Weak-Labeled Active Learning With Conditional Label Dependence for Multilabel Image Classification

Jian Wu, Shiquan Zhao, Victor S. Sheng, *Senior Member, IEEE*, Jing Zhang, Chen Ye, Pengpeng Zhao, and Zhiming Cui

***Abstract*—Multilabel image classification has been a hot topic in the field of computer vision and image understanding in recent years. To achieve better classification performance with fewer labeled images, multilabel active learning is used for this scenario. Several active learning methods have been proposed for multilabel image classification. However, all of them assume that either all training images have complete labels or label correlations are given at the beginning. These two assumptions are unrealistic. In fact, it is very difficult to obtain complete labels for each example, in particular when the size of labels in a multilabel dataset is very large. Typically, only partial labels are available. This is one type of "weak label" problem. To solve this weak label problem inside multilabel active learning, this paper proposes a novel solution called AE-WLMAL. AE-WLMAL explores conditional label correlations on the weak label problem with the help of input features and then utilizes label correlations to construct a unified sampling strategy and evaluate the informativeness of each example-label pair in a multilabel dataset for active sampling. In addition, a pruning strategy is adopted to further improve its computation efficiency. Moreover, AE-WLAML exploits label correlations to infer labels for unlabeled images, which further reduces human labeling cost. Our experimental results on seven real-world datasets show that AE-WLMAL consistently outperforms existing approaches.**

***Index Terms*—Multilabel active learning, weak label, image classification, label correlation, conditional label dependence.**

I. Introduction

ACTIVE learning is a machine learning technique that iteratively selects the most informative examples to query labels from oracle to optimize the performance of a learned model. It has been widely used in the multimedia research community to reduce human labeling effort. A general introduction and survey of the most recent developments applying active learning in the multimedia research community can be found in [1]. It considers two application domains: image and video annotation [2], [3] and content-based image retrieval [4], [5]. In this article, we focus on image classification for building an image classifier with a good generalization ability.

In traditional image classification, each image can only be classified into one class, which is well-known as single label classification, including binary and multi-class classification [6], [7]. However, in the real world, images are always simultaneously associated with multiple labels according to their contents [8]–[12]. It is known that in order to train a good model for multi-label image classification, we should have enough images with high-quality labels. However, it is not unusual that we do not have enough such images in many real-world applications [13]. Furthermore, the cost of acquiring high-quality labels for images is very high, in particular when images are associated with a great deal of labels [14], [15]. For the sake of reducing the labeling effort and the corresponding cost, active learning methods have been widely used in this field [16]. The main purpose of active learning is to obtain the best model with the least labeling effort and cost. There are several studies on this topic [17], [18]–[25].

However, all of the previous studies on multi-label active learning ignore a grim fact: Not all labels are available [26], [27]. For example, if a label space for an image training set contains *sunset*, *beach*, *urban* and *field*, annotators should provide the labels on *sunset*, *beach*, *urban* and *field* for all images. However, in many cases, it is difficult to provide all labels. Thus, only a portion of labels are usually provided. As can be seen from Fig. 1, many images are only labeled with partial labels. In such a scenario, the training set contains many *weak labeled images* [27]–[29]. This paper focuses on the *weak label problem* in multi-label active learning for image classification.

As we know, if all labels are independent in a multi-label classification application, a multi-label problem can be solved easily by decomposing into a series of binary classification problems. However, it is common for the labels in a multi-label problem to be correlated. Such a decomposition ignores the correlations among labels. In recent years, researchers have realized the

Manuscript received September 3, 2015; revised January 3, 2016, July 15, 2016, and December 28, 2016; accepted January 7, 2017. Date of publication January 11, 2017; date of current version May 13, 2017. This work was supported in part by the Natural Science Foundation of China under Grant 61402311 and Grant 61603186, in part by the Jiangsu Province Colleges and Universities Natural Science Research Project under Grant 13KJB520021, and in part by the U.S. National Science Foundation under Grant IIS-1115417. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Winston Hsu.

J. Wu, S. Zhao, C. Ye, P. Zhao, and Z. Cui are with the Institute of Intelligent Information Processing and Application, Soochow University, Suzhou 215006, China (e-mail: jianwu@suda.edu.cn; 20134227054@stu.suda.edu.cn; 20134227041@stu.suda.edu.cn; ppzhao@suda.edu.cn; szzmcui@suda.edu.cn).

V. S. Sheng is with the Department of Computer Science, University of Central Arkansas, Conway, AR 72035 USA (e-mail: ssheng@uca.edu).

J. Zhang is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: jingzhang.cs@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2017.2652065

images	*sunset*	*beach*	*urban*	*field*
x1	1	0	?	?
x2	1	?	0	1
x3	1	0	1	0
x4	?	?	0	1
x5	0	1	?	0
x6	?	1	?	?

Fig. 1. Example of the weak label problem.

importance of label correlations in multi-label classification. Research results show that label correlations play a significant role in improving the performance of multi-label classifiers [30]–[33]. In this paper, we study how to explore label correlations in a weak label setting and construct a unified sampling strategy based on the captured label correlations for active sampling. Specifically, we propose a novel solution called AE-WLMAL, which designs an effective sampling strategy under a weak label scenario for multi-label active learning. It has the following three characteristics.

First, it studies the weak label problem using problem transformation as the base approach for building multi-label classifiers, which is more suitable for weak label scenarios.

Second, it uses conditional label dependence to explore label correlations on the weak label problem with the help of input features of data. The label correlations are utilized to construct a unified sampling strategy to evaluate the informativeness of each example-label pair and determine which unknown label needs to be acquired from oracle. Meanwhile, it can avoid acquiring those labels that can be inferred from the obtained labels based on label correlations. These labels inferred with high confidence can be directly used in multi-label active learning, which is the third characteristic that will be discussed next.

Third, it combines human labeling with automatic labeling together to propose an integrated framework to further reduce the human labeling cost, which can use obtained labels to infer some unknown labels with label correlations.

The rest of the paper is organized as follows. In the next section, we briefly introduce the concept of multi-label active learning and related work. In Section III, we propose a new label correlation exploring method on the weak label problem and a computation efficiency optimization method for obtaining conditional label dependence. In Section IV, we propose a novel approach, a weak labeled multi-label active learning method with automatic labeling for image classification. In Section V, we investigate the performance of our novel active learning method on four image datasets and three non-image datasets and compare it with recent closely related active learning methods. Finally, we conclude our work and discuss further work in Section VI.

II. Related Work

Active learning is one of the most common methods in machine learning. Given a large pool of unlabeled examples, an active learner can iteratively select the most informative example from the pool to query an oracle or a human annotator for its true labels. It has been widely used in many specific areas, such as image classification, text categorization, and automatic annotation of multimedia content. The research results on multi-label active learning can be divided into two major categories: Example based sampling strategies and example-label pair based sampling strategies. The example based sampling strategies take each example as a sampling unit and measure the informativeness of each complete example. However, the example-label pair based strategies take each example-label pair as a sampling unit and measure the informativeness of each example and the informativeness of each label of this example. Therefore, an intelligent example-label pair based sampling strategy is able to reduce labeling effort to a great extent.

There are several articles on active learning for multi-label image classification. Tang *et al.* [18] proposed semantic-gap-oriented active learning for multi-label image annotation, which incorporates the semantic gap measure into an information minimization-based example selection strategy. Li *et al.* [19] proposed two multi-label active learning strategies, a max-margin prediction uncertainty strategy and a label cardinality inconsistency strategy and also integrated the two methods together in an adaptive way to obtain better results. Wu *et al.* [20] proposed a prototype of an example-label pair based batch mode active learning method for multi-label image classification, which selects the most uncertain example-label pair for reducing labeling effort. The above three methods ignore the interdependences between two labels. Qi *et al.* [21] proposed a 2DAL method, in which the example-label pair is used as the sampling unit and the mutual information of the label in this example-label pair across all other labels is used to measure the informativeness of this example-label pair. A labeled training set constructed by 2DAL is with incomplete labels. The problem created by incomplete labels was handled by integrating the unlabeled data to yield the marginal distribution of the labeled data. As we know, when a training set is small, predictions from a model built from the training set are inaccurate. Therefore, estimating label correlations based on such a model is not a good idea. Zhang *et al.* [22] proposed an example-label pairs based sampling approach by incorporating uncertainty with given static label constraints. According to our knowledge, it is not available for updating the label correlations. Bang *et al.* [23] proposed a batch mode active learning method, which selects a batch of example-label pairs by incorporating label correlations with uncertainty in each iteration of active learning. This method uses rule mining to explore label dependences. However, it is not suitable for the weak label problem. Ye *et al.* [25] uses chi-square statistics to evaluate label correlations and considers positive relationships and negative ones simultaneously. It is also not suitable for the weak label problem. In fact, as the active learning process proceeds, we gradually have more and more examples with complete or partial labels. With these examples and their requested labels, we can dynamically update the label correlations. Then, we can select the most effective example-label pairs to request their labels. In short, there exists a large problem: *There is no effective way to start active learning with weak labeled images or to obtain label correlations iteratively under the weak label problem.*

III. Label Dependence Exploration on the Weak Label Problem

In this section, we propose a new label correlation exploring method, in which label correlations are explored from labels and the feature space of examples. Then, we obtain a correlated label set for each label based on the label correlations. After this, we propose a novel example-label based multi-label active learning sampling strategy, which selects the most informative example-label pairs by integrating uncertainty with information gain over the correlated labels. Furthermore, to improve efficiency, a pruning strategy is adopted.

