STRATEGY OF COMBINING RANDOM SUBSPACE AND DIVERSIFIED ACTIVE LEARNING IN CBIR

Fang Wang   Zhenfeng Zhu   Yao Zhao

Institute of Information Science, Beijing Jiaotong University, Beijing 100044, P. R. China

ABSTRACT

Generally speaking, several aspects related to relevance feedback based CBIR include what means should be adopted for approximate semantic description of image content, what strategies be applied to sample labeling in feedback and what relevance model would be built for online discrimination. Using random sampling strategy, we construct a set of random subspaces for learning multiple intrinsic descriptions of image content, with each of which stable component classifier can be trained. To enhance the generalization capability of relevance model, the diversified active learning is carried out by collecting more informative samples, i.e. those samples spreading around decision boundary dispersedly. The final favorable performance also contributes to the application of ensemble scheme on individual component classifier.

***Index Terms*—**Content Based Image Retrieval, Random Subspace, Active Learning

1. INTRODUCTION

With the rapid growth of digital images, especially the emergence of large-scale image collections, the problem of how to manage and index the ever-growing volume of images efficiently has attracted many academic researchers' attentions as well as some commercial companies. To overcome the difficulty of manual annotation，an alternative scheme, content-based image retrieval (CBIR) was proposed in the early 1990's, which makes use of low level visual features instead of keywords to represent images. A thorough overview of current techniques with respect to CBIR was presented in [1].

However, after years of research, the retrieval accuracy is still far from users' expectations. It's mainly because of the well known 'semantic gap' between low-level visual features and high-level semantic concepts. In order to narrow down the gap, much research work has been developed. Relevance feedback (RF), which originally developed for information retrieval, has been shown to provide dramatic performance improvement [2, 3]. With the interaction from user during relevance feedback, information of which images he or she thinks are relevant to the query is interactively feed backed to the system.

Concentrating on the main issues involved with description of image content, strategy of online sample labeling, and online relevance model learning in relevance feedback based CBIR, we investigate the strategy of combining random subspace and diversified active learning in this paper. Unlike the random subspace approaches in [4,5], the intrinsic multiple description of image content are learnt offline in light of a set of random sample subspaces, to which the graph-based unsupervised learning technique[6] is applied. In relevance feedback, the online relevance model with ensemble of multiple component classifiers trained by virtue of each kind of intrinsic description is built to strengthen the decision capability. Instead of online sample labeling based on conventional active learning [7] for triggering out next round of relevance feedback, we propose a diversified strategy to label those samples spreading around the decision boundary dispersedly, which will bear more information for boosting the online learning of relevance model.

2. FLOWCHART OF THE PROPOSED STRATEGY

The flowchart of the proposed scheme for CBIR is shown in Fig.1, which is composed of two stages: offline semantic mining for multiple intrinsic descriptions of image content and online feedback based relevance model learning.
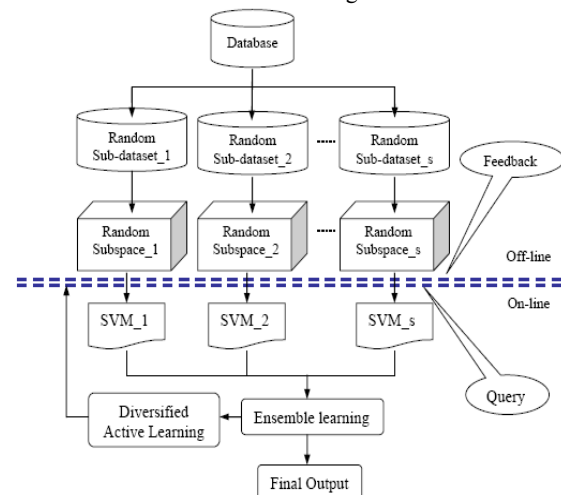

Figure 1 the flowchart of the proposed framework

In the off-line stage, the task of semantic mining for multiple intrinsic descriptions of image content is implemented based on Orthogonal Neighborhood Preserving Projection [6] (*ONPP*) by using random sub-sampling scheme.

For each kind of intrinsic description, a corresponding component *SVM* classifier is trained with diversified active learning strategy being involved in the on-line learning during feedback, and what is more, ensemble technique is applied to individual *SVM*s to enhance the discrimination of relevance model. In addition, for the initial round of feedback, sample labeling based on random grouping, instead of traditional *k-NN*, is considered to grasp user's multiple query intentions as nicely as possible.

