Synthetic Pattern Generation for Imbalanced Learning in Image Retrieval

Luca Piras, Giorgio Giacinto

Department of Electrical and Electronic Engineering
University of Cagliari
09123 Piazza D'armi, Cagliari, Italy
luca.piras@diee.unica.it, giacinto@diee.unica.it

Abstract

Nowadays very large archives of digital images are easily produced thanks to the wide availability of digital cameras, that are often embedded into a number of portable devices. One of the ways of exploring an image archive is to search for similar images. Relevance feedback mechanisms can be employed to refine the search, as the most similar images according to a set of visual features may not contain the same semantic concepts according to the users' needs. Relevance feedback allows users to label the images returned by the system as being relevant or not. Then, this labelled set is used to learn the characteristics of relevant images. As the number of images provided to users to receive feedback is usually quite small, and relevant images typically represent a tiny fraction, it turns out that the learning problem is heavily imbalanced. In order to reduce this imbalance, this paper proposes the use of techniques aimed at artificially increasing the number of examples of the relevant class. The new examples are generated as new points in the feature space so that they are in agreement with the local distribution of the available relevant examples. The locality of the proposed approach makes it quite suited to relevance feedback techniques based on the Nearest-Neighbor (NN) paradigm. The effectiveness of the proposed approach is assessed on two image datasets and comparisons with editing techniques that eliminate redundancies in non-relevant examples are also reported.

Keywords: Imbalanced Learning, Small Sample-Size, Artificial Pattern Injection, Image Retrieval, Relevance Feedback

1. Introduction

As the years go by, it is ever-easier to have access to a ever-greater amount of electronically archived images. As a consequence, there is an increasing need for tools enabling the semantic search and retrieval of images. The use of metadata associated to images solves the problems only partly, as the process of assigning metadata to images is not trivial, is slow, and closely related to the persons who performed the task. This is especially true for retrieval tasks in very high dimensional databases, where images exhibits high variability in semantic. It turns out that the description of image content tends to be intrinsically subjective and partial, and the search for images based on keywords may fit users' needs only partially. To this end, the analysis of image content is performed not only by human experts, but also by the use of automatic tools aimed

at producing similarity measurements, and image annotations (Lew et al., 2006; Li and Wang, 2008; Datta et al., 2008).

The main reason for the difficulty in devising effective image retrieval tools is caused by the vast amount of information conveyed by images, and the related subjectivity of the criteria to be used to assign labels to images (Lew et al., 2006; Smeulders et al., 2000; Datta et al., 2008). In order to capture such subjectivity, image retrieval tools may employ relevance feedback (RF) techniques. RF can be formulated as a $(1 + x)$ classification problem, where 1 represent the class of relevant images, and x represents the classes of non-relevant images, the number of such classes being unknown (Zhou and Huang, 2003; Datta et al., 2008; Tao et al., 2006a). The problem can be further formulated by resorting to the nearest case paradigm, as each relevant image as well as each non-relevant image can be considered as individual cases or instances against which the images of the database should be compared. This formulation of the problem arises from two considerations: i) non-relevant images clearly belong to multiple classes, and ii) the class of relevant images may be actually made up of distinct clusters of images in the low-level feature space (Giacinto, 2007).

One of the most severe problem in the design of the relevance feedback mechanism, is the imbalance between the number of samples of the class the user is interested in, and all other images of the database that share some characteristics with that class. For example, when SVM are employed for relevance feedback, typically a lower bound on the number of relevant and non relevant feedback examples is fixed in order to avoid the small sample size problem (Tao et al., 2006b).

In the machine-learning literature the imbalance problem has been widely investigated, and solutions based either on under-sampling the majority class, or on over-sampling the minority class have been proposed (He and Garcia, 2009). However, it is easy to see that these solutions may produce a distortion of the *real* distribution of the classes.

In the image retrieval domain, some solutions proposed so far involve the reduction of the set of images that are non-relevant to user's interest by the creation of bootstrap samples from the set of non-relevant images (Tao et al., 2006b). This solution is computationally quite expensive, as an ensemble of classifiers has to be created, and the choice of the most appropriate set of parameters for the various parts of the systems is far from being a trivial task.

In this paper, we address the imbalance problem by creating new artificial patterns from the relevant images at hand by exploiting Nearest-Neighbor relations between the available samples. To avoid creating noisy patterns, the new patterns are constrained to lie in a region of the feature space where it is more likely to find points related to relevant images. The exploitation of the Nearest-Neighbor information accounts for the manifold where relevant images are assumed to lie. Two techniques for generating artificial patterns are investigated. One technique is based on the Synthetic Minority Over-sampling Technique (SMOTE), while the other technique is inspired by the "directed noise injection" technique proposed for improving multilayer perceptron training (Chawla et al., 2002; Skurichina et al., 2000). In our opinion, the latter technique is more apt to be used in image retrieval tasks, and thus we propose an original version of the "directed" paradigm that we named Direct Pattern Injection (DPI). Some preliminary experiments have been recently reported by the authors, and the related results showed the effectiveness of the DPI approach (Piras and Giacinto, 2010b). Recently, some other papers addressed the imbalance problem by proposing the generation of artificial patterns (Thomee et al., 2008). However, while those papers, propose the generation of synthetic images for the sake of receiving informative feedback from the user, in this paper new patterns are generated in the feature space to improve the performances of learning mechanisms. Thus, no visual representation of the patterns

is produced.

This paper is organized as follows. Section 2 briefly reviews the main approaches proposed in the machine leaning domain to address the imbalanced problem. In particular, the two approaches that are investigated in this paper are described. Section 3 describes the RF technique based on the Nearest-Neighbor paradigm that has been used in the experiments. Section 4 shows the integration of the proposed approaches in the learning process whereas Section 5 illustrates some heuristics that are proposed to choose the parameters related to the technique based on the "directed" paradigm. Experimental results on two image datasets are reported in Section 6. Reported results show that the proposed approaches allows improving the performances of RF mechanisms based on Nearest-Neighbor. In particular, the DPI approach outperforms SMOTE as well as approaches aimed to reduce the redundancies in non-relevant images. Conclusions are drawn in Section 7.

