Multiple-Instance Active Learning for Image Categorization*

Dong Liu[1], Xian-Sheng Hua[2], Linjun Yang[2], and Hong-Jiang Zhang[3]

[1] School of Computer Science and Technology, Harbin Institute of Technology, 150001, Harbin, Heilongjiang, China
[2] Microsoft Research Asia, 100190, Beijing, China
[3] Microsoft Advanced Technology Center, 100190, Beijing, China
dongliu@hit.edu.cn, xshua@microsoft.com, linjuny@microsoft.com, hjzhang@microsoft.com

Abstract. Both multiple-instance learning and active learning are widely employed in image categorization, but generally they are applied separately. This paper studies the integration of these two methods. Different from typical active learning approaches, the sample selection strategy in multiple-instance active learning needs to handle samples in different granularities, that is, instance/region and bag/image. Three types of sample selection strategies are evaluated: (1) selecting bags only; (2) selecting instances only; and (3) selecting both bags and instances. As there is no existing method for the third case, we propose a set kernel based classifier, based on which, a unified bag and/or instance selection criterion and an integrated learning algorithm are built. The experiments on *Corel* dataset show that selecting both bags and instances outperforms the other two strategies.

Keywords: Multiple-Instance Learning, Active Learning, Bag & Instance mixture selection, Image Categorization.

1 Introduction

There has been significant work on applying multiple-instance (MI) learning to image categorization [1,2]. Key assumptions of these works are that each image is represented as a bag which consists of segmented regions as instances and a bag receives a particular label if at least one of its constitutive instances possesses the label. In MI learning, it is difficult to predict the labels of instances given the bags' labels. This difficulty is so-called MI ambiguity. Active learning is also a widely applied method in image categorization as it can significantly reduce the human cost in labeling training images. In its setting, the learner has access to a large pool of unlabeled images and selects the most valuable images for manual annotation, such that the obtained training set is more effective. Although promising performance on image categorization has been reported using these two methods, they are generally applied separately.

* This work was performed when Dong Liu was visiting Microsoft Research Asia as an intern.

B. Huet et al. (Eds.): MMM 2009, LNCS 5371, pp. 239–249, 2009.

In this paper, we study the integration of multiple-instance learning and active learning in image categorization application. A first attempt on MI active learning [3] proposes a strategy that selectively labels certain portion of instances from positive training bags. It firstly trains a MI logistic regression classifier, then computes MI uncertainty of each instance in positive training bags using current classifier's prediction probability and selects the instance with the highest uncertainty as the next querying sample. After the corresponding instance label is provided by the oracle, it updates the training set by adding a new singleton bag containing only a copy of the queried instance. The MI classifier is retrained on the expanded training set with mixed-granularity labels and hence the MI classifier's performance can be improved.

We argue that whereas the above method works well for MIL by selecting only instances of labeled bags for labeling, MIL based image categorization can also benefit from selectively labeling some unlabeled bags. For MI learning, labeling instance and labeling bag provide different information to the classifier in different cases. For example, if the chosen bag is labeled positive, then it provides less information compared with labeling a positive instance in that bag, because in this case the label of the bag can be inferred from the label of the instance, but it is difficult to get the instance label through the bag label. On the other hand, if the chosen bag is labeled as negative, it provides more information then labeling any instance in that bag. Although labeling instance and labeling bag can benefit the MI classifier, it remains a problem which is the most effective labeling way for MI active learning, labeling bags, labeling instances, or labeling bags & instances simultaneously.

In this paper, we conduct a comparative study on three different MI active learning sample selection strategies for image categorization, including selecting bags only, selecting instances only, and selecting bags and instances simultaneously. As there is no existing method for the third case, we need to build a mixed bag and/or instance sample selection framework to perform MI active learning.

To develop such a unified MI active learning approach, two crucial problems need to be solved. One is a unified MI classifier which can classify instances as well as bags. The other is the sample selection criterion, which can maximally reduce the classification error by selecting the most valuable samples for labeling.

