

# Multi-Instance Multi-Label Active Learning\*

Sheng-Jun Huang and Nengneng Gao and Songcan Chen

College of Computer Science & Technology, Nanjing University of Aeronautics & Astronautics  
 Collaborative Innovation Center of Novel Software Technology and Industrialization  
 {huangsj, gaonn, s.chen}@nuaa.edu.cn

## Abstract

Multi-instance multi-label learning (MIML) has achieved success in various applications, especially those involving complicated learning objects. Along with the enhancing of expressive power, the cost of annotating a MIML example also increases significantly. In this paper, we propose a novel active learning approach to reduce the labeling cost of MIML. The approach actively query the most valuable information by exploiting diversity and uncertainty in both the input and output spaces. It designs a novel query strategy for MIML objects specifically and acquires more precise information from the oracle without additional cost. Based on the queried information, the MIML model is then effectively trained by simultaneously optimizing the relevance rank among instances and labels. Experiments on benchmark datasets demonstrate that the proposed approach achieves superior performance on various criteria.

## 1 Introduction

In traditional supervised learning, an object is represented by one instance, and associated with one class label. However, in many real applications, such a learning framework is less effective to model the complicated objects. For example, a scene image may be simultaneously relevant to mountain, lake, trees, etc. If we simply extract one instance to represent it, some useful information may get lost. An alternative approach is to segment the image into multiple regions with relative clear semantics, and extract one instance from each region. Or in text categorization tasks, an article may be annotated with multiple labels. To fully exploit the content with multiple topics, it would be more effective if we represent each paragraph with one instance rather than extract only one instance from the whole article. To deal with such complicated objects, multi-instance multi-label learning was proposed in [Zhou and Zhang, 2006].

In multi-instance multi-label learning, every example is represented with a bag of multiple instances, and annotated

with multiple class labels to express its semantics. MIML has been successfully applied to various tasks, including image classification [Zhou and Zhang, 2006; Zhang and Zhou, 2008; Feng and Xu, 2010], text categorization [Zhang and Zhou, 2008], gene function prediction [Li *et al.*, 2012], relationship extraction [Surdeanu *et al.*, 2012] and video understanding [Xu *et al.*, 2011], etc.

It is well known that in supervised learning, sufficient labeled training examples are needed to get an effective model. However, in real cases, it is usually difficult and costly to manually label data examples. It thus becomes a very important task to train a strong model with as few labeled data as possible. Active learning is a main approach to overcome this challenge. It actively selects the most important instances and query their labels from the oracle, and expects to reduce the number of labeled examples required.

MIML provides a better framework for learning with complicated objects, yet its input and output spaces increase dramatically, make it rather difficult to train an effective model, which further implies that more training data are needed. On the other hand, because there are a large number of candidate labels in MIML, it becomes much more costly to annotated an example comparing to single-label learning. So active learning for MIML is highly desired to reduce the labeling cost.

Existing studies on active learning mainly focus on the single-instance single-label setting [Nguyen and Smeulders, 2004; Balcan *et al.*, 2007; Huang *et al.*, 2014b; Lin *et al.*, 2016]. Recently, there are a few efforts on designing active query strategies for multi-instance learning [Salmani and Sridharan, 2014] or multi-label learning [Li and Guo, 2013; Huang *et al.*, 2015]. However, these methods do not simultaneously exploit the information embedded in input and output space, and cannot directly applied to MIML setting. The method in [Retz and Schwenker, 2016] simply transforms MIML problem to single-instance representation and then directly employs traditional active learning method for label querying.

In this paper, we propose a multi-instance multi-label active learning approach, which can acquire the most helpful information at low cost with novel strategies for both selection and querying. Specifically, at each iteration of active learning, we firstly select the most valuable bag-label pair based on the diversity and uncertainty, and then pertinently query supervised information from the oracle. If the selected bag

\*This research was supported by JiangsuSF (BK20150754), NSFC (61503182), YESS and China Postdoctoral Science Foundation.

and label are relevant, then the oracle is required to also indicate the key instance among the bag that is most relevant to the label. In this way, the model can get more precise supervision. Note that the oracle need to identify the key instance when judging whether the bag and label are relevant or not; and thus this process does not bring additional cost. To utilize the queried information, we then proposed a MIML algorithm to optimize the relative rank of both instances and labels. For a bag, on one hand, the relevant labels will be ranked before irrelevant ones, and on the other hand, the key instance of a label will be ranked before other instances. Incorporating the active query mechanism with the joint rank optimization, we can query and fully utilize the most valuable supervised information to improve the learning performance. Experiments on benchmark datasets validated the effectiveness of the proposed approach.

The rest of this paper is organized as follows. Section 2 reviews related work. In Section 3, the proposed approach is introduced. Section 4 presents the experiments, followed by the conclusion in Section 5.