2. Learning from imbalanced data

The problem of imbalanced learning has attracted a number of researchers in the machine-learning field (He and Garcia, 2009). As a matter of fact, the vast majority of learning algorithms suffers from imbalanced sets, as they typically produce a poor representation of the minority class. This problem heavily affects content-based image retrieval tasks due to the ever-increasing availability of very large archives of digital images. The main difficulty basically arises from the wide variety of classes in the repositories. As a consequence, the number of images that exhibits similar semantic content with respect to a given query image is a tiny fraction of the whole repository, and the fraction of images available for training purposes is even smaller. Thus, the learning process where the goal is to retrieve images belonging to a target class, is usually imbalanced, as there are more chances of collecting images that do not belong to the target class, than chances of collecting target images (Zhou and Huang, 2001; Duin, 2004; He and Garcia, 2009; Tao et al., 2006a).

In the machine learning literature, a number of techniques have been proposed to address the problem of learning from imbalanced data. A thorough review of the literature can be found in the recent paper by He and Garcia (He and Garcia, 2009). Basically the imbalance between data classes can be reduced by either undersampling the majority class or by oversampling the minority class. These tasks can be performed either randomly or by resorting to some heuristics. Random undesampling may cause the loss of informative data samples, while random oversampling may distort the distribution of the minority class in the feature space. In addition random oversampling may be of little effect depending on the learning algorithm at hand, as it just produces replicas of existing data points. To overcome these limitations, a number of heuristics have been devised aimed at discarding some patterns of the majority class, and producing brand new minority patterns according to the distribution of the data points of the minority class.

A different way to address this problem, is to approach the learning task by taking into account just the minority class. This approach, sometimes called one-class classification, aims at learning the characteristics of the minority class (Tax, 2001). This approach is quite attractive for image retrieval tasks, and it has been used in some settings (Chen et al., 2001).

Another approach involves cost-sensitive learning that introduces different costs of misclassification for the minority and majority class, in order to bias the learning mechanism. Biased learning has been proposed to implement RF mechanisms in image retrieval (Tao et al., 2006a).

To the best of our knowledge the use of techniques for generating synthetic patterns of the minority class has not yet been proposed in the context of image retrieval tasks. This paper

proposes a tailored version of the K-NN directed pattern injection technique (Skurichina et al., 2000) to improve the performance of RF mechanisms. We also investigated the use of SMOTE as it is the most popular technique for oversampling the minority class (Chawla et al., 2002), Both techniques are based on the creation of new patterns in a feature space F with p components, by exploiting K-NN relations.

2.1. Directed Pattern Injection (DPI)

This technique has been originally proposed to inject noisy samples for improving neural network learning in small sample size problem (Skurichina et al., 2000). This technique generates new synthetic samples by taking into account all the patterns in a local region of the feature space defined by the neighborhood around one *reference* pattern belonging to the minority class. Synthetic patterns are generated by a linear combination of the directions defined by the *reference* pattern and its neighbors in F belonging to the minority class, i.e.,

$$\mathbf{x}_{syn} = \mathbf{x}_{ref} + \lambda \sum_{\mathbf{x}_i \in NN\left(\mathbf{x}_{ref}\right)} \xi_i \cdot \left(\mathbf{x}_i - \mathbf{x}_{ref}\right) \tag{1}$$

where $NN\left(\mathbf{x}_{ref}\right)$ denotes the set of Nearest-Neighbors of $\mathbf{x}_{ref}$, the weights ξ_i are drawn from a normal distribution with zero mean and unit variance, and λ is a normalization factor. This formula implicitly assume that the combination of directions generated by one point and its neighbors can generate a new pattern belonging to the class of interest. The validity of this assumption has not been proven formally, and it relies on the observation that if the data of the minority class locally lies on a low-dimensional manifold of the original feature space, then the generated data will lie on the same manifold.

We observed that this line of reasoning is also shared by some techniques aimed at discovering non-linear manifold embeddings (Roweis and Saul, 2000). These techniques compute the low-dimensional embedding by requiring the preservation of local neighboring relations. For a given pattern $\mathbf{x}_k$ these relations can be represented in terms of the weights of the linear combination of its neighboring patterns

$$\begin{aligned} \mathbf{x}_{rec} = \sum_{\mathbf{x}_i \in NN(\mathbf{x}_k)} w_i \cdot \mathbf{x}_i \quad = \quad \mathbf{x}_k + \sum_{\mathbf{x}_i \in NN(\mathbf{x}_k)} w_i \cdot \mathbf{x}_i - \mathbf{x}_k = \\ = \mathbf{x}_k + \sum_{\mathbf{x}_i \in NN(\mathbf{x}_k)} w_i \cdot \mathbf{x}_i - \sum_i w_i \cdot \mathbf{x}_k = \mathbf{x}_k + \sum_{\mathbf{x}_i \in NN(\mathbf{x}_k)} w_i \cdot (\mathbf{x}_i - \mathbf{x}_k) \end{aligned} \tag{2}$$

where $\sum_i w_i = 1$.

The weights are computed so as to minimize the so-called reconstruction error, i.e., the error occurring if the pattern $\mathbf{x}_k$ is represented by $\mathbf{x}_{rec}$, i.e., the linear combination of its neighbors. We argue that if a real pattern can be approximated by a combination of its Nearest-Neighbors, then the combination of the Nearest-Neighbors of a given pattern can produce synthetic patterns that can be deemed to belong to the minority class with high probability.

