Neurocomputing 168 (2015) 829–845

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Modeling user preferences in content-based image retrieval: A novel attempt to bridge the semantic gap

Esther de Ves [a,*], Guillermo Ayala [b], Xaro Benavent [a], Juan Domingo [a], Esther Dura [a]

[a] University of Valencia, Department of Informatics, Spain
[b] University of Valencia, Department of Statistics and Operations Research, Spain

ARTICLE INFO

Article history:
Received 12 March 2014
Received in revised form
10 March 2015
Accepted 3 May 2015
Communicated by Xiaofei He
Available online 20 May 2015

Keywords:
Semantic gap
Proportional odds model
Information retrieval
Relevance feedback
Content-based image retrieval

ABSTRACT

This paper is concerned with content-based image retrieval from a stochastic point of view. The semantic gap problem is addressed in two ways. First, a dimensional reduction is applied using the (pre-calculated) distances among images. The dimension of the reduced vector is the number of preferences that we allow the user to choose from, in this case, three levels. Second, the conditional probability distribution of the random user preference, given this reduced feature vector, is modeled using a proportional odds model. A new model is fitted at each iteration. The score used to rank the image database is based on the estimated probability function of the random preference. Additionally, some memory is incorporated in the procedure by weighting the current and previous scores. Also, a novel evaluation procedure is proposed in this work based on the empirical commutative distribution functions of the relevant and non-relevant retrieved images. Good experimental results are achieved in very different experimental setups and tested in different databases.

1. Introduction

Content based image retrieval is the process by which a system automatically selects a set of images from a possibly very large collection that match a user's preference, expressed either in words or as a visual query (showing the system one or a small sample of images that meet the user's intention). Most of the collections are not semantically annotated by textual labels and consequently the selection of relevant images is based only on visual features. Indeed, in most cases, low level visual features related with color, texture, etc., and less commonly mid-level features based on regions are extracted and compared with the query. As is universally acknowledged, the main challenge in the design of these systems is how to bridge the semantic gap between low level representation, mostly in the form of a vector of numerical features, and high level semantic representation of the user's intention expressed either textually or as a visual query. This goal was already introduced in publications as early as [6], [38], [37], or [27].

1.1. Previous work

The following review of previous works will focus on certain issues of CBIR systems, namely feature vector dimensionality reduction, query movement and/or expansion, combination of subspaces of features and image ranking; these issues are directly related to our contribution which will be explained in detail in Section 2.

Some ideas that have been previously applied by other researchers to reduce the semantic gap involve the reduction of the dimensionality of the feature vector, since it is assumed that the new dimensions of the reduced space have some kind of semantic significance. Common approaches rely on linear transformations, mostly by Principal Component Analysis (PCA) such as [32] that retain a certain amount of variation. Other ideas use Support Vector Machines (SVM) such as [35] or [36], which are known to cope well with the problem of high dimensionality with respect to the available data set size (the "curse of dimensionality"). A less common approach to dimensionality reduction uses a non-linear transformation based on the projection on subspaces of smaller dimension defined by the nearest neighbors of each point [34]. On the contrary, other authors prefer to make a previous classification of the training set to extract a small set of representatives and use some sort of distance to the elements of this set as features. The advantage is that a high reduction of the dimensionality can be achieved without compromising the system's effectiveness so

* Corresponding author.
E-mail addresses: esther.deves@uv.es (E. de Ves), Guillermo.Ayala@uv.es (G. Ayala), Xaro.Benavent@uv.es (X. Benavent), Juan.Domingo@uv.es (J. Domingo), Esther.Dura@uv.es (E. Dura).

http://dx.doi.org/10.1016/j.neucom.2015.05.041
0925-2312/

much, the work proposed in [26] being highly relevant. Other examples are [18] in which a Radial Basis Function is used to apprehend the topological structure of the semantic space, and [13] in which the structure is being learned during the relevance feedback process. The main drawback of some of these approaches is that the semantic meaning of some features, or groups of them, is completely hidden which may cause what the user perceives as erratic behavior during the feedback process.

The search carried out to satisfy the query is helped by the user's feedback through interaction using a graphical user interface, an approach known as relevance feedback which has been routinely used in recent years, [28] being a classical widely cited work.

