

Using forests to classify digital images

Classificazione di immagini digitali mediante l'uso di foreste

Carla Brambilla¹, Claudio Cusano², Raimondo Schettini²

¹ Istituto di Matematica Applicata e Tecnologie Informatiche, CNR, Milano
e-mail: carla@mi.imati.cnr.it

² DISCo, Università Bicocca, Milano

Riassunto: Questo lavoro illustra i risultati ottenuti utilizzando insiemi di classificatori ad albero, detti foreste, in un problema di classificazione automatica di immagini digitali. Il problema affrontato è quello della classificazione automatica di fotografie digitali come foto d'interno, d'esterno o primi piani. I risultati ottenuti mostrano che, rispetto all'uso di singoli alberi, l'utilizzo delle foreste produce un notevole miglioramento nell'accuratezza della classificazione, e anche che l'accuratezza può essere ulteriormente migliorata introducendo nel processo di classificazione una regola di rigetto.

Keywords: bagging, classification, digital images, trees

1. Introduction

A forest is an ensemble of trees obtained manipulating somehow a training set and running the training process on the different sets thus obtained. When trees are classification trees the forest is a classifier itself. The use of forests have proved quite successful in overcoming the overfitting due to instability, a phenomenon inherent in the hierarchical nature of the construction process of trees, thereby improving generalization accuracy. Indeed, even small changes in data may result in a very different series of splits, and this clearly affects classification results (Hastie *et al.*, 2001).

In this work we summarize our experience of using forests in the problem of automatically classifying digital photographs as indoor, outdoor or close-up. The photographs are described by low-level features, that is, features which can be automatically computed without any a priori knowledge of the content of the scene depicted.

The automatic classification of digital photographs in semantic categories is an unresolved challenge in the multimedia and imaging communities and a number of practical applications call for further research on this problem. Annotating photographs with semantic labels, for example, is very useful for adopting the most appropriate strategies of color adjustment in image processing, and can greatly facilitate the organization and retrieval of databases.

To date there have been few efforts to automatically classify digital photographs and no one has used forests. An experience of using forests in a problem of classifying digital images of different nature is described in Schettini *et al.* (2002).

2. Classification methodology

As base classifiers we have chosen trees because they make a powerful use of high dimensionality and conditional information, and this is very important in problems characterized by a great level of complexity, as the one we addressed. In the problem of distinguishing among indoor, outdoor, and close-up photographs, indeed, a precise mapping from a set of low-level features to image semantics is not possible, and the assumption that photographs with similar semantics share the same pictorial characteristics may not always hold.

Trees were constructed according to CART methodology, for which we refer to Breiman *et al.* (1984). We used equal prior probabilities and equal misclassification costs, and the Gini diversity index as measure of node impurity.

To build forests we used bagging (Breiman, 1996). According to this, the different trees of a forest are formed by making bootstrap replicates of the training set, and using these as new training sets. To combine the classification results produced by the single trees we applied the majority vote rule.

To further refine the classification process, we also included in it a rejection option (Jain *et al.*, 2000). Thus photographs that are to be classified are not compulsorily assigned to one of the designed classes and ambiguous photographs, that is, photographs that might be labeled differently by different observers, are likely to be rejected. Moreover, classification accuracy for non-ambiguous photographs is improved. The rejection rule we experimented is very simple: it states that a classification result obtained by means of the majority vote is rejected if the percentage of trees of the forest that contributes to it is less than a given threshold, which is called rejection threshold. This rule is constant over the feature space.

3. Experimental results

The photographs we used for our experiments were downloaded from the web, or acquired by a scanner or by digital cameras. Altogether they were 9000. Half of them was set aside to form a test set and the results we present here refer to this set. To describe the images we used about 70 low-level features relating to color, texture, edges distribution and composition of the image. All the features are quantitative.

At first we built 10 trees using different subsets of the training set, and then used each tree to classify the 4500 photographs of the test set. Each subset contained 3000 photographs randomly chosen from the training set in such a way that the three classes were equally represented. Each of the 10 trees was the most accurate tree within the sequence of subtrees obtained applying the pruning process. To select it, the accuracy of the subtrees was evaluated classifying the 1500 photographs not used for the training process. Table 1 shows the average results obtained on the test set using the single trees.