## 2 Related Work

Multi-instance multi-label learning has attracted much research attention due to its success in various applications [Zhou and Zhang, 2006; Zhou *et al.*, 2012]. The MIMLfast algorithm proposed in [Huang *et al.*, 2014a] optimizes the relevance ordering of labels for each training bag, and shows promising results on both effectiveness and efficiency. The model training part of this paper is similar to this method, however, our method can exploit the key instance information, and simultaneously optimizes the rank for both instances and labels, which is beyond the capacity of MIMLfast.

Active learning selectively queries the most valuable information from the oracle and aims to train an effective model with least queries. The key task in active learning is to design a proper strategy such that the queried information is most helpful for improving the learning model. There have been many active learning methods proposed under traditional setting [Settles, 2009]. Some of them prefer to query labels for the most informative instances [Balcan *et al.*, 2007], while informativeness can be estimated with different criteria, such as uncertainty, expected error reduction, etc. Some other methods prefer to query labels for representative instances [Nguyen and Smeulders, 2004], where representativeness can be estimated based on clustering structure or density. Recently there are some studies that try to consider both informativeness and representativeness for query selection, and achieved significant performance improvement [Huang *et al.*, 2014b].

While most active learning researches focus on traditional setting, there are a few works to extend the ideas to multi-instance [Settles *et al.*, 2007; Zhang *et al.*, 2010; Salmani and Sridharan, 2014] or multi-label learning [Li *et al.*, 2004; Hung and Lin, 2011; Tang *et al.*, 2012; Li and Guo, 2013; Wu *et al.*, 2014]. The method proposed in [Huang and Zhou, 2013] exploits both uncertainty and diversity to select the most valuable instance and label. This idea inspires our criterion for bag-label selection in this paper. However, all these

studies are focusing on either multi-instance or multi-label learning, and cannot be directly applied to MIML setting. The method in [Retz and Schwenker, 2016] is specifically designed based on MIMLSVM. It firstly degenerates the bags to single-instance representation and then directly employ traditional active learning method for label querying, which does not truly exploit the characteristics of MIML tasks.

## 3 The Method

In active learning, we usually have a small set of initial labeled data  $D_l$  and a large set of unlabeled examples  $D_u$ . The algorithm iteratively query the label of one or several selected example in  $D_u$ , and add them into  $D_l$  to update the model. This process is repeated until the performance is satisfied or the labeling cost reaches a predefined budget. We denote by  $(X_i, Y_i)$  a labeled MIML example, where  $X_i = \{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{im_i}\}$  is the  $i$ -th bag that consists of  $m_i$  instances, and each instance  $x_{ij}$  is a  $d$ -dimensional feature vector.  $Y_i = [y_{i1}, y_{i2}, \dots, y_{iK}]^T$  is the label vector for  $X_i$ ,  $y_{ik} = 1$  if the bag  $X_i$  is relevant to the  $k$ -th label, and  $y_{ik} = -1$  otherwise. We denote by  $U(X)$  the set of labels that have not been queried for the bag  $X$ . A bag  $X \in D_u$  if and only if  $|U(X)| > 0$ .

In multi-label learning, it has been validated that query one label rather than all labels of one instance at each time is more effective [Huang *et al.*, 2015]. Following this query type, one can easily adapt it to MIML setting by selecting a bag-label pair and querying whether they are relevant at each iteration of active learning. However, such a query type may lead to waste of labeling cost for MIML tasks. For example, when we query the oracle whether an image is relevant to the label 'dog', the oracle will of course identify the region corresponding to dog in the image before feedback. That means without further cost, we can ask the oracle to feedback more precise information in addition to bag-label relevance, i.e., indicate the key instance that trigger the queried label. Based on this observation, we propose a novel query type for MIML active learning.

We will introduce the active learning algorithm with two steps. In the first step, we design a criterion to select the most valuable bag-label pair; and in the second step, the oracle decides the relevance of the pair and give pertinent feedback. Inspired by the multi-label active learning method proposed in [Huang and Zhou, 2013], we consider both diversity and uncertainty to select the most valuable bag-label pair. In detail, we define a measure as follows to estimate the diversity and uncertainty of a bag  $X_i$ :

$$g(X_i) = \frac{\left| \sum_{k=1}^K I[\hat{y}_{ik} > 0] - \frac{1}{N_l} \sum_{j=1}^{N_l} \sum_{k=1}^K I[y_{jk} > 0] \right|}{\max\{\xi, K - \text{card}(U(X_i))\}}, \quad (1)$$

where  $\hat{y}_{ik}$  is the prediction on the  $k$ -th label for bag  $X_i$ ,  $N_l$  is the number of fully labeled bags,  $\xi \in (0, 1)$  is a constant to avoid the zero divisor,  $\text{card}(\cdot)$  calculates the set size and  $I[\cdot]$  is the indicator function which returns 1 if the argument is true and 0 otherwise. The numerator measures the inconsistency between the number of predicted positive labels on  $X_i$  and the average number of positive labels on the labeled set. A large

inconsistency implies that the prediction on this bag may be less reliable, or in other words, the model is less uncertain on this bag. The denominator counts how many labels have been queried for the bag. Obviously, bags with smaller value of denominator are preferred because we expect the queries cover diverse bags rather than densely located on a few bags. In summary, we will select the bag  $X^*$  with maximum value of  $g(X^*)$ .