The unified MI classifier is realized using standard SVM with a MI normalized set kernel. By treating instances as singleton bags which comprise only one instance, the MI normalized kernel can be employed to estimate the similarity between bags and bags, instances and instances, bags and instances. To select the most informative samples, regardless of bags or instances, a MI informativeness measure that takes uncertainty, novelty and diversity into consideration is proposed to query unlabeled samples from the querying pool for labeling.

The experiments are conducted on *Corel* dataset. By comparing three different sample selection strategies: selecting instances, selecting bags, and selecting the mixture of bags & instances, we can conclude that: (1) selecting the mixture of bags & instances performs the best in most of the active learning querying rounds;

(2) selecting instances performs better than selecting bags when selected samples are few while selecting bags performs better when more samples are selected.

The rest of this paper is organized as follows. Section 2 presents a framework for MI active learning. Section 3 describes the experiments to evaluate three different sample selection strategies for MI active learning on benchmark CBIR dataset. Section 4 concludes the paper.

2 MI Active Learning

In this section, two critical problems in MI active learning, including the unified bag/instance classifier and the sample selection criterion, are addressed. In order to construct a MI classifier, which can classify bags (images) as well as instances (image regions) in a unified fashion, a MI set kernel is adopted, as described in section 2.1. With the MI set kernel, standard SVM is employed as the classifier for both bags and instances. In section 2.2, the MI sample selection criterion is proposed to maximize the performance of the classifier. The MI active learning framework is presented in section 2.3.

2.1 MI Set Kernel

To build a unified MI classifier which can classify instances as well as bags, we need to estimate the similarity between bags and bags, instances and instances, bags and instances. To this end, we adopt a particular MI kernel called normalized set kernel [4] in the learning of SVM. A kernel on sets can be derived from the definition of convolution kernels and can be formally represented as follows,

$$k_{set}(B, B^{'}) = \sum_{x \in B, x' \in B'} k(x, x^{'}) \tag{1}$$

where $k(.,.)$ is any valid kernel function defined on instances, B and $B^{'}$ are two bags with x, $x^{'}$ are the corresponding instances. If the cardinalities of bags vary considerably, bag with large cardinalities will dominate the set kernel estimation. To overcome this problem, a natural normalization is given

$$k_{nset}(B, B^{'}) = \frac{k_{set}(B, B^{'})}{\sqrt{k_{set}(B, B)}\sqrt{k_{set}(B', B')}} \tag{2}$$

To estimate the kernel between the bag B and the instance b, we can regard the instance as a singleton bag which consists of only one instance. Thus the kernel can be defined as follows,

$$k_{nset}(B, b) = k_{nset}(B, \{b\}) \tag{3}$$

where B is the bag and b is the instance.

Accordingly, the kernel between two instances can also be defined in the same way,

$$\begin{aligned} k_{nset}(b,b^{'}) &= k_{nset}(\{b\},\{b^{'}\}) \\ &= \frac{k(b,b^{'})}{\sqrt{k(b,b)}\sqrt{k(b^{'},b^{'})}} \end{aligned} \tag{4}$$

which is degenerated to be the usual normalized kernel.

Once the kernel between bags and bags, instances and instances, bags and instances, is defined, by employing standard SVM, a unified multiple-instance classifier, which can predict the label of bags as well as instances, is constructed.

2.2 MI Sample Selection Criterion

In active learning, the most "informative" samples to the classifier learning should be selected for labeling by oracles firstly, so as to maximally reduce the classification error. Some heuristic rules, such as uncertainty and diversity, are proposed to approximate the "informativeness" measure. In this paper, we approach the "informativeness" by taking multiple measures including uncertainty, novelty, and diversity, into consideration. As a result, a MI sample selection criterion is proposed by fusing the multiple measures.

