

RSSVM-based Multi-Instance Learning for Image Categorization

Daxiang LI

School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications
Xi'an, 710121, China

Abstract

Focusing on the problem of natural image categorization, a novel multi-instance learning (MIL) algorithm based on rough set (RS) attribute reduction and support vector machine (SVM) is proposed. This algorithm regards each image as a bag, and low-level visual features of the segmented regions as instances. Firstly, a collection of "visual-words" is generated by Gaussian mixture model (GMM) clustering method, then based on the fuzzy membership function between instance and "visual-word", a fuzzy histogram is computed to represent bag. As a result, every bag is transform into a single sample, which converts MIL problem to a standard supervised learning problem. Finally, RS method is used to reduce the redundant features in the fuzzy histogram, and then standard SVM classifiers are trained for image categorization. Experimental results on the COREL image set show that this algorithm is robust, and the performance is superior to other key existing MIL algorithms.

Keywords: Multi-instance learning; Image categorization; Attribute reduction; Support vector machine.

1. Introduction

Semantic image categorization is bridge the gap between the image's low-level visual features and the high-level semantic, and then labeling an image into predefined semantic categories according to the main object or the scene type. On the Internet or personal computer, with the sharp increase of digital photography, automatic image categorization method becomes increasingly important [1]. In order to establish the link between images and semantics, we must extract the global visual features (color, texture & shape, etc.) or intermediate semantic features [2] or key-points feature [3] of the images, and then combine them with supervised learning methods (i.e. SVM), to carry on image scene classification.

When we carry on image classification under the framework of supervised learning, we face the following two problems [4]: (1) The problem of feature extraction,

which is how to represent the high-level semantics of image. Because the existence of "semantic gap", the semantics of image have fuzziness and ambiguity, and the relation among different semantics is complexly, therefore the semantic represent method is different with the general image vision characteristic represent method, so to seek one kind of effective image semantics represent model is very important. (2)The problem of training samples annotation. Before a learning machine can perform classification, it needs to be trained first, and training samples need to be accurately labeled, and the labeling process can be both time consuming and error-prone [3]. Aiming at these problems, based on rough set (RS) attribute reduction and support vector machine (SVM), a novel multi-instance learning (MIL) algorithm is proposed to perform the image classification. Because MIL allows for coarse labeling at the image level, instead of fine labeling at region level, it can significantly improve the efficiency of image categorization.

The remainder of the paper is organized as follows. In section 2, we introduce recent work related to MIL. Section 3 provides the details of our proposed method, including the method of creating visual vocabulary, the method of computing fuzzy histogram and the method of rough set attribute reduction. Thereafter, experimental results on COREL data set are presented in Section 4, followed by the conclusions in Section 5.

2. Related Work

Multiple instance learning (MIL) has become an active area of investigation in machine learning since it was first put forward for drug activity prediction [5]. In the MIL problem, the training samples are regard as bags, where each bag consists of multiple instances, and each instance is represented as a feature vector. According to the original MIL definition, a bag is labeled as positive if at least one of its instances is positive, and it is labeled as negative if all of its instances are negative. The goal of a MIL algorithm is

to generate a classifier that will classify unseen bags correctly.

During the past decade, many multi-instance learning algorithms have been presented, including axis-parallel rectangles [5], Diverse Density (DD) [6], EM-DD [7], k-nearest neighbor [8], multi-label multi-instance learning (MLMIL) [9], and neural network algorithm [10] et al. It is difficult to list all existing methods. Here, we main focus on methods based on the SVM, which have been highly successfully used in many machine-learning problems. Andrews et al. [11] modified the SVM formulation, and presented mi-SVM and MI-SVM algorithms, however, unlike the standard SVM, they lead to non-convex optimization problems that suffer from local minima. Then Gellar et al. [12] applied deterministic annealing to solve this non-convex optimization problem, and this method could find better local minima of the objective function. Gartner et al. [13] designed kernels directly on the bags, using a standard SVM to solve MIL problem, as the instance labels were unavailable, and the MI kernel implicitly assumed all instances in a bag to be equally important, obviously, the assumption made by the MI kernel was very crude. Aiming at this problem, Kwok et al. [14] designed marginalized multi-instance kernels by considering that the contribution of different instances can be different. Chen et al. [15] have proposed the DD-SVM method, which employed the DD method to learn a set of instance prototypes and then mapped the bags to a new feature space based on the instance prototypes. Zhou et al. [16] proposed the MissSVM method by regarding instances of negative bags as labeled examples, while those of positive bags as unlabeled examples. Recently, Chen et al. [17] also devised a new algorithm called Multi-Instance Learning via Embedded Instance Selection

(MILES) to solve multiple instance problems.