After deciding the selected bag  $X^*$ , we then employ the measure in Eq. 2 to estimate the informativeness of a label  $y$  with regard to  $X^*$ :

$$h(X^*, y) = |f_y(X^*) - f_{y_0}(X^*)|, \quad (2)$$

where  $f_y$  is the prediction function for label  $y$ , and  $y_0$  is a dummy label to separate relevant and irrelevant labels in a ranked label list. So  $h(X^*, y)$  is estimating how close the prediction on  $y$  is to the decision boundary. It is usually assumed that a prediction close to the decision boundary is more uncertain. So we select the most uncertain label  $y^*$  with smallest value of  $h(X^*, y^*)$ .

Next, for the selected bag-label pair  $(X^*, y^*)$ , the oracle decides whether they are relevant. If the answer is NO, then the oracle just returns to the algorithm that  $y^*$  is a negative label of bag  $X^*$ . And if they are relevant, then the oracle also tells which instance triggers the label  $y^*$ . We call this instance as key instance, and denote it by  $\mathbf{x}^*$ . In other words, among the instances in bag  $X^*$ , the key instance  $\mathbf{x}^*$  is most relevant to label  $y^*$ .

With such a query type, the learning system can acquire more precise supervised information. Unfortunately, there is no existing MIML algorithm can directly utilize such supervision. Next, we will propose a new MIML algorithm to exploit the information at both bag and instance level. In [Huang *et al.*, 2014a], the MIML classification problem is transformed into a label ranking problem by minimizing a rank loss defined on labels. Inspired by this method, we extend the rank loss to exploit the relevance order in both the input and output space. For a bag, on one hand, we minimize the rank loss on labels to rank all relevant labels before irrelevant ones; and on the other hand, we minimize the rank loss on instances to rank the key instance before others with regard to the corresponding label. Formally, as in [Huang *et al.*, 2014a], we define the prediction function of the  $k$ -th label on instance  $\mathbf{x}$  as:

$$f_k(\mathbf{x}) = \mathbf{w}_k^\top W_0 \mathbf{x}, \quad (3)$$

where  $W_0$  is a  $b \times d$  matrix that maps the original  $d$ -dimensional feature vector to a shared space, and  $w_k$  is the linear prediction function for label  $y_k$ . Based on the classical definition of multi-instance learning that a bag is positive if it contains at least one positive instance, we can easily get the prediction on a bag by picking the maximum prediction among its instances as follows:

$$f_k(X) = \max_{\mathbf{x} \in X} f_k(\mathbf{x}). \quad (4)$$

As in [Huang *et al.*, 2014a], the rank loss in the label space with regard to bag  $X$  and its relevant label  $y$  can be defined as:

$$\epsilon(X, y) = \sum_{i=1}^{R(X, y)} \frac{1}{i}, \quad (5)$$

where  $R(X, y)$  counts how many irrelevant labels are ranked before  $y$  based on the prediction values, i.e.,

$$R(X, y) = \sum_{\bar{y} \in \bar{Y}} I[f_{\bar{y}}(X) > f_y(X) - 1]. \quad (6)$$

Note here  $\bar{Y}$  is the set of all irrelevant labels of  $X$ , and we penalize  $R(X, y)$  with a margin 1.

Similarly, we can define the rank loss in the instance space as:

$$\delta(X, \mathbf{x}^*, y) = \sum_{i=1}^{P(X, \mathbf{x}^*, y)} \frac{1}{i}, \quad (7)$$

where  $P(X, \mathbf{x}^*, y)$  counts the number of instances that are ranked before the key instance  $\mathbf{x}^*$  based on the predictions on  $y$ , i.e.,

$$P(X, \mathbf{x}^*, y) = \sum_{\mathbf{x} \in X} I[f_y(\mathbf{x}) > f_y(\mathbf{x}^*) - 1], \quad (8)$$

To train an effective model, our target is to minimize  $\epsilon(X, y)$  and  $\delta(X, \mathbf{x}^*, y)$  on the whole training data, such that all labels and instances can be correctly ranked according to their relevance order. We employ stochastic gradient descent (SGD) to minimize the loss functions. Specifically, assuming the model variables at the  $t$ -th iteration of SGD are  $W_0^t$ ,  $w_k^t$  ( $k = 1 \cdots K$ ), and the current bag is  $X$ , while  $y$  and  $\bar{y}$  is one relevant and one irrelevant label of  $X$ , respectively. If the current model ranks  $\bar{y}$  before  $y$ , then it will induce a loss on the triplet  $(X, y, \bar{y})$ :

$$\mathcal{L}(X, y, \bar{y}) = \epsilon(X, y) |1 + f_{\bar{y}}(X) - f_y(X)|_+, \quad (9)$$

where  $|q|_+ = \max\{q, 0\}$ . Similarly, if the current model does not rank  $\mathbf{x}^*$  as the most relevant instance, we randomly sample one instance  $\mathbf{x}'$  before  $\mathbf{x}^*$ , and it will induce the loss:

$$\mathcal{L}'(\mathbf{x}^*, \mathbf{x}', y) = \delta(X, \mathbf{x}^*, y) |1 + f_y(\mathbf{x}') - f_y(\mathbf{x}^*)|_+, \quad (10)$$