Uncertainty. Tong [5,6] proposed a SVM active learning sample selection criterion from the perspective of version space, aiming at selecting the unlabeled samples which can provide most valuable information for the retraining of current SVM classifier. The basic idea is to find the unlabeled sample which results in the maximal reduction of the version space. An efficient implementation of this idea is to select the sample which is the closest to the SVM hyperplane in the kernel space. In other words, the samples with the largest $1-|f(x)|$ are selected for labeling by oracles, where $f(x)$ is the prediction score of the sample x by the SVM classifier, as defined in the below from the dual view,

$$f(x) = \sum_{i=1}^{l} \alpha_i k_{nset}(x, x_i) + b \tag{5}$$

where α_i is the coefficient and b is the offset.

The measure is called uncertainty in that the sample with larger $1-|f(x)|$ is closer to the classification boundary $f(x)=0$ and can be regarded the more uncertain to the prediction.

The uncertainty measure can be defined as follows

$$u(x) = 1 - |f(x)| \tag{6}$$

Where x is a unlabeled sample and f is the decision function.

Novelty. One intuitive assumption in the sample selection is that the samples will be selected with less chance if they are similar to the existing training data. In other words, the samples which are more novel to the training data should be selected with higher probability. The novelty criterion aims to select the samples with minimum overlapping with the existing training samples and enforces that the redundancy in the training samples is minimized.

The novelty measure can be defined as follows,

$$d(x_j) = 1 - max_{1 \leq i \leq l} k_{nset}(x_i, x_j) \tag{7}$$

where x_j is the unlabeled sample and x_i is an existing training sample and l refers to the number of samples in the training set.

Diversity. In this paper, the MI active learning problem we would like to address is batch-mode active learning, i.e., we select multiple samples at one time and query for their labels before putting them into the training set. As shown in [7], the redundancy among the selected unlabeled samples at each query round needs to be reduced to maximally utilize the human labeler's labor. Thus the sample set, which comprises multiple unlabeled samples, with little inter similarity will be preferred to be selected together for labeling. The diversity among the sample set can be estimated by averaging similarities among the samples. Given a set of n selected unlabeled samples $U = \{x_1, x_2, \cdots, x_n\}$ that contains either bags or instances, the redundancy of the $(n+1)th$ sample x_{n+1} with them can be formulated as

$$r(x_{n+1}) = 1 - \sum_{i=1}^{n} k_{nset}(x_{n+1}, x_i)/n \tag{8}$$

MI Informativeness. Now we propose a MI sample selection criterion called MI informativeness based on the above three criteria. According to the theoretical analysis [8], it's more rational to use the diversity measure as a weight of the novelty measure than linearly combine them. Thus we weight the novelty with the diversity, and then linearly combine it with the uncertainty as

$$MI\_Informativeness(u_i) = \lambda \times u(u_i) + (1 - \lambda) \times d(u_i) \times r(u_i) \tag{9}$$

where λ is the trade-off parameter to adjust the individual importance of each criterion, u_i is the unlabeled sample. With the proposed MI informativeness sample selection criterion as defined in Equation (9), the most informative samples which are deemed to maximally benefit the classification learning are selected for labeling and then added to the training set.

2.3 A MI Active Learning Framework

Based on the MI set kernel and the MI informativeness sample selection criterion presented above, a MI active learning framework is proposed and summarized in algorithm 1. Three variant sample selection strategies can be applied in our MI active learning framework based on the type of query sample pool, including bags, instances, and bags & instances.

Algorithm 1. A MI Active Learning Framework

Require: L,initial training data;
$U = \{u_1, u_2, \cdots, u_N\}$,initial pool of samples to be selected for labeling;
n, the number of samples to be queried at each round;
m, the number of rounds.
Ensure: f,the trained classifier.
1. Initialization. Train the initial SVM classifier using the initial training data L, $f = SVM\_Train(L)$. The kernel is defined as Equation (2), (3), and (4).
2. Active Learning.
Repeat m rounds:

a. $S = \Phi$.
b. Repeat until $|S| = n$

- $x = argmax_{u_i \in U} MI\_Informative(u_i)$.
- $S = S \bigcup \{x\}$
- $U = U - \{x\}$

c. $L = L \bigcup S$
d. Retrain the SVM classifier using the new training set L. $f = SVM\_Train(L)$.

