **Opening Session**

09:05 - 10:35 MORNING SESSION 1

Cluster-Based Graphs for Conceiving Dialog Systems Jean Leon Bouraoui, Vincent Lemaire

Characterising the Difference and the Norm Between Sequence Databases Frauke Hinrichs, Jilles Vreeken

Summarising Event Sequences Using Serial Episodes and an Ontology Kathrin Grosse, Jilles Vreeken

10:40 - 11:00 Coffee & Tea Break

11:00 - 12:30 MORNING SESSION 2

14:00 - 15:40 Making Efficient Use of a Domain Expert'S Time in Relation Extraction Linara Adilova, Sven Giesselbach, Stefan Rueping

> Combining Syntactic and Sequential Patterns for Unsupervised Semantic Relation Extraction Nadège Lechevrel, Kata Gábor, Isabelle Tellier, Thierry Charnois, Haifa Zargayouna. Davide Buscaldi

Correcting Linguistic Training Bias in an FAQ-Bot Using LSTM-VAE Mayur Patidar, Puneet Agarwal, Lovekesh Vig, Gautam Shroff

12:30 - 12:40 Closing Session

CONGRESS HALL 4

14:00 - 19:00

KnowMe 2017 KNOWledge Discovery from Mobility & Transportation Systems

Organizers : Luis Moreira-Matias, NEC Laboratories Europe, Germany Roberto Trasarti, KDD Lab ISTI-CNR, Italy Rahul Nair, IBM Research Ireland

Webpage

http://kdd.isti.cnr.it/knowme.eu/

The recent technological advances on telecommunications create a new reality on mobility sensing. Nowadays, we live in an era where ubiquitous digital devices are able to broadcast rich information about human mobility in real-time and at high rate. Such fact exponentially increased the availability of large-scale mobility data which has been popularized in the media as the new currency, fueling the future vision of our smart cities that will transform our lives. The reality is that we just began to recognize significant research challenges across a spectrum of topics. Consequently, there is an increasing interest among different research communities (ranging from civil engineering to computer science) and industrial stakeholders on build knowledge discovery pipelines over such data sources. However, such availability also raise privacy issues that must be considered by both industrial and academic stakeholders on using these resources.

This workshop intends to be a top-quality venue to bring together transdisciplinary researchers and practitioners working in topics from multiple areas such as Data Mining, Machine Learning, Numerical Optimization, Public Transport, Traffic Engineering, Multi-Agent Systems, Human-Computer Interaction and Telecommunications, among others. The ultimate goal of this venue is to evaluate not only the theoretical contribution of the data driven methodology proposed in each research work, but also its potential deployment/impact as well as its advances with respect to the State-of-the-Art/State-of-the-Practice in the domains of the related applications.

Invited talk BIG DATA FOR CITY SCIENCE: THE POWER HUMAN MOBILITY NETWORKS Fosca Giannotti

ISTI-CNR, Italy

KnowMe 2017 Program at a glance

14:00 - 15:40 **SESSION 1 Big Data for City Science: the Power Human Mobility Networks** Fosca Giannotti On Evaluating Floating Car Data Quality for Knowledge Discovery Spatio-Temporal Profiling of Public Transport Delays Based on Large Scale Vehicle Positioning Data from GPS in Wroclaw 15:40 - 16:00 Coffee & Tea Break 16:00 - 18:00 **SESSION 2** Traffic Patterns Recognition Through Big Data: Case Study in Thessalonik (Demo) Car Sharing Through the Data Analysis Lens Low-Dimensional Multi-Task Representation of Bivariate Histogram Data from Truck On-**Board Sensors** Mining Smart Card Data for Travelers Mini Activities Automatic Recognition of Public Transport Trips from Mobile Device Sensor Data and Transport Infrastructure Information

18:00 - Closing

14:00 - 14:05

Opening

14:00 - 17:40

DARE 2017 Data Analytics for Renewable Energy Integration

Organizers : Wei Lee Woon, Masdar Institute, UAE Zeyar Aung, Masdar Institute, UAE Oliver Kramer, University of Oldenburg, Germany Stuart Madnick, Massachusetts Institute of Technology, USA

Webpage

http://dare2017.dnagroup.org/

Climate change, the depletion of natural resources and rising energy costs have led to an increasing focus on renewable sources of energy. A lot of research has been devoted to the technologies used to extract energy from these sources; however, equally important is the storage and distribution of this energy in a way that is efficient and cost effective. Achieving this would generally require integration with existing energy infrastructure.

