Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2018) 26: 593 – 604

doi:10.3906/elk-1703-319

DEMIAL: an active learning framework for multiple instance image classification using dictionary ensembles

Gökhan KOÇYİĞİT, Yusuf YASLAN*
Computer Engineering Department, Faculty of Computer and Informatics Engineering, İstanbul Technical University, İstanbul, Turkey

Received: 27.03.2017 • **Accepted/Published Online:** 13.11.2017 • **Final Version:** 26.01.2018

Abstract: In many machine-learning applications, each data point can be represented as a set of instances that create multiple instance learning (MIL) problems. Due to the structure of images, different regions can be interpreted as instances. Thus, multiple instances can be obtained for each image, which makes image categorization a MIL problem. With abundant unlabeled image data, this MIL problem can be solved using active learning algorithms. Active learning is a framework that utilizes unlabeled data in which labeling samples is a labor-intensive and expensive task. Although many effective MIL active learning methods have been developed, most of the existing algorithms do not take into account classifier and feature representation. In this work, we develop DEMIAL (dictionary ensembles multiple instance active learning), a multiple instance active learning method that utilizes sparse feature representation and classifier ensemble techniques. In the proposed active learning framework, we employ dictionary learning and compare uncertainty- and entropy-based instance selection techniques. Experimental results show that classifier ensembles benefit from active learning and the DEMIAL algorithm outperforms the kernel-based multiple instance active learning framework.

Key words: Multiple instance learning, active learning, dictionary learning, sparse coding, classifier ensembles

1. Introduction

Multiple instance learning (MIL) is a machine-learning framework proposed by Dietterich et al. for the prediction of drug molecule activity [1]. In the MIL framework, each item of data contains a set of instances that forms a bag. The data labels are associated with the bags. For each class label, the bag is considered positive if at least one instance of the bag is positive. If all the instances are negative, the bag label is considered negative. MIL has become widely used in many applications, including drug activity prediction, image classification [2], retrieval [3], and text categorization [4]. For the image categorization problem, separated regions of an image are related to instances and the image itself becomes a bag, thus making image categorization an MIL problem [3].

Many different MIL algorithms have been proposed in the literature. These algorithms can fall under three different categories based on whether they work on instance space (IS), bag space (BS), or embedded space (ES) [5]. In IS, the classifiers are trained on the instance level and bags are classified based on instances. Some of the instance space algorithms are axis-parallel rectangle [1], diverse density (DD) method [6], and the EM-DD algorithm [7]. However, in some applications, only bag labels and instance features are available, which makes the MIL problem more challenging. To solve this problem, BS algorithms are proposed. In BS methods, the

*Correspondence: yyaslan@itu.edu.tr

classifiers are trained on the bag level and these algorithms deal with varying numbers of instances at each bag. While the bags are nonvector entities, one has to define a function that compares two bags in the dataset [8,9]. On the other hand, ES calculates the bag features explicitly with an appropriate mapping function. Since the set of instance vectors are converted into a single feature vector, one can train commonly used classifiers such as SVM and DD algorithm combined with SVM, as in [3]. Chen et al. [2] have proposed the MILES algorithm, in which each bag feature is mapped into a single feature space using similarities between prototype vectors. Alternatively, Song et al. [10] used sparse coding to map features into sparse feature space and obtained a classifier ensemble-based MIL algorithm, named the sparse coding and classifier ensemble-based multi-instance learning (SCCE-MIL) algorithm.

Sparse representations have lately gained a lot of attention from the image processing and signal processing community. This is because signals (image, audio, etc.) can be well approximated by the linear combination of a few elements (called atoms) of a redundant basis (called a dictionary). Recent publications show that it can lead to state-of-the art results in image restoration, denoising etc. [11,12]. On the other hand, nowadays it is easy to obtain a vast amount of unlabeled data. In order to use this abundant data in a supervised learning framework, each data point has to be labeled manually, which can be an overwhelming task. Instead of labeling all the unlabeled data, one can "smartly" select some of them using an active learning framework. Active learning is a well-known and widely used framework for many machine-learning problems where obtaining data labels is difficult and expensive. In this framework, an active learner is allowed to choose the most informative unlabeled data and ask (query) its label from the oracle. There are some studies that use an active learning framework to solve MIL problems. Settles et al. proposed a multiple instance (MI) active learning method in which the classifier queries instance-level labels from the selected bags [13]. Similarly, Fu et al. adopted a Fisher information matrix-based query strategy for bag-level active learning [14]. Moreover, Liu and Dong et al. proposed a query strategy for MIL by querying bags only, instances only, and a mix of these strategies [15]. Li et al. applied active learning on sparse graph representations for image annotation [16]. Recently, MIL was combined with active learning using an instance selection that exploits one class classification model. The performance of the system was evaluated for computer-aided detection of tuberculosis and pixel classification had improved [17]. Similarly, an MI active learning algorithm that exploits multicriteria decision making is proposed in [18]. However, in the literature, most of the active learning algorithms deal with the query strategy; on the other hand, the performance of an algorithm also depends on classifiers and feature representations.

