Image representation, annotation and retrieval with predictive clustering trees

Ivica Dimitrovski[1], Dragi Kocev[2], Suzana Loskovska[1], and Sašo Džeroski[2]

[1] University of Ss Cyril and Methodius, Skopje, Macedonia
[2] Jožef Stefan Institute, Ljubljana, Slovenia
ivica.dimitrovski@finki.ukim.mk, Dragi.Kocev@ijs.si,
suzana.loshkovska@finki.ukim.mk, Saso.Dzeroski@ijs.si

Abstract. In this paper, we summarize our work on using the predictive clustering framework for image analysis. More specifically, we have used predictive clustering trees to generate image representations, that can then be used to perform image retrieval and/or image annotation. We have evaluated the proposed method for performing image retrieval on general purpose images [6], and annotation of general purpose images [5], medical images [3] and diatom images [4].

Keywords: image representation, image retrieval, image annotation, multi-target prediction, predictive clustering

1 Introduction

The overwhelming increase in the amount of available visual information, especially digital images, has brought up a pressing need to develop efficient and accurate systems for image representation, retrieval and annotation. Most such systems for image analysis use the bag-of-visual-words representation of images. However, the computational bottleneck in all such systems is the construction of the visual codebook, i.e., obtaining the visual words. This is typically performed by clustering hundreds of thousands or millions of local descriptors, where the resulting clusters correspond to visual words. Each image is then represented by a histogram of the distribution of its local descriptors across the codebook.

The major issue in retrieval systems is that by increasing the sizes of the image databases, the number of local descriptors to be clustered increases rapidly: Thus, using conventional clustering techniques is infeasible. While existing approaches are able to solve the efficiency issue, a part of the discriminative power of the codebook is sacrificed for this. Considering this, we propose to construct the visual codebook by using predictive clustering trees (PCTs) [1], which can be constructed and executed efficiently and have good predictive performance.

PCTs are a generalization of decision trees towards the task of structured output prediction, including multi-target regression, (hierarchical) multi-label classification and time series prediction. Moreover, the definition of descriptive, clustering and target attributes is flexible thus facilitating the learning of both unsupervised and supervised trees. Furthermore, to increase the stability of the model, we propose to use random forests of PCTs [7]. We create a random forest of PCTs that represents the codebook, i.e., is used to generate the image representation.

The images represented with the bag-of-visual-words can then be used to perform image retrieval and/or annotation. In the former, the indexing structure for performing the retrieval is the same structure representing the codebook – the random forest of PCTs. We evaluate the proposed bag-of-visual-words approach for image retrieval on five benchmark reference datasets. The results reveal that the proposed method produces a visual codebook with superior discriminative power and thus better retrieval performance while maintaining excellent computational efficiency [6].

Additional complexity of image annotation arises from the complexity of the labels used for annotation: Typically, an image depicts more than one object, hence more than one label should be assigned for that image. Moreover, there might be some structure among the labels, such as a hierarchy of labels. To address this additional complexity, we learn ensembles of PCTs to exploit the potential relations that may exist among the labels. We have evaluated this approach on three tasks: multi-label classification of general purpose images [5], and hierarchical multi-label classification of medical images [3], as well as diatom images [4]. The results of the evaluation show that we achieve state-of-the-art predictive performance.

The remainder of this paper is organized as follows. We next briefly present the predictive clustering framework. We then outline the method for constructing image representations. Finally, we describe the evaluation of this approach, first for image retrieval and then for image annotation.

2 Predictive clustering framework

Predictive Clustering Trees (PCTs) [1] generalize decision trees and can be used for a variety of learning tasks including different types of prediction and clustering. The PCT framework views a decision tree as a hierarchy of clusters: the top-node of a PCT corresponds to one cluster containing all data, which is recursively partitioned into smaller clusters while moving down the tree. The leaves represent the clusters at the lowest level of the hierarchy and each leaf is labeled with its cluster's prototype (prediction). One of the most important steps in the PCT algorithm is the test selection procedure. For each node, a test is selected by using a heuristic function computed on the training examples. The heuristic used in this algorithm for selecting the attribute tests in the internal nodes is the reduction in variance caused by partitioning the instances. Maximizing the variance reduction maximizes cluster homogeneity and improves predictive performance.

In this work, we used three instantiations of PCTs for the tasks of multi-target regression (MTR), multi-label classification (MLC) and hierarchical multi-label classification (HMC). For the MTR task, the variance is calculated as the sum of the normalized variances of the target variables. For the MLC task, we used the sum of the Gini indices of the labels, while for the HMC task, the variance is calculated by using a weighted Euclidean distance that considers the hierarchy of the labels. The prototype function returns as a prediction the tuple with the mean values of the target variables, calculated by using the training instances that belong to the given leaf.