Recently, besides these supervised multi-instance learning algorithms, many multiple-instance semi-supervised (SSL) learning algorithms have been presented also. Rahmani et al. [18] combined a modified version of DD with graph-based semi-supervised algorithms, first presented a graph-based semi-supervised MIL algorithm, named MISSL, and applied it in object-based image retrieval. However, the work of Rahmani gave a solution to MISSL by combining MIL with SSL algorithms; it could not show significant improvements compared with traditional supervised CBIR methods. One main reason may be that the constructed bag-level graph had a bias towards bags with more instances. Wang et al. [19] presented a graph-based multi-instance learning (GMIL) algorithm, unlike the loosely coupled manner in MISSL, they considered the multi-instance and semi-supervised property simultaneously, and gave a directly solution to the MISSL problem by minimizing a cost function of the regularization framework, which explicitly took into account three kinds of data.

3. The Proposed Approach--RSSVM-MIL

Transforms the multi-instance bag into a single sample, so that standard supervised learning methods (i.e. SVM method) can be used to solve the MIL problem directly, such as DD-SVM [15] and MILES [17] algorithms, have become a kind of very effective MIL method. The common idea is: Constructs a concept space, and use a non-linear mapping function embedded each bag into a point in this space, so each multi-instance bag be transformed into a single sample, then the standard SVM method has been used to solve MIL problem. Unfortunately, these methods

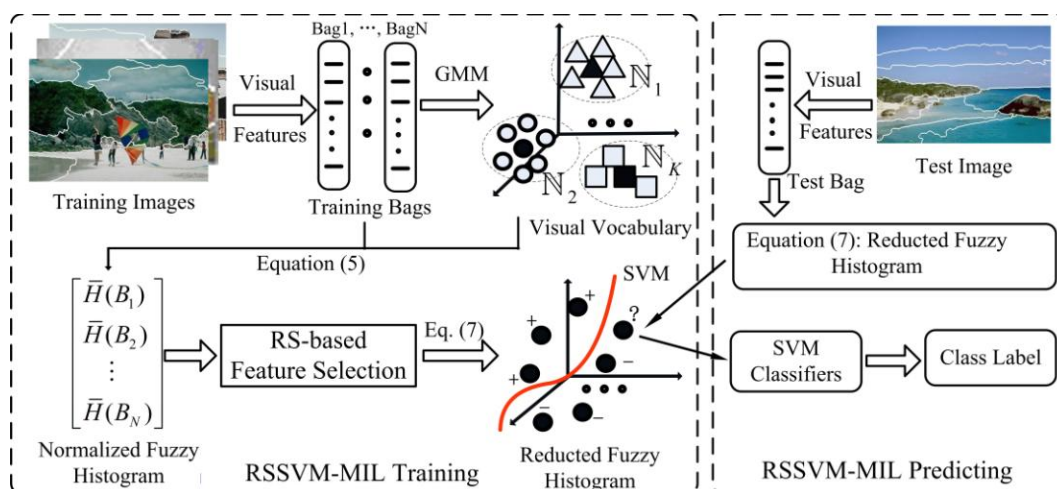


Fig.1 The framework of the proposed RSSVM-MIL algorithm

exist the following problems: (1)The construction of the concept space and the definition of the projection function are done by experience, lack of theoretical basis; (2)Not only the efficiency of the concept space constructs is low, but also the dimension of the projection feature is high, therefore, the efficiency of these algorithms are very low. Motivated by DD-SVM [15] and MILES [17], in this paper, we present a “visual-word” based method to convert the MIL problem into a standard supervised learning problem, and then combine with RS (rough set) attribute reduction and SVM, a novel MIL algorithm, named RSSVM-MIL, is proposed. The framework of the proposed algorithm is shown in Figure 1.

3.1 Creating Visual Vocabulary

In the proposed approach, all images are first segmented into homogeneous regions by one image segment method, and the low-level visual features (i.e. color, texture, and shape) are extracted for every region. In the MIL framework, each image is regarded as a bag, and the visual features of the segmented regions as instances. We denote the i -th positive bags as B_i^+ , and the j -th instance in this bag as x_{ij}^+ , the bag B_i^+ consists of n_i^+ instances $x_{ij}^+ \in \mathbb{R}^D$, $j=1,2,\dots,n_i^+$, where \mathbb{R}^D denote D -dimensional instance feature space. Similarly, B_i^- , x_{ij}^- , and n_i^- represent the i -th negative bag, the j -th instance and the number of instances within bag, respectively, the number of positive (negative) bags is denoted as l^+ (l^-), for the sake of convenience, when we line up all instances in all training bags together, and re-index these instances as

$$InstSet = \{x_i \mid i=1,\dots,n\} \quad (1)$$

Where $n = \sum_{i=1}^{l^+} n_i^+ + \sum_{i=1}^{l^-} n_i^-$.

To create a visual vocabulary, we use the unsupervised training method of Gaussian mixture model (GMM) clustering. Given a set of instance and assume that we know there are K clusters, then the density $p(x)$ of a Gaussian mixture distribution is a convex combination of K D -dimensional normal densities or components [20, 21],

$$p(x) = \sum_{k=1}^K w_k \mathbb{N}(x \mid \mu_k, \Sigma_k) \quad (2)$$

for $x \in \mathbb{R}^D$, where w_k, μ_k, Σ_k are the mixing coefficient, mean and covariance matrix respectively of the k -th component. Now, given a set of i.i.d. observations or data points $InstSet = \{x_i \mid i=1,\dots,n\}$, the maximum likelihood (ML) estimate is used to obtain the parameters of each component $\mathbb{N}(x \mid \mu_k, \Sigma_k)$. In this paper, each component $\mathbb{N}(x \mid \mu_k, \Sigma_k)$, denoted as \mathbb{N}_k , is regarded as a “visual-

word”, and all the components are regarded as “visual vocabulary”.

