PATTERNQUEST: LEARNING PATTERNS OF INTEREST USING RELEVANCE FEEDBACK IN MULTIMEDIA INFORMATION RETRIEVAL

Yimin Wu and Aidong Zhang

Department of Computer Science and Engineering,
State University of New York at Buffalo,
Buffalo, NY 14260, USA
{yiminkwu,azhang}@cse.buffalo.edu

ABSTRACT

In this paper, we present a PatternQuest framework to learn the patterns of interest (i.e., the distribution patterns of positive objects) using classification methods and relevance feedback. To improve the performance of multimedia retrieval, our PatternQuest first employs an efficient feature selection method to extract a low-dimensional feature subspace. With the feature selection, PatternQuest can effectively alleviate the curse of dimensionality for learning-based relevance feedback. To effectively discover patterns of interests in the feature subspace, we propose a multiresolution pattern discovery (MPD) approach, which trains an online pattern classification method known as adaptive random forests to filter negative objects, from the neighborhood of the query to the global scope, in a fine to coarse way. With MPD, our PatternQuest method can iteratively capture the patterns of interest with small training data from the user's feedback. We have carried out extensive experiments on an image database (with 31,438 COREL images) to demonstrate the effectiveness and robustness of our method.

1. INTRODUCTION

In recent years, the rapid increase in the volume of multimedia data collections motivates the research in multimedia information retrieval. In multimedia data collections, multimedia objects (such as images and video clips) are indexed with high-dimensional feature vectors (i.e., points). To successfully retrieve objects of interest, it is critical to capture their distribution patterns (which we term as *patterns of interest*). However, the mapping between patterns of interest and high-level semantic concepts (such as animals, buildings and plants, etc.) are often highly nonlinear, due to the subjective nature of information retrieval and the ambiguity of multimedia content. To address this issue, a promising method is to employ pattern classification methods [1, 2, 3, 4, 5, 6], which can learn patterns of interest from a set of samples.

Nonetheless, in multimedia retrieval, the application of pattern classification methods face the following challenges. First, due to the high-dimensionality of multimedia feature spaces, most classification methods face *the curse of dimensionality* [7], which degrades both classification accuracy and efficiency. Second, classification methods have to learn patterns of interest with small training data, because it is very tedious and time-consuming to label a large volume of training data. To address these issues in a principled way, we present a *PatternQuest* framework to discover the patterns of interest using pattern classification methods and relevance feedback. Our contributions are summarized as follows.

To alleviate the curse of dimensionality, we employ an *efficient feature selection* method (EFS) [8] to extract a low-dimensional feature subspace (from the original feature space). By extracting a subspace, our method not only removes the noises from the feature spaces, but also effectively improves the efficiency of the learning machine.

In the feature subspace, our method learns patterns of interest with a novel multi-resolution pattern discovery approach (MPD), which trains an online pattern classifier known as *adaptive random forests* (ARF) [6, 9] to classify database objects as positive or negative. ARF is chosen due to its proven effectiveness on learning with small training data [6, 9] (from relevance feedback). In our method, ARF is trained as a filter to filter negative objects from the feature space. To improve the effectiveness of the filtering operation, our MPD trains ARF with samples from both the neighborhood of the query and the centroids of the database: samples from the neighborhood manifest finer distribution information near the query, while samples from centroids of the database present a coarser sketch about the global distribution patterns. By training ARF with samples from both the above-mentioned sources, our method uses ARF to filter negative objects, from the neighborhood of the query to global scope, in a fine to coarse way. Thus, our approach can gradually capture the patterns of interest with small training data from the user's feedback. Experiments

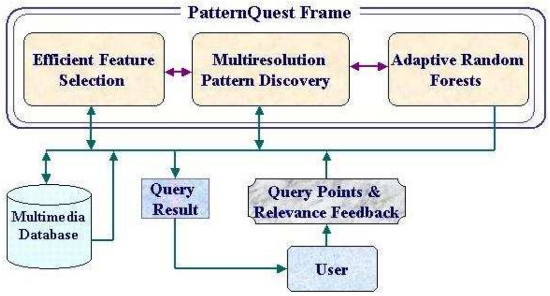

Fig. 1. An illustration of our PatternQuest framework.

on a large-scaled image database demonstrate the effectiveness and robustness of our presented approach.