2.2. SMOTE

SMOTE is one of the most widely known mechanisms to oversample the minority class by generating synthetic patterns. For each pattern of the minority class $\mathbf{x}_k$, new patterns are

generated by taking into account its K Nearest-Neighbors in F one at a time. Let $\mathbf{x}_i$ be one of the K Nearest-Neighbors, then a synthetic pattern $\mathbf{x}_{syn}$ is generated by the following formula

$$\mathbf{x}_{syn} = \mathbf{x}_k + \alpha \cdot (\mathbf{x}_i - \mathbf{x}_k) \tag{3}$$

where α is a random number in the range $[0, 1]$. Typically, the patterns in the neighborhood used to generate the synthetic patterns are chosen randomly.

3. Nearest-Neighbor Approach for Relevance Feedback in Image Retrieval

Recent works on outlier detection and one-class classification pointed out the effectiveness of Nearest-Neighbor approaches to identify objects belonging to the target class, while rejecting all other objects (Breunig et al., 2000; Tax, 2001). This approach is suited to cases when it is difficult to produce a high-level generalization of a class of objects. Thus, for each object, its likelihood of belonging to the target class is estimated locally, i.e. in terms of its nearest neighbors. According to this formulation, we proposed to estimate the relevance of an image by computing a score related to the distance from the nearest image belonging to the target class, and the distance from the nearest image belonging to a different class (Giacinto, 2007). This score is further combined to another score related to the distance of the image from the region of relevant images. The final score is thus computed as follows:

$$rel(\mathbf{x}) = \left(\tfrac{n/t}{1+n/t}\right) \cdot rel_{BQS}(\mathbf{x}) + \left(\tfrac{1}{1+n/t}\right) \cdot rel_{NN}(\mathbf{x}) \tag{4}$$

where n and t are the number of non-relevant images and the total number of images retrieved after the latter iteration, respectively. The two terms rel_{NN} and rel_{BQS} are computed as follows:

$$rel_{NN}(\mathbf{x}) = \frac{\|\mathbf{x} - NN^{nr}(\mathbf{x})\|}{\|\mathbf{x} - NN^{r}(\mathbf{x})\| + \|\mathbf{x} - NN^{nr}(\mathbf{x})\|} \tag{5}$$

where $NN^r(\mathbf{x})$ and $NN^{nr}(\mathbf{x})$ denote the relevant and the non-relevant Nearest Neighbor of $\mathbf{x}$, respectively, and $\|\cdot\|$ is the metric defined in the feature space at hand,

$$rel_{BQS}(\mathbf{x}) = \frac{1 - e^{1 - d_{BQS}(\mathbf{x}) / \max_i d_{BQS}(\mathbf{x}_i)}}{1 - e} \tag{6}$$

where e is the *Euler's number*, i is the index of all images in the database and d_{BQS} is the distance of image $\mathbf{x}$ from a modified query vector computed according to the Bayes decision theory (Bayes Query Shifting, BQS) (Giacinto and Roli, 2004). The modified query vector is computed on the line connecting the mean vectors of relevant and non-relevant images, and it is located on the side of the line where it is more likely to find relevant images. The distance from this modified query vector is called when very few relevant images are retrieved, to avoid that high values of the relevance score are obtained for images that are just dissimilar to non-relevant images (see Equation 4).

4. Image Retrieval with Synthetic Feature Vectors

The techniques presented in Section 2 can be used to provide a solution to the imbalance problem in image retrieval tasks. As the new patterns are not real new images but synthetic

samples generated in the feature space according to the available samples, we tailored the technique to image retrieval tasks, so that the new patterns provide additional information actually embedded into the available samples.

The choice of the number of artificial patterns to be created is not a trivial task. In fact, if the number is too large, there is the concrete risk to add noise to the dataset, thus producing a distortion of the *real* distribution of images. For these reasons, we propose to constrain the generation of new patterns so that the ratio between images that belong to the class of interest, and those that do not, is equal to a predefined ratio $1 : m$, where $m = 2, \ldots, 5$. In cases when the ratio in the training set exceeds the above ratio, then the artificial generation of patterns is not executed at all.

In the case of DPI, for each pattern $\mathbf{x}_{ref}$ and its K nearest patterns, it is possible to generate potentially an infinite number of synthetic patterns by varying the coefficients $\xi_1, \ldots, \xi_k$. These patterns lie in a region of the feature space defined by the reference pattern and its K neighbors. Further details on the implementation of the proposed technique will be provided in the following section.

On the other hand, the number of synthetic patterns that can be generated by SMOTE is limited by the number of available samples. If the number of samples of the minority class is T, then the maximum number of synthetic images that can be generated is equal to $T^2 - 1$. This can be a limiting factor in cases in which the dataset is heavily unbalanced. For example, if the number of patterns of the target class is equal to 2, and the number of patterns belonging to other classes is equal to 20, thus it follows that the total number of synthetic images that can be generated is equal to 3. If the goal is to generate synthetic patterns so that the ratio between patterns of the target class and patterns belonging to other classes is equal to $1 : 2$, then SMOTE in its original formulation does not allow attaining this goal. In addition, the patterns generated by SMOTE lie on the segments connecting pairs of nearest points of the target class, while DPI generates patterns in a region defined by the K nearest neighbors.

The value of K, i.e., the number of Nearest-Neighbors considered for the creation of artificial patterns, should not be large to avoid taking into account pattern that are actually far from the reference point (Skurichina et al., 2000). For example, the authors of SMOTE suggests the use of $K = 5$, and the patterns actually used to generate the synthetic patterns are drawn randomly from the neighborhood (Chawla et al., 2002).

The proposed approaches are particularly suited to be used with learning techniques based on the Nearest-Neighbor paradigm. In fact, new patterns are generated according to a linear combination of directions depending on the distribution of the K Nearest-Neighbors of a reference image, while the coefficients of the combination are random numbers. Starting from the assumption that the patterns of the same class lie in a manifold of the feature space, these techniques permit to identify the subspace determined by the K Nearest-Neighbors through the linear combination of known patterns and to generate the new ones in it.