3.2 Computing Fuzzy-Histogram

Let $\Omega = \{\mathbb{N}_k : k=1,2,\dots,K\}$ be the “visual vocabulary” obtained by GMM clustering method. If we regard every “visual-word” \mathbb{N}_k as a concept point, and the fuzzy membership function between instance x and \mathbb{N}_k is defined as:

$$f(\mathbb{N}_k, x) = p(x \in \mathbb{N}(x \mid \mu_k, \Sigma_k) \mid x) \\ = \frac{1}{(2\pi)^{D/2} |\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)\right) \quad (3)$$

Therefore, we define bag’s fuzzy histogram $H(B_i)$ for a bag $B_i = \{x_{ij} \mid j=1,2,\dots,n_i\}$, as:

$$\begin{cases} H(B_i) = [s(\mathbb{N}_1, B_i), s(\mathbb{N}_2, B_i), \dots, s(\mathbb{N}_K, B_i)] \\ s(\mathbb{N}_k, B_i) = \sum_{j=1}^{n_i} f(\mathbb{N}_k, x_{ij}) \end{cases} \quad (4)$$

where $k=1,2,\dots,K$. $s(\mathbb{N}_k, B_i)$ can be interpreted as a fuzzy frequency of “visual-word” \mathbb{N}_k in a bag B_i , which is determined by all the instances in this bag. Note that, before use the fuzzy histogram vector $H(B_i)$, we must carry out the following normalization:

$$\begin{cases} \bar{H}(B_i) = [\bar{s}_1, \bar{s}_2, \dots, \bar{s}_K] \\ \bar{s}_k = s(\mathbb{N}_k, B_i) / \sum_{t=1}^K s(\mathbb{N}_t, B_i) \end{cases} \quad (5)$$

where $k=1,2,\dots,K$. By Equation (5), every bag will be represented by a K -dimensional normalized fuzzy histogram. As a result, the bag is transformed into a single sample. If a bag is positive, the corresponding sample is labeled +1, while labeled as -1, so the MIL problem is transformed into a standard supervised learning problem.

3.3 An Example

Similar to MILES algorithm [17], we also formulate a multiple-instance problem where each instance is generated by one of the following two-dimensional probability distributions: N1-N([0, 0], I), N2-N([3, 3], I), N3-N([-3, 3], I), N4-N([-3, -3], I) and N5-N([3, -3], I), where $\mathbb{N}([x, y], I)$ denotes the normal distribution with mean $[x, y]$ and identity covariance matrix I . Each bag comprises at mean 5 instances, a bag is labeled positive if it contains instances from at least two different distributions among N1 and N2. Otherwise, the bag is negative. Using this model, we generated 50 positive bags and 50 negative bags with a total of 500 instances. After all the data normalized to [0, 1], Fig. 2(A) depicts all the instances on a two-dimensional plane. Note that instances from negative bags mingle with

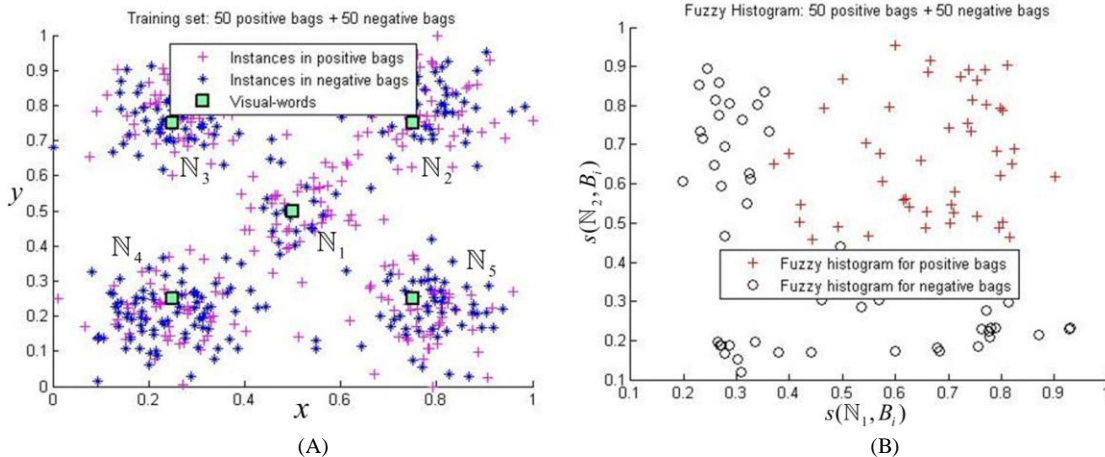


Fig. 2 Two-dimensional visualization of the instances and bags. (A) Instances and visual-words on a two-dimensional plane after data normalized. (B) The normalized fuzzy histogram computed use two manually selected visual-words.

those from positive bags because a negative bag may include instances from any one, but only one, of the distributions N_1 and N_2 .