A. Problem Transformation

Before presenting problem transformation of multi-label classification, we first briefly describe what multi-label classification is. In machine learning, multi-label classification is a variant of traditional single-label classification. Unlike single-label classification, which makes a prediction for an unseen instance whether it belongs to only one label of a potential label set, multi-label classification makes predictions for an unseen instance whether it belongs to each label in a multi-label set. Assume that $X = R^d$ represents a d dimension example space, $Y = y_1, y_2, \ldots, y_T$ represents a multi-label set that contains T labels. Given a labeled training set $D = \{(x_i, Y_i) | 1 \leq i \leq m\}$, where $x_i \in X$ and $Y_i \in Y$, multi-label classification is used to learn a multi-label classifier $h : x \rightarrow Y$.

There are two typical frameworks for multi-label classification: *problem transformation* and *algorithm adaptation* [34]. *Problem transformation* decomposes a multi-label classification problem into several independent binary classification problems. *Algorithm adaptation* adapts existing algorithms to deal with multi-label data directly. Because of the loss of partial labels, *algorithm adaptation* is not easily adaptable to the multi-label classification problem with partial labels. Therefore, we will study multi-label classification with partial labels using *problem transformation* as the base approach for building multi-label classifiers.

An example of problem transformation is shown in Fig. 2(a). That is, given a multi-label classification problem with T labels, T sub-training sets are decomposed from an initial multi-label training set, and then T single label classifiers are built from each sub-training set to make predictions for future unseen examples.

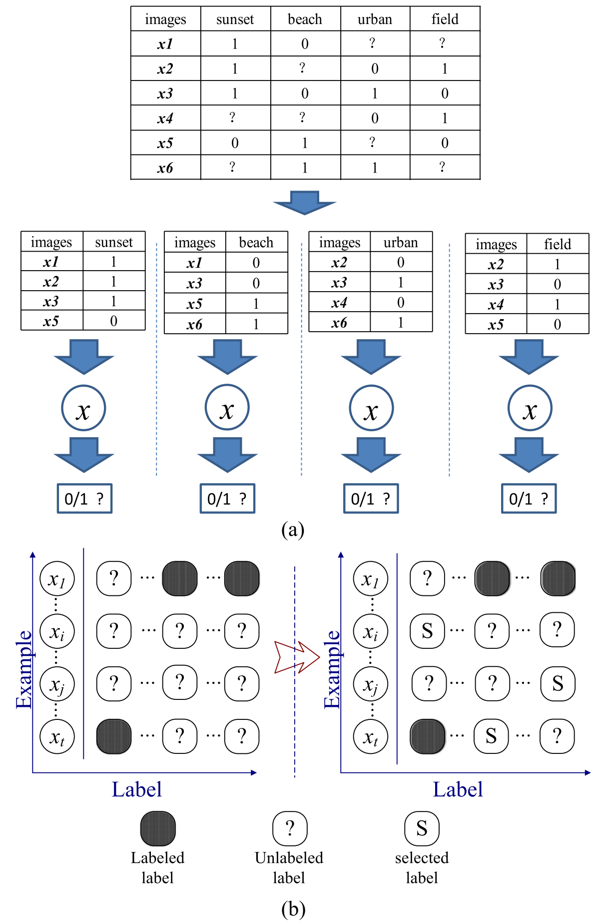


Fig. 2. (a) Example of problem transformation. (b) Illustration of the example-label pair-based active learning process.

B. Conditional Label Dependence

As is known, when a multi-label problem is transferred into several single label problems, label correlations are neglected during the transformation process. However, studies have shown that the correlations between labels should not be overlooked. Thus, to compensate for this deficiency, we develop a pair-wise method for mining label dependence under a scenario of partial labels.

Two types of dependence should be distinguished when analyzing label dependence in multi-label classification [35], i.e., conditional and unconditional label dependence. Conditional label dependence captures the dependence between labels from a specific set of examples. Unconditional label dependence is a global static dependence that is independent of concrete observation. In the weak label scenario, it is impossible to directly obtain global unconditional label dependence. Thus, we explore the conditional label dependence among the label set.

Unlike unconditional label dependence, conditional label dependence is associated not only with external labels but also the feature space of examples, which is used in multi-label learning algorithm Classifier Chains [36]. In this paper, conditional label dependence is used to derive associated label sets for each label in multi-label classification.

The conditional dependence of each pair of labels can be estimated by evaluating the gain from exploiting this dependence for binary classification of each label pair. In other words, we can follow the definition of conditional independence of a pair of labels. As is known, if the pair of labels is conditionally independent, the predictions for either one of the two labels from a probability-based classification model trained on a regular feature space should be the same as the predictions for this label from a probability-based classification model trained on the same feature space augmented by another label of this pair.

Suppose that a multi-label dataset has T labels. It has $T \times (T - 1)/2$ different pairs of labels. We need to estimate the dependence of each pair of labels. To estimate the dependence of each pair of labels (denoted as y_i and y_j), we need to train four binary classifiers. Two classifiers are trained only on the original feature space with each label as the target. We use h_{Oi} to denote the classifier with y_i as the target and h_{Oj} denoting the classifier and y_j as the target. We also train another two classifiers denoted as h_{ij} and h_{ji}. h_{ij} is trained on the original feature space augmented by the label y_i with the label y_j as the target, and the corresponding classifier h_{ji} is trained on the original feature space augmented by the label y_j with the label y_i as the target.

Using k-fold cross-validation, we can obtain the accuracy of each classifier. Obviously, if the accuracies h_{ij} and h_{ji} are both significantly higher than their corresponding classifiers h_{Oj} and h_{Oi}, we can view the two labels y_i and y_i as conditionally dependent.

The statistical significance of the average accuracy difference between each pair of classifiers (a pair of h_{ij} and h_{Oj}, another pair of h_{ji} and h_{Oi}) is determined using a paired t-test

$$t = \frac{\bar{A}_1 - \bar{A}_2}{s_{A_1A_2} \cdot \sqrt{\frac{2}{n}}} \tag{1}$$

$$s_{A_1A_2} = \sqrt{\frac{1}{2}(s_{A_1}^2 + s_{A_2}^2)} \tag{2}$$

here $S_{A_1}^2$ and $S_{A_1}^2$ are the unbiased estimators of the variances of the two accuracies. $\bar{A}_1$ is the average accuracy of the first classifier, and $\bar{A}_2$ is the average accuracy of the second classifier.

Fig. 3 illustrates the procedure of obtaining the conditional label dependence of a pair of labels. Note that our research focuses on multi-label active learning with partial labels. That is, there are missing labels in the original training set. As Fig. 3 shows, there are three labels (i.e., y_1, y_2, y_3) for this multi-label dataset. Thus, there are three pairs of labels, i.e., y_1 and y_2, y_1 and y_3, y_2 and y_3. We explain in detail how to estimate the dependence between y_1 and y_2. The procedure for estimating the dependence of other pairs of labels is the same.

Before we train four corresponding classifiers for the label pair y_1 and y_2, we need to prepare the training sets for each classifier. As the original training set has missing labels and either label y_1 or y_2 will be the target in its four corresponding classifiers, we filter the examples that have missing values for either label y_1 or y_2. Then, we obtain an intermediate dataset denoted as D_{12} (the most left sub-table, containing y_1 and y_2) as shown in Fig. 3. Based on the intermediate dataset D_{12}, we prepare four training sets for building four corresponding classifiers, i.e., h_{O1}, h_{O2}, h_{12}, h_{21}. According to k-fold cross-validation of the four corresponding classifiers, we can obtain the statistical dependence of the pair-wised labels y_1 and y_2 using (1). The conditional label dependence for other pairs of labels is calculated in the same way.

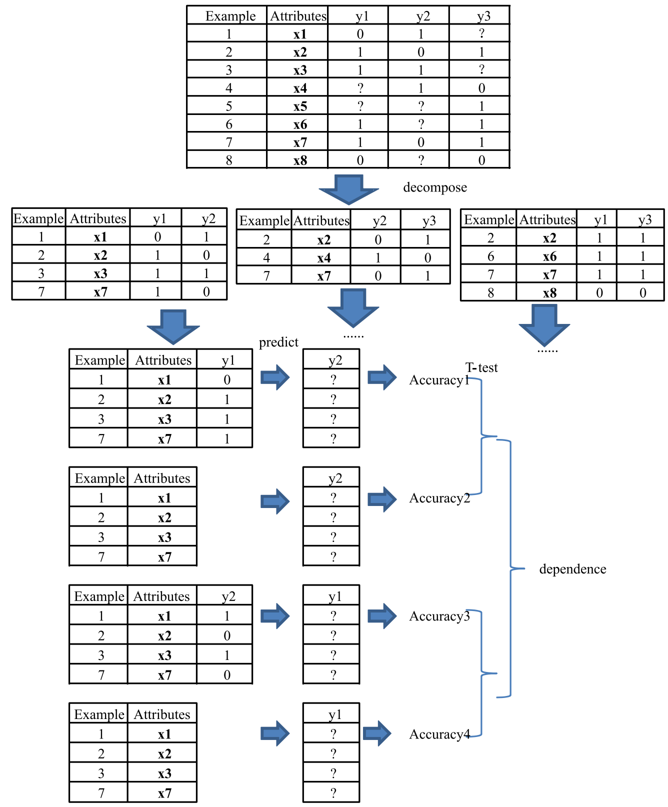
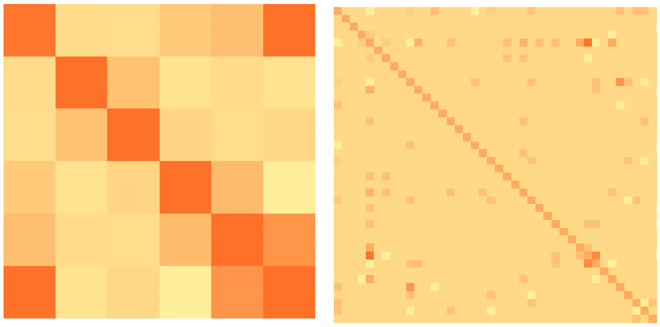


Fig. 3. Illustration of the procedure of obtaining the conditional label dependence of a pair of labels.

