CLUSTERING-BASED SUBSPACE SVM ENSEMBLE FOR RELEVANCE FEEDBACK LEARNING

Rongrong Ji, Hongxun Yao, Jicheng Wang, Pengfei Xu,[1]Xianming Liu
Visual Intelligence Laboratory, Harbin Institute of Technology, Harbin, P. R. China, 150001

ABSTRACT

This paper presents a subspace SVM ensemble algorithm for adaptive relevance feedback (RF) learning. Our method deals with the case that user's relevance feedback examples are usually insufficient and overlapped together in feature space, which decreases the learning effectiveness of RF classifiers. To enhance classification efficiency in such case, multiple SVMs are learned by clustering-based training set partition, each of which fits its cluster-specific sample distribution and gives labeling regressions to test samples that fall within this cluster. To adapt features to sample distribution within each cluster, AdaBoost feature selection is conducted onto pyramid Haar of H&I bands in HSI space. In AdaBoost, we evaluate the feature discriminative ability by an entropy-based uncertainty criterion, based on which an *Eigen* feature subspace is constructed in cluster-specific SVM training. Finally, regression results of multiple SVMs are probabilistic assembled to give the final labeling prediction for test image. We compare our cluster-based cascade SVMs (CSS) RF method in COREL 5,000 database with: 1. Single SVM; 2. Active Learning SVM [5]; 3. Bootstrap Sampling SVM [7]. The superior experimental results demonstrate the efficiency of our algorithm.

Index Terms— image retrieval, relevance feedback, data clustering, classifier sampling, SVM, classifier ensemble

1. INTRODUCTION

Content-based image retrieval (CBIR) aims at retrieving images based on their visual contents. It is a long-standing research hot spot over the past decade [1]. However, performances of state-of-the-art CBIR systems are still unsatisfactory due to the semantic gap between image visual contents and human's image perception.

To bridge the semantic gap, relevance feedback (RF) [2] was introduced into CBIR as an online learning strategy to reveal user retrieval intention. Many methods are adopted in relevance feedback learning, such as query vector movement (QVM) and query expansion (QE), both of which are derived from traditional text retrieval. Recently, there is an increasing research interest [1,2,6,7,8,9,10] in using machine learning methods for RF learning, in which the manually-labeled positive & negative images are treated as training examples to learn a probabilistic binary classifier, based on which the prediction probabilities of image labeling are adopted to refine similarity ranking.

The demand of human labeling in relevance feedback restricts the acquirement of sufficient training examples for classifier learning. Generally speaking, there are two crucial issues to design an effective RF classifier:

1. *Training Sample Distribution*: Since the retrieval results are visually similar to the query example, their distributions usually overlapped with each other in feature space. In this case, single classifier usually faces the difficulty to effectively partition such overlapping data using only insufficient and asymmetric labels.

2. *Visual Feature Selection*: RF algorithms should choose good features for classifier learning. Such features should congregate perceptually similar images while partition perceptually diverse ones. To adapt such demand in RF, learning-based feature selections methods [8] is required.

To deal with the first issue, adaptive sampling strategies are proposed in literature [5-7]. Two representative strategies includes: active learning sampling [5-6] and Bootstrap sampling [7] (both of which usually adopt SVM for transductive classification). In active learning sampling [5-6], samples that are close to classification hyper-plane are adaptively chosen to train new SVM, which selects the most "informative" samples instead of randomly sampling. Such selection can be evaluated by entropy-based uncertainty lost to halve the version space at each round RF. In Bootstrap sampling based relevance feedback learning, by sample bootstrapping and classifier aggregation, [7] used randomly-selected training sets to overcome the drawbacks of sample insufficiency and sample asymmetry. However, how to better deal with the overlapped training samples is still an open problem.

To deal with the second issue, boosting feature selection is proposed in literature [8, 9]. Tieu [8] adopted AdaBoost in RF learning. He constructed a "highly selective" feature set from a high-dimensional feature space. However, the computational efficiency of [8] is low since over 46,000 features are extracted from each image. Hertz [9] proposed a DisBoost algorithm to learn a distance function by boosting binary classifiers with a confidence interval in the product space. However, over-fitting problem would happen in DisBoost algorithm when training weak classifiers with insufficient and overlapped feedback samples.