Ideally, all the instances should be clustered into 5 categories, then the center points of each data model are “visual-words”, as shown in Figure 2 (A) at the green “□” position, denoted as $N_1 \sim N_5$, and the fuzzy histogram of each bag should be a 5-dimensional vectors by Equation (6). Now, we only use the “visual-word” of N_1 and N_2 as a reduced visual vocabulary, the fuzzy histogram of each bag was shown in Figure 2(B).

Because, each positive bag must contain at least one instance from the N_1 and N_2 simultaneously, so the fuzzy frequency between each positive bag and the “visual-word” N_1 and N_2 are high simultaneously. While a negative bag would not include instance from the N_1 and N_2 simultaneously, so the fuzzy frequency between each negative bag and the “visual-word” N_1 and N_2 cannot are great simultaneously. So it is very easy to classify positive bag and negative bag by SVM method, which also can be seen from Figure 2 (B) clearly.

3.4 RSSVM-MIL Algorithm

It can be seen form Section 3.3, as long as we correctly select two “visual-words” to construct a visual vocabulary from the five “visual-words”, it will be able to distinguish positive bag and negative bag. It also shows that simply using the GMM method to create visual vocabulary, there must be much redundant information in the fuzzy histogram, in order to improve the efficiency of SVM training and classification accuracy, rough sets (RS) method is used to

reduce the redundant features that in fuzzy histogram firstly.

3.4.1 Algorithm design for feature selection

The motivation of rough set based feature selection is to select a minimal attribute subset, which has the same characterizing power as the whole attribute set, and without any redundant attribute. In other words, the dependency of the selected attributes is the same as that of the original attributes. Moreover, the dependency will decrease if any selected attribute is deleted. There are two key problems in constructing a feature selection algorithm. One is how to evaluate the selected features; the other is how to search for a good feature subset. In this paper, because the projection features are numerical attributes, so we adopt the neighborhood rough set model for attribute reduction, which has been described in the reference [22-24]. we will introduce them in the following.

Because the significance of an attribute can be used to evaluate the goodness of selected features, so given a decision system $\langle U, C \cup D, N \rangle$, $B \subseteq C$, $a \notin B$, the significance of an attribute is [23-25]

$$Sig(a, B, D) = \gamma_{B \cup a}(D) - \gamma_B(D) \quad (6)$$

where $\gamma_B(D) = |POS_B(D)|/|U|$ is the dependency degree of D to B .

It is a combinational optimization problem to find all of the reducts. There are $2^{|C|}$ combinations of attribute subsets. It is not practical to search all of the reducts in $2^{|C|}$ combinations. Fortunately, in practice, we usually just require one of the reducts to train a classifier, and we do not much care whether the reduct is the minimal one. Then

a tradeoff solution can be constructed, a forward greedy search algorithm for attribute reduction can be formulated as follows [22-24].

Algorithm 1: Feature selection algorithm.

Input: $\langle U, C \cup D \rangle$, ε and δ (control the size of the neighborhood);

Output: *reduct* red .

1: $\Phi \rightarrow red$; // red is the pool to contain the selected attributes

2: For each $a_i \in C - red$

3: Compute $SIG(a_i, red, D) = \gamma_{red \cup a_i}(D) - \gamma_{red}(D)$

4: end

5: select the attribute a_k satisfying

$$SIG(a_k, red, D) = \max_i(SIG(a_i, red, D))$$

6: If $SIG(a_k, red, D) > 0$

7: $a_k \cup red \rightarrow red$

8: go to step2

9: else

10: return red

11: end if

3.4.2 RSSVM-MIL

Because there is massive redundancy information in the fuzzy histogram of each bag inevitably, so in this paper, we regard each bag's fuzzy histogram as condition attributes, its concept label as decision feature, and then use **Algorithm 1** to reduce the redundant attribute. Let N'_i be the "visual-word" corresponds to the attribute retained after RS feature selection, so a *reduced visual vocabulary*, denoted as $\bar{\Omega} = \{N'_1, N'_2, \dots, N'_M\}$, $M \ll K$, is obtained. The reduced fuzzy histogram can be computed in the same way as Equation (4) and (5), is:

$$\begin{cases} \phi(B_i) = \bar{H}(B_i) = [\bar{s}_1, \bar{s}_2, \dots, \bar{s}_M] \\ \bar{s}_k = s(N'_k, B_i) / \sum_{t=1}^M s(N'_t, B_i) \end{cases} \quad (7)$$

where $k = 1, 2, \dots, M$. By $\phi(B_i)$, every bag is transformed into a single samples, so bag-level SVM classifier can be trained use these samples. The maximum margin formulation of the MIL problem is given as the following quadratic optimization problem [15]:

$$\begin{cases} \max L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j K(\phi(B_i), \phi(B_j)) \\ s.t. \sum_{i=1}^N y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, N. \end{cases} \quad (8)$$

Where $K(x_1, x_2)$ is a kernel function, $C > 0$ is the penalty parameter of the error sample, to control the trade-off

between accuracy and regularization, $N = l^+ + l^-$. The bag classifier is then defined by α^* as

$$label(B) = sign(\sum_{i=1}^N y_i \alpha_i^* K(\phi(B_i), \phi(B)) + b^*) \quad (9)$$

Where b^* is chosen so that

$$y_j (\sum_{i=1}^N y_i \alpha_i^* K(\bar{\phi}(B_i), \bar{\phi}(B_j)) + b^*) - 1 = 0 \quad (10)$$

For any α_j^* with $C > \alpha_j^* > 0$.