C. A Correlated Label Set for Each Label

As is known, a label may not correlate to all other labels in a multi-label dataset. We investigate the label dependence on multi-label image datasets and show our analysis on two datasets in Fig. 4(a) and 4(b). Fig. 4(a) illustrates the dependence of all pairs of labels on the image dataset *Scene*, which has six labels. Fig. 4(b) shows the dependence of all pairs of the first 40 labels on the image dataset *NUS_WIDE*. Note that *NUS_WIDE* has 81 labels in total. To see the illustration clearly, we only show the label dependence of all pairs of the first 40 labels.

Both Fig. 4(a) and 4(b) show that a label may be correlated to some labels but not all other labels. Among the correlated labels, the degree of correlation is different. Thus, it is necessary to distinguish its correlated labels for a specific label in a multi-label dataset. The degree of dependence of each pair of labels is shown in Fig. 4(a) and 4(b). The deeper the color, the greater the dependence is. In multi-label active learning, it is obvious that we should pay more attention to the pairs of labels that have



(a) (b)

Fig. 4. Dependences for all pairs of labels on the dataset *Scene* and the dataset *NUS_WIDE*.

greater dependence; when we request one label for a specific example from the oracle, we automatically obtain the possible information for the other label in the pair if this pair of labels is completely dependent.

In addition, when a multi-label dataset has a large number of labels, the computation cost for evaluating the effectiveness of each example-label pair is very high. To reduce the computational cost and promote the efficiency of multi-label active learning, it is necessary to choose a few highly dependent labels as the correlated label set for each label in a multi-label dataset. The correlated label set for each label can be obtained by setting a threshold over the conditional dependence calculated following (1) between this label and all others. The threshold can be set by looking up the t-statistical table. As is known, the higher the threshold, the smaller the correlated label set is.

D. Information Gain Over the Correlated Label Set

After having obtained the correlated label sets for all labels in a multi-label dataset, we introduce the information gain over the correlated label sets into the multi-label active learning sampling process. Considering an active learning task for multi-label image classification, each image is associated with multiple labels. Our goal is to learn a classifier for each label: $\hat{p}_i = \hat{p}_i(y_i|x), i = 1, 2, ..., T$. Again, we focus on multi-label active learning with partial labels. That is, each training example is associated with T labels, but not all labels have available values.

During the active learning process, it is necessary to evaluate the informativeness of each example-label pair. Before discussing how to evaluate the informativeness of each example-label pair, we introduce some notations that are used later. We use $UL(x_j)$ to denote the set of unknown labels of a specific example x_j, $L(x_j)$ to denote all labeled labels of this example x_j, and $CL(y_i)$ to denote the correlated label set of the label y_i. Note that obtaining the correlated label set $CL(y_i)$ for the label y_i was discussed in the previous subsection.

To evaluate the informativeness of each example-label pair, we need to estimate the information gain over the correlated label set $CL(y_i)$ for an unlabeled label y_i in $UL(x_j)$. This information gain is estimated by KL divergence (Kullback-Leibler Divergence), which is defined as follows:

$$Info\_gain(x_j, y_i) = \sum_{y_t \in CL(y_i)} \sum_{l \in 0,1} \hat{p}_i(y_i = l|x_j) \times \ln \frac{\hat{p}_i(y_i = l|x_j)}{\hat{p}_t(y_t = l|x_j)}. \tag{3}$$

As indicated previously, we will propose an example-label based multi-label active learning algorithm. That is, in each iteration of active learning, we need to choose the most informative example-label pairs to request labels from the oracle. Evaluating the informativeness of each example-label pair is the key of active learning for multi-label classification. It is obvious that both the uncertainty of an example-label pair and its information gain obtained from (3) are the important factors for constructing an evaluation criterion. Since entropy is a common method for measuring uncertainty, the informativeness of an example-label pair (x_j, y_i) can be evaluated as follows:

$$\begin{aligned} E_{x_{ji}} &= uncertainty + Info\_gain \\ &= -\sum_{l \in 0,1} \hat{p}_i(y_i = l|x_j) \ln \hat{p}_i(y_i = l|x_j) \\ &\quad + \sum_{y_t \in CL(y_i)} \sum_{l \in 0,1} \hat{p}_i(y_i = l|x_j) \ln \frac{\hat{p}_i(y_i = l|x_j)}{\hat{p}_t(y_t = l|x_j)} \end{aligned} \tag{4}$$

where $\hat{p}(y_i|x_j)$ denotes the posterior probability of label y_i of example x_j belonging to the positive or negative class produced by its multi-label learning model. If the correlated label set $CL(y_i)$ of label y_i is empty, (4) will reasonably degenerate to consider uncertainty only.

After integrating the uncertainty and information gain into the evaluation of the informativeness of an example-label pair, we can use it as a selection criterion to select the most informative example-label pairs as follows:

$$x^*_{WLMAL} = \arg\max_{x_j \in UD, i \in [1,...,T]} E_{x_{ji}} \tag{5}$$

where *UD* denotes the unlabeled dataset. Note that this is our conference paper [37], called WLMAL.

E. Computation Efficiency Optimization

As noted above, the process of obtaining conditional label dependence is the most important step. However, it is time consuming to evaluate conditional label dependence, in particular when the label space is very large. To improve the efficiency of this evaluation process, we adopt a pruning strategy before the process of obtaining conditional label dependence.

Our existing studies indicate that each label is always correlated with some labels but not all of them. Calculating conditional dependence among two uncorrelated labels is useless. Only exploring conditional label dependence among correlated labels can greatly reduce computation time. Therefore, before exploring the conditional label dependence for each label, a correlation coefficient should be calculated for each pair of labels. Here, the Phi coefficient is adopted as the correlation coefficient, which is a measure of the association between two

TABLE I
STATISTICAL TABLE OF THE PAIR OF LABELS y_i AND y_j

	$y_i = 0$	$y_i = 1$	total
$y_j = 0$	n_{00}	n_{01}	n_{0*}
$y_j = 1$	n_{10}	n_{11}	n_{1*}
total	n_{*0}	n_{*1}	N

binary variables. It is similar to the Pearson correlation coefficient in its interpretation.

Given two labels y_i and y_j, we have a 2 × 2 table, as shown in Table I.

In the above table, $n_{00,}$ $n_{01,}$ n_{10} and n_{11} are non-negative counts of observations, $n_{0*} = n_{00} + n_{01}, n_{1*} = n_{10} + n_{11}$, and N is the total number of observations, i.e., $N = n_{0*} + n_{1*}$. The Phi coefficient that describes the correlation between y_i and y_j is defined as follows:

$$\phi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{0*}n_{1*}n_{*0}n_{*1}}}. \tag{6}$$

After the Phi coefficient of each pair of labels is calculated, a correlated label set of each label can be obtained if a determination criterion is set. After the correlated label set of each label is obtained, the process of voting committee constructions for each label and information measurement for each example-label pair is based on the correlated label set (not all other labels) of each label, so computation time is significantly reduced. To reach a better understanding, we provide the following explanation.

Given a multi-label dataset with T labels, without computation efficiency optimization, we know that there are $T \times (T - 1)$ auxiliary classifiers that need to be built to obtain the dependence matrix D, and to make the results more accurate, k cross-validation should be used for each classifier. Thus, the total times for computing conditional label dependence is $T \times (T - 1) \times k$, and the time complexity of this process is $O(T^2)$ when $k << T$ (note that k is a constant). However, as observed previously, each label is always correlated to some of the labels but not all the labels. Thus, there is no need to calculate such dependence on non-correlated labels. For a multi-label dataset, let us assume that each label is correlated with t labels on average. Then, we only need to build $T \times t$ auxiliary classifiers, and k cross-validation is still necessary. Thus, the total time that is necessary to compute conditional label dependence is $T \times t \times k$, and the time complexity of this process is $O(T)$ when $k << T$ and $t << T$. Again, k is a constant. Note that after the computation efficiency of conditional label dependence is optimized as above for WLMAL [37], we refer to it as E-WLMAL.

Here, the Phi correlation is only used for measuring whether there is a dependency between two labels. Note that we only need to determine whether the Phi coefficient is zero. If a Phi coefficient is zero, this label pair is pruned. Otherwise, we compute its label dependence.

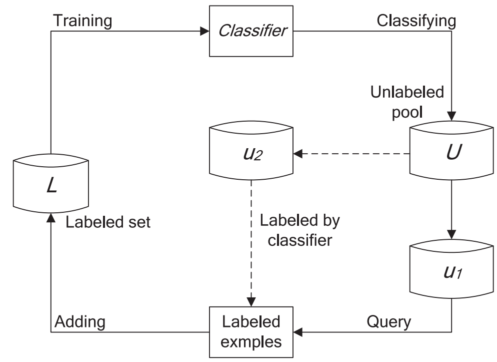


Fig. 5. Typical framework of semi-supervised active learning for single-label classification.