3. ENSEMBLE LEARNING BASED ON RANDOM SUBSPACE

3.1. Orthogonal Neighborhood Preserving Projections

The problem of dimensionality reduction appears in many fields including data mining, machine learning and computer vision. Its goal is to map the high dimensional samples to a lower dimensional subspace such that certain informative properties are preserved. As for CBIR, the reduced dimensionality subspace can be approximately taken as intrinsic semantic description of image content [8]. In this paper, we use a kind of projection-based linear dimensionality reduction technique ONPP，also known as one of graph-based unsupervised learning methods，to learn multiple approximate intrinsic descriptions of the image content on a set of random sample subspace. Base on the component classifier trained with each kind of intrinsic description，ensemble strategy is imposed so that the well known gap between the low level visual feature and high-level semantic concept can be narrowed down to some extent.

Actually, the main idea of *ONPP* is to seek an orthogonal mapping of a given database so as to best preserve a graph which describes the local and global geometry. As the linear mapping V between the input and the reduced spaces is explicit, it can map a query image to a lower dimensional space directly no matter whether or not the query image is in the database. The mapping is given as

$$Y = V^T \times X \tag{1}$$

where $V \in R^{m \times d}$, X is a dataset of m dimension and Y is an accurate representation of X, but of smaller dimension d .

A database consisting of n samples can be represented by a matrix $X = [x_1, x_2, \ldots, x_n] \in R^{m \times n}$, where, m is the dimension of each sample. *ONPP* works by first building an "affinity" graph for a certain sample. The basic assumption is that each data sample along with its k nearest neighbors approximately lies on a locally linear manifold. Hence, each data sample x_i is reconstructed by a linear combination of its k nearest neighbors. The reconstruction errors are measured by minimizing the objective function

$$\varepsilon(W) = \sum_i \| x_i - \sum_j \omega_{ij} x_j \|_2^2 \tag{2}$$

The weight ω_{ij} represents the linear coefficient for reconstructing the sample x_i from its neighbors $\{ x_j \}$.

It assumes that the same weights, which reconstruct the sample x_i by its neighbors in the high dimensional space, will also reconstruct its lower dimensional representation $y_i \in R^d$ by its corresponding neighbors in the low dimensional space. In order to obtain $Y = [y_1, y_2, \ldots, y_n] \in R^{d \times n}$, *ONPP* employs the objective function

$$\phi(Y) = \sum_i \| y_i - \sum_j \omega_{ij} y_j \|_2^2 \tag{3}$$

In this case the weights W are fixed and we need to minimize the above objective function. Note that $\phi(Y)$ can be rewritten by Fun. 4 and *ONPP* imposes the additional constraint that the columns of V are orthogonal, i.e. $V^T V = I$, so V can be determined by computing the $d+1$ eigenvectors of $M = X(I-W^T)(I-W)X^T$ associated with smallest eigenvalues.

$$\phi(Y) = \| Y - YW^T \|_F^2 = \| Y(I-W^T) \|_F^2$$
$$= tr[Y(I-W^T)(I-W)Y^T] \tag{4}$$
$$= tr[V^T X(I-W^T)(I-W)X^T V]$$

3.2. *ONPP* Based Online Ensemble of Component Classifier

The purpose of ensemble learning is to build a learning model which integrates a number of diverse base learning models, so that the integrated model can exhibit better generalization performance on application to a particular data set than any of the individual base models. Ensemble learning techniques have demonstrated powerful capacities in improving upon the classification or retrieval accuracy of a base learning algorithm. It has been pointed that an ensemble of classifiers is more accurate than any of its members if the classifiers in the ensemble are both accurate and diverse [9].Varying the feature subsets used by each member can help to promote the diversity [10].