In this paper, we have extended our previous work [19] and propose to use an MI active learning method that employs classifier ensembles trained on sparse representations of the data from multiple learned dictionaries. We have used 3 different active learning query strategies, namely informativeness, uncertainty, and entropy measures. In addition to COREL 1000 and COREL 2000 [2] datasets, we have extended our experimental result using support vector machines (SVM), decision tree (DT), and multilayer perceptron (MLP) algorithms with Elephant, Fox, and Tiger [4] datasets. Since the informativeness measure can only employ SVM, we have implemented DT and MLP classifiers using entropy and uncertainty measures. The results show that the proposed algorithm outperforms the kernel-based MI active learning framework.

2. Background on dictionary learning and MIL active learning

2.1. Dictionary learning and sparse coding

The goal of sparse coding is to represent the input vectors as a weighted linear combination of a small number of basis vectors. By doing this transformation, input data are converted into a high-level representation that reflects

the patterns of the basis vectors. Formally, given a signal $x \in \Re^n$ and a dictionary $D = [d_1, ..., d_k] \in \Re^{n \times k}$, the sparse representation of the signal can be stated as

$$\min_{\alpha} \|\alpha\|_0 \ \ s.t. \ \ x = D\alpha, \tag{1}$$

where $\|\alpha\|_0$ is the ℓ_0 pseudonorm of the coefficient vector $\alpha \in \Re^k$, the number of nonzero elements. This is a very simple and intuitive measure of the sparsity of a vector, counting the number of nonzero elements in it. However, this minimization is an NP-hard problem and therefore cannot be solved in polynomial time. One of the most common approximations is to relax the ℓ_0 pseudonorm with ℓ_1-norm as

$$\min_{\alpha} \|x - D\alpha\|_2^2 + \lambda \|\alpha\|_1 , \tag{2}$$

where λ is a parameter that balances the trade-off between reconstruction error and sparsity. This is a convex optimization problem that can be solved by the LARS-Lasso algorithm [20].

It is important to decompose the data as sparsely as possible with an appropriate dictionary, which should be learned from the data themselves [21]. Given the set of signals $X = \{x_i | x \in \Re^n\}$, the goal is to learn a dictionary $D = [d_1, ..., d_k] \in \Re^{n \times k}$ such that each signal x_i can be well approximated as a sparse linear combination of its atoms d_i:

$$\min_{D, \{\alpha_i\}_{i=1...m}} \sum_{i=1}^{m} \|x_i - D\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \tag{3}$$

This dual optimization problem can be solved as a convex problem by fixing the dictionary and sparse code the data and then update the dictionary. This minimization problem has been modeled using, for example, K-SVD [22] and online dictionary learning [11] methods in the literature. Once the sparse representations of the signals are found, one can use the reconstruction error to classify a data sample or train a classifier from the sparse representations. We used dictionaries with different sizes to extract different sparse representations of a feature vector to obtain classifier ensembles.

2.2. Classifier ensembles

The intuition of the ensemble method is that if each classifier makes different errors upon the training features, then the combination of these classifiers can reduce the total error regarding the single classifier. In order to obtain better classification accuracy than a single classifier, the classifiers in the ensemble should be diverse enough, which can be achieved in several ways. The most popular method is to use different training datasets to train individual classifiers. In this paper, we obtain the diversity of the classifiers using different sparse representations of the features. Alternatively, different classifiers can be combined to improve classification accuracy. Another approach is to use different training parameters for classifiers [23]. In this paper, we also employed majority voting to obtain the final decision of the ensembles.