The challenge of renewable energy integration is inherently multidisciplinary and is particularly dependent on the use of techniques from the domains of data analytics, pattern recognition and machine learning. Examples of relevant research topics include the forecasting of electricity supply and demand, the detection of faults, demand response applications and many others. This workshop will provides a forum where interested researchers from the various related domains will be able to present and discuss their findings.

DARE 2017 Program at a glance

14:00 - 14:05	Welcome Address
14:05 - 14:25	Solar Energy Forecasting and Optimization System for Efficient Renewable Energy Integration Diana Manjarres, Ricardo Alonso, Sergio Gil, Itziar Landa
14:30 - 14:50	Gradient Boosting Models for Photovoltaic Power Estimation Under Partial Shading Conditions Nikolaos Nikolaou, Efstratios Batzelis, Gavin Brown
14:55 - 15:15	Multi-Objective Optimization for Power Load Recommendation Considering User'S Comfort Jaroslav Loebl, Helmut Posch, Viera Rozinajova
15:20 - 15:40	Minimizing Grid Interaction of Solar Generation and DHW Loads In nZEBs Using Model-Free Reinforcement Learning Adhra Ali, Hussain Kazmi
15:40 - 16:00	Coffee & Tea Break
16:00 - 16:20	Improving Time-Series Rule Matching Performance for Detecting Energy Consumption Patterns Maël Guillemé, Laurence Rozé, Véronique Masson, Cérès Carton, René Quiniou, Alexandre Termier
16:25 - 16:45	Identifying Representative Load Time Series for Load Flow Calculations Janosch Henze, Tanja Kneiske, Martin Braun, Bernhard Sick
16:50 - 17:10	Scalable Gaussian Process Models for Solar Power Forecasting Astrid Dahl and Edwin Bonilla
17:15 - 17:35	NWP Ensembles for Wind Energy Uncertainty Estimates Alejandro Catalina Feliu, Jose Dorronsoro
17:35 - 17:40	Closing

Papers to Be Presented As Posters

An Approach for Erosion and Power Loss Prediction of Wind Turbines Using Big Data Analytics Dina Fawzy, Sherin Moussa, Nagwa Badr

Probabilistic Wind Power Forecasting By Using Quantile Regression Analysis Mehmet Baris Ozkan, Umut Guvengir, Dilek Kucuk, Ali Unver Secen, Serkan Buhan, Turan Demirci, Abdullah Bestil, Ceyda Er, Pinar Karagoz

Wind Speed Forecasting Using Statistical and Machine Learning Methods: A Case Study in the UAE $\,$

Khawla Al Dhaheri, Wei Lee Woon, Zeyar Aung

TUTORIAL

Deep Learning for Computer Vision Applications: Robotics & Driving

Organizers : Anelia Angelova, Google Research / Google Brain, USA Sanja Fidler, University of Toronto, Canada

Webpage

https://sites.google.com/site/deeplearningvision/home

Deep Learning methods have become ubiquitous for computer vision tasks. This tutorial will focus on recent advances in deep learning for vision applications in robotics and autonomous vehicles. The tutorial will start with basic Deep Learning techniques and will highlight state-of-the-art methods in the three major topics in computer vision: classification, detection and segmentation. Then the tutorial will continue with more concrete methods and their applications, e.g. in scene understanding, 3D analysis, perception for robotics and autonomous driving. The goal of the tutorial is to focus on relevant techniques, which are of significant impact to real-world applications, and which will benefit the broader Machine Learning community.

09:00 - 10:40	SESSION 1
	PART 1 : Deep learning for Computer Vision
	PART 2 : Deep learning for Perception in Robotics
10:40 - 11:00	Coffee & Tea Break
11:00 - 12:40	SESSION 2
	PART 3 : Deep Learning for 3D
	PART 4 : Deep Learning for Perception in Autonomous Driving

TUTORIAL

Machine Learning with Fossil Data: Analyzing Environmental & Climate Change

Organizier : Indre Žliobaitė, University of Helsinki, Finnland

Webpage

http://www.zliobaite.com/fossils-tutorial

Global fossil databases have been growing rapidly in the last decade. They aggregate and accumulate findings and knowledge that palaeobiologists acquired over many years. These datasets are big data in their essence - compiled from different sources, to an extent subjective, include specific biases and uncertainties, data sparseness and quality varies over time and space. In addition, to understand relations between organisms and climate high volume and large velocity satellite observations some into play that require scalability in computing. Databases of this kind offer an excellent ground for interdisciplinary machine learning research.

This tutorial will outline research questions that could be addressed using computational methods, discuss characteristics of fossil data and computational tasks for machine learning and data mining, overview existing computational approaches, and discuss what more could be done from the machine learning and data mining perspective.