Let $X_i, i = 1, \ldots s$ be a set of random sample subspaces from X . With applying *ONPP* on each X_i, the intrinsic description of image content can be obtained. Later, in the on-line stage, stable component $SVM(i)$ classifier is trained in view of the ith class of intrinsic description. Consequently, to enhance the description power of online relevance model, ensemble technique is exploited. During each round of feedback，we simply combine each of component *SVM* classifier with the coefficients $\omega_i = 1/s$ as in Fun. 5, for which the major reason is out of considerations of computational efficiency and the sample subspaces being randomly sampled from the dataset. In fact, the popular Adaboost algorithm, which iteratively evaluates ω_i for corresponding $SVM(i)$, can be favorably introduced into our case leaving out of complexity account.

$$Output = \sum_{i=1}^{s} \omega(i) * SVM(i) \tag{5}$$

4. STRATEGY OF DIVERSIFIED ACTIVE LEARNING

Active learning is an effective training data sampling method studied in machine learning research [11].Generally speaking, an active learning method selects a small number of samples from the unlabeled sample database according to a certain criterion and then requests users to label them. Thereafter, the classifier is retrained with these newly training samples.

However, in practice, most of active learning methods empirically apply the "*K* Nearest Neighbors (*k*-NN)" criterion or "Closest-to-Boundary" to choose samples for labeling. The major limitation of these algorithms is that the samples selected may not be the most informative ones. For example, if two samples that are close to each other are selected as training samples, they probably belong to the same semantic concept, and then provide the classifier with similar information. Clearly, the performance of classifier can't get improved under such circumstances.

According to [12], the classifier with larger margin has a better performance in terms of the generalization error. Therefore, in this case, we maximize the margin of the SVM classifier in each round by selecting the most informative training samples which distribute

dispersedly across the whole database. In particular, the random grouping strategy is taken to carry out the diversified active learning.

4.1. Initial feedback based on random grouping

As previously mentioned, for the initial feedback in CBIR, k-NN criterion gives the query image the nearest neighbors based on some distance measures，i.e. user can label both the positive and negative examples among the top k nearest neighbors to the query instance. But one of drawbacks for such initial feedback is that the user's multiple query intentions can't be grasped promptly, which will make the following online learning procedure lack of generalization capability for further feedback.

To avoid confusion of illustration with Sec.3, let D denotes the image database as X, on which the retrieval action will be conducted. Firstly, we randomly divide D into p sub-dataset, $D_i, i = 1，\ldots p$. Hence, for each sub-database D_i, the top k nearest neighbors $\tilde{D}_i$ to the query instance are returned and furthermore $p \times k$ samples can be collected to compose the candidate pool D_c for being labeled by users.
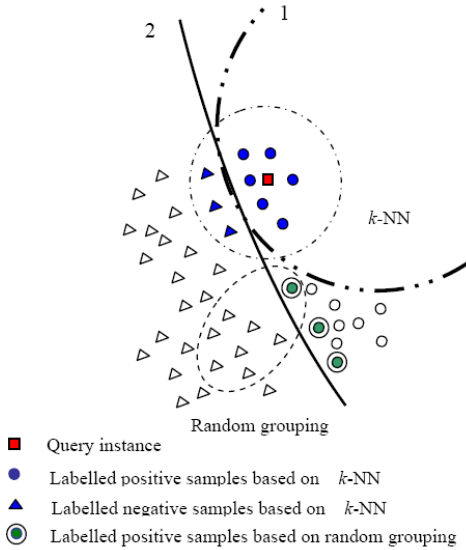

Figure 2 Random grouping for diversified sample labeling

For a large scale of database of miscellaneous images, generally, only a small part of it can take the same semantic concept as the user's retrieval intention. Hence, the strategy of random grouping can make those positive images that are far away from the query in distance space with similar semantic concept win more chance to be labeled, i.e. the users' multiple query intentions can be grasped as nicely as possible, which make the positive training samples more diversified and enhance the generalization ability of online learning for next round of feedback.

The advantage of applying random grouping in the initial feedback over k-NN is shown in Fig 2. The labeled training samples picked out by k-NN method lie in a compact cluster, and hence they are less informative for the later online learning. The broken line 1 simulates the hyper plane of a binary *SVM* classifier, for which training samples are selected by k-NN strategy. We can see that there are many mistakenly classified samples because the training samples aren't dispersive enough. Contrastively, real line

2 shows stronger generalization ability of *SVM, w*hich can thanks to the random grouping scheme forcing the labeled samples to spread more globally.

4.2. Diversified SVM active learning

In traditional SVM based relevance feedback, those samples with higher output score are provided to user for labeling from the second round of feedback. But as pointed in [7], such choice for labeling is only from the view of the accuracy of classification and may not be helpful to improve the performance of the system.