This paper presents a clustering-based subspace SVM ensemble algorithm to overcome above two problems by a unified framework. Form interaction interface aspect, our strategy differs from state-of-the-art RF procedures: Based on image query, we firstly extend the most similar images by nearest neighbor expansion. Then a *C-Mean* clustering process is conducted to divide these images into visually similar groups. Subsequently, users' relevance feedback is demanded within each cluster to label images of this cluster

Authorized licensed use limited to: Harbin Institute of Technology. Downloaded on November 27, 2008 at 03:27 from IEEE Xplore. Restrictions apply.

as training examples. From learning algorithm aspect, in each image cluster, the *Eigen* feature subspace is *boosted* from a base feature set using an entropy-based uncertainty criterion, based on which cluster-specific SVM is trained to give labeling prediction for examples that fall within this cluster. For a test images in the database, the regression results of these subspace SVMs are probabilistic intergrated to output a final decision, in which the voting probability is based on the distance between test example and these clusters. These probabilities are treated as visual distances for refined similarity ranking (Fig.1).

In particular, two above issues are addressed as follows:

1. We train cluster-specific SVMs in their best-fit feature subspace to follow the distribution of training samples. We refine traditional SVM training by clustering-based sample selection. Comparing to passive sampling, in the case that training samples are overlapped, our method can adaptively fit data distribution by conjunctive local approximation and hence can save computational time.

2. Based on *Eigen* feature subspace boosting as well as adaptive SVM regression assembling, we construct a clustering-driven, subspace-selective *semantic space* for perceptually similar ranking.

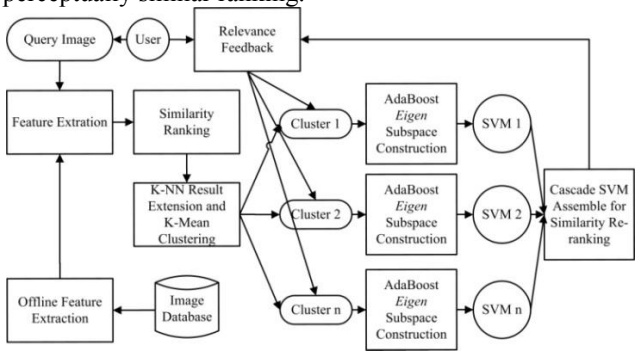

Figure 1: Clustering-Based Subspace SVM Ensemble Framework

2. CLUSTERING-BASED SUBSPACE SVMS

Based on initial query results, similar to Chen [9], we partition the most similar returning images into C clusters. The top m returning results are extended to include $k \times m$ images (k images for each cluster) in the database by nearest neighbor extension. Subsequently, a *C Means* clustering process is conducted to cluster images into C clusters, in which we use *Minimal Description Length* (*MDL*) criterion [10] to determine C.

Based on cluster-specific relevance feedbacks of user, we present a two-step learning strategy to train SVM in each cluster: 1). an entropy-based criterion is proposed to boost a discriminant feature subspace using AdaBoost [13]. 2). using "*Eigen*" subspace, we train SVM in this cluster, and utilize its regression results to produce the probabilistic labels of images that fall within this cluster.

As the first step, for a cluster that contains several images from above $k \times m$ images, suppose the positive and negative samples are Ip_i ($i = 1$ to n_p) and In_j ($j = 1$ to n_n) respectively. We construct a weak classifier for each feature using an entropy-based criterion as follows:

The distances between each image and labeled images (positive and negative) in each cluster are calculated as:

$$Dis_{i,j}^{Pos} = \| F_{i,j}^{l} - F_{PosCenter,j}^{l} \| \quad (1)$$

$$Dis_{i,j}^{Neg} = \| F_{i,j}^{l} - F_{NegCenter,j}^{l} \| \quad (2)$$

In each cluster, entropy-based K-L divergence is adopted to evaluate the target representation of each feature as:

$$Entropy_{i,j}^{k,l} = \frac{Dis_{i,j}^{Pos}}{Dis_{i,j}^{Pos} + Dis_{i,j}^{Neg}} \log\left(\frac{Dis_{i,j}^{Pos}}{Dis_{i,j}^{Pos} + Dis_{i,j}^{Neg}}\right) + \frac{Dis_{i,j}^{Neg}}{Dis_{i,j}^{Pos} + Dis_{i,j}^{Neg}} \log\left(\frac{Dis_{i,j}^{Neg}}{Dis_{i,j}^{Pos} + Dis_{i,j}^{Neg}}\right) \quad (3)$$

$$Entropy_{j}^{k,l} = \frac{1}{(n_p + n_n)} \left(\sum_{i=1}^{n_p} Entropy_{i,j}^{k,l} + \sum_{i=1}^{n_n} Entropy_{i,j}^{k,l}\right) \quad (4)$$

where $F_{i,j}^{\ l}$ is the j^{th} feature of i^{th} image in l^{th} cluster; Dis_i^{pos} and Dis_i^{neg} are the distances of image i to the positive and negative center within each image cluster respectively. $Entrop^{k,l}{}_{i,j}$ is the entropy of j^{th} feature in i^{th} image of the l^{th} subspace of l^{th} cluster and $Entrop^{k,l}{}_{j}$ is the overall entropy of j^{th} feature in the l^{th} subspace of l^{th} cluster.