IV. WEAK LABELED MULTILABEL ACTIVE LEARNING WITH AUTOMATIC LABELING

A typical multi-label active learning consists of three important parts, such as the learning engine, sampling strategy and annotators. Among them, the sampling strategy has been studied to a great extent in previous studies. In this paper, we first propose an effective sampling strategy on the weak label problem. Moreover, to further reduce the labeling cost, an integration of human labeling and automatic labeling is also studied from the aspect of the annotator.

In this section, we first present a typical semi-supervised framework for single-label active learning. Then, a novel multi-label active learning framework for weak labeled image classification (called AE-WLMAL) is put forward, which combines E-WLMAL with automatic labeling. With this framework, the complete algorithm is presented later.

A. A Typical Framework of Semi-Supervised Active Learning

As is known, self-training is a useful technique in semi-supervised learning [38], which has been introduced into single-label active learning for reducing human annotation efforts [39]. During self-training, an initial classifier is obtained from scarce labeled training examples first, and then this classifier is used to perform classification for unlabeled examples. Generally, unlabeled examples with the highest certainty are added to the current training set along with their predicted labels. Next, the classifier is re-trained on the expanded training set. The framework of typical semi-supervised active learning for single-label classification is shown as Fig. 5.

Intuitively, the probability of introducing an error label is minimal if an example is selected from the most explicit results under a current classifier. However, from the perspective of information content, an example with the highest certainty always contains the least information. Thus, the effect of adding these examples to the training set on the current classifier is very limited. During the process of active learning, examples containing the most information are the most uncertain examples that are sampled from the unlabeled example pool. How can we infer the labels for the most uncertain examples? In this paper, we introduce conditional label dependence into the framework of

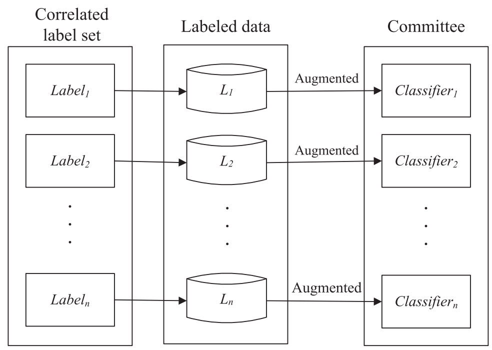


Fig. 6. Construction of an augmented classifier committee.

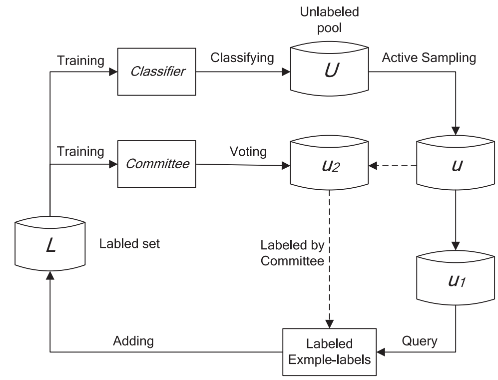


Fig. 7. Framework of our semi-supervised active learning for multilabel classification with MV.

semi-supervised active learning and propose an automatic labeling method with MV, i.e., majority voting, which is discussed next.

B. Our Novel Framework With Automatic Labeling

Let us engage in a discussion regarding the feasibility of introducing automatic labeling. Semi-supervised learning is a well known learning method. It faces the same problem as active learning. That is, labeled data are limited, but unlabeled data are abundant. Semi-supervised learning does not require manual intervention and predicts the labels of the image only by using the trained classifier. Obviously, there is a challenge, i.e., ensuring prediction accuracy when semi-supervised learning is used for automatic labeling. To solve this thorny problem, an ingenious way based on the obtained label correlations is considered in this paper.

Before introducing our automatic labeling method, we perform further analysis of the conditional label dependence. As observed in Section III-C, after obtaining the conditional dependence matrix, a correlated label set for each label can be determined. In Section III-D, we utilized the correlated label set of each label to obtain its information gain. We call this correlated label set *condep-set*. As is known, if a classifier is built on the original feature augmented by each label in *condep-set*, the performance of this classifier is always more superior than the classifier built on the original feature only. Based on this idea, we propose making full use of *condep-set* to construct an ensemble of classifiers, which form an effective voting committee for the to-be-label label. Obviously, this action further augments predictive performance and ensures the quality of automatic labeling.

Here, we exploit *condep-set* to infer the labels for unlabeled images. Supposing that the *condep-set* of $Label_t$ is $Label_1, \ldots, Label_n$, first, n classifiers $Classifier_1$, $\ldots, Classifier_n$ are built on the original feature space augmented by each label in *condep-set*. As discussed previously, the classification accuracy of $Label_t$ can be improved by these augmented classifiers. Then, we construct a voting committee for label $Label_t$, as shown in Fig. 6.

After having constructed a committee of classifiers for each label, we can make use of these committees to predict the labels for the most informative examples selected by our active sampling. Majority voting is a decision rule that makes predictions according to the majority. We automatically annotate label $Label_t$ if it has high voting consistency. That is, our active sampling does not need to query these labels from the oracle, so it can significantly reduce human annotation efforts. The framework of our novel method with MV is shown in Fig. 7.

As shown in Fig. 7, during each iteration of active learning, after having selected the most informative unlabeled examples, we divide them into two sets. One set can be classified by the current constructed committees, and the other set is labeled by human experts. In the next subsection, we provide the details of our semi-supervised active learning algorithm AE-WLMAL.

C. Complete Algorithm

Now, it is time to briefly show the complete framework of our algorithm AE-WLMAL in Algorithm 1.

AE-WLMAL is an iterative active learning process. It first evaluates the informativeness of each example-label pair. Then, it chooses the example-label pair to acquire its label, which has the highest informativeness. After the label of this example-label pair is obtained, this example-label pair is put into the training set. With the updated training set, AE-WLMAL updates the conditional label dependence with that example-label pair. Again, AE-WLMAL iteratively updates the conditional label dependence during its active learning process. Thus, it can guarantee that the obtained label dependence dynamically fits the current training set. This is unlike all previous methods. In addition, according to our knowledge, it is the first method to use conditional label dependence to obtain label correlations with weak labels. We investigate its performance by comparing it with previous methods in the next section.

V. EXPERIMENTS

In this section, we evaluate our proposed AE-WLMAL by conducting experiments on four image multi-label datasets (i.e., *Scene*, *NUS-WIDE*, *Corel5k*, and *Corel16k*). To further investigate the versatility of AE-WLMAL, we also conduct experiments on three non-image multi-label datasets (i.e., *Genbase*, *Emotions* and *Yeast*). All of these datasets can be downloaded from MULAN [40]. The characteristics of these datasets are shown in Table II.

Algorithm 1: AE-WLMAL

Input:
The labeled example-label set *LD*;
The unlabeled data set *UD*;
The size of label set *T*.
Procedure:
Build multi-labelBuild multi-labelclassifier model θ based on *LD* by *problem transformation*;
Calculate ϕ values for all label pairs;
Exploiting conditional label dependence on *LD* according to ϕ values;
Derive correlated label sets *CL* for all labels;
while stopping criterion is not met
  for $i = 1$ to *T*
    for each example-label pair x_{ji} of x_j in *UD*
      Compute the informativeness of each example-label pair x_{ji} using (4);
    end for
    Select the most informative example-label pair x_{ji} using (5);
    if the selected example-label pair x_{ji} can be classified by current committees with a high voting consistency
      Annotate this example-label pair x_{ji} with its classified label;
    else
      Request the label for the selected example-label pair;
    $UD = UD \backslash x_{ji}$;
    $LD = LD \cup x_{ji}$;
End for
Rebuild the classifier model θ based on *LD*;
Update the conditional label dependence based on *LD* and then update *CL* for all labels;
End while
Output: The final classifier model θ;

TABLE II
CHARACTERISTICS OF THE SEVEN DATASETS USED IN OUR EXPERIMENTS

Datasets	Domains	Instances	Labels	Features
Scene	image	2407	6	294
NUS-WIDE	image	269648	81	128
Corel5k	image	5000	374	499
Corel16k	image	13811	161	500
Genbase	biology	662	27	1186
Emotions	music	593	6	72
Yeast	biology	2417	14	103

We compare our AE-WLMAL and E-WLMAL with the following approaches in our experiments.

WLMAL – an example-label pair-based sampling method proposed in our conference paper [40], which proposes a solution for the weak label problem on multi-label active learning for image classification.

CSMAL – an example-label pair-based sampling method proposed in [25], which uses chi-square statistics to evaluate label correlations.

MTAL – a method proposed in [22], which takes output constraints into account during the active sampling process.

LMAL – the closest related method proposed in [20], which selects a batch of example-label pairs with the most uncertainty.

Adaptive – an adaptive method proposed in [19], which integrates the uncertainty and the label cardinality together.

TDAL – a method proposed in [21], which selects the example-label pairs based on the informativeness measurement, in terms of mutual information.