14:00 - 14:40	SESSION 1
	Analysis of the fossil record: research questions, challenges and trends
	Fossil Data and Fossil Databases
14:40 - 14:50	Short Break
14:50 - 15:40	SESSION 2
	Overview of Machine Learning Tasks and Existing Approaches
15:40 - 16:00	Coffee & Tea Break
16:00 - 16:40	SESSION 3
	Modeling Relationships Between Organisms and their Environments
16:40 - 16:50	Short Break
16:50 - 17:40	SESSION 4
	Tracking communities over time and space
	Analysis of Macroevolution Processes



ORGANIZATION

GENERAL CHAIRS

Sašo Džeroski

Michelangelo Ceci

Jožef Stefan Institute Slovenia

Univeristy of Bari Italv

PROGRAM CHAIRS

Michelangelo Ceci

Univeristy of Bari Italy

Aalto University

Jaakko

JOURNAL TRACK CHAIRS

Kurt **Driessens**

Maastricht University

The Netherlands

Dragi Kocev

Slovenia

Jožef Stefan Instiute

Marko Robnik-Šikonja

Univeristy of Ljubljana

Myra Spiliopoulu

Celine

KU Leuven Kulak

Vens

Belaium

Magdeburg University Germany

APPLIED DATA SCIENCE TRACK CHAIRS

Yasemin Altun

Google Research Switzerland

Kamalika Das

USA

NASA Ames Research Center

Yahoo! USA

Taneli

Hollmén Finland

Slovenia

Mielikäinen

Ljupčo

Slovenia

Todorovski

Univeristy of Liubliana

NECTAR TRACK CHAIRS

Donato Malerba

Jerzy Stefanowski

University of Bari Italy

Poznan University of Technology Poland

DEMONSTRATION TRACK CHAIRS

Jesse Read

Marinka Žitnik

École Polytechnique France

Stanford University USA

WORKSHOPS & TUTORIALS CHAIRS

Nathalie Japkowicz Panče Panov

American University USA Jožef Stefan Institute Slovenia

EU PROJECTS FORUM CHAIRS

Petra Kralj Novak

Nada Lavrač

Jožef Stefan Institute Slovenia Jožef Stefan Institute Slovenia

PHD FORUM CHAIRS

Tomislav Šmuc

Bernard Ženko

Rudjer Bošković Institute Croatia Jožef Stefan Institute Slovenia

DISCOVERY CHALLENGE CHAIR

Dino Ienco

IRSTEA - UMR TETIS France

AWARDS COMMITTEE

Peter Flach

Rosa Meo

Indrė Žliobaitė

University of Bristol UK University of Torino Italy

University of Helsinki Finland

PRODUCTION & PUBLIC RELATIONS CHAIRS

Dragi Kocev

Nikola Simidjievski

Jožef Stefan Institute Slovenia

Jožef Stefan Institute Slovenia
LOCAL ORGANIZATION

lvica Dimitrovski

Tina Anžič

CHAIR

Ss. Cyril & Methodious Univerisy, Macedonia Jožef Stefan Institute Slovenia

SPONSORSHIP CHAIRS

Albert Bifet

Panče Panov

Télécom ParisTech France

Jožef Stefan Instiute Slovenia

PROCEEDINGS CHAIRS

Jurica Levatić

Gianvito Pio

Jožef Stefan Instiute Slovenia University of Bari Italy Mili Bauer

Jožef Stefan Institute Slovenia

Gjorgji Madjarov

Ss. Cyril & Methodious Univerisy, Macedonia

ECML PKDD 2017 Conference brochure

Editorial board:

Nikola Simidjievski Dragi Kocev Panče Panov

Design and pre-press: Nikola Simidjievski

Photography: Nikola Simidjievski

Print: Ri-grafika, Skopje Macedonia

-

The editing of the conference brochure was closed on September 10, 2017. Changes to the program made after this date are not reflected in its contents.

September 2017, Ljubljana, Slovenia



Deutsche	Post	DHL
Group		



SILVER SPONSORS





Be part of progress

Deloitte.

\Orchestrating a brighter world





BRONZE SPONSORS



IEEE/CAA JOURNAL OF AUTOMATICA SINICA

PUBLISHING PARTNER & SPONSOR



LANYARDS SPONSOR

PHD FORUM SPONSOR



IBM Research

AWARDS & INVITED TALKS SPONSORS



Deloitte.

ORGANIZING INSTITUTIONS











"Ss. Cyril and Methodius" University in Skopje FACULTY OF COMPUTER SCIENCE AND ENGINEERING

