Textual Query Based Image Retrieval

Patil Priyanka N
PG Student At Shriram Institute ,Paniv India.
Email id -priyankapatiln49@gmail.com

Prof. Yevale Ramesh S
Shriram Institute,Paniv India
Email id:ryevale@gmail.com

Prof. Dhainje Prakash B
Vice Principal At Shriram Institute,Paniv India

.
Dr. Deshmukh Pradeep K
Principal at Shriram Institute, Paniv India.
Email id:principalsietc@gmail.com

Abstract- As digital cameras becoming popular and mobile phones are increased very fast so that consumers photos are increased. So that retrieving the appropriate image depending on content or text based image retrieval techniques has become very vast. Content-based image retrieval, a technique which uses visual contents to search images from large scale image databases according to users interests, has been an active and fast advancing research area semantic gap between the low-level visual features and the high-level semantic concepts. Real-time textual query-based personal photo retrieval system by leveraging millions of Web images and their associated rich textual descriptions. Then user provides a textual query. Our system generates the inverted file to automatically find the positive Web images that are related to the textual query as well as the negative Web images that are irrelevant to the textual query. For that purpose we use k-Nearest Neighbor (kNN), Decision stumps, and linear SVM, to rank personal photos. For improvement of the photo retrieval performance, we have used two relevance feedback methods via cross-domain learning, which effectively utilize both the Web images and personal images.

Keywords-*Image Retrival,Consumer photos.cross domain,textual query.*

*****

I. Introduction

Advances in data storage and image acquisition technologies have evolved the creation of large image datasets. In order to deal with these data, it is necessary to develop appropriate information systems to efficiently manage these collections.The current image retrieval systems are successful in retrieving images, using keyword based approaches. However, they are incapable to retrieve the images which are context sensitive and annotated inappropriately. Content-Based Image Retrieval (CBIR) aims at developing techniques that support effective searching and browsing of large image repositories, based on automatically derived image features. In these systems, image processing algorithms are used to extract feature vectors that represent image properties in form such as color, texture, and shape. The tasks performed by CBIR works in two stages as Pre-processing and Feature extraction stages. In Pre-processing stage removal of noise and enhancement of some object features takes place which are relevant to understanding the image is performed. Image segmentation is also performed to separate objects from the image background. In Feature Extraction stage, features such as shape, colour, texture etc. are used to describe the content of the image. This feature is generated to accurately represent the image in the database. The colour aspect can be achieved by the techniques like moments, histograms etc. The texture aspect can be achieved by using transforms or vector quantization. In this approach, it is possible to retrieve images similar to one chosen by the user.The current CBIR systems suffer from the semantic gap. Text-based image retrieval uses traditional database techniques to manage images. Through text descriptions, images can be organized by topical or semantic hierarchies to facilitate easy browsing based on standard Boolean queries. However, since automatically generating descriptive texts for a wide spectrum of images is not feasible, most text-based image retrieval systems require manual annotation of images. Annotating images [3] manually is a hectic task and expensive for large image databases, and is often subjective, context-sensitive and incomplete. Searching for interesting and useful images from among the enormous number of images available generally relies either on content-based image retrieval using visual image features or text based search using text queries to search for images based on textual annotations of the images. Many techniques have been developed for text-based information retrieval and they proved to be highly successful for indexing and querying web sites. In this work, we propose a real-time textual query-based retrieval system which directly retrieves the desirable personal photos without undergoing any intermediate image annotation process. A preliminary version of this work appeared in. There are two challenges. The first one is how to use the textual queries to directly retrieve unlabeled consumer photos that are not associated with any semantic textual metadata. We propose using the massive and valuable social media data as the training data for retrieving the raw consumer photos using textual queries. Our work is motivated by the advances in Web 2.0 and the recent advances of Web data-based image annotation techniques. Everyday, rich and massive social media data are posted to the Web, and Web images generally accompanied by rich contextual information such as tags, categories, titles, and comments. A typical CBIR solution as shown in figure 1requires the construction of an image descriptor, which is characterized by firstly an extraction algorithm to encode image features into feature vectors and then similarity measure to compare two images. The similarity measure is a matching function, which gives the degree of similarity for a given pair of images as represented by their feature vectors. The CBIR has many applications areas as architectural design, education,

 ISSN: 2321-8169
1001 - 1004

commerce, military, medical diagnosis, mining, biomedical and web image classification. All most of The CBIR research is finding technique to measure the performance for retrieving more similar image from the image databases of retrieval. Based on this there are following algorithms including k-Nearest Neighbor (kNN), decision stump ensemble, and linear SVM, to rank the photos in the personal collections. Observing that the total number of negative Web images is much larger than the total number of positive Web images, we randomly sample a fixed number of negative samples and combine these samples with the positive samples for training decision stump ensemble and SVM classifiers. Similarly to, the whole procedure is repeated multiple times by using different randomly sampled negative Web images and the average output from multiple rounds is finally used for robust consumer photo retrieval. The second research challenge in this work is the semantic gap between the low-level visual features and the high-level semantic concepts. To bridge the semantic gap, relevance feedback has been frequently used to help acquire the search intention from the user and further improve the retrieval performance.

