KNSC: A NOVEL LOCAL CLASSIFICATION METHOD FOR MULTIMEDIA SEMANTIC ANALYSIS

Kun Tao[1,2], Shouxun Lin[1] and Yongdong Zhang[1]

1 Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
2 Graduate School of the Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT

The local classification methods try to simplify the complex global modeling problem by decomposing it into a set of local classification sub-problems, which is a potential key to overcome the semantic gap in multimedia content analysis. In this paper we proposed a Sample-Balancing Clustering segmentation method and an effective local classification framework named K-Nearest Sub-classifiers (KNSC). In KNSC the final prediction is an ensemble of the predictions made by K nearest local classifiers. We experimentally compare the effect of different sub-domain segmentation methods, different types of sub-classifiers and different classification/ensemble strategies. The applications on semantic analysis of TRECVID data show the good performance of our method.

Index Terms— local classification, semantic analysis

1. INTRODUCTION

The semantic gap in multimedia data analysis is a widely recognized challenge, which has severely obstructed the popularization of semantic-based multimedia retrieval and other applications. The commonly used semantic concepts can be roughly categorized into objects, scenes and events. For most concepts, when we observe their positive samples, we can find that they show diversiform visual (or audio etc.) appearance. Analyzing such diversity in the perspective of pattern classification, it means the samples are dispersive in feature space.

For example, when we observe the images with the concept “Street”, we find that the streets own different building structures and appearance colors. Moreover, the images might be photographed from different viewpoints at different time. The vehicles and people in the street are also uncertain factors. Correspondingly, when we observe the Color-Histogram feature space, we can find the dominant color of positive samples may be gray, red or many other colors. In other feature spaces such as Edge-Histogram or Texture-Cooccurrence we can also find such dispersiveness.

The sample dispersiveness greatly increased the complexity and difficulty of classification modeling .We can hardly find out an existing statistical model that can deal with such data perfectly. The Gaussian Mixture Model might satisfy such requirement theoretically, but it is difficult to practice because of the parameter selection problem. As the traditional solving methods have met great challenge, some new thinking such as local classification must be imported.

The core idea of local classification is decomposing the complex global modeling problem into some local sub-problems. Based on this principle, some previous works have been published. An early exploration in [1] discussed the strategy that fit models of different complexity on sub-domains of the input space, in which each cluster is learned as an intermediate concept. Cevikalp et al. [2] deal with the classification problems based on the idea of manifold, and build decision surfaces for the subspaces. It provided a good theoretical support to the value of local classification. Zhang et al. [3] combine the KNN and SVM, and proposed a SVM-KNN method for visual category recognition. But they use the KNN as a prior stage of getting SVM sub-classifier, which limits its applications in great-scale data set. In [4] a modified version of Growing Neural Gas is introduced to find the topology and create clusters dynamically. The combination of Neural Networks and local SVMs results in good performance. A further discussion on multimodal combination of multiple local classifiers is also presented in [5].

Above works have proved the power of local classification methods, but we can still find many aspects that can be discussed more deeply. In this paper, we focus on two of these aspects: 1) how to define and segment the sub-domains; 2) can we learn from the comparison between Nearest Neighbor method and K Nearest Neighbor method, and use more neighboring sub-classifiers to assist the classification. Through the comparison experiments, we find that choosing an overlapped sub-domain segmentation and using the ensemble of multi sub-classifier can improve the classification performance effectively. Based on above conclusion, a novel local classification framework is proposed.

2. THE DESCRIPTION OF ALGORITHM

We consider that a local classification method is decided by three components: the sub-domain segmentation algorithm, the sub-classifier and the classification/ensemble strategy. Each of them can be selected out of many existing candidate algorithms. In sub-domain segmentation step, we proposed a Sample-Balancing Clustering (SBC) algorithm. Gaussian Naive Bayes is selected as the default sub-classifier. Finally, we take an ensemble of k-nearest sub-classifiers as our classification strategy. The details of the algorithms is introduced as follows:

2.1. Sub-domain Segmentation

The sub-domain segmentation is a decisive step for a local classification method. A generalized definition of sub-domain is a connected region in the input space. Sub-domains might be overlapped or not. But for the sake of classification, it is better that the sub-domains can be defined by prototypes. In this paper, we use the Sample-Balancing Clustering to find out overlapped sub-domains. As a contrast, a density-based K-means method [6] is also used int experiments as the paradigm of non-overlapping clustering methods.