Then, to minimize Eqs. 9 and 10, we update the model variables as follows:

$$W_0^{t+1} = W_0^t - \gamma_t \sum_{i=1}^{R(X^*, y^*)} \frac{1}{i} (\mathbf{w}_{\bar{y}}^t \mathbf{x}^{*\top} - \mathbf{w}_y^t \mathbf{x}^{*\top}) - \gamma_t \sum_{i=1}^{P(X^*, \mathbf{x}^*, y)} \frac{1}{i} (\mathbf{w}_{\bar{y}}^t \mathbf{x}'^\top - \mathbf{w}_y^t \mathbf{x}^{*\top}), \quad (11)$$

$$\mathbf{w}_y^{t+1} = \mathbf{w}_y^t + \gamma_t \sum_{i=1}^{R(X^*, y^*)} \frac{1}{i} W_0^t \mathbf{x}^* - \gamma_t \sum_{i=1}^{P(X^*, \mathbf{x}^*, y)} \frac{1}{i} W_0^t (\mathbf{x}' - \mathbf{x}^*), \quad (12)$$

$$\mathbf{w}_{\bar{y}}^{t+1} = \mathbf{w}_{\bar{y}}^t - \gamma_t \sum_{i=1}^{R(X^*, y^*)} \frac{1}{i} W_0^t \mathbf{x}^*. \quad (13)$$

In the above formulations,  $\gamma_t$  is the learning rate. Note that after updating  $W_0$ ,  $\mathbf{w}_y$  and  $\mathbf{w}_{\bar{y}}$ , each column of them is normalized to ensure that their  $\ell_2$  norms will be upper bounded by a constant  $C$ .

---

**Algorithm 1** The MIML-AL Algorithm

---

```

1: Input:
2:    $D_l$ : a small set of initially labeled examples;
3:    $D_u$ : the pool of unlabeled data for active selection;
4: Initialize:
5:   Train a initial MIML model  $f$  on  $D_l$ ;
6: Repeat:
7:   Make predictions with  $f$  for bags in  $D_u$ ;
8:   Calculate  $g(X)$  for all bags  $X \in D_u$ ;
9:   Select the bag  $X^* = \operatorname{argmax} g(X)$ ;
10:  Calculate  $h(X^*, y)$  for all labels  $y \in U(X^*)$ ;
11:  Select the label  $y^* = \operatorname{argmin} h(X^*, y)$ ;
12:  Query whether  $X^*$  is relevant to  $y^*$ ;
13:  Feedback with the key instance  $\mathbf{x}^*$  if relevant;
14:  Update the model  $f$  according to Eqs. 11 to 13;
15:  Remove  $y^*$  from  $U(X^*)$ ;
16:  Move  $X^*$  from  $D_u$  to  $D_l$  if  $|U(X^*)| = 0$ ;
17: Until stop criterion reached.

```

---

The pseudo-code of the proposed approach is summarized in Algorithm 1. At first, a MIML model is trained on the initially labeled set  $D_l$ , then the model is used to make predictions for all bags in  $D_u$ . Based on the predictions, the criterion  $g(X)$  is calculated according to Eq. 1, and the bag  $X^*$  with maximum value of  $g$  is selected. Then the criterion  $h(X^*, y)$  is calculated for each  $y \in U(X^*)$  according to Eq. 2, and the label  $y^*$  with minimal value of  $h$  is selected. After that, the relevance of  $X^*$  and  $y^*$  is queried. If they are relevant, the key instance is also indicated by the oracle, and the model is updated according to Eqs. 11 to 13. At last,  $y^*$  will be marked as a queried label of  $X^*$ , and  $X^*$  will be moved from  $D_u$  to  $D_l$  if it is already fully labeled. These steps are repeated until certain stop criterion reached, for example, the performance is good enough or the labeling cost reaches a predefined budget.

## 4 Experiments

To examine the effectiveness of the proposed method, the key instance information should be available to simulate the oracle. Among the public available MIML datasets, there are four datasets, i.e., *MSRC* [Winn *et al.*, 2005], *Letter Frost*, *Letter Carroll* and *Bird Song* [Briggs *et al.*, 2012] where the labels of instances are available. We thus perform the experiments on these datasets. There are in all 26 candidate labels for *Letter Frost* and *Letter Carroll*, while 23 labels for *MSRC* and 13 labels for *Bird Song*. On average, each bag has 2.5/3.6/3.9/2.1 relevant labels on the four datasets, respectively.

For each dataset, we randomly sample 20% of bags as the test data, and the rest as the unlabeled pool for active selection. And at the beginning, 5% of the unlabeled data will be randomly selected and initially labeled. This small labeled set is used to train the initial MIML model. After each query, we update the model and examine the performance on the test set. We repeat the random data split for 30 times and report the average results.