The DPI technique further depends on the values assigned to a number of free parameters, that are, the image $\mathbf{x}_{ref}$ used as a reference, the number of Nearest-Neighbors considered, the number of artificial feature vectors that are generated, and the scale factor λ. These values cannot be a priori chosen. On the other hand, tuning all the free parameters in order to find the optimal configuration is a difficult and computationally expensive search task. We thus selected these parameters according to some heuristics that are reported in the following section.

5. Choice of the Parameters for the DPI technique

There are at least two parameters of DPI that need to be properly adjusted, namely the values of the parameters $\xi_1, \ldots, \xi_k$, and the choice of the reference point $\mathbf{x}_{ref}$ used to generate the artificial patterns. While the proposed approach generates new vectors only in some directions of the feature space (see Equation (1)), the generated patterns may still lie outside the region explored, i.e., the region defined by the known neighborhood of the reference point. In our opinion, it is too risky to generate patterns outside that region, as we have no information available about the distribution of images. This can happen depending on the random values of the coefficients $\xi_1, \ldots, \xi_k$. To avoid this risk, we propose to constrain the creation of new patterns in the region delimited by the nearest and the farthest known image of interest w.r.t. the reference image $\mathbf{x}_{ref}$ used in equation (1).

Different choices for the reference point can be investigated. In an image retrieval problem the reference point could be selected as being **the pattern associated to the query image**, as the user asked for images similar to the query. While this can be a reasonable choice, it can also exhibit some drawbacks, as its representation in the low-level feature space may not reflect its representativeness w.r.t. the images the user considers as being relevant. In other words, the so-called *semantic gap* between user perception of similarity and its representation in the low-level feature space may suggest to use a different point in the feature space as a query vector. As an alternative, we propose to use **the mean vector of all the known images of the target class** as the reference point, thus taking into account the distribution of the images of the class of interest in the feature space. This choice, with respect to the first one, takes into account all available information. In addition, this choice allows creating synthetic patterns that lie in the region where we actually observed images of interests.

Other options have been also investigated, and the experimental results have been reported in (Piras and Giacinto, 2010a). In order to keep the system simple to implement, by reducing the number of parameters which affects the final performance of the system, we decided to use the mean vector of all images belonging to the target class as the reference point.

6. Experimental Results

6.1. Datasets

Experiments have been carried out using two datasets, namely the Caltech-256 dataset, from the California Institute of Technology[1], and the Microsoft Research Cambridge Object Recognition Image Database[2] (in the following referred to as MSRC). The first dataset consists of 30607 images subdivided into 257 semantic classes (Griffin et al., 2007), while MSRC contains 4135 images subdivided into 17 main classes, each of which is further subdivided into subclasses, for a total of 32 classes. Three different kind of features have been extracted, namely the *Tamura* features (Tamura et al., 1978) (18 components), the *Scalable Color* descriptor (ISO/IEC:15938-3:2003, 2003) (64 components), and the *Color and Edge Directivity Descriptor* (*Cedd*, 144 components) (Chatzichristofis and Boutalis, 2008). The open source library LIRE (Lucene Image REtrieval) has been used for feature extraction (Lux and Chatzichristofis, 2008).

[1] http://www.vision.caltech.edu/Image_Datasets/Caltech256/

[2] http://research.microsoft.com/downloads

6.2. Experimental Setup

In order to test the performances, 500 query images have been randomly extracted for each dataset, covering all the classes. The top twenty best scored images for each query are returned to the user. Relevance feedback is performed by labelling images belonging to the same class of the query as relevant, and all other images in the top twenty as non-relevant. It is worth noting that at each round of relevance feedback, the user is asked to label twenty brand new images never seen before.

Performances are evaluated in terms of retrieval precision and recall. Precision is measured by taking into account the top twenty best scored images at each iteration, regardless they have been already labelled by the user. The recall takes into account all the relevant images retrieved so far, including the images labelled by the user to providing the feedback to the system. The recall is evaluated against the minimum between the number of relevant images in the dataset, and the total number of images evaluated by the user (i.e., the maximum number of relevant images that can be actually retrieved). Finally, in order to evaluate the improvement attained by the generation of synthetic patterns (DPI and SMOTE) w.r.t. the Nearest-Neighbor (NN) relevance feedback technique, the following improvement measure has been computed: $(performance_X - performance_{NN}) \cdot (performance_{NN})^{-1}$, where X is either SMOTE or DPI, and *performance* is either the precision or the recall measure. In order to choose the most suitable values of the parameters discussed in Section 5, a number of preliminary experiments have been performed (Piras and Giacinto, 2010b). Accordingly, we computed the normalization parameter $\lambda = \frac{1}{K}\left(\xi_1^2 + \xi_2^2\right)^{-2}$, and created new synthetic patterns at each iteration by taking in account only the information from the last iteration. On the other hand, the relevance feedback mechanism takes into account all the images retrieved so far, and all the synthetic images generated so far. Synthetic patterns have been created so that the final ratio between relevant and non-relevant images retrieved at the current iteration is equal to $1 : 2$. Figure 1 shows the average number of generated synthetic patterns at each iteration for the Caltech (C), and MSRC (M) dataset, respectively. As we expected, when the dataset contains a large number of classes (e.g., the Caltech dataset), then the set of retrieved images typically contains very few relevant images, and a large number of synthetic patterns are created.

We chose different values of K for the DPI and SMOTE approaches. In the case of DPI, we used a value of K equal to 2 (Piras and Giacinto, 2010b). In the case of SMOTE, the value of K depends on the number of synthetic patterns needed to attain the ratio between relevant and non-relevant images equal to $1 : 2$.

For comparison purposes, relevance feedback has been also computed by a SVM classifier with an *RBF* kernel whose variance has been estimated at each iteration through a 5-fold cross-validation procedure.