The following is the algorithm description of the Sample-Balancing Clustering:

1) Get the train data set $X = \{x_1, x_2, \ldots, x_n\}$

2) Calculate the average distance d_{avg} of every two points' distance

3) Take d_{avg} as the sphere radius, count the neighboring sample number ns_i of each x_i, get a set $NS = \{ns_1, ns_2, \ldots ns_n\}$

4) Sort the set NS in ascending order, we define that the members in NS correspond to those un-assigned samples which have never been included by any existing cluster.

5) Calculate the average value ns_{avg} of NS as an expected value of the size of each cluster

6) Find the first sample ns_i in NS, taking the corresponding x_i as the center $center_k$ of a new cluster C_k

7) Take d_{avg} as the initial radius, search for a property radius r_k that can make the neighboring sample number of C_k is between $0.8 * ns_{avg} \sim 1.2 * ns_{avg}$

8) Delete the neighboring samples with too low density $ns_i < 0.5 * ns_{center_k}$ or with too high density $ns_i > 5 * ns_{center_k}$, the left samples are considered to be the members of C_k

9) Find out the samples of C_k who is included in NS, delete them from NS

10) If $NS = NULL$, stop clustering; otherwise jump to step 6

In above algorithm, each cluster is described by a center point and a corresponding radius. For a newcome sample x_j, its similarity to a cluster C_k is defined as

$$s_{jk} = r_k / distance(x_j, center_k) \quad (1)$$

In low density regions the clusters own big radius, and in high density regions they own small radius. Such scale-adjustable segmentation can balance the clusters' sizes and improve the performance of sub-classifiers. The sparse samples will not be ignored, and the clusters which are too big for modeling will also be avoided.

The SBC searches for new clusters from low density regions to high density regions. The ns_i is an important attribute, it is a substitution of local sample density. Each time when a new cluster is built up, it commonly include some assigned samples and some un-assigned samples. As the un-assigned samples belong to the regions with higher density, the sphere of a new cluster should be able to cover more un-assigned samples, which makes the algorithm converges faster. The step 8 is to reduce the density unbalance of clusters caused by overquick expansion, which might affect sub-classifier modeling. Fig. 1 shows a demonstrative illustration of SBC searching.

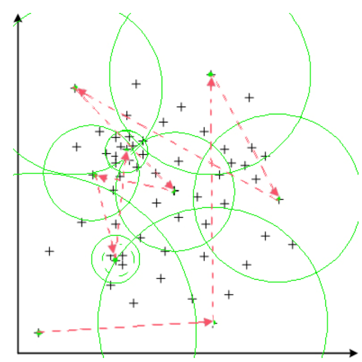

Fig. 1. A demonstrative illustration of SBC searching

2.2. Local Classifiers and the Ensemble

There are many different kinds of classifiers which can be selected as the local sub-classifier. In this paper, two sophisticated method are selected for contrasting: SVM and Gaussian Naive Bayes (GNB). In following experiments, we can see that GNB performs better than SVM. So we selected GNB as our default sub-classifier.

There are two points which must be mentioned: First, when we are training SVM models in a feature space, we use a uniform cost function parameter C for all sub-classifiers. The C was chosen for a global SVM classifier through cross-validation. Second, as it's very hard to build model of high-dimension Gaussian distribution, we have to work on the assumption of dimension-independence.

We regard that neighboring sub-domains have some mutual influences no matter they are overlapped or not. Most latent data structures in feature space can hardly be captured by a single artificially defined sub-domain, but they can be shared by several neighboring sub-domains. Therefore the sample distributions of those neighboring sub-domains can provide important reference information for the classification of a testing sample. As we know, the K Nearest Neighbor method is more robust than Nearest Neighbor method. So we can also use the ensemble of K nearest sub-classifiers. Furthermore, it is obviously reasonable when we are using overlapped clusters because there is great possibility that a sample is included in several clusters. We use formula (2) to calculate the ensemble prediction. $p_k(x_j)$ is the prediction of the *k*–th nearest sub-classifier, and the weight value s_{jk} is the similarity between testing sample x_j and cluster C_k :

$$p(x_j) = \sum_{k=1}^{K} s_{jk} * p_k(x_j) \qquad (2)$$

In our experiments $K = 5$. No matter the sub-domains are overlap or not, the KNSC performs better than using a single nearest sub-classifier because more information is referred.