In the experiments, the following methods are compared:

- MIML-AL: the proposed approach;
- MIML-ABRI: which is a degenerated version of our method. The difference is that the oracle randomly feed-back an instance as the key instance when the queried bag and label are relevant.
- MIML-RBAI: which is another degenerated version of our method. The difference is that this method randomly selects the bag and label for querying, and the key instance will be given if the bag is relevant to the label.
- MIML-AUDI: which adapts the method in [Huang and Zhou, 2013], which is a state-of-the-art multi-label active learning method, to select bag-label pairs.
- MIML-TMLA: which is the method proposed in [Retz and Schwenker, 2016], which applies multi-label active learning strategy after transform the MIML examples to single-instance representations.

We compare with MIML-ABRI to examine the effectiveness of utilizing the precise supervised information; and compare with MIML-RBAI to examine the effectiveness of the bag-label selection strategy. Note that after querying, the MIML model is trained with the same method, although some of them does not have the precise information. For MIML-AL, we fix the parameters  $b = 200$ ,  $C = 10$  for all datasets.

We evaluate the performance on six measures: hamming loss, one error, average precision, micro-precision, micro-recall and micro-F1, which are commonly used criteria in MIML [Zhou *et al.*, 2012]. For hamming loss and one error, a smaller value indicates a better performance, while for the other four criteria, the larger the better.

Figures 1 to 4 plot the performance curves as the number of queries increases on *MSRC*, *Letter Frost*, *Letter Carroll* and *Bird Song* respectively. The red line represents the proposed approach. It is shown that the proposed method MIML-AL is superior to other methods on different measure.

We observe that the performances of MIML-RBAI and MIML-TMLA are relatively poor. For MIML-RBAI, the random selection of bag-label pairs may lead less informative queries. For MIML-TMLA, the result is probably caused by the query type it used. As disclosed in [Huang *et al.*, 2015], it is usually more effective to query one label for an example rather than query all the labels at a time. Because the labels are usually correlated with each other and thus some redundant information may queried by MIML-TMLA.

When comparing MIML-AL with MIML-ABRI, the proposed method is always better on all measures and all datasets. This fully validates that the MIML model can be improved by utilizing the key instance information to optimize instance rank. MIML-AL also beats MIML-RBAI for most cases, which validates that the bag-label selection strategy based on diversity and uncertainty is effective. At last, the superiority of MIML-AL over MIML-AUDI tells that the key instance bring some helpful and precise information without additional cost. These observations demonstrates that our method does query more helpful information from the oracle and fully utilize them to improve the learning performance.

On *Letter Frost* and *Letter Carroll* datasets, we notice that at the beginning stage of active learning, MIML-RBAI and