Finally, the detailed steps of RSSVM-MIL algorithm can be summarized as follows

Algorithm 2: RSSVM-MIL Algorithm

1. RSSVM-MIL Training

Input: A set of labeled bags D and K value;

Output: reduced visual vocabulary $\bar{\Omega}$ and SVM classifier (α^*, b^*) ;

Initialize: Set $S = \Phi$;

Step 1: Line up all instances in all labeled bags together, denoted as $InstSet$, then use GMM clustering method, obtain a collection of "visual-word";

Step 2: For $\forall B_i \in D$

Calculate its fuzzy histogram $\bar{H}(B_i)$ by Equation (5), and add $(\bar{H}(B_i), y_i)$ to S , here y_i is the label of B_i 's.

Step 3: In the S , use **Algorithm 1** to reduce its redundant information, and obtain a reduced fuzzy histogram feature set \bar{S} , then according to the attribute retains in \bar{S} , a reduced visual vocabulary $\bar{\Omega}$ is obtained;

Step 4: Train a standard SVM classifier (α^*, b^*) use the training sample within \bar{S} ;

2. RSSVM-MIL predicting

Let B be an unlabeled bag. Firstly, use Equation (7) to extract its reduced fuzzy histogram $\bar{\phi}(B)$, then use SVM classifier (α^*, b^*) (Eq. (9)) to predict its label.

4. Experiments and Analysis

We evaluate the proposed method based on some publicly available benchmarks: the MUSK¹ data sets for drug activity prediction and the COREL set for image classification. In RSSVM-MIL method, the Libsvm2.88 software is used to train SVM Classifiers, and the Gaussian kernel, $K(x, z) = \exp(-g \|x - z\|^2)$ is used. One-against-rest strategy is employed for multi-class tasks, and the winner of all SVM Classifiers decides the final predicted class label. In **Algorithm 1**, we set $\varepsilon = 0.8$ and $\delta = 0.2$, to determine the parameters g (Gaussian kernel radius) and C (penalty factor) in SVM, we conduct a 2-fold cross-validation on the

training set use the grid.py tools provided in the Libsvm software package.

4.1 Drug Activity Prediction

The MUSK data sets, including MUSK1 and MUSK2, are benchmark data sets for MIL. Both data sets consist of descriptions of molecules. Specifically, a bag represents a molecule, and instances in a bag represent low-energy conformations of the molecule. Each instance (or conformation) is defined by a 166-dimensional feature vector that describing the surface of a low-energy conformation. MUSK1 contains 47 musk molecules (positive bags) and 45 similar non-musk molecules (negative bags). MUSK2 contains 39 musk molecules and 63 non-musk molecules. A total of 72 molecules are shared between MUSK1 and MUSK2. MUSK1 contains approximately 5.17 instances per bag and MUSK2 contains 64.69 instances per bag.

Three parameters K (GMM clustering), g and C (SVM) need to be specified for RSSVM-MIL. Firstly, we fixed $K=100$, then to search parameters g and C use grid.gy tool, we found that $g=0.5$ and $C=16$ gave the minimum 2-fold cross-validation error on MUSK1, and $g=4.768 \times 10^{-4}$ and $C=1.677 \times 10^3$ gave the best 2-fold cross-validation performance on MUSK2, then we fixed these values in the subsequent experiments. The average accuracy and 95% confidence interval of the results over 10 runs of 10-fold cross-validation are reported in Table 1. We also list some other results on the same data sets for comparison.

Table 1: Comparison of averaged 10-fold cross-validation accuracies on MUSK data sets.

<i>MIL Algorithms</i>	<i>MUSK1</i>	<i>MUSK2</i>
RSSVM-MIL	90.2	86.3
DD-SVM	85.8	91.3
MILES	84.8	84.9
MI-SVM	77.9	84.3
mi-SVM	87.4	83.6
DD	88.9	82.5
GMIL-M	91.2	84.2
MissSVM	87.6	80.0
IAPR	92.4	89.2

From the Table 1, we can see that the RSSVM-MIL method achieves performance comparable with the state-of-the-art results, and the performance of our methods on MUSK1 is better than DD-SVM [15] and MILES [17] methods, but inferior to IAPR [5] and GMIL-M [19] methods. The reasons maybe: The IAPR was designed especially for drug activity prediction but not for general MIL tasks, and

the parameters to maximize the test set (not training set) performance, so it is not a fair comparison. Furthermore, GMIL-M is a semi-supervised learning algorithm, and uses the information of the all unlabeled bags to train classifier.

