Precision-Oriented Active Selection for Interactive Image Retrieval.
Philippe-Henri Gosselin, Matthieu Cord

To cite this version:

HAL Id: hal-00520305
https://hal.archives-ouvertes.fr/hal-00520305
Submitted on 22 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Active learning methods have been considered with an increased interest in the content-based image retrieval (CBIR) community. Those methods used to be based on classical classification problems, and do not deal with the particular characteristics of the CBIR. One of those characteristics is the criteria to optimize, for instance the error of generalization for classification, which is not the most adapted to CBIR context. Thus, we introduce in this paper an active selection which chooses the image the user should label such as the Mean Average Precision is increased. The method is smartly combined with previous propositions, and lead to a fast and efficient active learning scheme. Experiments on a large database have carried out in order to compare our approach to several other methods.

1. INTRODUCTION

Human interactive systems has attracted a lot of research interest in recent years, especially for content-based image retrieval systems. Contrary to the early systems, focused on fully automatic strategies, recent approaches introduce human-computer interaction [1, 2].

Starting with a coarse query, the interactive process allows the user to refine his request as much as necessary. Many kinds of interaction between the user and the system have been proposed [3], but most of the time, user information consists of binary annotations (labels) indicating whether or not the image belongs to the desired category.

In this paper, we focus on the retrieval of concepts within a large document collection. We assume that a user is looking for a set of documents, the query concept, within an existing document database. The aim is to build a fast and efficient strategy to retrieve the query concept.

Performing an estimation of the query concept can be seen as a statistical learning problem, and more precisely as a binary classification task between the relevant and irrelevant classes [4]. The relevant class is the set of documents within the query concept, and the irrelevant class the set of documents out of the query concept.

Active learning have been introduced with success in CBIR context in order to deal with the interaction between the user and the system [5]. However, a lot of active strategies consider CBIR as a pure classification problem, and thus are not fully adapted to the special characteristics of this context. For instance, we have shown in a previous paper [6] that the few training data and the imbalance of the classes lead to a noisy boundary, which quality is an important factor in active classification.

In this paper, we focus on another characteristic of CBIR : the active selection criterion. Active classification aims at minimizing the error of classification. However, this criterion is not the most representative of user satisfaction. Thus, we propose to select images the user should label using common metric of CBIR, the Precision.
it has revealed being the most efficient [5, 7]. The result is a function $f_y(x_i)$ which returns the relevance of each image $x_i$, according to the examples $(x_i, y_i), i \in I$.

**Correction.** We add an active correction to the boundary in order to deal with the few training data and the imbalance of the classes (the size of the searched category used to be very small against the number of irrelevant images). Details about this method can be found in [6].

**Selection.** The main purpose of this paper comes in this part. In the case where the user is not satisfied with the current classification, the system selects a batch of images the user should label. The selection may be such as the labeling of those images will give the best performances. We divide the selection in three steps.

The first step aims at reducing the computational time, by pre-selecting some hundreds of pictures which may be in the optimal selection set. We propose to pre-select the closest pictures to the (corrected) boundary. This process is computed very fast, and the uncertainly-based selection method have proven their interest in CBIR context.

The second step is the computation of the selection criterion. In active classification, the criterion is the minimization of the error of classification (or risk). In these cases, the risk is computed for each classification function $f_y(t(x_i))$, which is trained with the label $t(x_i)$ of an unlabeled image $i \notin I$ added to current training set $(y)$. The selected image $i^*$ is the one which minimizes the risk:

$$i^* = \arg\min_{i \notin I} \text{risk}(f_y(t(x_i)))$$

The main difficulty of this task is the fact that the label $t(x_i)$ is unknown, and an estimation is required. This estimation is replaced with a cost function that we denote by $g_y(x_i)$, and including the pre-selection, the problem can be written as:

$$i^* = \arg\min_{i \in I} g_y(x_i)$$

The cost function is, for instance, the distance to the boundary for the SVM active method [5]: $g_y(x_i) = |f_y(x_i)|$. In this paper, we introduce a cost function $g_y(x_i)$ in this scheme, aiming at maximizing the Average Precision.

The third step of active selection computes the batch selection. As we focus on real-time application, we use a fast method close to the angle diversity [8]. The method selects $g$ images using the previously computed cost $g_y(x_i)$, and return the set $I^*$ of image indexes the user should label:

$$I^* = \{\}$$

forall $l \in [1..g]$:

$$i^* = \arg\min_{i \notin I^*} (g_y(x_i) + \max_{j \in I^*} s(x_i, x_j))$$

$$I^* = I^* \cup \{i^*\}$$

endfor

where $s(x_i, x_j)$ is the similarity (using the kernel) between image $x_i$ and image $x_j$.