The rest of this paper is organized as follows. Section 2 presents the structure of our PatternQuest framework. We then introduce our multiresolution pattern discovery method in Section 3. Empirical results are given in Section 4. Finally, we conclude in Section 5.

2. PATTERNQUEST FRAMEWORK

In this section, we present our *PatternQuest* framework for discovering patterns of interest using relevance feedback.

Figure 1 demonstrates the following three main components of our PatternQuest framework:

- **Efficient Feature Selection (EFS)**. EFS [8] is a novel feature selection method, which measures feature importance with an information- theory-based criterion. Our extensive experiments [8] showed that EFS can effectively alleviate the curse of dimensionality for classifying multimedia data.

- **Multiresolution Pattern Discovery (MPD)**. MPD is a hierarchical pattern discovery method. We will introduce MPD in the next section.

- **Adaptive Random Forests (ARF)**. ARF [6] adapts the composite classifier known as random forests for relevance feedback. It runs 2-3 times faster than the regular random forests, while achieving comparable retrieval performance against the latter.

In multimedia retrieval, our PatternQuest interactively learns the patterns of interest as follows. In the beginning, our method runs an initial retrieval and returns the k-NN (i.e., k nearest neighbors) of the query to the user. It then iterates through the following steps to improve the initial retrieval results:

1. Asks the user to provide feedback by labeling retrieval results as positive or negative;
2. Extracts the feature subspace with EFS;
3. Calls MPD, which trains an ARF to filter negative objects;
4. Returns K-NN of the query from objects that are classified as positive (by ARF); and updates the query as the centroid of positive objects.

After the above steps, the user can choose to start a new iteration by providing more training samples, or provide another query to start a new retrieval process.

3. MULTIRESOLUTION PATTERN DISCOVERY

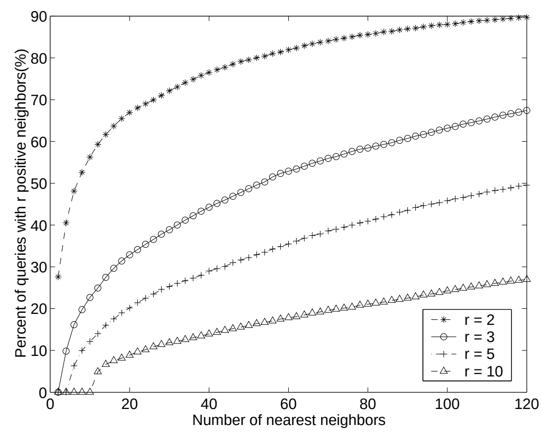

Fig. 2. Percentage of queries with r positive neighbors in their neighborhood, the size of which increases from 1 to 120.

In order to capture the user's subjective query concept, most relevance feedback methods do not assume any preknowledge about the data. To initialize relevance feedback, they often prompt a set of nearest neighbors (of the query) for the user to annotate. For this learning scheme, a critical issue is that the machine needs to start its learning process with small positive training data. As an illustration, we demonstrate the necessity for handling small positive training data with our test database (cf. Section 4). From the database, we choose 5,172 images from 44 semantic classes (such as rose, butterfly and falls, etc.) as queries. For each query, we perform a nearest neighbor search. In Figure 2, we present the percentage of queries that have the given number of positive objects in its nearest neighbors. We can see from Figure 2 that only about 50% queries have 5 positive images [1] in its top 120 nearest neighbors. If the user only checks the top 20 nearest neighbors, only 20% queries have 5 positive images, while about 33% queries have only

[1]In our experiments, only images from the same semantic class are counted as positive.

one positive image (which is the query itself). Hence, we can draw the following observation:

Observation 1 *In our test database, if each image from our query set is used (as a query by the user) with equal probability, in most cases, the learning machine has to initialize the relevance feedback with 1-5 positive training samples.*

As shown in Figure 3, in the first few iterations, the positive training samples can not well manifest the patterns of interest, so we may not be able to train a classifier as an *oracle* about the class of multimedia objects. Hence, the key issue (in initial iterations) is to effectively enlarge the positive training data. To address this issue, our method exploits the fact that negative samples are more readily available than positive ones. Instead of training an adaptive random forest (ARF) as an *oracle*, our approach employs it as a filter to filter negative objects. To improve the effectiveness of the filtering operation, we present a multiresolution pattern discovery method (MPD), which trains ARF to filter negative objects in a hierarchical way.