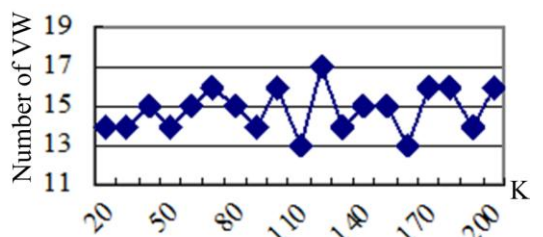
4.2 Image Categorization

To evaluate our method in image categorization problem, we applied it on the COREL data set[15], a widely used standard benchmark data set for image categorization. The data set consists of 2000 images in JPEG format with size 256×384 or 384×256 . There are a total of twenty different categories, each containing 100 images. The twenty categories(labeled form 0 to 19) are: Africa people and villages, beaches, buildings, buses, dinosaurs, elephants, flowers, horses, mountains and glaciers, food, dogs, lizards, fashion models, sunset scenes, cars, waterfalls, antique furniture, battle ships, skiing, and desserts. In this data set, each image is pre-segmented into around 5 patches. Color, texture and shape features have already been extracted for each patch, and form a set of 9-dimension feature vectors (Chen and Wang, 2004). In our experiments, these features are exactly normalized to the range from 0 to 1. Treat each image as a bag and each patch in this image as an instance in this bag. The experimental routine described in MILES [17] is adopted here, in detail, the original data set is used as two data sets. The first one (i.e. **Corel 1k**) used the first ten categorizes' images while the second (i.e. **Corel 2k**) used all the categories' images. For each data set, images within each category are in experiment is repeated for 5 times for 5 random splits, and the average accuracy as well as 95% confidence interval was reported.

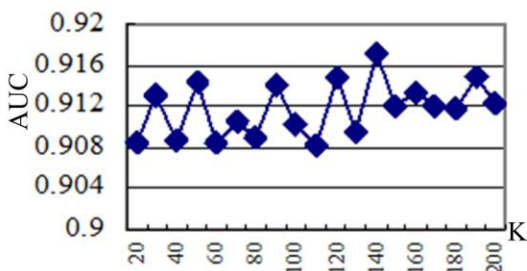
4.2.1 Sensitivity to the number of K

When create visual vocabulary, one important parameter K needs to be predefined in GMM clustering method. In order to test the sensitivity of RSSVM-MIL to the number of K , we chose K from 20 to 200 with step size 10, randomly selected 50 images from Category 7 (Horses) as positive images, and other 50 negative images are randomly selected as negative images from the other 19 categories. Experiment is repeated for 5 times for 5 random select. The average number of the "visual-words" after RS reduction is shown in Figure 3(A), and the average AUC (the area under ROC curve) values are show in Figure 3(B).

As seen in Figure 3(B), the K values have little effect to the accuracy of Classification, and the average AUC values almost maintained at about 0.91. The reason is: the bigger K is, the more "visual-word" will be extracted by GMM, but after RS reduction, the remaining "visual-word" almost is equal. This indicates that when K is bigger, the more redundant information will exist in the fuzzy histogram, but



(A) Number of “visual-word” after RS reduction



(B) Different K value impact to the classification accuracy

Fig 3 Different K value impact to the performance of the RSSVM-MIL algorithms

the redundant information will be reduced by the RS method, it can be seen from Figure 3(A). This experiment indicates that RSSVM-MIL is robust to K values.

4.2.2 Multi-Class Categorization Accuracy

In following experiments, we set $K=100$, and first report the confusion matrix of the RSSVM-MIL method in Table 2 based on **Corel 1K** (Category 0 to Category 9). where each row lists the average percentages of images in a specific category classified to each of the 10 categories, therefore, the numbers on the diagonal show the classification accuracy for each category and off-diagonal entries indicate classification errors. Table 2 reveals that RSSVM-MIL works well on most categorizes, but the largest errors are errors between Category 1 (i.e. Beach) and Category 8

(i.e. Mountains and glaciers): 28.1% Beach images were misclassified as Mountains and glaciers while 13.5% Mountains and glaciers images were misclassified as Beach. This phenomenon has appeared in DD-SVM and MILES. These high classification errors are due to the fact that many images of these two categories contain semantically related and visually similar regions, such as those corresponding to sky, mountain, river, lake and ocean. Figure 4 show some mislabeled images take form these two categories.



Fig.4 some images are misclassified between Beaches and Mountains

Based on **Corel 1K** and **Corel 2K**, the overall classification accuracy of RSSVM-MIL ($K=100$) compared with other existing multi-instance learning algorithms was reported in Table 3, including MILES [17], DD-SVM [15], MI-SVM [11] and MissSVM [16], et al.. The numbers listed are the average classification accuracies (in percent) over 5 random test sets and the corresponding 95 percent confidence intervals. The Corel 1k data set contains images from Category 0 to Category 9. The Corel 2k data set contains images from all 20 categories. Training and test sets are of equal size.