The subsequent step was the construction of 50 forests of increasing size. As in the experiments with single trees, each tree in each the forests was the most accurate among the sequence of subtrees obtained by applying pruning. To do so each bootstrap replicate of the training set was subdivided in two subsets, one used to build a large initial tree and perform the pruning, the other to select the best tree in the sequence. For all the three classes, forests of more than 20 trees achieved an accuracy that was nearly stable and the

Table 1: *Confusion matrix obtained on the test set with single trees (average values).*

		Predicted Class		
		Indoor	Outdoor	Close-up
True Class	Indoor	0.787	0.117	0.096
	Outdoor	0.085	0.817	0.098
	Close-up	0.079	0.077	0.844

improvement over the accuracy achieved using single trees was quite significant. Table 2 shows the results obtained using the forest of 25 trees. The Table shows also (in brackets) the results obtained applying the rejection option to that forest, with 30% rejection rate.

Table 2: *Confusion matrix obtained on the test set with the forest of 25 trees. In brackets the results obtained applying the rejection option with 30% rejection rate.*

		Predicted Class		
		Indoor	Outdoor	Close-up
True Class	Indoor	0.881 (0.962)	0.050 (0.017)	0.069 (0.021)
	Outdoor	0.027 (0.018)	0.902 (0.963)	0.071 (0.015)
	Close-up	0.054 (0.032)	0.061 (0.014)	0.885 (0.954)

The misclassified photographs were generally photographs either overexposed or underexposed, or with a background that provides little information about the class to which the images belong. We consider acceptable the misclassification of some close-up images as indoor or outdoor, and viceversa: it simply reflects the overlapping between the close-up class and the other categories.

Figure ?? shows how the rejection rate depends on the rejection threshold. Looking at the Figure we see that when the rejection threshold is set at its maximum value, the rejection rate is about 40% for all the three classes. This means that about 60% of the photographs of the test set were classified with the high confidence that derives from a unanimous vote. The rejection rate of 30% corresponds to a rejection threshold of about 95%, once again for all the three classes. Clearly, the rejection rate that is acceptable is an application-dependent parameter. In the retrieval framework, for example, the above result is very interesting: for databases of millions of photographs the possibility of an automatic and reliable classification of 70% of images signifies a considerable reduction in the manual effort.

4. Conclusions

The results obtained are very promising and lead to think that forests can be a well suited tool for dealing with the difficult problem of identifying semantic classes of digital photographs simply on the basis of low-level features. It is important to recall that the accuracies we reported refer to the test set alone and are not averaged on the entire databases, which include the training sets. We believe that the performance would scale well to different, databases, but a number of issues deserve further work. First of all,

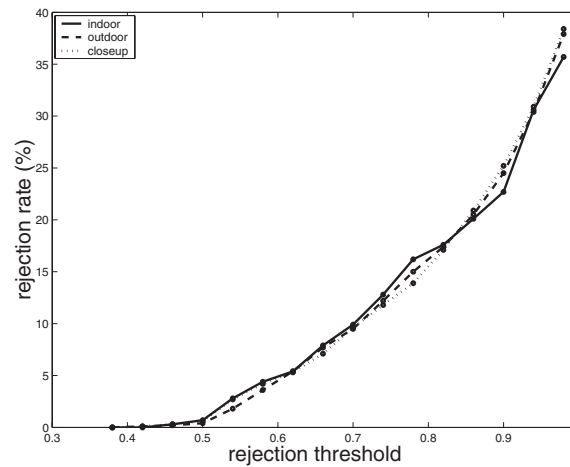


Figure 1: Graph of the rejection rate vs. the rejection threshold for the three classes.

it would be very important to identify suitable interpretational tools that allow to carry out a thorough analysis of the role played by the different features in the classification processes. Forests, indeed, are not as easily interpretable as single trees.

Some aspects of CART methodology not exploited in the present study might be very useful in future experiments: qualitative features could be dealt with as easily as quantitative features, and different misclassification costs and prior probabilities could be taken into account. Moreover, since the methodology allows for the presence of missing values both in the training set and in new cases to be classified, we could take advantage of ancillary information, such as image captions, which may not always be available.

In the future we plan to experiment also using the random forests proposed by Breiman (2001) and forests obtained by the random subspace method (Ho, 1997), whose construction is based on random selection of features subsets.

References

- Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1984) *Classification and Regression Trees*, Wadsworth and Brooks/Cole.
- Breiman L. (1996) Bagging predictors, *Machine Learning*, 24, 123–140.
- Breiman L. (2001) Random forests, *Machine Learning*, 45, 5–32.
- Hastie T., Tibshirani R., Friedman J.H. (2001) *The Elements of Statistical Learning*, Springer.
- Ho T.K. (1997) The random subspace method for constructing decision forests, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20, 832–844.
- Jain A., Duin R., Mao J. (2000) Statistical pattern recognition: a review, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 4–37.
- Schettini, R., Brambilla C., Ciocca G., Valsasna A., De Ponti M. (2002) A hierarchical classification strategy for digital documents, *Pattern Recognition*, 35, 1759–1769.