Regarding the process of feedback, there are several ways of incorporating the information provided by the user. Roughly, these can be classified as techniques based on query re-weighting, query expansion and query movement. Two recent compact summaries of these classifications, and a comparative study, are [29,24]. The first group (query re-weighting) changes the weights assigned to each feature, or group of them, based on the user's choices. A typical example can be seen in [8]. This is most commonly done by altering the weights of a pseudo-Euclidean metric which is used to calculate distances between the query image and all the images in the database, as in [36]. On the other hand, query expansion proceeds by adding more images to the original query taken from those the user marks as positive during the feedback process and finally, the query movement acts by changing the original query proposed by the user, on the understanding that it was intrinsically ambiguous and that it will be refined by the users themselves through their own choices during the feedback process. A substantial difference between query expansion and query movement, pointed out by [21], is an underlying assumption assumed in query movement: that the relevant images form a uni-modal cluster in the used feature space. This can leave out entire collections of images that would be classified as relevant if shown to the user (false negatives). On the contrary, query expansion techniques aim to admit a multimodal query which is usually the case, especially for complex semantic requests, but they have the drawback of giving a higher number of false positives (images given by the system as relevant, but which are not). An interesting example of a clever combination of both methods is [21].

Both query expansion and query movement can be considered as ways of learning the user's preferences. Differences can be established according to how this information is used across successive iterations. Most methods only use the choices of the last iteration, assuming that former ones are implicitly incorporated into the current state, but more complex methods may take into account the whole history of the search (user's log). Interesting examples are [15,31].

An important point in CBIR systems is how to rank the images in the database to show them to the user. Ranking by distance to the query is the most obvious choice, but if the query is multi-objective (which is always true in query expansion techniques) some global measure of ranking must be used. There are examples based on post-retrieval clustering [23] or on rank aggregation [25]. An experimental comparison of some of these methods can be found in [17].

Finally, a less treated but important point in CBIR systems is the system's ability to rank the images and show them to the user in a reasonable time. This is compulsory if several iterations must be performed to attain a result of sufficient quality. Obviously, the key points to be considered are the computational cost of the evaluation of the similarity index chosen for a given image, the cost of ranking and the total number of images in the database. In our experience, the most important point is the database size and, close to this, the evaluation cost. Many of the published experiments work with small databases (around 1000 images), or medium-size ones (up to 100,000 images) with the highly relevant exception of [9] which evaluates its algorithm in a 100-million image database using a similarity caching system.

1.2. Contribution

The main differences between the previously cited works and the current work in each of the aforementioned issues are as follows:

The relationship between low-level features and high-level preferences (reduction of the semantic gap) will be approached by using generalized linear models, in short, GLM [20]. The use of GLMs requires either a relatively large number of images evaluated by the user, or the reduction of the dimensionality of the low level feature vector. We have opted for the second approach: indeed, a significant reduction of the dimension of the feature vector is done by using a new procedure that relies on a previously evaluated matrix that contains the distance between every pair of images in the database. Once the dimensionality has been reduced, the GLMs can be applied. In particular, an accumulated proportional odds model will be used.

Regarding the feedback process, what we change in each iteration are the coefficients of a generalized linear model that links a weighted Mahalanobis distance to the query components with the probability of each image being similar to the query; this can be seen as a sophisticated way of query re-weighting. Our system does not carry out query movement (the query keeps all the original images), but it does query expansion (the images marked as relevant as long as other images given by the model in successive iterations are added to the query) with the particularity that the images visited (seen by the user, but not explicitly marked) are classified and used, as well, for the current iteration.

With respect to the ranking procedure, we decided that, since the images can be classified into three categories (relevant, neutral and non-relevant), good ranking can be built by the weighted addition of probabilities of belonging to the first two classes.

Finally, in order to accelerate the search we use a pre-calculated table of distances, model fitting with the provided data and model evaluation on the whole database are not critical since the generalized linear model is expressed as a simple formula.

2. Methodology

As stated before, we are concerned with the retrieval of images within large databases by using stochastic modeling. In particular, the random preference of the user given the low level features of the image is the event to be modeled. This in turn involves the choice of appropriate low level features, the reduction of their dimensionality so that a sound model can be fitted and the ranking of the results based on the evaluation of the model and taking into account the user's feedback.

Although the original motivation of this paper was concerned with Content Based Image Retrieval (CBIR) systems, many aspects of the proposed methodology could be applied to searching in other types of databases with minor modifications.

The general methodology involves seven steps:

1. Selection of an appropriate set of low level visual features and their computation for all the images of the database, and for all the images in the given query. This will generate a feature vector associated with each image that will be generically denoted as $\boldsymbol{x}$. The feature vector will be considered as compound by several sub-vectors, each of them containing semantically related features (for instance summaries of color histograms or texture descriptors).