3 PCTs for image representation

The proposed method for constructing the visual codebook is as follows. First, we randomly select a subset of the local (SIFT) descriptors from all of the training images [8]. Next, the selected local descriptors constitute the training set used to construct a PCT. For the construction of a PCT, we set the descriptive attributes (i.e., the 128 dimensional vector of the local descriptor) to be also target and clustering attributes. Note that this feature is a unique characteristic of the predictive clustering framework. The PCTs are computationally efficient: it is very fast to both construct them and use them to make predictions. However, tree learning is unstable, i.e., the structure of the learned tree can change substantially for small changes in the training data [2]. To overcome this limitation and to further improve the discriminative power of the indexing structure, we use an ensemble (i.e., random forest) of PCTs. The overall codebook is obtained by concatenating the codebooks from each tree.

4 PCTs for image retrieval

In the proposed system, a PCT (or a random forest of PCTs) represents the search/indexing structure used to retrieve images similar to query images. Namely, for each image descriptor (i.e., each training example used to construct the PCTs), we keep a unique index/identifier. The identifier consists of the image ID from which the local descriptor was extracted coupled with a descriptor ID. This indexing allows for faster computation of the image similarities.

We have evaluated the proposed improvement of the bag-of-visual words approach on three reference datasets and two additional datasets of 100K images and 1M images, comparing it to two state-of-the-art methods based on approximate k-means and extremely randomized tree ensembles. The results from the experimental evaluation reveal the following. First and foremost, our system exhibits better retrieval performance by 6-8% (mean average precision) than both competing methods at the same efficiency. Additionally, the increase of the number of local descriptors and number of PCTs used to create the indexing structure improve the retrieval performance of the system.

5 PCTs for hierarchical annotation of images

We first use our system for multi-label annotation of general purpose images [5]. We compare the efficiency and the discriminative power of the proposed approach to the literature standard of using k-means clustering. The results reveal that our approach is much more efficient in terms of computational time (24.4 times faster) and produces a visual codebook with better discriminative power as compared to k-means clustering. Moreover, the difference in predictive performance increases with the average number of labels per image.

Next, we evaluate the performance of ensembles of PCTs for HMC (bagging and random forests) on the task of annotation of medical images using the hierarchy from the DICOM header [3]. The experiments on the IRMA database show that random forests of PCTs for HMC outperform SVMs for flat classification. The average difference is 17 points for the ImageCLEF2007 and 20 points for the ImageCLEF2008 dataset (a point in the hierarchical evaluation measure roughly corresponds to one completely misclassified image). Additionally, the random forests are the fastest method; they are 10 times faster than bagging and 5.5 times faster than the SVMs.

Finally, for the task of hierarchical annotation of diatom images, by using random forests of PCTs for HMC, we obtained the best results on the different variants of the ADIAC database of diatom images [4]: The obtained predictive power of our method was in the range 96-98%. More specifically, we outperformed a variety of methods for annotation that use SVMs, bagged decision trees and neural networks. Finally, we used these annotations in an on-line annotation system to assist taxonomists in identifying a wide range of different diatoms.

Acknowledgments

We would like to acknowledge the support of the European Commission through the project MAESTRA - Learning from Massive, Incompletely annotated, and Structured Data (Grant number ICT-2013-612944).

References

1. Blockeel, H., Raedt, L.D., Ramon, J.: Top-down induction of clustering trees. In: Proc. of the 15th Int. Conf. Machine Learning. pp. 55–63. Morgan Kaufmann (1998)
2. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
3. Dimitrovski, I., Kocev, D., Loskovska, S., Dzeroski, S.: Hierarchical annotation of medical images. Pattern Recognition 44(10-11), 2436–2449 (2011)
4. Dimitrovski, I., Kocev, D., Loskovska, S., Dzeroski, S.: Hierarchical classification of diatom images using ensembles of predictive clustering trees. Ecological Informatics 7(1), 19–29 (2012)
5. Dimitrovski, I., Kocev, D., Loskovska, S., Dzeroski, S.: Fast and efficient visual codebook construction for multi-label annotation using predictive clustering trees. Pattern Recognition Letters 38, 38–45 (2014)
6. Dimitrovski, I., Kocev, D., Loskovska, S., Dzeroski, S.: Improving bag-of-visual-words image retrieval with predictive clustering trees. Inf. Sci. 329, 851–865 (2016)
7. Kocev, D., Vens, C., Struyf, J., Džeroski, S.: Tree ensembles for predicting structured outputs. Pattern Recognition 46(3), 817–833 (2013)
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)