3. EXPERIMENTS AND DISCUSSION

We use the TRECVID 2008 High-level Feature Extraction data set for our comparing experiments [7]. There are 39674 key-frame pictures, and the labels of 20 concepts are also provided. 2/3 of the pictures are randomly selected as training data, the others are testing data. Six different features are used here: Color-Correlogram (CC), Color-Histogram (CH), Color-Moments (CM), Edge-Histogram (EH), Texture-Cooccurrence (TC) and Texture-Wavelet (TW) [8].

In this part, the combinations of different sub-domain segmentation algorithms, different sub-classifiers and different classification/ensemble strategies are investigated. The sub-domain segmentation is based on SBC and K-means (KM). The sub-classifiers include SVM and GNB. The classification/ensemble is based on simple Nearest Sub-Classifier (NSC) and K Nearest Sub-Classifier (KNSC). There are 8 combination modes, each method is named by its steps in turn, such as NSC-GNB-KNSC. A global model based on SVM (G-SVM) is selected as the baseline.

We first make a general survey on the average precision of different methods. The average precision is calculated for each feature by averaging the result of all 20 concepts. From Fig. 2 we can see that the 8 local classification methods perform better than global SVM in most conditions. We can also find that the SBC-GNB-KNSC is the best. It's a natural result because it always selects the better method in each step. Such issue is supported by the following analysis.

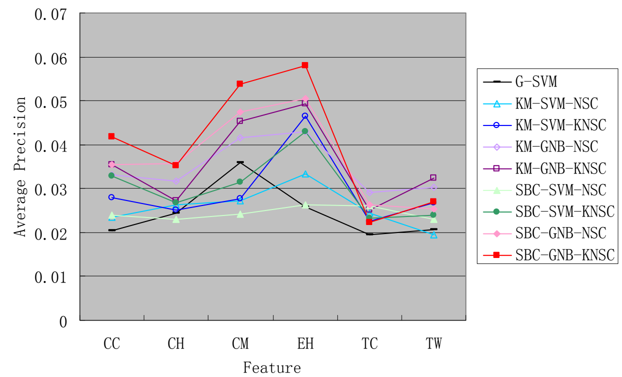

Fig. 2. The average precisions of 9 methods

Then we compare the 8 local classification methods pair by pair and try to discuss how different strategies affect the classification result. For each method, the mean value of the 6 average precisions of different features is named as Mean Average Precision (MAP). The radar chart is used here in order to give a more intuitional comparison of MAPs. For example, the Fig. 3 can help us to contradistinguish the effect of KM and SBC. The four combination modes of other two step act as the four data axis, and the distance for data points to origin point indicates the MAP value.

3.1. KM vs. SBC

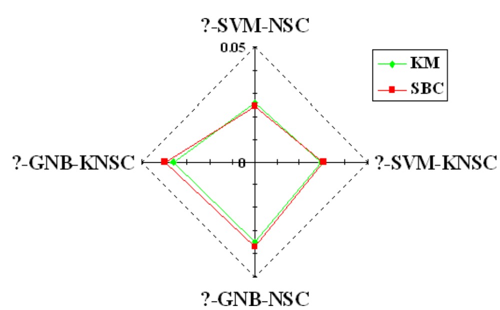

Fig. 3. The comparison of KM and SBC

As we can see in Fig. 3, the selection of clustering methods has little effect on SVM-based methods, but SBC performs better than K-means when they are combined with GNB sub-classifiers. This result is decided by the different operating principle of two classification methods. The SVM model is decided by the support vectors beside the decision boundaries, which is robust to outlier samples. The GNB needs to calculate the mean and variance of samples, which makes it more sensitive to outliers. Thus, GNB can get more profit from the sample-balancing segmentation of SBC.

3.2. SVM vs. GNB

The radar chart of Fig. 4 shows that the GNB performs better than SVM significantly when they act as the sub-classifiers. We attribute this result to the over-fitting problem. Each sub-classifier is trained only by the samples in a small sub-domain. When we train SVM as sub-classifiers, it is very hard to find a proper cost function C for each sub-classifier which can ensure its generalization capability when it is applied on the whole data space. In contrast, decomposing the data space is more beneficial to GNB. The Gaussian model can not deal with complex distributions, but it's still a good tool to simulate the local distribution.