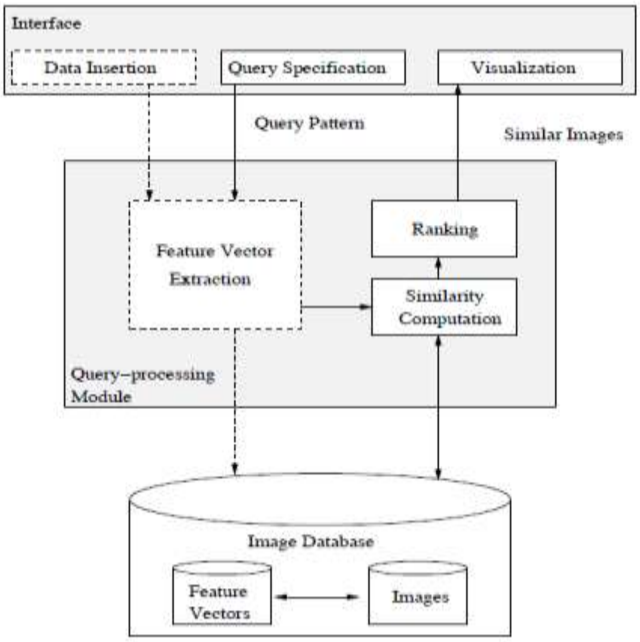

Fig 1: Typical architecture of a content-based image retrieval system.

II. RELATEDE WORK

A. Proposed Framework

As fig.2. shows the proposed architecture of image retrieval. The first module is automatic Web image retrieval, in which first interprets the semantic concept of textual queries by a user. Based on the semantic concept and WordNet, the sets of relevant and irrelevant Web images are retrieved from the Web image database using the inverted file method. The second module then uses these relevant and irrelevant Web images as a labeled training set to train classifiers and these classifiers are used to retrieve potentially relevant consumer photos from personal collections. Our scheme is a general textual query-based image retrieval framework, and therefore, any classifiers can be used. To further improve the retrieval performance, relevance feedback and cross-domain learning techniques are employed in the last module to refine the image retrieval results.