We treat evaluation results above as the weak classifier of this feature, based on which we adopt AdaBoost [13] to select top N most discriminative features to construct an *Eigen* feature set. In this *Eigen* feature subspace, a SVM is trained for this cluster to fit its cluster-specific sample distribution.

3. SUBSPACE SVMS ENSEMBLE FOR RF

Different from traditional SVM ensemble methods [4, 12] which usually adopt simply classifier voting in labeling prediction, we combine regression results from subspace SVMs by probabilistic integration to give the final labeling prediction. In each cluster, we adopt SVM regression results instead of binary Positive/Negative labeling to classify the unlabeled images that falls within this cluster. To address the margin problem, we distribute each test image to multiple clusters by a distance-based probability as:

$$\Pr o_{i,j} = \log \frac{Dis_{\min}}{Dis(i,j)} \quad (5)$$

where *Dis(i, j)* is the distance between i^{th} image and j^{th} cluster center in original feature space, divided by the variance of this cluster; Dis_{min} is the minimum distance between i^{th} image and all cluster centers.

We normalize SVM regression results into [-1, 1] and probabilistically combine them to produce our final prediction about the labeling of each image for ranking.

Tab.1 presents our subspace SVM ensemble algorithm. By subspace SVMs ensemble, the final classification hyper-plane can be considered as a clustering-based, subspace-mixture hyper-plane. Dealing with the training examples overlapping problem and conjunct subspace SVMs by local

Authorized licensed use limited to: Harbin Institute of Technology. Downloaded on November 27, 2008 at 03:27 from IEEE Xplore. Restrictions apply.

margin approximation, our method is more adaptive to users' query concept.

Table 1: Relevance feedback ranking by subspace SVM ensemble

Input: Positive training samples $S+$, Negative training samples $S-$, Images in the database I_1, I_2, …,I_n; Image Cluster C_1, C_2,…,C_k from the former RF iteration; And their corresponding SVM: SVM_1, SVM_2,…, SVM_n.

For each image I_i in the database

{

For each cluster C_j

{

Use Eq.5 to **output** the cluster belonging probability of I_i.

Utilize its corresponding SVM: SVM_j to **give** the probabilistic regression labeling of I_i. **Output** the probability ranking results of j^{th} cluster by:

$$n_{Selected} = \arg\min\left(\frac{1}{n_{Ck}}\sum_{i=1}^{n_{Ck}}\left(\sum_{m=1}^{n_F}(F_j^m - F_i^m)^2\right)^{\frac{1}{2}}\right) \quad (10)$$

}

Integrate regression results of all n clusters for I_i as the final probabilistic positive/negative labeling:

$$Labeling_{final} = \sum_{j=1}^{n} Labeling_j \quad (7)$$

}

Output: **Select** top n most nearest images to users

We briefly compare our **C**lustering-based **S**ubspace **S**VM (CSS) with bootstrap sampling SVM [7] and active learning SVM [5] as follows. The experimental comparison is further presented in Section.4.

CSS vs. Bootstrap Sampling SVM: Tao [7] proposed a bootstrap-based random sampling algorithm to address the problems of sample insufficiency and sample asymmetry in relevance feedback learning. He utilized an asymmetry bagging strategy to randomly partition negative samples into several overlapped subsets to train SVMs, based on which multiple SVMs are assembled by sequential aggregation.

In our method, based on cluster-based user labeling in RF interaction, cluster-specific training sets are gain to facilitate classifier to find a good partition in each cluster. Furthermore, we use *Eigen* subspace instead of random subspace [7]. Comparing to [7], our merits lie in:

1. CSS elaborately selects cluster-specific labels for training. It fits sample distribution better by local approximation and global conjunction. Since the time cost of SVM training is exponential proportional to the volume of training set, the computational efficiency of our method is far better in online RF scenario.

2. CSS adaptively use less discriminant features in SVM training to avoid the *small sample learning* problem, which is more appropriate than randomly feature selection in [7].

CSS vs. Active Learning Sampling SVM: In active learning sampling, there are two commonly-used strategies:

1. Maytal [11] adopted sample density estimation for training sample selection: uncertain regions in feature space would be densely sampled while determinative regions would be sparsely sampled. 2. Tong [5] proposed a criterion to adaptively select the sampling points that closest to the classification hyper-plane for SVM training in relevance feedback.