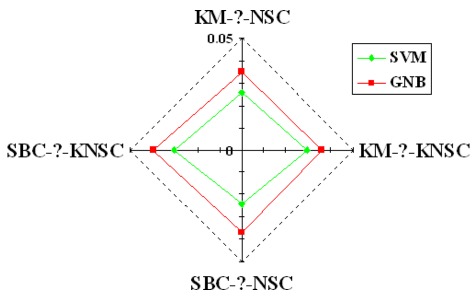

Fig. 4. The comparison of SVM and GNB

3.3. NSC vs. KNSC

Fig. 5 is the comparison between NSC and KNSC. The KNSC is the winner. Such result is a strong proof of our previous argument: the ensemble of multi neighboring sub-classifiers is better than a single sub-classifier.

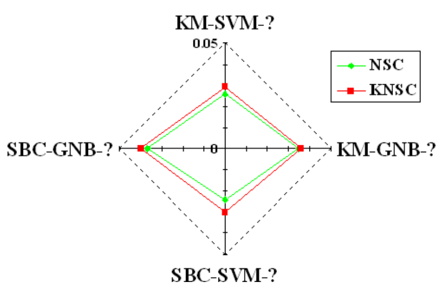

Fig. 5. The comparison of NSC and KNSC

4. CONCLUSION

In this paper, we proved that the local classification method is a very effective means to resolve the modeling problem of multimedia semantic analysis. Our work concerns on three important aspects of local classification method: sub-domain segmentation, sub-classifier and classification/ensemble. Through the pair by pair comparison, eight different combinations of algorithms are tested. Some preliminary conclusions can be made: 1) comparing to non-overlapping segmentation, overlapped segmentation can make sub-classifiers more robust, 2) Gaussian Naive Bayes is a better choice for sub-classifier than SVM, 3) the ensemble of K nearest sub-classifiers is better than using single sub-classifier. Certainly, above works is only a very primary exploration, there are many other methods which can be discussed. In our feature works, we plan to test some other clustering methods such as CURE, which can handle non-spherical shapes. The effect of using other distance metric or sub-classifier will also be discussed.

5. ACKNOWLEDGEMENT

This work was supported in part by the National Basic Research Program of China (973 Program, 2007CB311100), the National High Technology and Research Development Program of China (863 Program, 2007AA01Z416), National Nature Science Foundation of China (60873165、60802028), the Beijing New Star Project on Science & Technology (2007B071), and Co-building Program of Beijing Municipal Education Commission.

6. REFERENCES

[1] R. Vilalta, M Achari and C.F. Eick, "Piece-Wise Model Fitting Using Local Data Patterns", 16th European Conference on Artificial Intelligence(ECAI), Spain, 2006

[2] H. Cevikalp, D. Larlus, M. Douze and F. Jurie, "Local Subspace Classifiers: Linear and Nonlinear Approaches", IEEE Workshop on Machine Learning for Signal Processing, Greece, pp. 57-62, Aug. 2007

[3] H. Zhang, A.C. Berg, M. Maire and J. Malik, "SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition", IEEE Conf. on Computer Vision and Pattern Recognition, IEEE Computer Society, Vol. 02, pp. 2126-2136, 2006

[4] R. M. Rodriguez-Pena, B. Perez-Sanchez and O. Fontenla-Romero, "A Novel Local Classification Method using Growing Neural Gas and Proximal Support Vector Machines", International Joint Conference on Neural Networks(IJCNN), pp. 1607-1612, 2007

[5] H. Cevikalp and R. Polikar, "Local Classifier Weighting by Quadratic Programming", IEEE Trans. on Neural Networks, Vol. 19, No. 10, pp. 1832-1838, October 2008

[6] Y. X. Xie, Y. N. Hu, etc., "An efficient indexing algorithm of clustering supporting QBE image retrieval for large image database", Mini-Micro Systems, vol. 10, pp. 1229-1233, 2001 (in Chinese).

[7] TREC Video Retrieval Evaluation, http://www-nlpir.nist.gov/projects/trecvid/

[8] Arnon Amir, Janne Argillander, Murray Campbell, et al., "IBM Research TRECVID-2005 Video Retrieval System", NIST TRECVID-2005 Workshop, Gaithersburg, MD, November 2005.