2.3. Multiple instance active learning and query strategies

In the MIL framework, the data consist of a set of bags $B_i = \{\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iK_i}\}$, where each bag B_i has a bag label y_i and instance features α_{ij}. For binary classification, B_i is considered positive when at least one instance α_{ij} is positive; otherwise B_i is negative. Note that for each bag the number of instances (K_i) is not constant, which makes the MIL problem challenging. In our problem, the data set consists of labeled set $B^L = \{B_1, B_2, ..., B_M\}$ and unlabeled set $B^U = \{B_1, B_2, ..., B_N\}$.

As mentioned in Section 1, one way of classifying the bags is to convert the MIL problem into a single-instance learning problem. There has been a lot of research for converting a bag matrix into the best representative bag features. One of the methods used to convert MIL into a single-instance learning problem is the set kernel method [15]:

$$k_{set}(B_i, B_j) = \sum_{\alpha \in B_i, \alpha' \in B_j} k(\alpha, \alpha'), \tag{4}$$

where $k(.,.)$ is any kernel function on instances and $B_i, B_j \in B^L$ have instances α and α', respectively. In order to be consistent with all of the data, the normalized set kernel is used:

$$k_{nset}(B_i, B_j) = \frac{k_{set}(B_i, B_j)}{\sqrt{k_{set}(B_i, B_i)}\sqrt{k_{set}(B_j, B_j)}} \tag{5}$$

Instead of using the original feature representation, we used sparse coding to compute the kernel sets in the proposed algorithm. Using this set kernel method, one can train a classifier that employs kernels such as SVM.

Liu et al. [15] proposed an informativeness measure as a query strategy. An informativeness measure selects the data based on three selection criteria and combines them to select the most informative one so that the classification error is reduced significantly. The first criterion is the uncertainty, in which the unlabeled data are assessed from the perspective of the version space, aiming at selecting the most uncertain one. For instance, in SVM implementation, unlabeled data that are the closest to the SVM hyperplane in the kernel space are selected due to the uncertainty criteria, as in

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

where

$$f(B_j) = \sum_{i=1}^{M} \tilde{\alpha}_i k_{nset}(B_j, B_i) + b \tag{7}$$

is the dual view of the SVM classifier. Note that $B_j \in B^U$ and $\tilde{\alpha}_i$ is the Lagrangian multiplier of the support vectors.

The second criterion is the novelty measure, in which the unlabeled data that are similar (closer) to the training data are less likely to be selected. In other words, unlabeled data that are more novel to the training data should be selected and defined as

$$d(B_j) = 1 - \max_{1 \le i \le M} k_{nset}(B_i, B_j), \tag{8}$$

where $B_j \in B^U$ and $B_i \in B^L$.

The third criterion is diversity. Selecting unlabeled data samples that are too close to each other does not give much information. The diversity among the unlabeled data can be estimated by averaging similarities among the samples. The diversity measure of the jth date sample in the unlabeled set B^U would be as follows:

$$r(B_j) = 1 - \sum_{i=1, i \ne j}^{N} k_{nset}(B_j, B_i)/N - 1 \tag{9}$$

Thus, informativeness of a given unlabeled data B_j can be calculated as [15]

$$Informativeness(B_j) = \lambda \times u(B_j) + (1 - \lambda) \times d(B_j) \times r(B_j), \tag{10}$$

where λ is the trade-off parameter to adjust the criteria values. After the evaluation, we choose the unlabeled data that have the maximum informativeness value.

The techniques above do not exploit the advantage of the classifier ensemble. Moreover, using the set kernel method in MIL framework only works if the classifier uses a kernel, as in the case of SVM. In order to train different nonlinear classifiers, we have to use different mapping functions. Since we extract sparse representations of the dataset, we can use the pooling function as a mapping function. Pooling functions summarize the feature distribution of the data into a statistical representation [8]. In that sense, different pooling techniques give different signal statistics. As in sparse representation, max pooling function gives the maximum contribution of the given atom in the learned dictionary. In addition, the max pooling function has been empirically justified by many algorithms used for image categorization [12]. Max pooling function can be defined as

$$B_{ik} = \max\{\,|\alpha_{i1k}|\,, |\alpha_{i2k}|\,, ..., |\alpha_{ijk}|\,, ...\}\,, \tag{11}$$