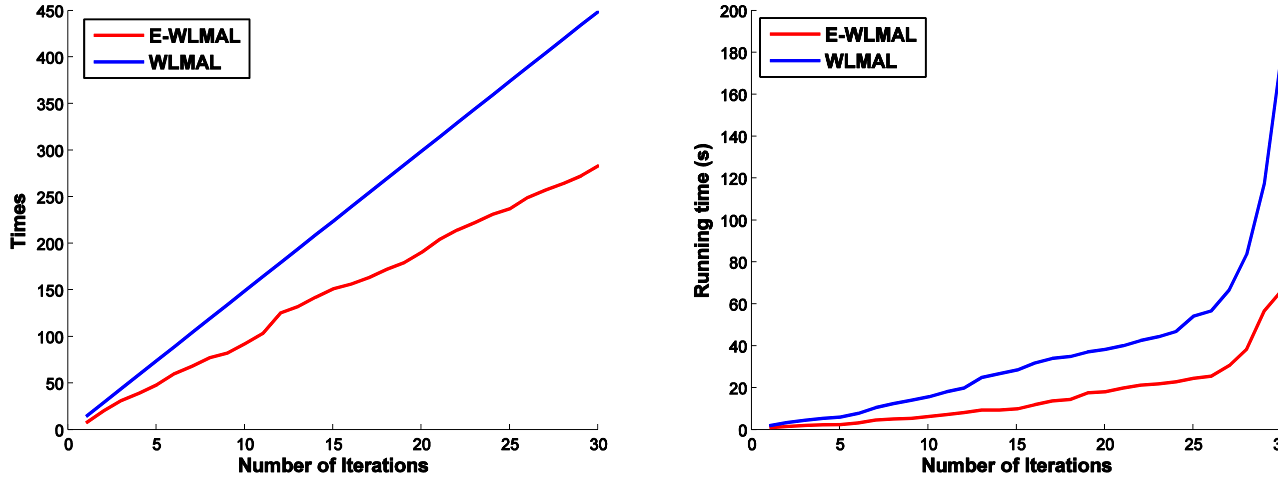
In our experiments, BRkNN [34] is used as the base learner, which is a typical model of *problem transformation*. It is a widely used multi-label learning algorithm and has been implemented in MULAN [39], an open source Java library for multi-label learning. For each dataset, we first randomly divide an entire dataset into three partitions: an initial labeled set, an unlabeled pool and a test set. For each dataset, we randomly select 30% instances as the test set, and the rest works as the unlabeled pool after a few examples are chosen as the initial labeled set. For each image dataset, we randomly choose 10 examples as the initial training set, and for each non-image dataset, we choose 30 examples as the initial training set. Considering the size of all datasets and the time consumption of our experiments, we set the number of active learning iterations as 100 for the dataset flags and as 400 for the other datasets. For comparison purposes, we select *T* example-label pairs in each iteration for seven example-label pair based methods (i.e., AE-WLMAL, E-WLMAL, WLMAL, CSMAL, MTAL, LMAL and TDAL), and for the Adaptive method, we select one complete example to request all labels in each iteration. Three evaluation metrics are used in our experiments, i.e., Accuracy, Macro-F1 and Micro-F1. Each experiment is repeated 10 times, and the average experimental results are shown in the following figures.

We first present the computation efficiency optimization of E-WLMAL and then analyze our experimental results on four image datasets and three non-image datasets.

A. *Investigation of Computation Efficiency Optimization*

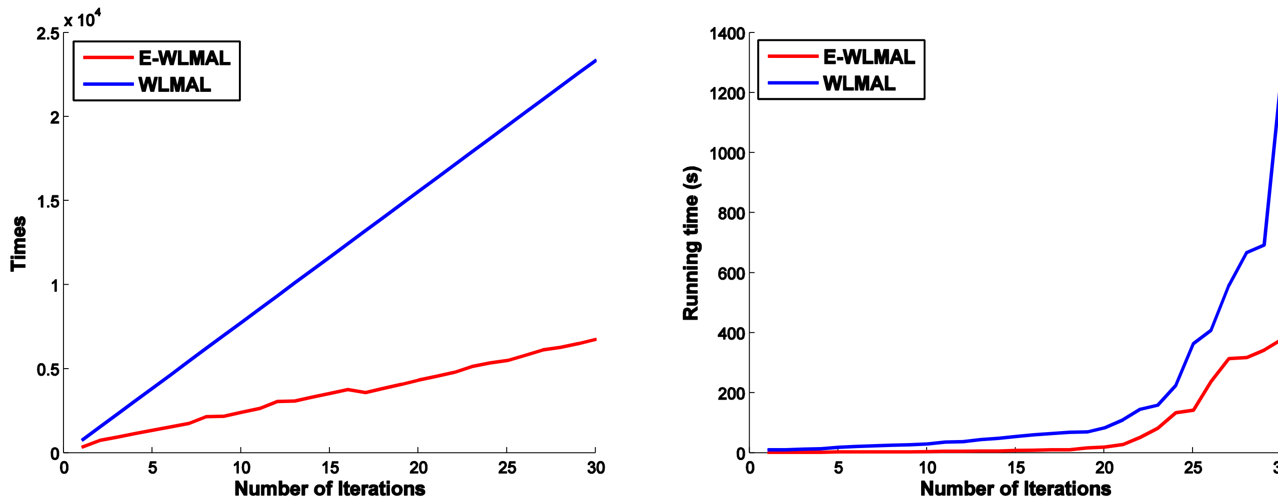
It is first necessary to investigate the computation efficiency optimization of AE-WLMAL and E-WLMAL. As is known, AE-WLMAL is an advanced version of E-WLMAL. AE-WALMAL has an extra component (automatic labeling). It is unfair to compare the computation time of AE-WLMAL with WLMAL directly. Therefore, we only compare E-WLMAL with WLMAL in our experiments.

We conduct our experiments on a workstation with Intel(R) Xeon(R) CPU E3-1270v3 3.50 Hz 8-core CPU, 16 GB RAM, Java 7 VM and 64 bit Windows 7 OS. We have a head-to-head comparison in terms of the total times for computing conditional label dependence as well as the total running time for computing conditional label dependence. Our experimental results on two image datasets *Scene* and *NUS_WIDE* are shown in Figs. 8 and 9, respectively. Note that the left-side sub-figure in Figs. 8 and 9 shows the total times for computing



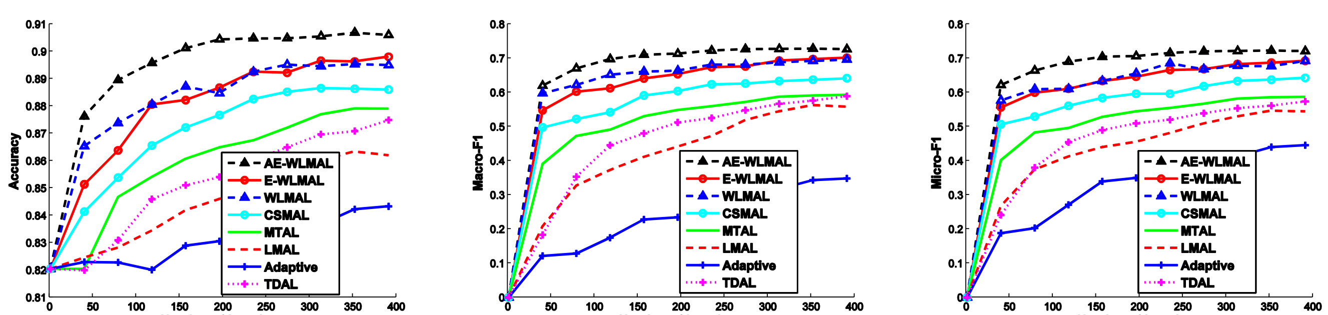
(a) (b)

Fig. 8. Efficiency improvement of conditional dependence on the dataset *Scene*. (a) Times. (b) Running time.



(a) (b)

Fig. 9. Efficiency improvement of conditional dependence on the dataset *NUS_WIDE*. (a) Times. (b) Running time.

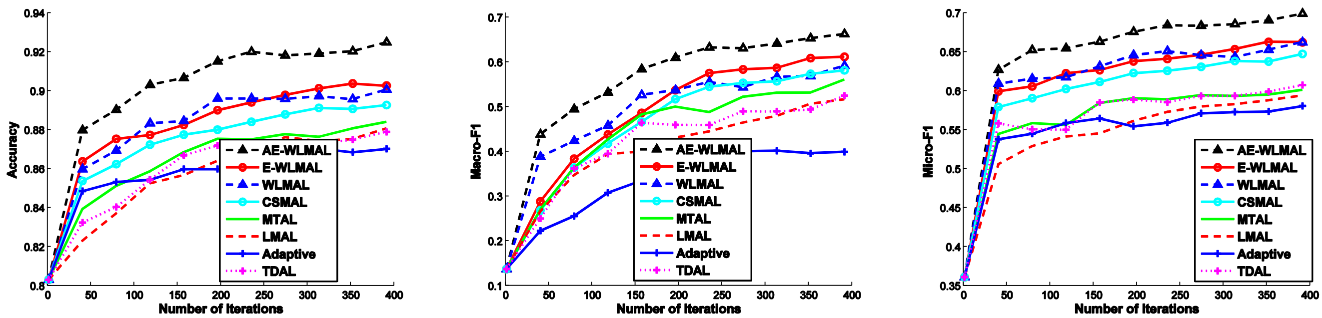


(a) (b) (c)

Fig. 10. Average results over 10 runs in terms of three evaluations on the dataset *Scene*. (a) Accuracy. (b) Macro-F1. (c) Micro-F1.