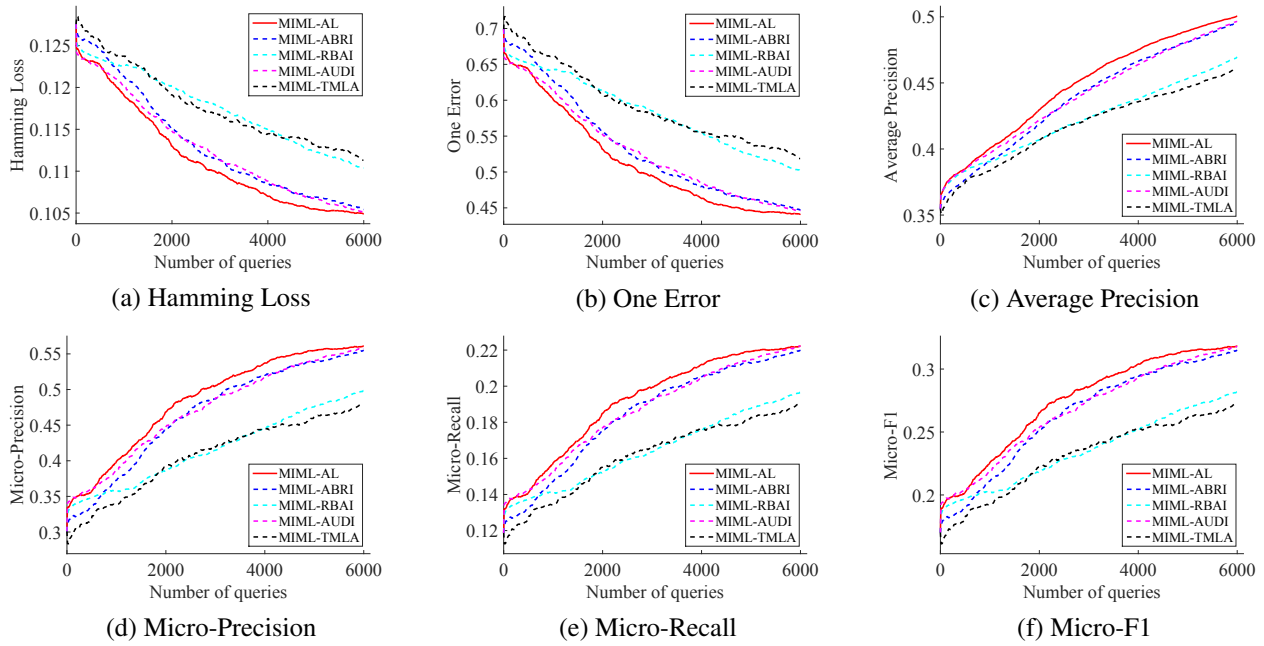


Figure 1: Performance comparison on MSRC

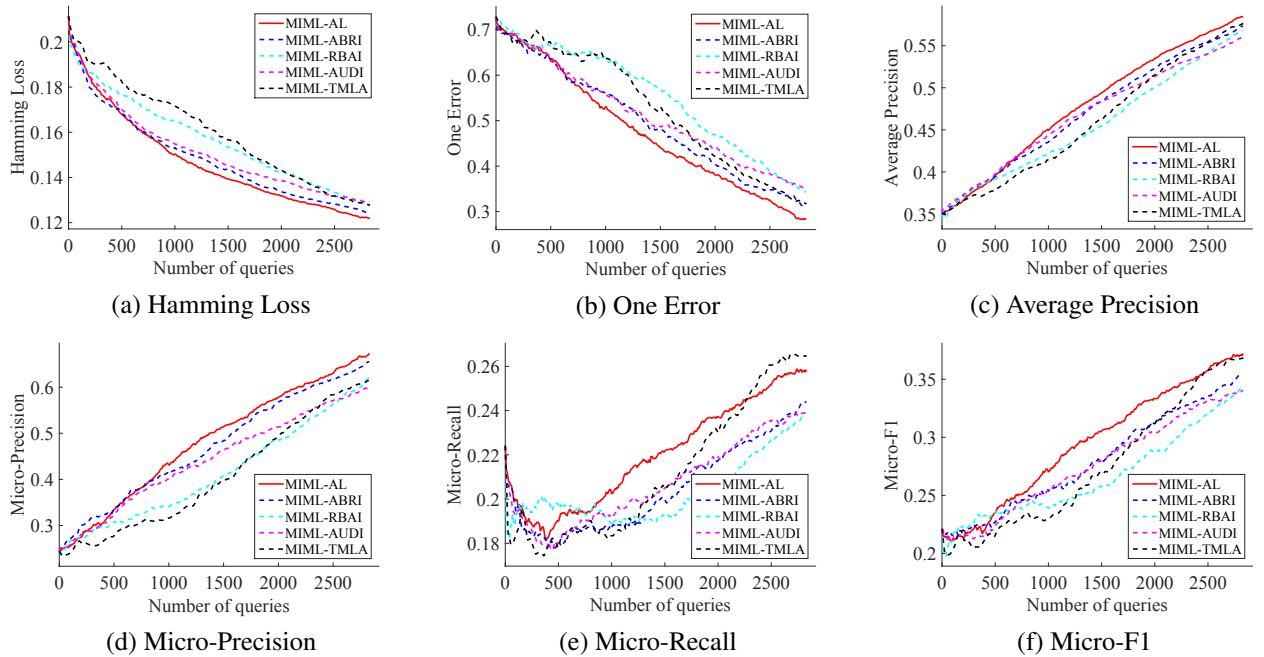


Figure 2: Performance comparison on Letter Frost

MIL-TMLA achieve better performance on Micro-Recall, while their performance on other measures are worse. One possible reason is that at the beginning stage, the model is less reliable, and these two approaches may predict too much positive labels. This can be also implied from their poor performance on hamming loss and one error.

## 5 Conclusion

To reduce the labeling cost of MIML tasks, we propose an active learning approach with novel contributions on bag-label pair selection, query type as well as model training. Specifically, to select the most valuable bag-label pairs, diversity and uncertainty in both the input and output spaces are si-