where B_{ik} is the kth element of the ith bag feature and α_{ijk} is the kth element of α_{ij} instance. Using the max pooling function, we convert the MIL problem into a single-instance learning problem. Furthermore, by using the set of classifiers, we can choose the best data by calculating uncertainty in the ensemble, which gives us the least certain data in the unlabeled set. Uncertainty measure can be given as

$$B^* = \underset{B_i \in B^U, i=1...N}{\arg\max}\; 1 - P_\theta(y|B_i), \tag{12}$$

where B^* is the selected unlabeled data and $P_\theta(y|B_i)$ is the posterior probability under the model θ. However, only considering the least certain data as a query strategy can be misleading. For example, if the least certain class label and the second- and third-least certain class labels are close enough, only considering the least probable class label would "throw away" the other class labels' information about the unlabeled data. As a more general uncertainty measure, we can choose entropy:

$$B^* = \underset{B_i \in B^U, i=1...N}{\arg\max} -\sum_j P_\theta(y_j|B_i) \log P_\theta(y_j|B_i) \tag{13}$$

After the choice of the most suitable unlabeled data, we query its label from the oracle and $\langle B^*, y^* \rangle$ tuple is added to the training set. Note that for binary classification problems this entropy-based query strategy will yield the same result as uncertainty strategies, but the entropy-based approach generalizes easily to probabilistic multilabel classifiers.

3. The proposed DEMIAL algorithm

Recently, Song et al. proposed a supervised MIL algorithm, namely SCCE-MIL, by applying sparse coding and obtained promising results using a SVM classifier on the COREL 1000 dataset [10]. In this paper, we combine the sparse coding and classifier ensemble strategy and propose an MI active learning algorithm named DEMIAL. Our algorithm extends the SCCE-MIL method using different base classifiers and an active learning framework with different query strategies. The block diagram of the proposed algorithm is given in Figure 1. The proposed algorithm initially learns a common dictionary matrix from the training instance features. Then, using the dictionary matrix, the sparse representations of the features are extracted. Our method uses ensembles of classifiers trained on sparse features generated from different sized dictionaries. Then, using these trained dictionaries, we calculate the sparse representations of the data. After that, we use the max pooling

technique to convert sparse representations of instances into a single bag representation for each dictionary size. Ensemble of base classifiers are obtained by training classifiers on various sizes of sparse bag features. Unlabeled data are also processed using the same dictionaries to extract sparse features. Each classifier is applied to the unlabeled data and their output values are evaluated using query strategies. Using the query strategies, the oracle is queried for the label of certain data samples. These labeled data samples are added to the training dataset and then dictionaries, sparse feature vectors, and classifiers are updated. This process is repeated for a certain number of iterations. The impact of the base classifiers is examined through the use of SVM, DT, and MLP algorithms. Detailed descriptions of the classifiers can be found in [24]. The computational complexity of the proposed method is highly dependent on dictionary learning and it has two major steps: 1) approximating sparse representation of the training signals and 2) updating the dictionary using the sparse representation. Two extensively used techniques related to dictionary learning are PCA and K-Means clustering [25]. Thus one can see dictionary learning as PCA with multiple subspaces or clustering a signal using multiple locations. On the other hand, the ensemble size constitutes the number of dictionaries that also increases the computation time to extract sparse features.