- Bags. Only the images in the test data set are selected for labeling, i.e. $u = \{B_i\}$,$i = 1, 2, \cdots, N$.
- Instances. Only regions in positive training images are selected for labeling, i.e. $u = \{b_i\}$,$i = 1, \cdots, N$ where b_i is a singleton bag that contains only a region in a positive training image.
- Bags & Instances. The pool u is composed of both testing images and regions of positive training images.

Our MI active learning framework mainly consists of two steps. The first step is to learn a MI normalized set kernel matrix using the initial training data L and train the initial SVM classifier. The second step is the active learning iterative procedure, in which we apply the proposed sample selection criterion to select the samples with maximum MI informativeness for labeling.

3 Experiments

In this section, we will evaluate three different sample selection strategies for MI active learning on *Corel* image data set. For convenience, we will name the bag and instance selecting strategy as mixed selecting strategy and the other two selecting strategies as bag only and instance only strategy, respectively. Firstly, we compare our active sample selection strategy with random sample selection strategy to confirm the effectiveness of MI active learning. Then we conduct experiments to evaluate the performances of three different sample selection strategies for MI active learning. We use *libSVM* [9] as kernel learner and RBF as basic kernel for normalized set kernel. The parameters γ and C for RBF

Fig. 1. Sample images from *Corel*

and SVM are determined through 10-fold cross validation. The parameter λ in Equation (9) is fixed to 0.7.

3.1 Experimental Testbed and Setup

We adopt the image data set with region labels in [10] as our test bed. This data set consists of 11 classes with 4002 natural scene images from *Corel*. These images are segmented using JSEG [11] into average 26 regions. Then 9-dimensional color moment in HSV color space and 20-dimensional Pyramid-structured wavelet texture are combined into a 29-dimensional region-level feature vector. The detail information of this data set is shown in Table 1. Some sample images form *Corel* are shown in Figure 1.

In the active learning experiments, the learner begins with 20 randomly drawn positive images and 20 random negative images as the initial training data while the remaining images are used as unlabeled test data. The samples are selected at each round from the unlabeled test images for labeling bag label, from the regions in the positive training images for labeling instance label, or both. The remaining samples in the test data are used for performance evaluation. The query batch size n is set to 10 and the query round m to 18 respectively.

The measure used to evaluate the performance of the active learning methods is AUROC [12], which is the area under the ROC curve. Five independent runs are conducted for each image class in the *Corel* data set and the results are averaged as the performance evaluation.

Following previous work on SVM active learning [13], we operate SVM using a kernel correction method which guarantees that the training set is linearly separable in kernel space [14]. This is done by modifying the kernel matrix K so that each diagonal element is added by a constant ξ and the all other elements remain the same. In all experiments in this paper, we fix $\xi = 4$ empirically.

3.2 Performance Evaluation

We firstly compare the average performance of MI active learning with that of random sample selection strategy on *Corel* dataset. From Figure 2, it can

Table 1. Detail Information For Corel Dataset

Concept	Image Num	Region Num	Concept	Image Num	Region Num
Water	1690	9257	Sky	3382	13540
Flower	251	1701	Mountain	1215	9809
Building	1852	19422	Rock	580	6573
Grass	1660	12820	Earth	953	7598
Animal	477	2699	Tree	2234	19454
Ground	553	1753	ALL	4002	104626