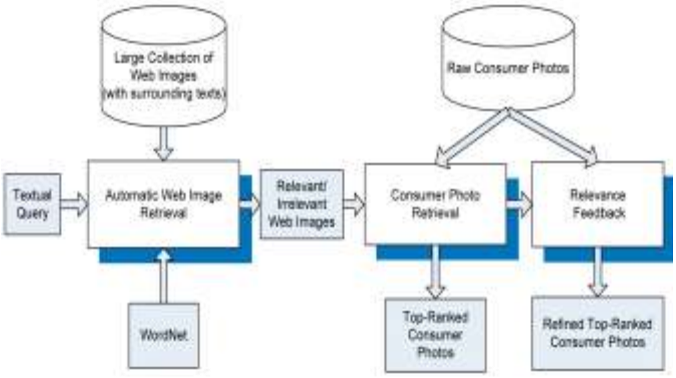

FIG. 2. TEXTUAL QUERY-BASED CONSUMER PHOTO RETRIEVAL SYSTEM

A. k-Nearest neighbor

Image classification approach, derived from the kNN classification strategy, which is particularly suited to be used when classifying images described by local features. Our proposal based on the possibility of performing similarity search between image local features. With the use of local features generated over interest points, we revised the single label kNN classification approach to consider similarity between local features of the images in the training set rather than similarity between images, opening up new opportunities to investigate more efficient and effective strategies. We will have our focus on that classifying at the level of local features we can exploit global information contained in the training set, which cannot be used when classifying only at the level of entire images, as for instance the effect of local feature cleaning strategies.The *k*-NN algorithm is a similarity-based learning algorithm that has been shown to be very effective for a variety of problem domains including text categorization. From a given test document the *k*-NN algorithm finds the k nearest neighbors among the training documents, and uses the categories of the k neighbors to weight the category members. The similarity score of each neighbor document to the test document is used as the weight of the categories of the neighbor document. If several of the k nearest neighbors belongs to same or shares category, then the per-neighbor weights of that category are added together, and the resulting weighted sum is used as the score of candidate categories. In image retrieval algorithms, retrieval is according to feature similarities with respect to the query, ignoring the similarities among images in database. To use the feature similarities information, this paper presents an application of k-means clustering algorithm for retrieving image. Combining the low-level visual features and high-level concepts, the proposed approach fully explores the similarities among images in database, using such clustering algorithm and optimizes the relevance results from traditional image retrieval system. It firstly clusters the similar images in the images database to improve the efficiency of images retrieval system. The results of experiments on the testing images show that the proposed approach can greatly improve the efficiency and performances of image retrieval, as well as the convergence to user's retrieval concept. We compute the average distance between photos and depending on average distance between images ranking is to be done k is the most important parameter in a text categorization system based on k-Nearest Neighbor algorithm (kNN).In the classification process, k nearest documents to the test one in the training set are determined firstly. Then, the predication can be made according to the category distribution among these k nearest neighbors. Generally the class distribution in the training set is uneven. Some classes may have more samples than others. Therefore, the system performance is very sensitive to the choice of the parameter k. And it is very likely that a fixed k value will result in a bias on large categories. To deal with these problems, we propose an improved kNN algorithm, which uses different numbers of nearest neighbors for different categories, than a fixed number across all categories. More samples will be used for deciding whether a test document should be classified to a category, which has more samples in the training set. Former experiments on Chinese text categorization show that our method is less sensitive to the parameter k than the traditional one, and it can properly classify documents belonging to smaller classes with a large k. The method is promising for some cases, where estimating the parameter k via cross-validation is not allowed. While using kNN algorithm, after k nearest neighbors are found, several strategies could be taken to predict the category of a test document based on them. But a fixed k value is usually used for all classes in these methods, regardless of their different distributions. Equation (1) and (2) below are two of the widely used strategies of this kind method.

$$y(dt)= \arg\max_k \sum_{xj \in kNN} y(xj,ck) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

$$y(dt)= \arg\max_k \sum_{xj \in kNN} sim(di,xj)\, y(xj,ck) \quad \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

where di is a test document, is one of the neighbors in the training set y(xj,ck), indicates whether xj belongs to class ck , and sim(di,xj) is the similarity function for di and . Equation (1) means that the predication will be the class that has the largest number of members in the k nearest neighbors whereas equation (2) means the class with maximal sum of similarity will be the winner. The latter is considered as better than the former and used more widely.

C. Decision Stump

Decision Stump is a one level decision tree. It is a weak learner as it is based on simple binary decisions. Thus the Decision Stump is normally integrated with boosting and bagging methods.As decision stump is basically a one level decision tree where the split at the root is based on a specific attribute and value pair. Decision trees represent a supervised approach of classification. A decision tree is a simple structure where non-terminal nodes represent tests on one or more attributes and terminal nodes represents decision outcomes. The ordinary tree consists of one root, branches, nodes and leaves. In the same way the decision tree consists of nodes which stand for circles, the branches stand for segments connecting the nodes. A decision tree is usually drawn from left to right or beginning from the root downwards, so it is easier to draw. The first node is a root. The end of the chain root - branch - node-...- node is called “leaf”. From each internal node may grow out two or more branches. Each node corresponds to a certain characteristic and the branches correspond with a range of values. These ranges of values must give a partition of the set of values of the given characteristic. After sampling, a decision $f_d(x)= h(s_d(x_d-ø_d))$ stump is learned by finding the sign $s_d \in \{\pm1\}$ and the threshold $ø_d \in R$ of the dth feature x_d of the input x such that the threshold $ø_d$ separates both classes with a minimum training error rate d on the smaller training set. For discrete output, h(x)is the sign function, that is, h(x)=1 if x>0, otherwise. For continuous output, h(x) can be defined as the symmetric sigmoid activation function,

$$\text{i.e., } h(x)=\frac{1-\exp(-x)}{1-\exp(+x)}$$

We observe that it is difficult to rank the consumer photos by using discrete output because the responses of many consumer photos are the same in this case. In this work, we therefore use the continuous output of h(x). The threshold $ø_d$ can be determined by sorting all samples according to the feature x_d and scanning the sorted feature values. In this way, the decision stump can be found efficiently. Next, the weighted ensembles of these decision stumps are computed for prediction, i.e.