Whereas, the SVM active learning method (SVM$_{active}$) tries to focus the user's intentions on images whose classification is difficult. It asks user to label samples closest to the SVM plane as training samples. In fact, SVM$_{active}$ does increase the classification accuracy, but the system performance can't be improved rapidly in a few round of feedback, i.e. the query intent of the user can't be grasped as soon as possible. The main reason for it is that the new incorporated training data from last round of feedback are lack of exploiting ability since the intervention of them won't change the hyper plane greatly between two consecutive rounds of feedback.

In order to boost the re-training course of SVM, in this paper, the similar random grouping strategy for user labeling as in Section 4.1 is proposed. In the j iteration of feedback ($j \geq 2$), let $\left\{D_i^j\right\}_{i=1\ldots p}$ be a random partition of image dataset D. And for each D_i^j, it is easy to find a sub-dataset $\tilde{D}_i^j$, which is composed of the top k closest images in D_i^j to the decision boundary via the ensemble of multiple component SVM classifiers as given by Fun.5 Thus a new candidate pool $D_c^j = \left\{\tilde{D}_i^j\right\}_{i=1\ldots p}$ for labeling by user can be constructed correspondingly. With the application of the random grouping strategy to the building of candidate pool for labeling, the new labeled samples will be distributed more sparsely and take stronger exploiting ability for new round of online learning.

5. EXPERIMENTAL RESULTS

We have evaluated the performance of our method using a general purpose image database consisting of randomly selected 6000 Corel images. The images are categorized into 60 groups, and each of the categories contains 100 images of essentially the same content. For low level visual image representation, 72-d color histograms in HSV space were extracted. We randomly selected 10 images in each category as queries. The retrieval precisions are averaged over the 600 queries, and the precision versus scope curve is used to evaluate the performance of various methods.

The kernel function in *SVM*s is the RBF kernel. Since the determination of related parameters with SVM is not our main focus in this work, we experimentally set the kernel radius $\sigma_s = 0.7$ for all cases. In addition, the intrinsic dimension is reduced to 20 and 5 random subspaces are constructed with 3000 randomly selected samples for each. Meanwhile, we divide the whole database into $p = 40$ random groups and in each group only the top 3 nearest neighbors of the query are returned to construct the candidate pool to be labeled by user. Thus the total number of candidates returned to users for being labeled is 120, from which 5 positive (generally less than 5) and 5 negative samples are simulated to be finally labeled by user in each round of feedback. If every page consists of 40 images, there are only 3

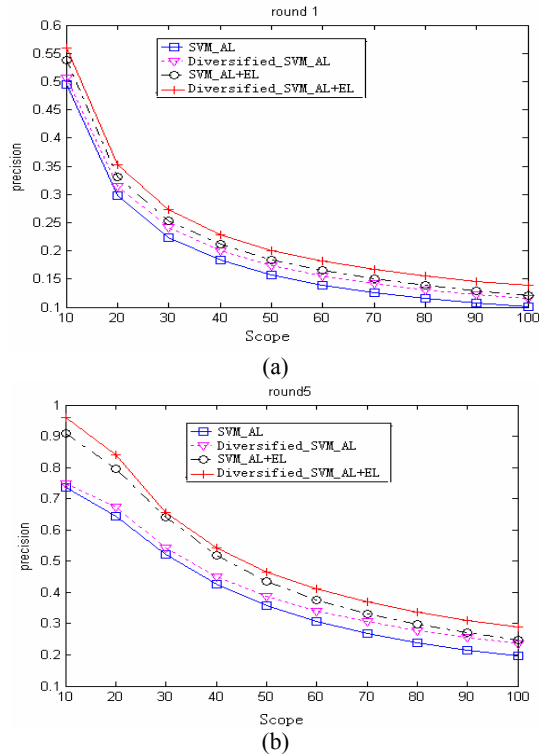pages for users to browse, which is consistent with the real application.