We also compared the performance of synthetic pattern generation with three undersampling techniques. One technique, referred to as Naive RF, selects for RF the best scored non-relevant images so that the number of non-relevant images is twice the number of relevant images. We also evaluated the performances of two other undersampling techniques based on the *Tomek links* (Tlinks) and the *Neighborhood Cleaning Rule* (NCL) (Batista et al., 2004), respectively. These two techniques exploit the Nearest-Neighbor paradigm to discard non-relevant images too close to the area where the relevant images lie. The Tomek links approach removes non-relevant images whose distance w.r.t. the nearest relevant image is smaller than the distance from their respective nearest non-relevant image. On the other hand, NCL considers each feedback image and its three Nearest-Neighbors. According to the majority rule, a non-relevant image is removed

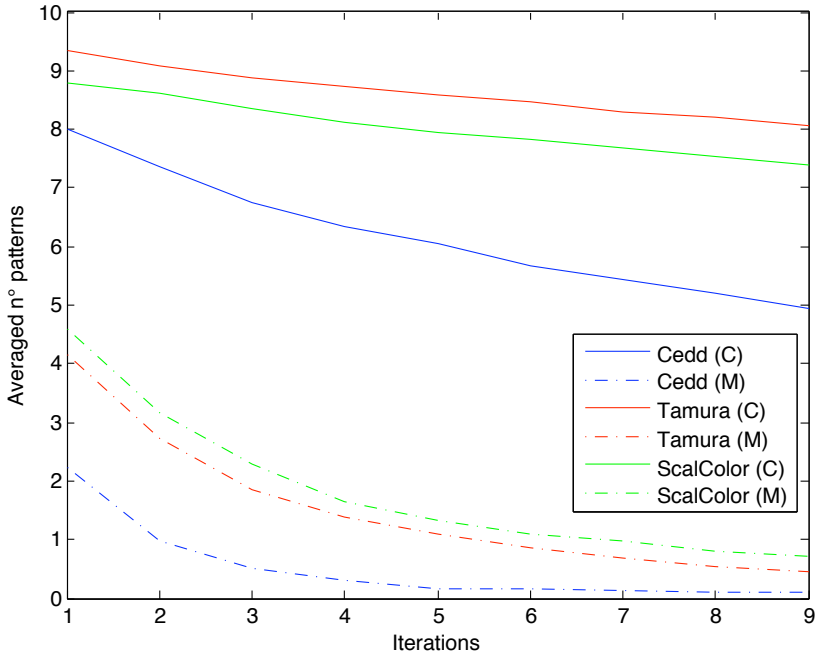

Figure 1: Average number of generated pattern at each iteration.

if at least two out of three of its Nearest-Neighbors are relevant images. Non-relevant images are also removed if they are the majority among the nearest neighbors of a relevant image.

Finally, as the three feature representations provide complementary information on image semantic, we also evaluated the performances attained by averaging the relevance scores computed separately for each feature representation by the Nearest-Neighbor technique (referred to as NN-Combination).

6.3. Results

Figures 2, 3, 4, and 5, show the performances in terms of precision and recall using the two datasets and the three feature representations. As we expected, the best performances on both datasets are attained by the *Cedd* representation, as it allows to better capture the different semantic of the classes in the dataset. On the other hand, both the *Tamura*, and the *Scalable Color* representations allow capturing only partially the semantic of the classes. This behavior can be easily seen by comparing the initial retrieval results on both datasets without relevance feedback. It can be also observed that the performance attained by the MSRC dataset are usually higher than those of the Caltech dataset. The reason of this behavior is related to the different semantic of the images contained in the two datasets, and to their subdivision into classes. The MSRC dataset is used for object recognition, and the vast majority of images contains just the object of interest, or at least a number of objects of the same kind. It turns out that the images contained in each class exhibit similar characteristics. On the other hand, the classes in the Caltech dataset comprise images whose visual content can be loosely related to the semantic of the class.

Reported results in Figures 2, 3, 4, and 5, show that the artificial generation of patterns allows improving the performance of relevance feedback in all the considered feature spaces. Figure 6 shows the average improvements in the three feature sets as described in 6.2. It can be seen that both DPI and SMOTE allow improving the precision, and the recall with respect to the "plain" Nearest-Neighbor relevance feedback technique. In addition, it can be also seen that DPI always outperforms SMOTE (Piras and Giacinto, 2010b). In particular, the main gap between DPI and SMOTE can be observed in the precision figure, where the difference in improvement between DPI and SMOTE is equal to 9% for the Caltech dataset, and 4% for the MSRC dataset. These results can be explained by the different technique employed by DPI and SMOTE in generating synthetic patterns. DPI generates patterns by exploiting the region defined by the mean vector of the patterns of the minority class, and its Nearest-Neighbors, whereas SMOTE generates patterns only on the segment connecting two neighboring patterns. It turns out that DPI allows for a better exploitation of the available information.

In the case of the Caltech dataset, and the *Cedd* feature representation, Figure 2 shows that the precision attained by the generation of synthetic patterns improves the performances attained by the "plain" Nearest-Neighbor relevance feedback technique, as well as with respect to the use of SVM. Reported results also show that all the available non-relevant images are useful to drive the search towards regions of the feature space where it is more likely to find relevant images, as the use of undersampling techniques (namely, NCL, Tlinks, and Naive RF) turned out to be not effective.

Finally, the precision of DPI is also higher than that attained by the combination of the three feature representations, starting from the third iteration. At the end of the ninth iteration, the improvement in precision attained by DPI is nearly equal to 3.5%. Thus it can be concluded that the generation of synthetic patterns in a feature space that is suited to the concepts the user is looking for can be more effective than the combination of information from different feature spaces. This aspect is particularly interesting from the point of view of time complexity. It is worth noting that while the combination of complementary feature representation may allow attaining improvements in performances, the computational overhead of the combination with respect to the most computational demanding feature space (i.e., the *Cedd* representation) is equal to 77%. On the other hand, the overhead of the proposed techniques based on the synthetic generation of patterns is equal to 15%. If we consider the recall reported in Figure 3, we can observe a behavior similar to the one seen in the case of the precision, apart that the performances of DPI and SMOTE are quite similar to each other, the DPI performing slightly better. In particular, significative improvements can be seen since the third iteration.