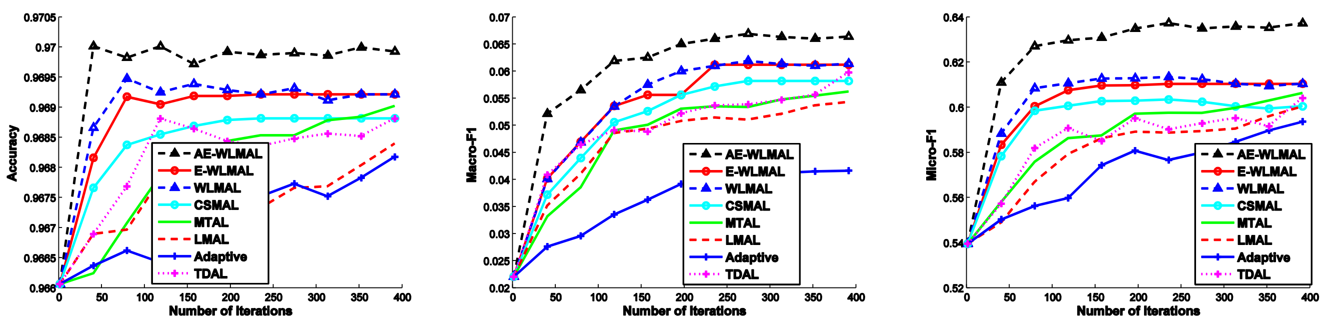
conditional label dependence, and the right-side sub-figure in Figs. 8 and 9 shows the total running time in seconds for computing conditional label dependence.

From Figs. 8 and 9, we can see that the computation efficiency optimization significantly reduces the total times of computing conditional label dependence, and the total running time is also significantly reduced after optimization is applied. As discussed in Section III-E, the time complexity of WLMAL is proportional to the number of labels in the multi-label set of a multi-label classification domain. Since the



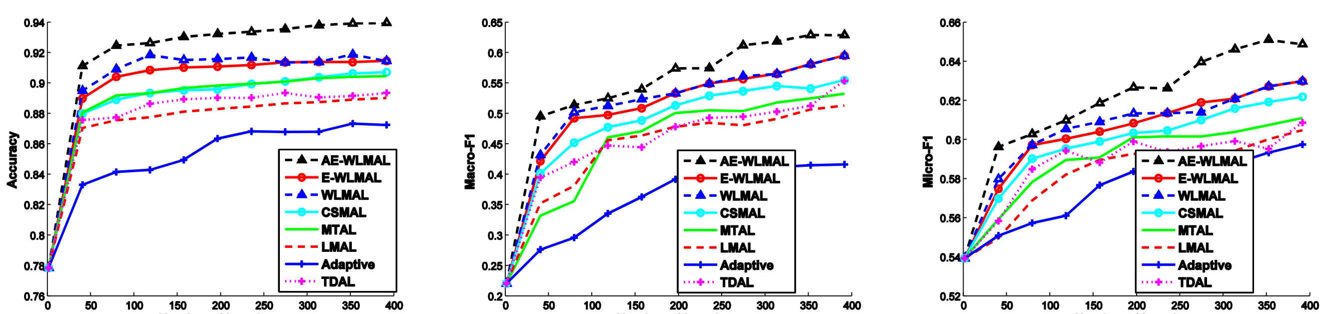
(a) (b) (c)

Fig. 11. Average results over 10 runs in terms of three evaluations on the dataset *NUS-WIDE*. (a) Accuracy. (b) Macro-F1. (c) Micro-F1.



(a) (b) (c)

Fig. 12. Average results over 10 runs in terms of three evaluations on the dataset *Corel5k*. (a) Accuracy. (b) Macro-F1. (c) Micro-F1.



(a) (b) (c)

Fig. 13. Average results over 10 runs in terms of three evaluations on the dataset *Corel16k*. (a) Accuracy. (b) Macro-F1. (c) Micro-F1.

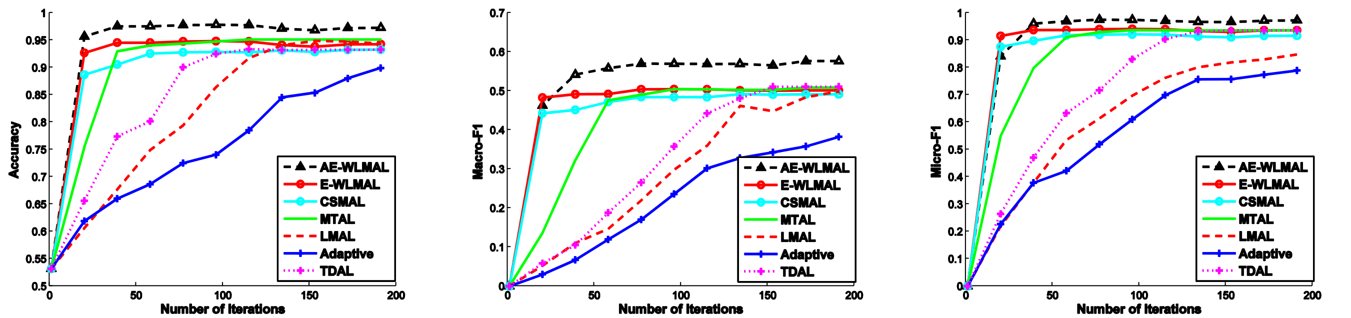
number of labels in the multi-label dataset *NUS_WIDE* (i.e., 81 labels) is much more than that of the multi-label dataset *Scene* (i.e., only six labels), both the total times and the total running time for computing the conditional label dependence on the dataset *NUS_WIDE* is much higher than that on the dataset *Scene*. Furthermore, we can see that the reduction of both the total times and the running time on the dataset *NUS_WIDE* is much higher than that on the dataset *Scene*. That is, the improvement in computation efficiency optimization is more significant when there are more labels in a multi-label domain. Time complexity reduction is very important in image recognition applications [41].

B. Classification Performance Investigation

In the above section, our experimental results showed that E-WLMAL can significantly reduce the computation workload of WLMAL. In this subsection, we conduct experiments to investigate the classification performance of both AE-WLMAL and E-WLMAL compared with the other six state-of-the-art multi-label active learning methods (i.e., WLMAL, CSMAL, MTAL, LMAL, Adaptive, and TDAL).

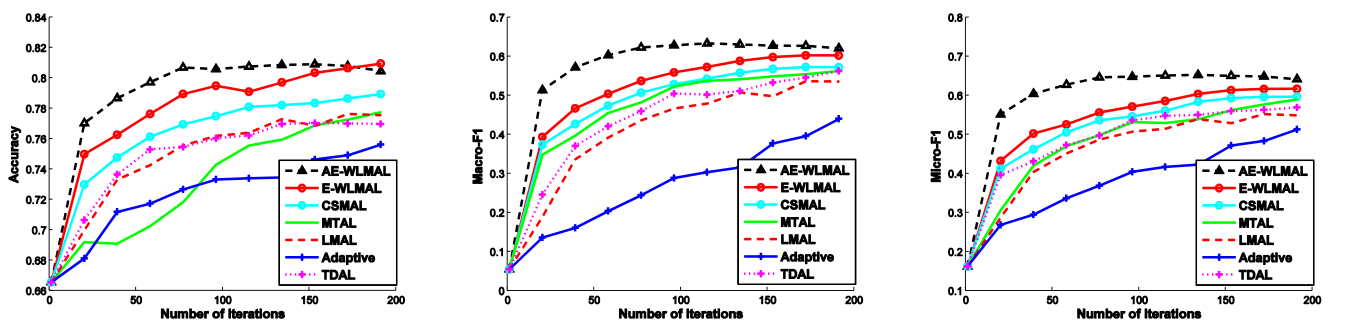
1) Results on Image Datasets: We conduct experiments on four image datasets (i.e., *Scene*, *NUS_WIDE, Corel5k* and *Corel16k*), which are shown in Table II. Note that *Corel5k* and *Corel16k* are two large image datasets that have more instances and labels. As shown in Table II, *Corel5k* has 3000 instances and 374 labels, and *Corel16k* has 13811 instances and 161 labels. These four image datasets are used to investigate the adaptability of all seven methods. In the experiments, the threshold of MV is set as 0.9 for AE-WLMAL.

The average experimental results for AE-WLMAL, E-WLMAL and the other six methods on four image datasets are shown in Figs. 10–13 in terms of three evaluation metrics.



(a) (b) (c)

Fig. 14. Average results over 10 runs in terms of three evaluations on the dataset *Genbase*. (a) Accuracy. (b) Macro-F1. (c) Micro-F1.



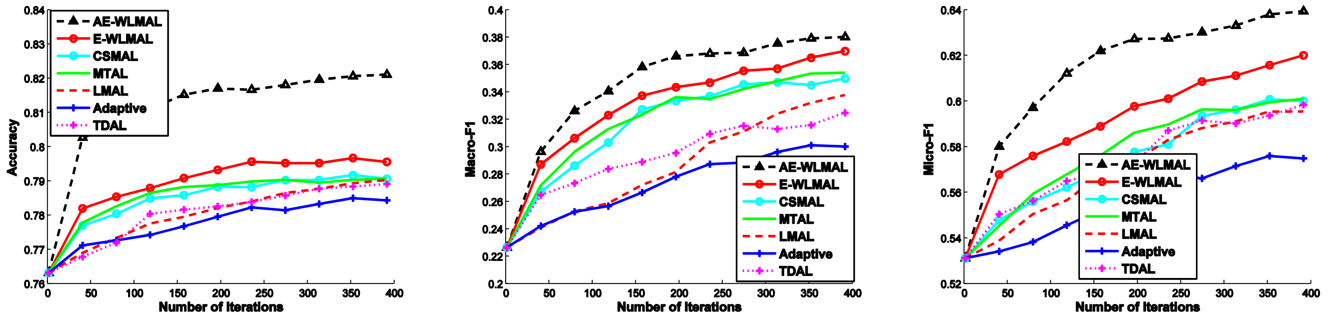
(a) (b) (c)

Fig. 15. Average results over 10 runs in terms of three evaluations on the dataset *Emotions*. (a) Accuracy. (b) Macro-F1. (c) Micro-F1.