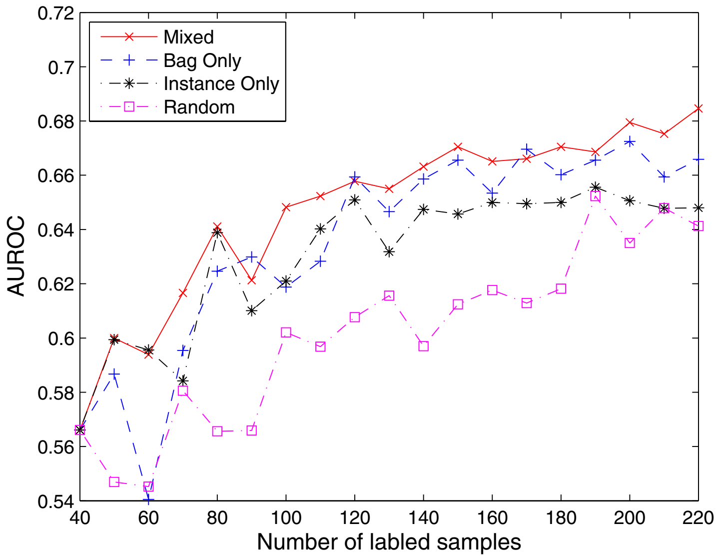

Fig. 2. The Performance of Different Querying Strategies

be observed that MI active learning methods perform significantly better than random sampling on all of the three sample selection strategies, including bags only, instances only, and the mixed bags & instances. This experimental result demonstrates that the proposed MI active learning framework is effective and helps to learn the classifier quickly.

We further conduct experiments to compare the bag & instance mixed selecting strategy with bag only and instance only selecting strategies. The average performances of the mixed selecting strategy, the bag only selecting strategy, and the instance only selecting strategy for MI active learning are shown in Figure 2. It's clear that the performance of the mixed selecting strategy is better than that of the bag only and instance only selecting strategies. We can also observe that instance selecting strategy performs better than bag selecting at the early query rounds. However, as the query rounds proceed as well as more and more samples are added into the training set, the bag selecting strategy shows better performance than instance selecting strategy.

Table 2. Average AUROC Improvement On *Corel*

Query Round	Random	Bag	Instance	Mixed
1	-0.019	+0.021	+0.033	***+0.034***
2	-0.021	-0.026	***+0.030***	+0.028
3	+0.015	+0.029	+0.018	***+0.050***
4	-0.001	+0.058	+0.072	***+0.075***
5	-0.001	***+0.064***	+0.044	+0.055
6	+0.036	+0.052	+0.055	***+0.082***
7	+0.031	+0.062	+0.074	***+0.086***
8	+0.042	***+0.093***	+0.085	+0.092
9	+0.050	+0.080	+0.066	***+0.089***
10	+0.031	+0.092	+0.081	***+0.097***
11	+0.046	+0.099	+0.080	***+0.104***
12	+0.052	+0.087	+0.084	***+0.099***
13	+0.047	***+0.103***	+0.083	+0.100
14	+0.052	+0.094	+0.084	***+0.104***
15	+0.086	+0.100	+0.090	***+0.103***
16	+0.069	+0.106	+0.085	***+0.113***
17	+0.082	+0.093	+0.082	***+0.109***
18	+0.075	+0.100	+0.082	***+0.119***

In Table 2 we summarize the learning curves in Figure 2 by reporting the average AUROC improvement over the initial MI classifier for each querying strategy. The values are averaged across all concepts in *Corel* dataset at each round. The winning strategy at each point is indicated with bold font. It's clear that the mixed selecting strategy improves leaner's performance mostly after a few query rounds, and the instance selecting strategy outperforms bag selecting at early rounds and then lose as query rounds increase.

3.3 Discussion

We can draw some conclusions from our experimental results.

- MI learner's performance can be improved if we select to label certain regions in positive training images. This may owe to the reduction of MI ambiguity in the training images when labels of some regions are provided to MI learner. However, the performance tends to improve gently as more and more regions are selected for labeling. This is reasonable since MI ambiguity cannot be further reduced with only limited training bags after MI learner reaches a certain performance. To further improve learner, more unlabeled bags are needed.
- Bag-only selecting strategy shows terrible performance at the early stage of active samples selection, which may imputes to the MI ambiguity in the training set. As MI ambiguity is solved at some extent after several querying rounds, adding more informative bags into the training set further helps to improve MI learner's generalization performance.