In the case of the *Tamura*, and the *Scalable Color* representations, we can observe a similar trend as far as the comparison of relevance feedback technique that exploit information on individual feature spaces are concerned. On the other hand, the combination of the three feature sets outperforms all other techniques, as it exploit the information from the *Cedd* representation. Thus, when the feature spaces are not effective for the task at hand, the generation of synthetic patterns is not competitive with respect to the performances attained by the combination of different features.

In the case of the MSRC dataset, it can be seen that using the *Tamura*, and the *Scalable Color* representations, the behavior is the same as the one seen in the Caltech dataset. In particular, the generation of synthetic patterns allows for performance improvements, the DPI providing the best performances in the precision (see Figure 4) since the first iteration. In the case of the recall measure (Figure 5), the improvement starts since the fifth iteration, and the performance attained by both DPI and SMOTE reaches the one attained by the combination of the three

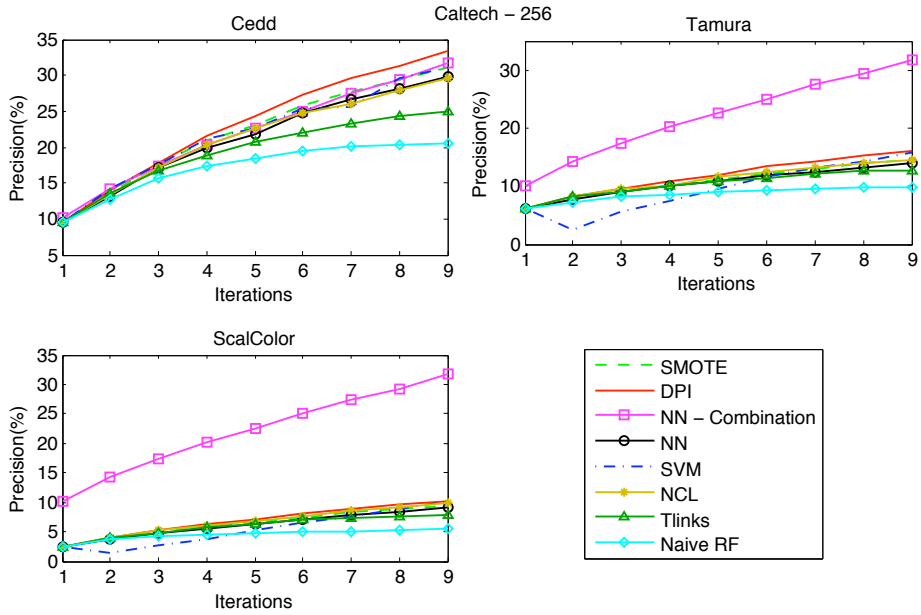

Figure 2: Caltech 256 Dataset - Precision for 9 rounds of relevance feedback.

feature sets at the ninth iteration. In the *Cedd* feature space, differently from the behaviour in the Caltech dataset, the precision attained by generating synthetic patterns is always slightly worse than the one attained by the combination of information from different feature spaces. On the other hand, the performance attained for the recall shows the effectiveness of the generation of synthetic patterns with respect to the "plain" Nearest-Neighbor relevance feedback, the SVM, the combination of different feature spaces, and the techniques based on undersampling.

7. Conclusion

In this paper we proposed a technique that address the imbalance problem in image retrieval tasks by generating synthetic patterns according to Nearest-Neighbor information. Reported results show that the proposed technique allows improving the performances not only with respect to the performances attained without artificial patterns, but also with respect to performance attained by combining different feature spaces. The computation overhead of the proposed technique is small if compared to the overhead due to the combined use of different feature spaces.

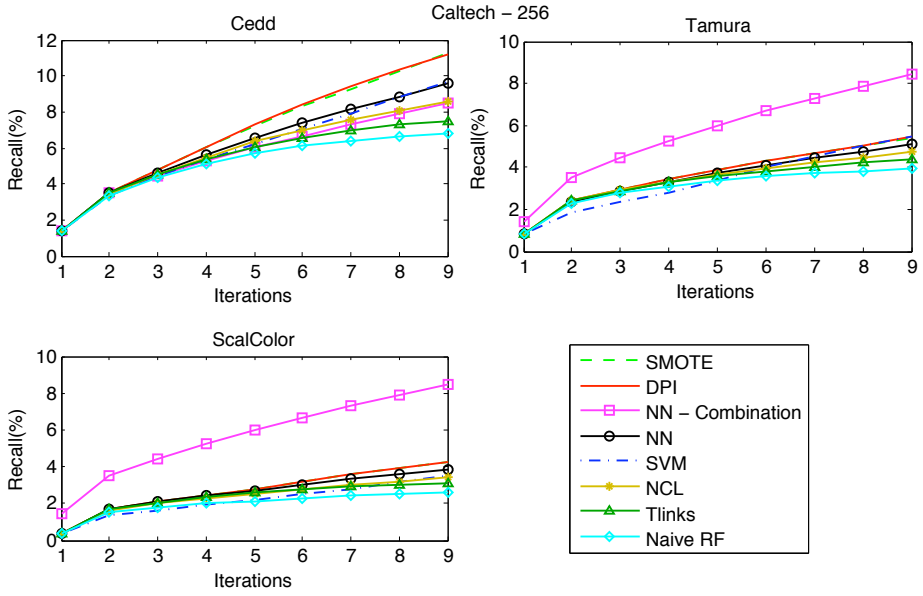

Figure 3: Caltech 256 Dataset - Recall for 9 rounds of relevance feedback.