**Feedback.** The user labels the selected images, and a new classification and correction can be computed. The process is repeated as many times as necessary.

3. ERROR OF GENERALIZATION VS MEAN AVERAGE PRECISION (MAP)

Active classification methods have been built to select elements which decrease the error of classification. However, in our interactive CBIR context, users are concerned by a similarity ranking of the database. An usual metric to evaluate this ranking is the Mean Average Precision\(^1\), i.e. the sum of the Precision/Recall curve.

Active classification methods has already proven their capacity to increase the Average Precision, whenever this is not what they aim at. In order to evaluate the two criteria, – classification error-based minimization and MAP-based maximization, we propose an experimentation using a ground-truth.

Active learning methods aim at optimizing some criterion, but in practice the true evaluation of the criterion is impossible, because we do not have the ground truth. Thus, active classification methods differ one to another by the way they estimate the error of classification. In this section, we propose to use two “perfect” active learner, with no estimation, in order to compare the two criteria. These methods are not designed for real application, and use the ground truth to compute the error of classification or the Average Precision.

\(^1\)cf. TREC VIDEO conference:
http://www-nlpir.nist.gov/projects/trecvid/
The results of those experiments are shown on Figure 2 and 3. Considering error of classification (cf. Fig. 2), the methods which directly optimize the error criterion is the most efficient. However one can see that maximizing the Average Precision also decreases the error. Considering the criteria that interest us (cf. Fig. 3), we can see a similar but inverted behavior. The most interesting result is the fact that, whenever the minimization of the error of classification increases the Average Precision, direct maximization of the Average Precision leads to a large increasing of performance, with a gain of 20%.

4. PRECISION-ORIENTED SELECTION

The results of the previous experiments lead us to the exploration of a precision-oriented selection. As an active classification method has to minimize the error of classification, the aim here is to select the images so that the Average Precision will be increased. Furthermore, we also need to perform an estimation of the criterion.

However, estimating the Average Precision is particularly difficult with the kind of samples we have chosen. As we work on generalist CBIR systems, we opted for binary labels, which are simple enough for any non-expert user. These labels are well adapted for classification, but are less suitable with the estimation of Average Precision: the criterion is based on a ranking of the database, but binary labels do not give any information in that way.

Some researchers propose to use true/false positives/negatives in order to solve the problem. However, in our context, the number of samples is so small that it becomes especially difficult to find such an information. In other words, the labels are so precious that each one will necessarily have a huge weight in the result.

Thus, we propose to stay in an active classification framework, and introduce a bias in the selection in order to support the increasing of the Average Precision. We weight the distance to the boundary by a factor $h_y(x_i)$:

$$g_y(x_i) = |f_y(x_i)| \times (1 - h_y(x_i))$$

The aim of introducing $h_y(x_i)$ is to support the picture $x_i$ so that, once labeled, the Average Precision is increased. In order to get this behavior, we propose to consider the subset of the labeled pictures. As we have the labels of all its images from the sub set, we can use it as ground truth. Thus, it becomes feasible to compute the Average Precision on this subset, without any estimation. We still need a ranking of this subset in order to compute the Average Precision. We propose to compute the similarity of an unlabeled image $x_i$ to all to labeled images, and then rank the labeled images according to these similarities. The resulting factor $h_y(x_i)$ is the Average Precision on the subset with this ranking. Note that we could use the ranking provided by the classifier, however, still because of the few training data and the binary labels, this approach that we have experimented is not efficient.

5. EXPERIMENTS

We experimented different active learning methods with the following protocol:

Database. The image database is an extract of 6,000 images of the COREL photo database. To get tractable computation for the statistical evaluation, we randomly selected 77 of the COREL folders, to obtain a database of 6,000 images.

Features. We use an histogram of 25 colors and 25 textures for each image, computed from a vector quantization.
6. CONCLUSION

In this paper, we showed the interest of active learning based on an Average Precision maximization criterion for interactive image retrieval. We introduced a selection method in that way, and combine it with other techniques in a global active learning scheme. The method has been validated through experiments and compared to reference active learning methods. The results show that it is a powerful tool to improve the performances of image category retrieval tasks.

7. REFERENCES