- Bag and instance mixed selecting strategy appears to be the most effective sample selection strategy. It integrates the advantages of both bag selecting and instance selecting strategies. At each query round, both informative instances and bags are selected for labeling. The selected instances help to reduce ambiguity and the selected bags help to improve learner's generalization ability.

4 Conclusion

In this paper, we study multiple-instance active learning with application to image categorization. A MI active learning framework that can query bags, instances or their combination is proposed. Within this framework, a comparative study is conducted on three MI active learning sample selection strategies for image categorization: selecting bags only, selecting instances only, selecting the mixture of bags and instances. Experimental results demonstrate that the mixed sample selection strategy outperforms the other two and appears to be the most effective sample selection strategy for MI active learning in image categorization.The main contribution of this work can be summarized as following:

- By proposing a set kernel based classifier, we build a unified bag and/or instance instance sample selection strategy and an integrated learning algorithm,which make learning with mixed granularities operable.
- We conduct comparative study for MI active learning on image categorization task and prove that the bag and instance mixed sample selection strategy would be the best suitable strategy for MI active learning in image categorization applications.
- We explore the integration of two widely successful learning methods in image categorization,based on which, a reasonable strategy to improve the performance of MI based image categorization is built.

In the future, we plan to test our multiple-instance active learning strategy on other data collections with region labels and also conduct further research to reveal in principle why bag and instance mixture sample selection strategy benefits the multiple-instance classifier most.

References

1. Maron, O., Ratan, L.: Multiple-Instance learning for natural scene classification. In: Proceedings of the 15th International Conference on Machine Learning, pp. 341–349 (1998)
2. Bunescu, R.C., Mooney, R.J.: Multiple Instance Learning for Sparse Positive Bags. In: Proceedings of the 24th International Conference on Machine Learning, pp. 105–112 (2007)
3. Settles, B., Craven, M., Ray, S.: Multiple-Instance Active Learning. In: Advances in Neural Information Processing Systems (NIPS), vol. 20, pp. 1289–1296 (2007)
4. Gärtner, T., Flach, P., Kowalczyk, A., Smola, A.: Multi-Instance Kernels. In: Proceedings of 19th International Conference on Machine Learning, pp. 179–186 (2002)

5. Tong, S., Koller, D.: Support Vector Machine Active Learning with Applications to Text Classification. Journal of Machine Learning Research 2, 45–66 (2001)
6. Tong, S., Chang, E.: Support Vector Machine Active Learning for Image Retrieval. In: Proceedings of the ninth ACM international conference on Multimedia, pp. 107–118 (2004)
7. Brinker, K.: Incorporating Diversity in Active Learning with Support Vector Machines. In: Proceedings of the 20th International Conference on Machine Learning, pp. 59–66 (2003)
8. Cohn, D., Ghahramani, Z., Jordan, M.: Active Learning with Statistical Models. Journal of Artificial Intelligence Research 4, 129–145 (1996)
9. Chang, C., Lin, C.: LIBSVM: a library for support vector machines (2001), `http://www.csie.ntu.edu.tw/~cjlin/libSVM`
10. Yuan, J., Li, J., Zhang, B.: Exploiting Spatial Context Constraints for Automatic Image Region Annotation. In: Proceeding of ACM Multimedia, pp. 595–604 (2007)
11. Deng, Y., Manjunath, B.S.: Unsupervised Segmentation of Color-Texture Regions in Images and Video. IEEE Trans. Pattern Anal. Mach. Intell. 23(8), 800–810 (2001)
12. Fawcett, T.: An introduction to ROC Analysis. Pattern Recognition Letters 27(8), 861–874 (2006)
13. Baram, Y., Yaniv, R.E., Luz, K.: Online Choice of Active Learning Algorithms. Journal of Machine Learning Researh 5, 255–291 (2004)
14. Taylor, J.S., Christianini, N.: On the Generalization of Soft Margin Algorithms. IEEE Transaction on Information Theory 48(10), 2721–2735 (2002)