(a)

(b)

Figure 3 (a) Average Accuracy –Scope curve in the first round of relevance feedback (b) Average Accuracy –Scope curve after 5 iterations of relevance feedback
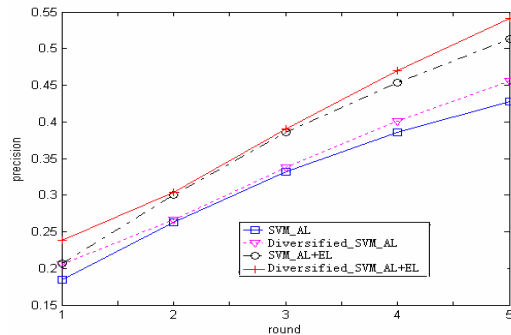

Figure 4 Average Accuracy-Number of iterations of relevance feedback with scope of 40

As shown in Figure 3 and Figure 4, the proposed diversified SVM active learning (Diversified_SVM_AL) and *ONPP* based random subspace ensemble learning (SVM_AL+EL) take on better retrieval performance than traditional *SVM* active learning (SVM_AL) [15] respectively in each round of feedback. Meanwhile, combining the two strategies achieves significant improvement over traditional (SVM_AL). Taking round 5 in Figure 4 as an example, the average precision of our approach (Diversified_SVM_AL+EL) within 40 (p(40)) is higher than SVM_AL by 12% . In addition, it is also obvious from fig.3 (a) that the direct application of *k*-NN in the initial feedback for SVM_AL shows inferior to the proposed random grouping based Diversified_SVM_AL, and the same situation happens to SVM_AL+EL compared with the Diversified_SVM_AL+EL.

6. CONCLUSIONS

In this paper, we investigate a novel method that combines random subspace based ensemble learning and diversified active learning for CBIR. Aiming at semantic mining for multiple intrinsic descriptions of image content, the random sample subspace technique is exploited. In feedback, we propose the strategy of diversified active learning to force the labeled images to spread more desperately, which will be helpful for the online learned component SVM classifier to take on stronger generalization capability. Meanwhile, the ensemble scheme is also considered to boost the discrimination power of relevance model. In addition, it shouldn't be underestimated that some factors, such as the intrinsic detention d , the number S of random sample subspaces and the number p of random partition on the dataset in diversified active learning, will clearly impact the performance of the proposed strategy, of which the determination needs to be with more attention in the future work.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] Ritendra Datta, Dhiraj Joshi, Jia Li and James Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," *ACM Computing Surveys*, vol. 40, no. 2, article 5, 2008

[2] Y. Rui et al, "Relevance feedback: a power tool for interactive content-based image retrieval," *IEEE Trans. on CSVT*, Vol.8,No.5, pp.644-655,1998.

[3] P.Hong, Q.Tian, and T.S.Huang, "Incorporate support vector machines to content-based image retrieval with relevant feedback," *Proc. of ICIP,* Vancouver, BC, Canada, 2000.

[4] D. C. Tao, X.O. Tang, X.L. Li, and X.D. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," IEEE Trans. on PAMI，Vol 28，No.7, pp.1088-1099

[5] W. Jiang, M.J Li, H.J. Zhang, J. Zhou, "Relevance feedback using random subspace method," *Procs of Int. Symposium on Circuits and Systems,* 2004

[6] E. Kokiopoulou, Y. Saad, "Orthogonal neighborhood preserving projections," *Proc. of Fifth IEEE International Conference on Data Mining* , 2005.

[7]S. Tong, E. Chang,"Support vector machine active learning for image retrieval ," *Proc. of the 9th ACM Int. Conf. On Multimedia,* pp.107-118, 2001.

[8] X.F. He, O. King, W. Y. Ma，"Learning a semantic space from user's relevance feedback for image retrieval," *IEEE Trans. on CSVT*, Vol.13, No.1, pp. 39- 48, 2003.

[9] L.K. Hansen, P. Salamon, "Neural network ensembles," *IEEE Trans. on PAMI*, Vol.12,No.10, pp. 993-1001,1990.

[10] D. W. Opitz, "Feature selection for ensembles," *Proc. of the 16th National Conf. on Artificial Intelligence*, pp. 379-384, 1999.

[11] D. D. Lewis and J. Catlett, "Heterogeneous uncertainty sampling for supervised learning," *Proc. of 11th ICML*, pp. 148-156, 1994.

[12] C. Campbell, N. Cristianini, A. Smola, "Query learning with large margin classifiers," *Proc. of 17th ICML*, pp. 111-118*, 2000*.