From these figures, we can make the following conclusions: 1) all example-label pair based methods perform better than the example based method (i.e., Adaptive) in multi-label active learning, although the example based method Adaptive enhances the performance of multi-label classification. This indicates that example-label pair based methods are superior to example-based methods. For this reason, researchers nowadays focus on taking an example-label pair as a sampling unit in multi-label active learning, which also results in the weak label problem. 2) Among the example-label pair based methods, the methods (i.e., AE-WLMAL, E-WLMAL, WLMAL, CSMAL, MTAL, and TDAL) that take label correlations into consideration during the active learning process perform better than the one (i.e., LMAL) that ignores label correlations. In recent years, researchers have realized the importance of label correlations in multi-label classification and multi-label active learning. The related studies show that label correlations play a significant role in both improving the performance of multi-label classifier and optimizing sampling strategies in multi-label active learning. This is confirmed according to the above experimental results. 3) Among the methods that consider label correlations in the active learning process (i.e., AE-WLMAL, E-WLMAL, WLMAL, CSMAL, MTAL, and TDAL), our method, AE-WLMAL, always performs the best on all four datasets in terms of the three evaluation metrics. E-WLMAL performs similarly to WLMAL. Both E-WLMAL and WLMAL perform better than the others (i.e., CSMAL, MTAL, and TDAL). MTAL is a sampling approach that incorporates uncertainty with given static label constraints. The label correlations are not updated during the active learning process. Obviously, different datasets correspond to different label correlations. Therefore, the adaptability of MTAL relies on specific datasets. It is not applicable in real-world situations. TDAL uses the mutual information of the label in an example-label pair across all other labels to measure the informativeness of this example-label pair. The labeled training set constructed by 2DAL is with incomplete labels. It handles the problem brought by incomplete labels by predicting based on a model built from a training set. Therefore, estimating label correlations based on such a model is not a good idea. CSMAL uses chi-square statistics to evaluate label correlations and considers positive relationships and negative ones simultaneously. It also requires that the labeled examples have full labels when mining label correlations. However, our methods (AE-WLMAL, E-WLMAL and WLMAL) can explore label correlations on the weak label problem and then consider the information gain from correlated label sets during the active learning process. The experimental results show its effectiveness.

2) Results on Nonimage Datasets: In order to investigate the universality of the two new methods AE-WLMAL and E-WLMAL, we conduct experiments on three non-image datasets (i.e., *Emotions*, *Genbase*, and *Yeast*), one music dataset and two biology datasets. As shown in Table II, *Emotions* has 6 labels, *Genbase* has 27 labels and *Yeast* has 14 labels. Among them, the feature dimension of *Genbase* is 1186 (the largest one), and that of *Emotions* is 72 (the smallest one). These three non-image datasets are used to investigate the universality of all five methods.

Our experimental results are shown in Figs. 14–16. From these figures, we can make the following conclusions.



(a) (b) (c)

Fig. 16. Average results over 10 runs in terms of three evaluations on the dataset *Yeast*. (a) Accuracy. (b) Macro-F1. (c) Micro-F1.

1) Although the example based method Adaptive performs better on yeast than on the other two datasets, it is still the worst among the seven methods. As a whole, all example-label pair-based methods perform better than the example based method (i.e., Adaptive) in multi-label active learning. 2) Among the example-label pair based methods, the methods that take the label correlations into consideration during the active learning process perform better than the one (i.e., LMAL) that ignores label correlations, although LMAL enhances the performance of multi-label classification. 3) Among the methods that consider label correlations in the active learning process, our method AE-WLMAL always performs the best on all three datasets, followed by the method E-WLMAL, in terms of the three evaluation metrics. This shows that considering information gain from correlated label sets during the active learning process, which is used in AE-WLMAL and E-WLMAL, is more effective. Again, that AE-WLMAL performs better than E-WLMAL is due to automatic labeling.

Figs. 14–16 show that the conclusions made in the last subsection are sustained. That is, our proposed method AE-WLMAL performs very well on non-image datasets. It can be directly applied to non-image datasets.

VI. Conclusion and Future Work

In this paper, we proposed a novel multi-label active learning method under the weak label scenario for multi-label image classification, i.e., AE-WLMAL, which integrates the uncertainty and information gain obtained from the correlated label sets to select the most informative example-label pairs. The procedure for obtaining the correlated label set for each label in a multi-label dataset exploits label dependence with input features. In contrast with existing methods, the proposed approach AE-WLMAL explores label correlations under the weak label scenario of multi-label classification, which not only considers how to obtain the correlated label sets by exploiting conditional label dependence but also iteratively updates the correlated label sets during the active learning process. In addition, a pruning strategy is adopted to improve computation efficiency. Furthermore, AE-WLAML exploits label correlations to infer labels for unlabeled images, which further reduces human labeling cost. Our experimental results show that AE-WLMAL performs better than existing methods.

There are several important issues that can be further investigated in the future.

1) Note that the important hypothesis of traditional active learning is that the labels are provided by experts. Note that we only consider the selection strategy for multi-label active learning but have not considered how to annotate the selected example-label pairs. One important research direction is to incorporate crowd computing into active learning. Crowd computing is referred to as problem solving using crowdsourcing, which was first proposed in [42]. In crowdsourcing, a group of people are asked to contribute to a task. One of the most remarkable applications of crowdsourcing is Amazon Mechanical Turk (MTurk), a marketplace for providing solutions to micro tasks. To collect labels for unlabeled images, each unlabeled image is submitted as a task to MTurk for people to provide labels. Its advantage is that plenty of cheap labels can be obtained in a short time; a disadvantage is that the obtained labels may have errors because of the different labeling abilities of labelers.
2) Active learning is a learning program that actively acquires extra information with an effort for a certain gain in building models. Both efforts and gains are quantified by cost on the same scale. Thus, the goal of active learning is to minimize total cost. It is interesting to study multi-label active learning in a cost-sensitive framework [43]–[45].
3) In addition, in many real-world applications, the features that present the images need to be extracted from the images. We will investigate the performance of different feature extraction methods in multi-label active learning for image classification [46]–[48].

References

[1] M. Wang and X. S. Hua, "Active learning in multimedia annotation and retrieval: A survey," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 2, 2011, Art. no. 10.

[2] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 606–617, Jun. 2011.

[3] Z. J. Zha *et al.* "Interactive video indexing with statistical active learning," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 17–27, Feb. 2012.

[4] C. Zhang and T. Chen, "An active learning framework for content-based information retrieval," *IEEE Trans. Multimedia*, vol. 4, no. 2, pp. 260–268, Jun. 2002.

[5] D. Picard, M. Cord, and A. Revel, "Image retrieval over networks: Active learning using ant algorithm," *IEEE Trans. Multimedia*, vol. 10, no. 7, pp. 1356–1365, Nov. 2008.

[6] A. J. Joshi, F. Porikli, and N. Papanikolopoulos, "Multi-class active learning for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2009, pp. 2372–2379.

[7] Y. Lin *et al.*, "Large-scale image classification: Fast feature extraction and SVM training," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 1689–1696.

[8] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recog.*, vol. 37, no. 9, pp. 1757–1771, 2004.

[9] G. Nasierding, G. Tsoumakas, and A. Z. Kouzani, "Clustering based multi-label classification for image annotation and retrieval," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2009, pp. 4514–4519.

[10] X. S. Xu, Y. Jiang, L. Peng, X. Xue, and Z. H. Zhou, "Ensemble approach based on conditional random field for multi-label image and video annotation," in *Proc. ACM Int. Conf. Multimedia*, 2011, pp. 1377–1380.

[11] Y. Pang, Z. Ma, Y. Yuan, X. Li, and K. Wang, "Multi-modal learning for multi-label image classification," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 1797–1800.

[12] R. S. Cabral, F. Torre, J. P. Costeira, and A. Bernardino, "Matrix completion for multi-label image classification," *Adv. Neural Inf. Process. Syst.*, vol. 201, pp. 190–198, 2011.

[13] X. Wen, L. Shao, Y. Xue, and W. Fang, "A rapid learning algorithm for vehicle classification," *Inf. Sci.*, vol. 295, no. 1, pp. 395–406, 2015.

[14] B. Settles, M. Craven, and L. Friedland, "Active learning with real annotation costs," in *Proc. Neural Inf. Process. Syst. Workshop Cost-Sensitive Learn.*, 2008, pp. 1–10.

[15] S. Vijayanarasimhan and K. Grauman, "What's it going to cost you?: Predicting effort versus informativeness for multi-label image annotations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2009, pp. 2262–2269.

[16] B. Settles, "Active learning literature survey," Univ. Wisconsin, Madison, WI, USA, Comput. Sci. Tech. Rep. 1648, 2010.

[17] X. Li, L. Wang, and E. Sung, "Multi-label SVM active learning for image classification," in *Proc. Int. Conf. Image Process.*, 2004, vol. 4, pp. 2207–2210.

[18] J. Tang, Z. J. Zha, D. Tao, and T. S. Chua, "Semantic-gap-oriented active learning for multi-label image annotation," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2354–2360, Apr. 2012.