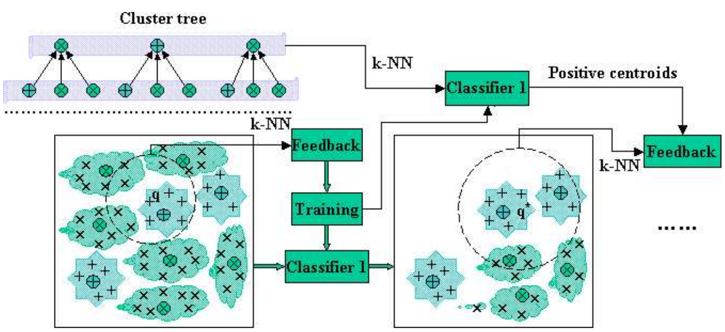

Fig. 3. Multiresolution pattern discovery, where plus/times sign denotes positive/negative points, and circled plus/times signs are centroids.

Figure 3 presents our multiresolution pattern discovery method. We cluster the multimedia database (offline) with the doubling algorithm [10]; and we maintain the clustering result with a hierarchical structure, which we term as *cluster tree*. In the tree, each node at a higher level is a centroid of some cluster at the next lower level. Figure 3 demonstrates that samples from higher levels of the tree present a coarser sketch about the global distribution patterns, while samples from lower levels and the neighborhood of the query manifest finer distribution information. By training the classifier with samples from all the above sources, our method filters negative objects, from the neighborhood of the query to global scope, in a fine to coarse manner. After the filtering operation, positive objects are more readily available in the neighborhood of the query (see Figure 3), so MPD can iteratively increase the size of positive training set by itself. Due to its similarity to multiresolution signal processing techniques (such as Wavelet), we borrow the term *multiresolution* and term our method as *multiresolution pattern discovery*.

4. EXPERIMENT

4.1. Experimental Configuration

We evaluate our approach on an image database with 31,438 Corel images. To search the database, we use 5,172 images from 44 semantic categories (such as tiger, rose and city, etc.) as queries. We employ precision to evaluate retrieval performance, where precision is the number of retrieved positive images over the total number of retrieved images. The average precision of all queries are used as the overall performance. To calculate precision, only images from the same semantic category as the query are counted as positive. For a comparison, we provide the performance of adaptive random forests (ARF) [6] under the same experimental conditions.

In our experiments, each image is represented by a 179-bin feature vector which consists of 5 image features. The first one is a 64-bin color coherence vector. The second one is the 9-bin color moments defined in L*a*b color space. The third one is a 10-bin wavelet-based texture feature [6]. The fourth one is the 64-bin edge coherence histogram [6]. The fifth one is a 32-bin Fourier shape descriptor[6].

4.2. User Interface

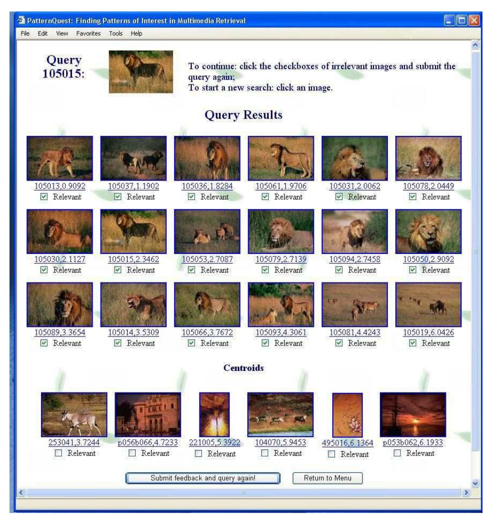

Fig. 4. User interface of our system.