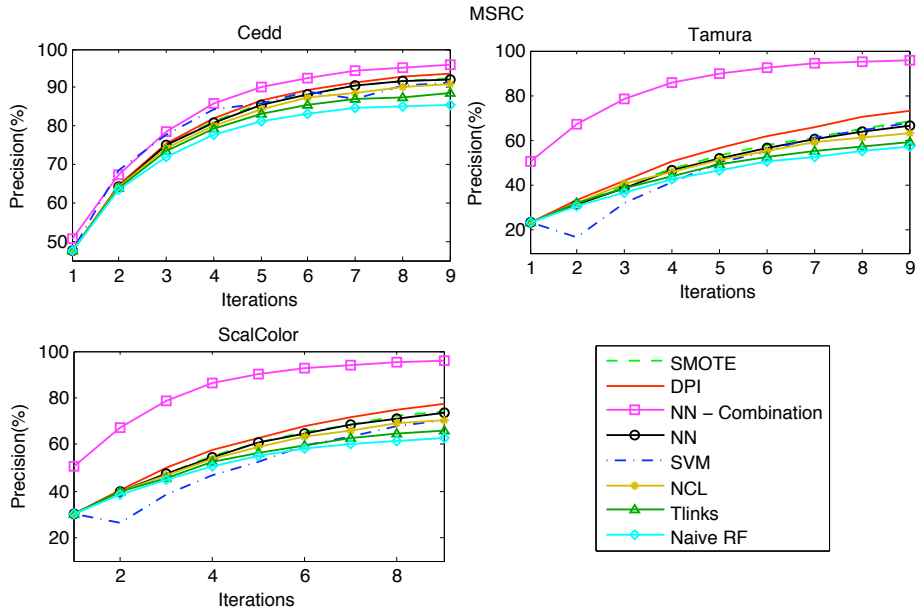

Figure 4: MicroSoft Research Dataset - Precision for 9 rounds of relevance feedback.

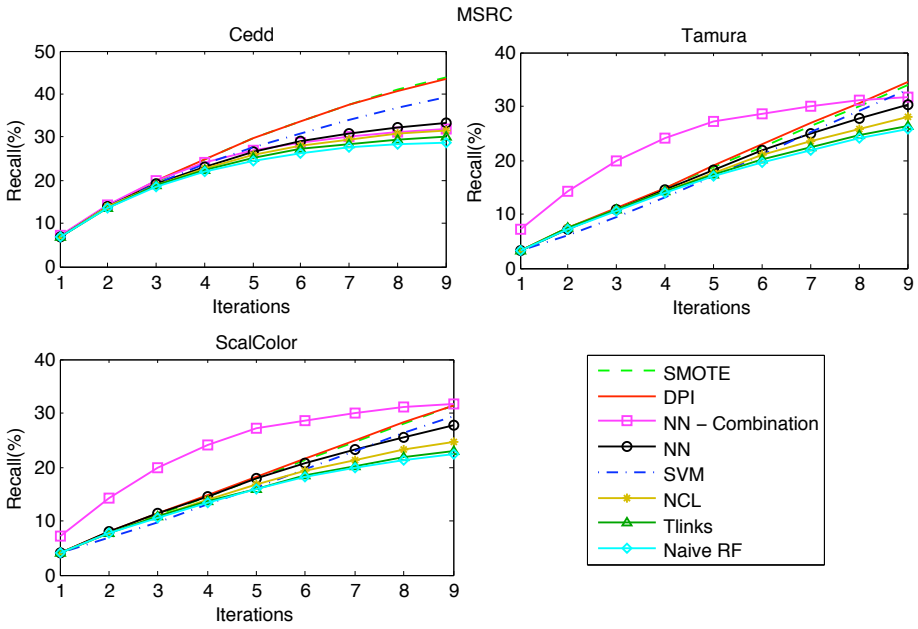

Figure 5: MicroSoft Research Dataset - Recall for 9 rounds of relevance feedback.