Comparing to active learning sampling [5, 11], CSS also merits in dealing with overlapping training data: If the data is generated from two heavily overlapping Gaussian distribution, the decision boundary would go right through the densest region, and consequently its corresponding SVM would perform poorly. In such scenario, CSS partition would be a more feasible solution. Furthermore, since CSS adaptively maps each image cluster to its fitness feature subspace, it can further avoid the negative effect of feature over fitting in SVM training.

4. EXPERIMENTAL RESULTS

Experimental Setup: In our method, 3-level pyramid Haar wavelet features are adopted to construct a hierarchical feature vector. Different from traditional methods which extract features from solely color component (I or H in HSI; R, G, B of RGB), we extract the Hue histogram from LL components in H-band Haar and the LH, HL and HH wavelet coefficient histogram in I-band Haar at each resolution. The 360-bin Hue histogram can be considered as the color representation of an image, while the 3×64-bin wavelet coefficient histogram can be regarded as its texture description. Subsequently, 552 features are constructed at each resolution, which form a 1656-dimentional feature base set in 3 resolutions for boosting selection. We offline extract these features in image database and construct a hierarchical color & texture representation for each image.

Our experiments are conducted on a subset of COREL image database, with over 5,000 general purpose images based on 50 CD-ROMs published by COREL Corporation. 500 images (10 of each) are randomly selected as query set. All our experiments are performed on a 3.0 GHz Intel® Pentium IV machine with 1 GB RAM. The algorithms are implemented using C++ in MS® Visual Studio environment.

We use precision-recall curve to evaluate our relevance feedback performance. Precision and recall are calculated using Eq.8, in which $n_{related}$ is the number of correctly returning images; n_{return} represents the total number of returning images; n_0 is the related image number in database:

$$\text{Precision} = \frac{n_{related}}{n_{return}} \qquad \text{Recall} = \frac{n_{related}}{n_0} \quad (8)$$

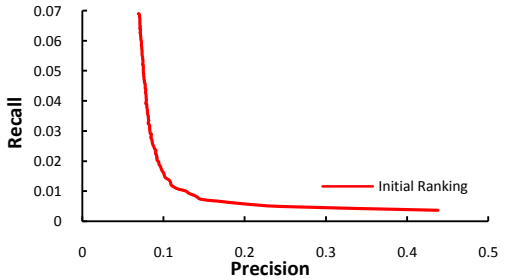

Figure 2: Initial Ranking by Proposed Color & Texture Features

Fig.2 presents the P-R curve of initial query results in top 100 images using our 1656-dimentional visual features. Consequently, by nearest neighbor extension of most similar m ranking result, *C Means* clustering is conducted (average

Authorized licensed use limited to: Harbin Institute of Technology. Downloaded on November 27, 2008 at 03:27 from IEEE Xplore. Restrictions apply.

C is 3-4 in our experiments), based on which user can label at most *10* images within each cluster as positive/negative.

Clustering-based Subspace SVM Ensemble (CSS) vs. Standard SVM: As presented in Fig.3, by clustering-based sample selection, proposed method (CSS) significant outperforms standard SVM learning method based on passive sampling. Besides Fig.3, we also discover that the performance enhancement is over 20% in top 40 images and over 14% in top 35 images for first RF; over 26% in top 50 images and over 18% in top 45 images for second RF.

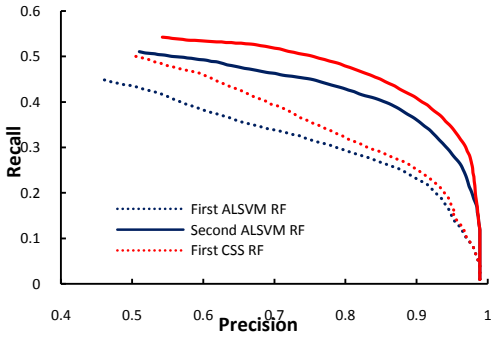

Figure 3: Relevance Feedback Comparison between Proposed CSS Method and Standard (Passive Sampling) SVM

Clustering-based Subspace SVM Ensemble (CSS) vs. Bootstrap Sampling SVM: As presented in Fig.4, compared to bootstrap sampling [7], our method is more efficient in first and second RF operations. Besides Fig.4, we also discover that the performance enhancement is over 10% in top 40 images and over 14% in top 45 images for first RF; over 16% in top 50 images and over 8% in top 45 images for second RF.