[19] X. Li and Y. Guo, "Active learning with multi-label SVM classification," in *Proc. AAAI Int. Joint Conf. Artif. Intell.*, 2013, pp. 1479–1485.

[20] J. Wu, V. S. Sheng, J. Zhang, P. P. Zhao, and Z. M. Cui, "Multi-label active learning for image classification," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 5227–5231.

[21] G. J. Qi, X. S. Hua, Y. Rui, J. Tang, and H. J. Zhang, "Two-dimensional active learning for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2008, pp. 1–8.

[22] Y. Zhang, "Multi-task active learning with output constraints," in *Proc. AAAI Conf. Artif. Intell.*, 2010, pp. 667–672.

[23] B. Zhang, Y. Wang, and W. Wang, "Batch mode active learning for multi-label image classification with informative label correlation mining," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Jan. 2012, pp. 401–407.

[24] B. Zhang, Y. Wang, and F. Chen, "Multilabel image classification via high-order label correlation driven active learning," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1430–1441, Mar. 2014.

[25] C. Ye *et al.*, "Multi-label active learning with chi-square statistics for image classification," in *Proc. ACM Int. Conf. Multimedia Retrieval*, 2015, pp. 583–586.

[26] S. J. Yang, Y. Jiang, and Z. H. Zhou, "Multi-instance multi-label learning with weak label," in *Proc. Int. Joint Conf. Artif. Intell.*, 2013, pp. 1862–1868.

[27] Y. Y. Sun, Y. Zhang, and Z. H. Zhou, "Multi-label learning with weak label," in *Proc. AAAI Conf. Artif. Intell.*, 2010, pp. 593–598.

[28] S. J. Yang, Y. Jiang, and Z. H. Zhou, "Multi-instance multi-label learning with weak label," in *Proc. AAAI Int. Joint Conf. Artif. Intell.*, 2013, pp. 1862–1868.

[29] G. Yu, G. Zhang, H. Rangwala, C. Domeniconi, and Z. Yu, "Protein function prediction using weak-label learning," in *Proc. ACM Conf. Bioinformat., Comput. Biol. Biomed.*, 2012, pp. 202–209.

[30] S. J. Huang and Z. H. Zhou, "Multi-label learning by exploiting label correlations locally," in *Proc. AAAI Conf. Artif. Intell.*, 2012, pp. 949–955.

[31] Y. Yu, W. Pedrycz, and D. Miao, "Multi-label classification by exploiting label correlations," *Expert Syst. Appl.*, vol. 41, no. 6, pp. 2989–3004, 2014.

[32] G. Tsoumakas, A. Dimou, E. Spyromitros, I. Kompatsiaris, and I. Vlahavas, "Correlation-based pruning of stacked binary relevance models for multi-label learning," in *Proc. 1st Int. Workshop Learn. Multi-Label Data*, 2009, pp. 101–116.

[33] Y. Yang *et al.*, "Web and personal image annotation by mining label correlation with relaxed visual graph embedding," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1339–1351, Mar. 2012.

[34] E. Spyromitros, G. Tsoumakas, and I. Vlahavas, "An empirical study of lazy multi-label classification algorithms," in *Artificial Intelligence: Theories, Models and Applications*. Berlin, Germany: Springer, 2008, pp. 401–406.

[35] K. Dembczynski, W. Waegeman, W. Cheng, and E. Hüllermeier, "On label dependence and loss minimization in multi-label classification," *Mach. Learn.*, vol. 88, pp. 5–45, 2012.

[36] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Mach. Learn.*, vol. 85, no. 3, pp. 333–359, 2011.

[37] S. Q. Zhao *et al.*, "Weak labeled multi-label active learning for image classification," in *Proc. ACM Int. Conf. Multimedia*, 2015, pp. 1127–1130.

[38] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis Lectures Artif. Intell. Mach. Learn.*, vol. 3, no. 1, pp. 1–130, 2009.

[39] S. C. H. Hoi, R. Jin, J. K. Zhu, and M. R. Lyu, "Semi-supervised SVM batch mode active learning for image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2008, pp. 1–7.

[40] G. Tsoumakas, E. Spyromitros-Xioufis, J. Vilcek, and I. Vlahavas, "Mulan: A Java library for multi-label learning," *J. Mach. Learn. Res.*, vol. 12, no. 7, pp. 2411–2414, 2011.

[41] B. Chen, "Color image analysis by quaternion-type moments," *J. Math. Imag. Vis.*, vol. 51, no. 1, pp. 124–144, 2015.

[42] J. Howe, "The rise of crowdsourcing," *Wired Mag.*, vol. 14, no. 6, pp. 1–4, 2006.

[43] V. S. Sheng, "Studying active learning in the cost-sensitive framework," in *Proc. Hawaii Int. Conf. Syst. Sci.*, 2012, pp. 1097–1106.

[44] B. Gu, V. S. Sheng, and S. Li, "Bi-parameter space partition for cost-sensitive SVM," in *Proc. 24th Int. Conf. Artif. Intell.*, 2015, pp. 3532–3539.

[45] B. Gu and V. S. Sheng, "A robust regularization path algorithm for ν-support vector classification," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2016.2527796.

[46] Z. Xia, X. Wang, X. Sun, Q. Liu, and N. Xiong, "Steganalysis of LSB matching using differences between nonadjacent pixels," *Multimedia Tools Appl.*, vol. 75, no. 4, pp. 1947–1962, 2016.

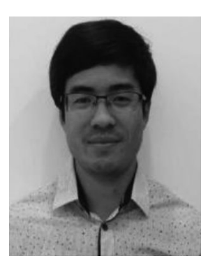
[47] Y. Zheng, B. Jeon, D. Xu, Q. M. Wu, and H. Zhang, "Image segmentation by generalized hierarchical fuzzy C-means algorithm," *J. Intell. Fuzzy Syst.*, vol. 28, no. 2, pp. 961–973, 2015.

[48] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 3, pp. 507–518, Mar. 2015.

Jian Wu received the M.Sc. and Ph.D. degrees in computer science from Soochow University, Suzhou, China, in 2004 and 2012, respectively.

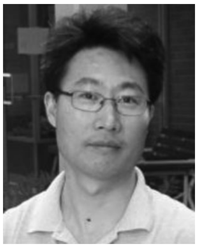
He is currently an Assistant Professor with the Institute of Intelligent Information Processing and Application, Soochow University. He is also currently a Postdoctoral Fellow at Washington University, St. Louis, MO, USA. He has authored or coauthored several papers about multilabel active learning and crowdsourcing in some prestigious journals, such as TKDE, *Information Sciences*, and conferences, such as ACM MM, ICMR, ICME, and CIKM. His research interests include computer vision, image processing, and data mining.

Prof. Wu is a PC Member of several international conferences and a Reviewer of several international journals.



Shiquan Zhao received the B.E. degree in computer science from Anhui Normal University, Wuhu, China, in 2013, and is currently working toward the M.S. degree at the Department of Computer Science and Technology, Soochow University, Suzhou, China.

His research interests include label correlation mining and semisupervised active learning.



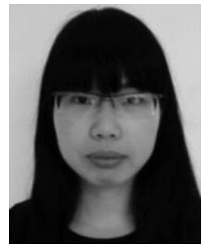
Victor S. Sheng (M'11–SM'14) received the M.Sc. degree from the University of New Brunswick, Fredericton, NB, Canada, and the Ph.D. degree from the University of Western Ontario, London, ON Canada, both in computer science, in 2003 and 2007, respectively.

He is an Associate Professor of computer science with the University of Central Arkansas, Conway, AR, USA, and the Founding Director of Data Analytics Laboratory, University of Central Arkansas. After receiving the Ph.D. degree, he was an Associate Research Scientist and NSERC Postdoctoral Fellow in information systems with the Stern Business School, New York University, New York, NY, USA. His research interests include data mining, machine learning, and related applications.

Prof. Sheng is a Lifetime Member of ACM. He is a PC Member for a number of international conferences and a Reviewer of several international journals. He was the recipient of the runner-up Best Paper Award from KDD'08 and the Best Paper Award from ICDM'11.

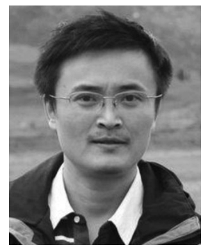
Jing Zhang received the B.E. degree from Anhui University, Hefei, China, in 2003, the M.S. degree in computer science from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2006, and the Ph.D. degree in computer science from Hefei University of Technology, Hefei, China, in 2015.

He is an Assistant Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. He has authored or coauthored several articles in some prestigious journals, such as TKDE, TCYB, JMLR, and *Information Sciences*, and conferences, such as AAAI, CIKM, and ASONAM. His research interests include machine learning and data mining.



Chen Ye received the B.E. degree in computer science in 2013 from Soochow University, Suzhou, China, where she is currently working toward the M.S. degree at the Department of Computer Science and Technology.

Her research interests include label correlation mining and multilabel active learning.



Pengpeng Zhao received the Ph.D. degree in computer science from Soochow University, Suzhou, China, in 2008.

He is an Associate Professor with the Department of Computer Science and Technology, Soochow University. His research interests include the study of data integration, spatial data processing, data mining, machine learning, and crowdsourcing.

Zhiming Cui is a Professor with the Institute of Intelligent Information Processing and Application, Soochow University, Suzhou, China. He has authored or coauthored a large number of papers in computer vision, image processing, and data mining. His research interests include deep web, computer vision, image processing, and pattern recognition.

Prof. Cui is an outstanding expert of Jiangsu Province, China. He presided four National Natural Science Foundations of China.