Figure 4 presents the user interface of our *PatternQuest* system. The current query is shown on the left top corner of the screen. The retrieval results are sequentially listed under the query. Under the results, we prompt classified-positive

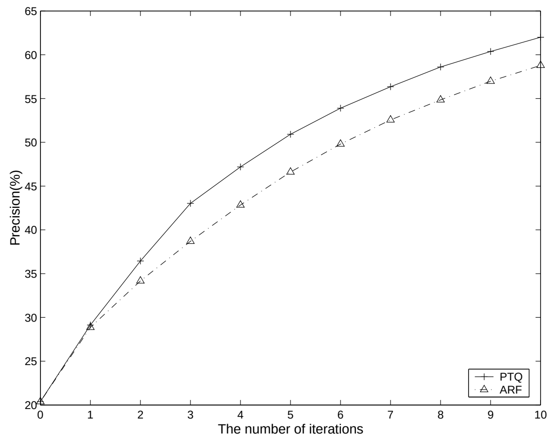

Fig. 5. Precision of our method (PTQ) and ARF.

centroids from the cluster tree (see Figure 3). These centroids are often negative images, which have similar visual features (such as: color, texture and shape) with the query. So, they are appropriate samples for training the filter (cf. Section 3). In our experiments, we generate a cluster tree with two levels, so we only ask the user to annotate several centroids in the first two iterations.

For each returned image, two numbers, which are separated by a comma, are displayed: the first one is the unique number of the image; and the second one is the distance between the image and the query. In the beginning, each image is assumed to be positive, while every centroid is assumed negative. The user then annotates the query results and centroids by enabling or disabling their *relevant* checkboxes. Afterwards, he or she can activate the *Submit* button to submit the relevance feedback. The user can start a new query at any time by clicking on a retrieved image or the *Return to menu* button.

4.3. Performance Evaluation

We run 10 iterations for each query. Figure 5 compares our method (PTQ) against ARF with precision measured over top-20 retrieval results. We can see from this figure that PTQ consistently outperforms ARF on precision. After a few iterations, PTQ improves retrieval precision over ARF by about 3%. By using a multiresolution pattern discovery approach, our PTQ filters negative objects globally from the first iteration, while ARF has to learn the global distribution patterns from multiple feedback iterations.

5. CONCLUSIONS

In this paper, we presented a PatternQuest framework to learn the patterns of interest using classification methods and relevance feedback. Our PatternQuest framework comprises the following three major components: (1) an efficient feature selection method (EFS); (2) a multiresolution pattern discovery approach (MPD); and (3) an online pattern classification method termed adaptive random forests (ARF). With EFS, MPD and ARF, our PatternQuest can iteratively capture the patterns of interest with small training data from the user's feedback. We have carried out extensive experiments on an image database (with 31,438 COREL images) to demonstrate the effectiveness and robustness of our method as compared against one of the state-of-the-art approaches [6].

6. REFERENCES

[1] N. Vasconcelos and A. Lippman, "A probabilistic architecture for contet-based image retrieval," in *IEEE Conference on CVPR*, 2000.

[2] Ying Wu, Qi Tian, and Thomas S. Huang, "Discriminant-em algorithm with application to image retrieval," in *IEEE Conference on CVPR*, 2000.

[3] Z. Su, S. Li, and H. Zhang, "Extraction of feature subspaces for content-based image retrieval using relevance feedback," in *Proc. of ACM Multimedia*, 2001.

[4] Yimin Wu and Aidong Zhang, "A feature reweighting approach for relevance feedback in image retrieval," in *Proc. IEEE Int'l Conf. on Image Processing (ICIP'02)*, 2002.

[5] Yimin Wu and Aidong Zhang, "Interactive pattern analysis for searching multimedia databases," in *Proc. 8th Int'l Workshop on Multimedia Information Systems (MIS'02)*, Oct.31-Nov.2 2002.

[6] Yimin Wu and Aidong Zhang, "Adaptive pattern discovery for interactive multimedia retrieval," in *IEEE Conf. on CVPR*, June 2003.

[7] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, San Deigo, California: Academic Press, Inc., 1990.

[8] Yimin Wu and Aidong Zhang, "Feature selection for classifying high-dimensional numerical data," *to appear in IEEE Conf. on CVPR*, June 2004.

[9] Yimin Wu and Aidong Zhang, "Interactive pattern analysis for relevance feedback in multime information retrieval," *ACM Multimedia Systems Journal (to apper)*, 2004.

[10] M. Charikar, C. Chekuri, T. Feder, and R. Motwani, "Incremental clustering and dynamic information retrieval," in *Proc. of ACM Symposium on Theory of Computing*, 1997.