In the image classification experiments based on COREL data set, as shown in Table 3, the performance of the

Table 2: The confusion matrix of RSSVM-MIL on Corel 1k (%)

	<i>Cat. 0</i>	<i>Cat. 1</i>	<i>Cat. 2</i>	<i>Cat. 3</i>	<i>Cat. 4</i>	<i>Cat. 5</i>	<i>Cat. 6</i>	<i>Cat. 7</i>	<i>Cat. 8</i>	<i>Cat. 9</i>
<i>Cat. 0</i>	78.1	2.8	2.5	1.1	0	5.2	1.1	1	2.1	6.1
<i>Cat. 1</i>	2.1	61.6	1.5	3.2	0	1.9	1.6	0	28.1	0
<i>Cat. 2</i>	4.1	8.8	68.5	4.9	0	3.3	0.6	0	8.2	1.6
<i>Cat. 3</i>	1.6	1.1	3.1	90	0	0.2	0	0.1	1.1	2.8
<i>Cat. 4</i>	0.1	0	0	0.2	99	0.3	0	0	0	0.4
<i>Cat. 5</i>	3.1	2.1	3.2	0	0.1	88.2	0	0.2	4.1	0
<i>Cat. 6</i>	0.1	0	1.1	0.3	0.5	0	96.1	0.3	0.5	1.1
<i>Cat. 7</i>	1.5	0.2	0.5	0.4	0.3	2	0.7	93.2	1	0.2
<i>Cat. 8</i>	0	13.5	1.5	4.3	0	0	0.7	0	78.8	1.2
<i>Cat. 9</i>	4.9	2.5	0	0.4	0	1.6	0.7	0.3	0	89.6

Note: The numbers are the average percentage of 5 times repeated experiments.

Table 3: Comparing the Image Categorization Accuracy Obtained Using RSSVM-MIL with Other MIL Methods.

<i>MIL Algorithms</i>	<i>Corel 1k</i>	<i>Corel 2k</i>
RSSVM-MIL	85.1 ± 1.1	70.2 ± 1.2
DD-SVM	81.5 ± 3.0	67.5 ± 0.8
MILES	82.6 ± 1.2	68.7 ± 1.4
MI-SVM	74.7 ± 0.6	54.6 ± 1.5
MissSVM	78.0 ± 2.2	65.2 ± 3.1

RSSVM-MIL algorithm is superior to other MIL algorithms. This reason is: Each "visual-word" extracted by GMM clustering method represents a group of image regions with similar visual characteristic, which has an explicit high-level semantic concept, and the fuzzy histogram reflects not only the probability of a image containing these "visual-words", but also the co-occurrence relationship of them in a image. It has the ability to represent the scene semantics or the simple semantics of a image simultaneously.

4.3 Speed

On average, the learning of each binary classifier using a training data set of 100 images (4.31 regions per image) take around 9 seconds of CPU time in matlab 2008 on a AMD 4200+ PC running the windows XP operating system. The time mainly spends in three aspects: (1) Using GMM method to create visual vocabulary; (2) RS attribute reduction. If K and n are the numbers of features and samples respectively, the worst case of computational complexity of reduction is $O(K^2 n \log n)$; (3) SVM training. The training set consists of 100 images, there are around 439 different instances in all bags, when $K=100$, GMM clustering spends around 3.86 seconds to extract "visual-word", RS attribute reduction spend 3.16 seconds (matlab code), and to train a binary SVM classifier took 1.72 seconds (C code). Compared with the DD-SVM and MILES algorithms in the same size of training set, the efficiency is improved significantly.

5 Conclusions

In this paper, we have proposed a new MIL algorithm for image classification, named RSSVM-MIL, which is based on RS attribute reduction and SVM. In order to transform bag to a single sample, we first presented a novel method to create visual vocabulary, and defined a fuzzy membership function to computer the fuzzy histogram, which converts MIL problem to a supervised learning problem. One technique based on RS attribute reduction is applied to reduce the redundant attributes in the fuzzy histogram, and then bag-level SVM classifiers are trained

for image classification. We evaluate the performance of RSSVM-MIL on the COREL data set, comparative experimental results show that this algorithm is feasible, and the performance is competitive with other state-of-the-art MIL algorithms.

Acknowledgements

The author acknowledges the support of the National Natural Scientific Youth Foundation (Grant 61003255), the Natural Science Research Foundation of Education Department of the Shaanxi Province (Grant 12JK0734, 12JK0504) and the Youth Research Foundation of the Xi'an University of Posts and Telecommunications (Grant 1090428), China.