$$f_s(x)=\sum \Upsilon_d\, h()s_d(x_d- ø_d)$$

D. Linear SVM

Decision stump enables classifier can effectively exploit both relevant and irrelevant Web photos. It is inefficient to use this classifier on a large consumer photo data set because all of the decision stumps need to be applied on every test photo in the testing stage. Support Vector Machines (SVM) have recently gained prominence in the field of machine learning and pattern classification. Linear SVM is the emerging technology which is very fast machine learning [4] algorithm for solving multiclass classification problems from large data sets. The SVM in particular defines the criteria for a decision surface i.e at maximum and far away from data point. Classification is achieved by realizing a linear or non-linear separation surface in the input space. A fast iterative algorithm for identifying the Support Vectors of a given set of points. Our algorithm works by maintaining a candidate Support Vector set. It uses a greedy approach to choose points for inclusion in the candidate set. When the addition of a point to the candidate set is blocked because of other points already present in the set we use a backtracking approach to prune away such points. To speed up convergence we initialize our algorithm with the nearest pair of points from opposite classes. We then use an optimization based approach to increment the candidate Support Vector set.

E. Relevance Feedback via Cross-Domain Learning

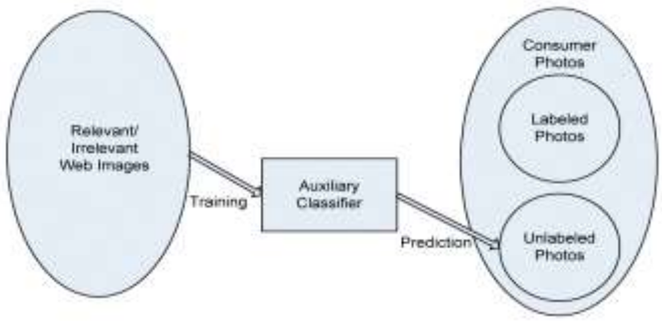

Fig 3: Relevance Feedback via Cross-Domain Learning

With Relevance Feedback method, we can obtain a limited number of relevant and irrelevant consumer photos from the user to refine the image retrieval results. In fig 3 results is seen when 3classifiers are applied on dataset then each algorithm gives the above result. That means the linear SVM retrieve’s most relevant images than other two algorithms.

ISSN: 2321-8169
1001 - 1004

III.CONCLUSION

By leveraging a large collection of Web data , we have proposed a real-time textual query-based personal photo retrieval system which can retrieve consumer photos without using any intermediate image annotation process. For a given textual query, our system can automatically and efficiently retrieve relevant and irrelevant Web images using the inverted file method and WordNet.

IV.ACKNOLEDGEMENT

This paper is based on Textual Query of Personal Photos Facilitated by Large-Scale Web Data IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 33, NO. 5, MAY 2011

V.REFERENCES

[1] M. Artae, M. Jogan, and A. Leonardis, “Incremental PCA for On-Line Visual Learning and Recognition,” Proc. Int’l Conf. PatternRecognition, 2002.

[2] J. Blitzer, M. Dredze, and F. Pereira, “Biographies, Bollywood,Boom-Boxes and Blenders: Domain Adaptation for Sentiment Classification,” Proc. Ann. Meeting Assoc. for Computational Linguistics, 2007.

[3] L. Cao, J. Luo, and T.S. Huang, “Annotating Photo Collections byLabel Propagation According to Multiple Similarity Cues,” Proc. ACM Conf. Multimedia, 2008.

[4] G. Cauwenberghs and T. Poggio, “Incremental and Decremental Support Vector Machine Learning,” Neural Information Processing Systems, MIT Press, 2000.

[5] C.-C. Chang and C.-J. Lin, “LIBSVM: A Library for Support Vector Machines,” http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001.

[6] S.-F. Chang, D. Ellis, W. Jiang, K. Lee, A. Yanagawa, A.C. Loui, and J. Luo, “Large-Scale Multimodal Semantic Concept Detection for Consumer Video,” Proc. ACM SIGMM Workshop Multimedia Information Retrieval, 2007.

[7] S.-F. Chang, J. He, Y. Jiang, A. Yanagawa, and E. Zavesky, “Columbia University/VIREO-CityU/IRIT TRECVID2008 High- Level Feature Extraction and Interactive Video Search,” Proc. NIST TRECVID Workshop, 2008.

[8] L. Chen, D. Xu, I.W. Tsang, and J. Luo, “Tag-Based Web Photo Retrieval Improved by Batch Mode Re-Tagging,” Proc. IEEE Conf.