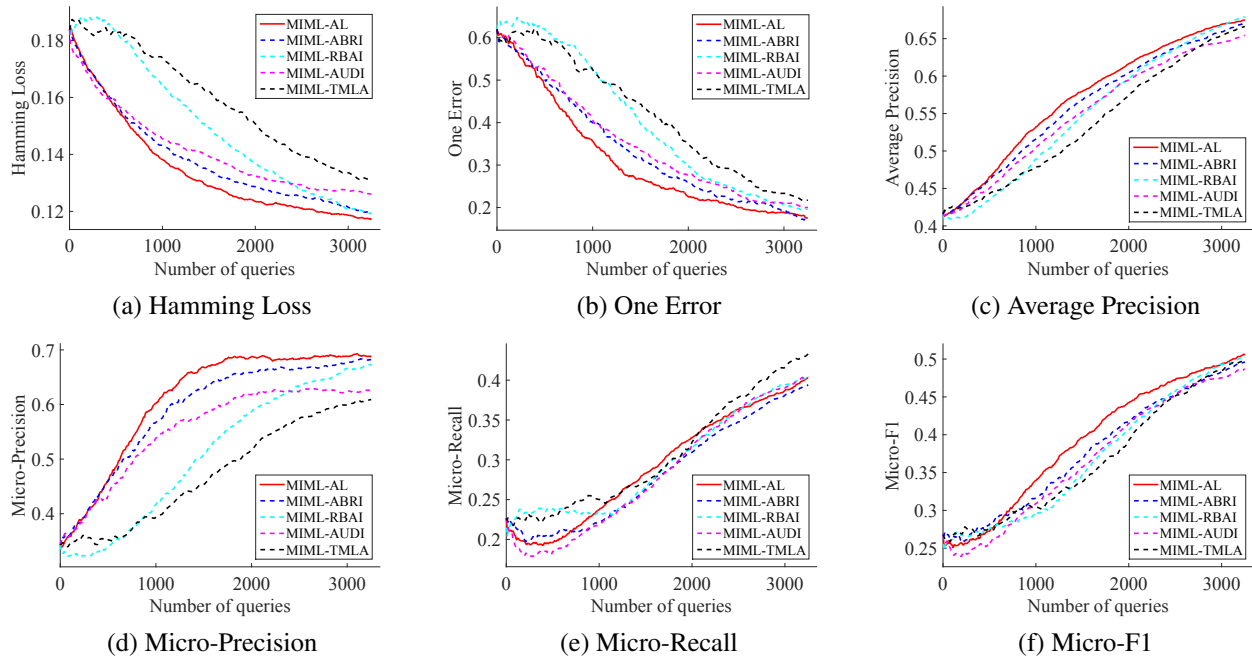


Figure 3: Performance comparison on Letter Carroll

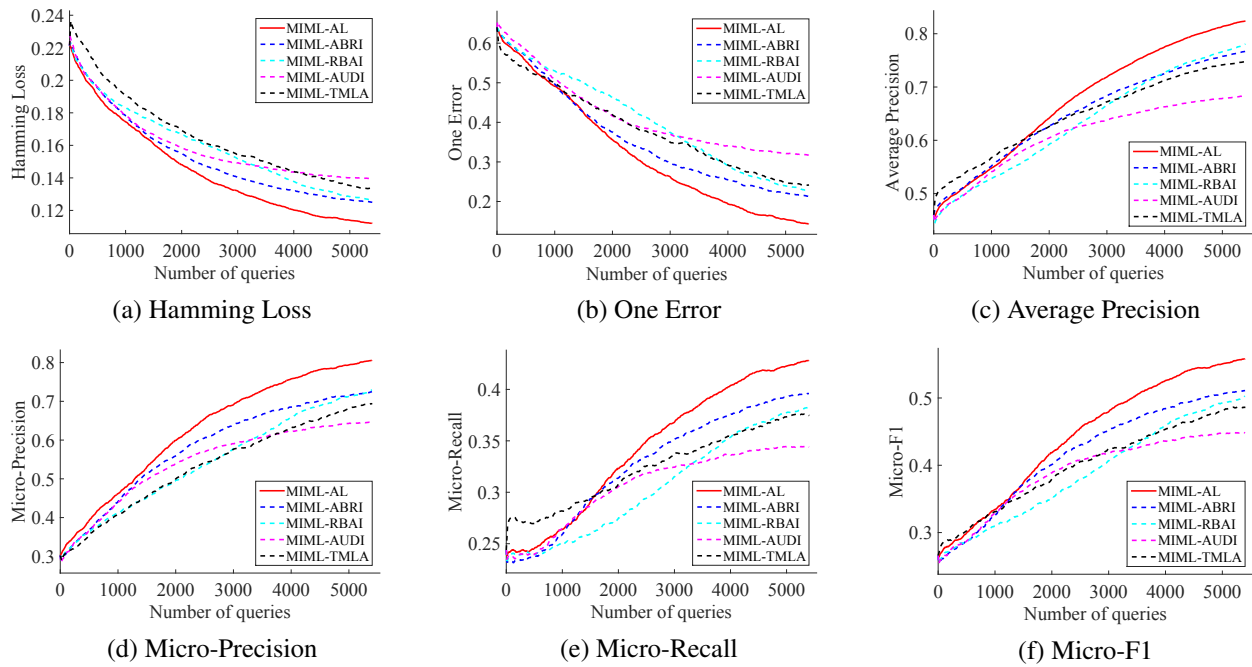


Figure 4: Performance comparison on Bird Song

multaneously considered; to query more precise information, the key instance is identified without additional cost when the bag and label is relevant; and to train the MIML model effectively, the queried information is fully exploited to optimize the relevance rank between both instances and labels. Experi-

ments on benchmark datasets show that the proposed method achieves superior performance on various criteria. In the future, we plan to test our method on more datasets and design other strategies for bag-label selection.