References

- [1] M. A. Helala, M. M. Selim, H. H. Zayed, "A Content Based Image Retrieval Approach Based On Principal Regions Detection", International Journal of Computer Science Issues, Vol. 9, No. 4, 2012, pp. 204-213.
- [2] A. Oliva, and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope", International Journal on Computer Vision, Vol. 42, No. 3, 2001, pp. 145-175.
- [3] J. Yang, Y. G. Jiang, G. H. Alexander, et al., "Evaluating bag-of-visual-words representations in scene classification", In Proceedings of the international workshop on Workshop on multimedia information retrieval, 2007, Vol. 1, pp. 197-206.
- [4] D. X. Li, J. L. Fan, D. W. Wang, et al., "Latent Topic Based Multi-instance Learning Method for Localized Content-Based Image Retrieval", Computers & Mathematics with Applications, Vol. 64, No. 4, 2012, pp. 500-510.
- [5] T. G. Dietterich, R. H. Lathrop, Lozano-Pérez T, "Solving the multiple instance problem with axis-parallel rectangles", Artificial Intelligence, Vol. 89, No. 12, 1997, pp. 31-71.
- [6] O. Maron, A. L. Ratan, "Multiple-instance learning for natural scene classification", in Proceedings of the 15th International Conference on Machine Learning, 1998, Madison, WI, pp. 341-349.
- [7] Q. Zhang, S. A. Goldman, "EM-DD: an improved multiple-instance learning technique", in: Advances in Neural Information Processing Systems 14, 2002, Cambridge, CA, pp. 1073-1080.
- [8] J. Wang, J. D. Zucke, "Solving the multiple-instance problem: a lazy learning approach", In: Proceedings of the 17th International Conference on Machine Learning, 2000, San Francisco, CA, pp. 1119-1125.
- [9] Z. H. Zhou, and M. L. Zhang, "Multi-instance multi-label learning with application to scene classification", Advances in Neural Information Processing Systems 19, 2007, pp. 1609-1616.

- [10] M. L. Zhang, and Z. H. Zhou, "A Multi-Instance Regression Algorithm Based on Neural Network", *Journal of Software*, Vol. 14, No. 7, 2003, pp. 1238-1242.
- [11] S. Andrews, T. Hofmann, I. Tsochantaris, "Multiple instance learning with generalized support vector machines", in *Proceedings of the 18th National Conference on Artificial Intelligence*, Edmonton, 2002, Canada, pp. 943-944.
- [12] P. V. Gehler, and O. Chapelle, "Deterministic Annealing for Multiple-Instance Learning", in *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics*, 2007, San Juan, Puerto Rico, pp. 123-130.
- [13] T. Gartner, P. A. Flach, A. Kowalczyk, et al., "Smola. Multi-instance kernels", in *proceedings of the 19th International Conference on Machine Learning*, 2002, Sydney, Australia, pp. 179-186.
- [14] J. T. Kwok, and P. M. Cheung, "Marginalized multi-instance kernels", In: *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2007, Hyderabad, India, pp. 901-906.
- [15] Y. X. Chen, J. Z. Wang, "Image categorization by learning and reasoning with regions", *Journal of Machine Learning Research*, Vol. 5, No. 8, 2004, pp. 913-939.
- [16] Z. H. Zhou, and J. M. Xu, "On the Relation Between Multi-Instance Learning and Semi-Supervised Learning", in *Proceedings of the 24th ICML*, Corvallis, Oregon June, 2007, pp. 1167-1174.
- [17] Y. X. Chen, J. B. Bi, and J. Z. Wang, "MILES: Multiple-Instance Learning via Embedded Instance Selection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 12, 2006, pp. 1931-1947.
- [18] R. Rahmani, and A. S. Goldman, "MISSL: multiple-instance semi-supervised learning", in *Proceedings of the 23rd international conference on Machine learning*, Pittsburgh, Pennsylvania, June, 2006, pp. 705-712.
- [19] C. H. Wang, L. Zhang, H. J. Zhang, "Graph-based multiple-instance learning for object-based image retrieval", In: *Proceeding of the 1st ACM international conference on Multimedia information retrieval*, 2008, Vancouver, British Columbia, Canada, pp. 156-163.
- [20] J. Farquhar, S. Szedmak, H. Meng, et al., "Improving 'bag-of-keypoints' image categorization: Generative models and PDF-kernels", *Technical Report*, University of Southampton, 2005.
- [21] F. Perronnin, "Universal and adapted vocabularies for generic visual categorization", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 7, 2008, pp. 1243-1256.
- [22] Q. H. Hu, Z. X. Xie, D. R. Yu, "Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation", *Pattern Recognition*, Vol. 40, No. 12, 2007, pp. 3509-3521.
- [23] Q. H. Hu, D. R. Yu, Z. X. Xie, "Neighborhood classifiers", *Expert Systems with Applications*, Vol. 34, No. 2, 2008, pp. 866-876.
- [24] Q. H. Hu, D. R. Yu, J. F. Liu, et al., "Neighborhood rough set based heterogeneous feature subset selection", *Information Sciences*, Vol. 178, NO. 18, 2008, pp. 3577-3594.

Daxiang LI received the B. S in electrical engineering in 1997, the M.S in 2005 and PhD in 2011 from Northw estern University (xi'an China) in computer softw are and theory. He is now a lecturer at the school of telecommunication and Information engineering, Xi'an University of Posts and Telecommunications. His areas of interest are image processing, machine learning, image semantic analysis and content based image retrieval (CBIR).