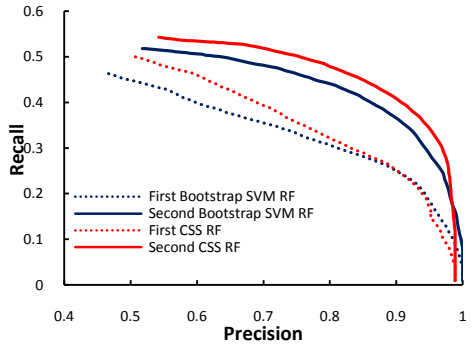

Figure 4: Relevance Feedback Comparison between Proposed CSS Method and Bootstrap (Random Sampling) SVM

Clustering-based Subspace SVM Ensemble (CSS) vs. Active Learning Sampling SVM: Fig.5 presents the RF performance comparison of proposed method and the SVM learning method based on active sampling, our method still outperforms ALSVM. Besides Fig.5, we also discover that the performance enhancement of CSS comparing to ALSVM is over 14% in top 40 images and over 8% in top 35 images for first RF; over 24% in top 50 images and over 12% in top 45 images for second RF.

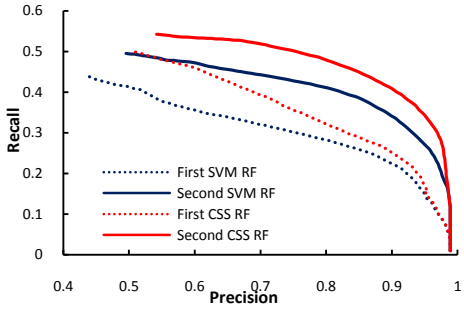

Figure 5: Relevance Feedback Comparison between Proposed CSS Method and Active Learning (Active Sampling) SVM

5. CONCLUSION

This paper presents a clustering-based subspace SVM ensemble framework for adaptive relevance feedback. Our contribution is two-folds: From human-computer interaction viewpoint, to better capture users' retrieval intension, we extend traditional relevance feedback to cluster-level interaction by nearest-neighborhood result extension and cluster-specific user labeling. From learning algorithm improvement aspect, we adopt AdaBoost feature selection to discover the *Eigen* feature subspace for each cluster, based on which cluster-specific SVMs are trained for labeling regression. By integrating SVM regression results, we can give probabilistic labels of images for similarity re-ranking. In experiments, our method outperforms bootstrap sampling SVM [7] and active learning sampling SVM [5] in COREL 5,000 dataset, which demonstrates our effectiveness.

Acknowledgement: This work is support by State 863 High Technology R&D Project of China (No. 2006AA01Z197).

6. REFERENCES

1. A. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, " Content-Based Image Retrieval at the End of the Early Years", *PAMI*, Vol.22, No.12, December 2000, pp.1349-1380.
2. Y. Gong, H. J. Zhang and T. C. Chua, "An image database system with content capturing and fast image indexing abilities," Int. Con. on Multimedia Computing and System, Boston, 14-19 May 1994, pp.121-130.
3. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *DMKD*, vol.2, no.2, pp.121-167, 1998.
4. HC. Kim, S Pang, HM Je, D Kim, S Yang Bang, "Constructing support vector machine ensemble," *Pattern Recognition*, 2003.
5. S. Tong, and E. Chang, "Support Vector Machine Active Learning for Image Retrieval," *ACM Multimedia*, 2001.
6. F. Jing, M. Li, H. Zhang, B. Zhang, "Entropy-based active learning with support vector machines for content-based image retrieval," *ICME*, 2004.
7. D. Tao, X. Tang, X. Li, X. Wu, "Asymmetric Bagging and Random Subspace for Support Vector Machines-Based Relevance Feedback in Image Retrieval," *PAMI*, vol.28, no.7, July, 2006.
8. K. Tieu, and P. Viola, "Boosting Image Retrieval," *IJCV*, 2004.
9. T. Hertz, A. Bar-Hillel, and D. Weinshall, "Learning Distance Functions for Image Retrieval," *ICPR*, 2004, vol.2, pp.570-577.
10. Y. Chen, J. Z. Wang, and Robert Krovetz, "CLUE: Cluster-Based Retrieval of Images by Unsupervised Learning," *IEEE Trans. on Image Processing*, Vol.14, No.8, 2005, pp.1187-1201.
11. J. Rissanen, "Modeling by shortest data description," *Automatica*, 14 (5): 465-471, 1978.
12. HC. Kim, S Pang, HM Je, D Kim, S Yang Bang, "Pattern Classification Using Support Vector Machine Ensembel," *ICPR*, 2002.
13. Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *ICML*, 1996.

Authorized licensed use limited to: Harbin Institute of Technology. Downloaded on November 27, 2008 at 03:27 from IEEE Xplore. Restrictions apply.