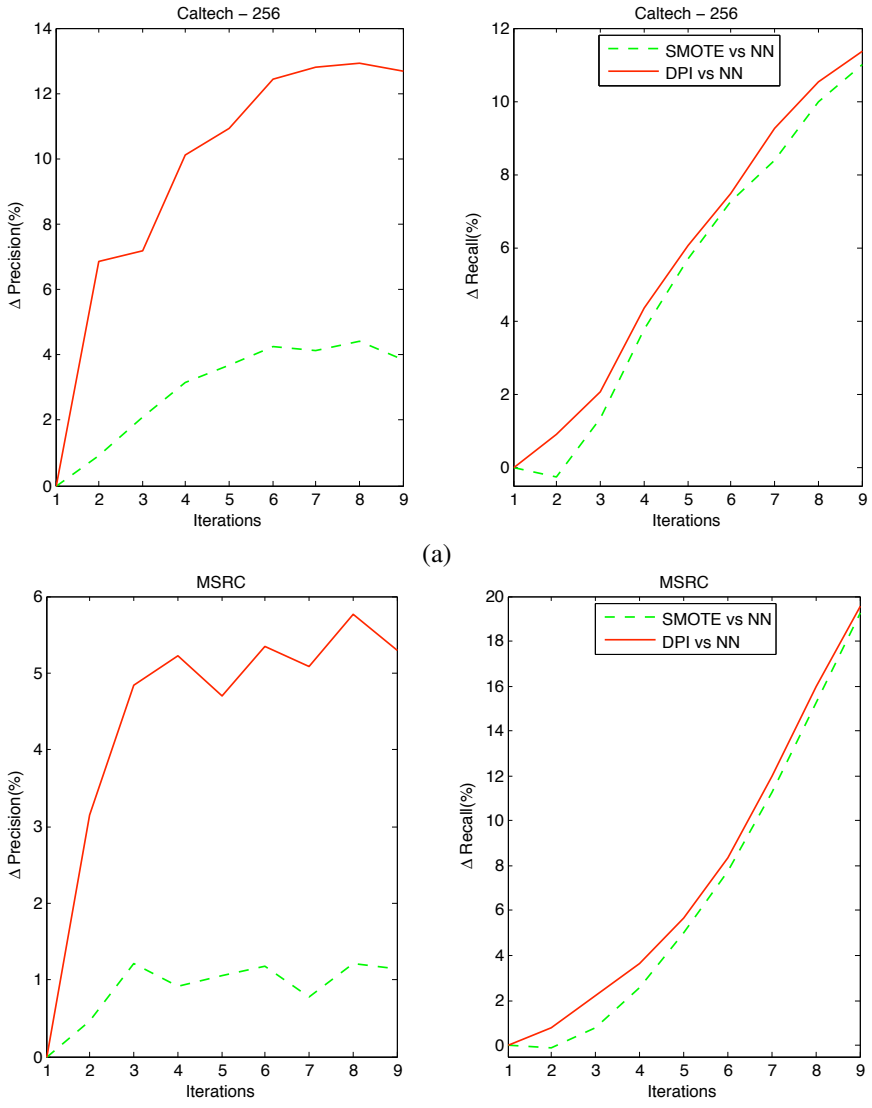

Figure 6: Performance Improvements of SMOTE and DPI against the Nearest-Neighbor technique for relevance feedback, averaged over the three feature spaces. (a) Caltech 256 Dataset - (b) Microsoft Research Dataset

References

Batista, G., Prati, R.C., Monard, M.C., 2004. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explorations 6, 20–29.

Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J., 2000. LOF: Identifying density-based local outliers, in: Chen, W., Naughton, J.F., Bernstein, P.A. (Eds.), SIGMOD Conference, ACM. pp. 93–104.

Chatzichristofis, S.A., Boutalis, Y.S., 2008. Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval, in: Gasteratos, A., Vincze, M., Tsotsos, J.K. (Eds.), ICVS, Springer. pp. 312–322.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. Smote: Synthetic minority over-sampling technique. J. Artif. Intell. Res. (JAIR) 16, 321–357.

Chen, Y., Zhou, X.S., Huang, T., 2001. One-class svm for learning in image retrieval, in: ICIP, pp. 34 –37 vol.1.

Datta, R., Joshi, D., Li, J., Wang, J.Z., 2008. Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys 40, 1–60.

Duin, R.P.W., 2004. Pattern recognition in almost empty spaces, in: WIC Winter Symposium, Eindhoven, Netherlands.

Giacinto, G., 2007. A nearest-neighbor approach to relevance feedback in content based image retrieval, in: CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval, ACM, New York, NY, USA. pp. 456–463.

Giacinto, G., Roli, F., 2004. Bayesian relevance feedback for content-based image retrieval. Pattern Recognition 37, 1499–1508.

Griffin, G., Holub, A., Perona, P., 2007. Caltech-256 Object Category Dataset. Technical Report 7694. California Institute of Technology.

He, H., Garcia, E., 2009. Learning from imbalanced data. Knowledge and Data Engineering, IEEE Transactions on 21, 1263 –1284.

Huijsmans, D.P., Sebe, N., 2005. How to complete performance graphs in content-based image retrieval: Add generality and normalize scope. IEEE Trans. Pattern Anal. Mach. Intell. 27, 245–251.

ISO/IEC:15938-3:2003, 2003. Information technology - Multimedia content description interface - Part 3: Visual.

Lew, M.S., Sebe, N., Djeraba, C., Jain, R., 2006. Content-based multimedia information retrieval: State of the art and challenges. ACM Trans. Multimedia Comput. Commun. Appl. 2, 1–19.

Li, J., Wang, J., 2008. Real-time computerized annotation of pictures. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30, 985 –1002.

Lux, M., Chatzichristofis, S.A., 2008. Lire: lucene image retrieval: an extensible java cbir library, in: MM '08: Proceeding of the 16th ACM international conference on Multimedia, ACM, New York, NY, USA. pp. 1085–1088.

Piras, L., Giacinto, G., 2010a. K-nearest neighbors directed synthetic images injection, in: Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 11th International Workshop on, pp. 1–4.

Piras, L., Giacinto, G., 2010b. Unbalanced learning in content-based image classification and retrieval, in: ICME, IEEE. pp. 36–41.

Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326.

Skurichina, M., Raudys, S., Duin, R.P.W., 2000. K-nearest neighbors directed noise injection in multilayer perceptron training. IEEE Trans. on Neural Networks 11, 504–511.

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R., 2000. Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. 22, 1349–1380.

Tamura, H., Mori, S., Yamawaki, T., 1978. Textural features corresponding to visual perception. IEEE Trans. Systems, Man and Cybernetics 8, 460–473.

Tao, D., Tang, X., Li, X., Rui, Y., 2006a. Direct kernel biased discriminant analysis: A new content-based image retrieval relevance feedback algorithm. IEEE Trans. on Multimedia 8, 716–727.

Tao, D., Tang, X., Li, X., Wu, X., 2006b. Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence 28, 1088–1099.

Tax, D.M., 2001. One-class classification. Ph.D. thesis. Delft University of Technology. Delft, The Netherlands.

Thomee, B., Huiskes, M.J., Bakker, E.M., Lew, M.S., 2008. Using an artificial imagination for texture retrieval, in: Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, pp. 1–4.

Wasikowski, M., wen Chen, X., 2010. Combating the small sample class imbalance problem using feature selection. Knowledge and Data Engineering, IEEE Transactions on 22, 1388 –1400.

Zhou, X.S., Huang, T.S., 2001. Small sample learning during multimedia retrieval using biasmap, in: CVPR (1), IEEE Computer Society. pp. 11–17.

Zhou, X.S., Huang, T.S., 2003. Relevance feedback in image retrieval: A comprehensive review. Multimedia Syst. 8, 536–544.