## References

- [Balcan *et al.*, 2007] M.-F. Balcan, A. Z. Broder, and T. Zhang. Margin based active learning. In *Proceedings of the 20th Annual Conference on Learning Theory*, pages 35–50, 2007.
- [Briggs *et al.*, 2012] F. Briggs, X. Z. Fern, and R. Raich. Rank-loss support instance machines for MIML instance annotation. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 534–542, 2012.
- [Feng and Xu, 2010] S. Feng and D. Xu. Transductive multi-instance multi-label learning algorithm with application to automatic image annotation. *Expert Systems with Applications*, 37(1):661–670, 2010.
- [Huang and Zhou, 2013] S.-J. Huang and Z.-H. Zhou. Active query driven by uncertainty and diversity for incremental multi-label learning. In *Proceedings of the 13th IEEE International Conference on Data Mining*, pages 1079–1084, 2013.
- [Huang *et al.*, 2014a] S.-J. Huang, W. Gao, and Z.-H. Zhou. Fast multi-instance multi-label learning. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, pages 1868–1874, 2014.
- [Huang *et al.*, 2014b] S.-J. Huang, R. Jin, and Z.-H. Zhou. Active learning by querying informative and representative examples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (10):1936–1949, 2014.
- [Huang *et al.*, 2015] S.-J. Huang, S. Chen, and Z.-H. Zhou. Multi-label active learning: query type matters. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, pages 946–952, 2015.
- [Hung and Lin, 2011] C.-W. Hung and H.-T. Lin. Multi-label active learning with auxiliary learner. In *Proceedings of the 3rd Asian Conference on Machine Learning*, pages 315–330, 2011.
- [Li and Guo, 2013] X. Li and Y. Guo. Active learning with multi-label svm classification. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, pages 1479–1485, 2013.
- [Li *et al.*, 2004] X. Li, L. Wang, and E. Sung. Multi-label SVM active learning for image classification. In *Proceedings of the 2004 International Conference on Image Processing*, pages 2207–2210, 2004.
- [Li *et al.*, 2012] Y.-X. Li, S. Ji, S. Kumar, J. Ye, and Z.-H. Zhou. Drosophila gene expression pattern annotation through multi-instance multi-label learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9(1):98–112, 2012.
- [Lin *et al.*, 2016] C. H. Lin, M. Mausam, and D. S. Weld. Re-active learning: Active learning with relabeling. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, pages 1845–1852, 2016.
- [Nguyen and Smeulders, 2004] H. T. Nguyen and A. W. M. Smeulders. Active learning using pre-clustering. In *Proceedings of the 21st International Conference on Machine Learning*, 2004.
- [Retz and Schwenker, 2016] R. Retz and F. Schwenker. Active multi-instance multi-label learning. In *Proceedings of the 2nd European Conference on Data Analysis*, pages 91–101, 2016.
- [Salmani and Sridharan, 2014] K. Salmani and M. Sridharan. Multi-instance active learning with online labeling for object recognition. In *Proceedings of the 27th International Flairs Conference*, 2014.
- [Settles *et al.*, 2007] B. Settles, M. Craven, and S. Ray. Multiple-instance active learning. In *Advances in Neural Information Processing Systems 20*, pages 1289–1296, 2007.
- [Settles, 2009] B. Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin-Madison, 2009.
- [Surdeanu *et al.*, 2012] M. Surdeanu, J. Tibshirani, R. Nallapati, and C. D. Manning. Multi-instance multi-label learning for relation extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 455–465, 2012.
- [Tang *et al.*, 2012] J. Tang, Z.-J. Zha, D. Tao, and T.-S. Chua. Semantic-gap-oriented active learning for multilabel image annotation. *IEEE Transactions on Image Processing*, 21(4):2354–2360, 2012.
- [Winn *et al.*, 2005] J. M. Winn, A. Criminisi, and T. P. Minka. Object categorization by learned universal visual dictionary. In *Proceedings of the 10th IEEE International Conference on Computer Vision*, pages 1800–1807, 2005.
- [Wu *et al.*, 2014] J. Wu, V. Sheng, J. Zhang, P. Zhao, and Z. Cui. Multi-label active learning for image classification. In *Proceedings of IEEE International Conference on Image Processing*, pages 5227–5231, 2014.
- [Xu *et al.*, 2011] X.-S. Xu, X. Xue, and Z.-H. Zhou. Ensemble multi-instance multi-label learning approach for video annotation task. In *Proceedings of the 19th International Conference on Multimedia*, pages 1153–1156, 2011.
- [Zhang and Zhou, 2008] M.-L. Zhang and Z.-H. Zhou. M3MIML: A maximum margin method for multi-instance multi-label learning. In *Proceedings of the 8th IEEE International Conference on Data Mining*, pages 688–697, 2008.
- [Zhang *et al.*, 2010] D. Zhang, F. Wang, Z. Shi, and C. Zhang. Interactive localized content based image retrieval with multiple-instance active learning. *Pattern Recognition*, 43(2):478–484, 2010.
- [Zhou and Zhang, 2006] Z.-H. Zhou and M.-L. Zhang. Multi-instance multi-label learning with application to scene classification. In *Advances in Neural Information Processing Systems 19*, pages 1609–1616, 2006.
- [Zhou *et al.*, 2012] Z.-H. Zhou, M.-L. Zhang, S.-J. Huang, and Y.-F. Li. Multi-instance multi-label learning. *Artificial Intelligence*, 176(1):2291–2320, 2012.