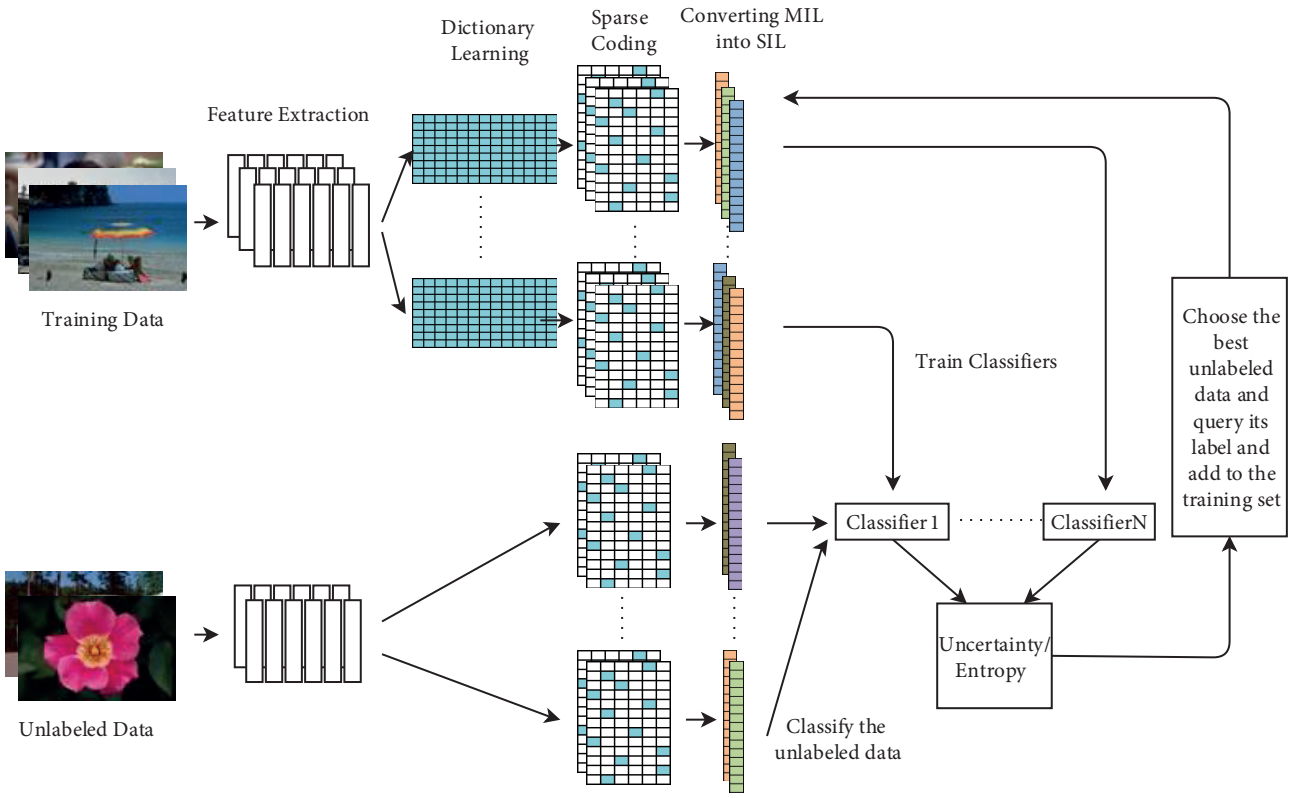

Figure 1. The block diagram of the proposed MIL active learning framework DEMIAL.

4. Experimental results

We conduct experiments on the Elephant, Fox, Tiger, and Corel 1000 and 2000 datasets. The Corel 1000 and 2000 datasets contain 10 and 20 categories of images, respectively, where each category has 100 different images. Some examples from the Corel 1000 dataset are given in Figure 2. Each image is in JPEG format with sizes of

384 × 256 or 256 × 384. To extract image features, each image is partitioned into nonoverlapping blocks of size 4 × 4 pixels. Each feature vector is composed of 9 features: three of them represent the average LUV color components in the block and three of them are the square root of energy values in the high frequency band of wavelet transform. The other three features are calculated for each region to describe the shape properties and they are normalized inertia of order 1, 2, and 3. Note that we have used the previously used features and detailed descriptions of these feature vectors can be found in [3]. The Fox, Tiger, and Elephant datasets[1] contain three separate sets of images that each have 200 images with 100 positive and 100 negative classes. All of the images are segmented into set of regions and for each region a 320-dimensional feature vector is extracted to represent its color, texture, and shape characteristics [4].

Figure 2. Example images from the Corel 1000 dataset.

In the experiments, the datasets are divided into training, unlabeled, and test dataset partitions. Initial training, unlabeled, and test datasets consist of 20%, 60%, and 20% of the whole dataset, respectively. Active learning is conducted using the unlabeled set. We repeat each experiment on 5-fold cross-validation and report the average classification accuracies. For sparse coding and dictionary learning, the Lasso algorithm in the SPAMS toolbox [11] is used with regularization parameter $\lambda = 0.05$ and dictionary sizes ranging from 10 to 400. SVM is implemented using the libsvm toolbox [26]. The proposed method is analyzed under different scenarios. Initially, we give the classification accuracy of the SCCE-MIL algorithm in a supervised classification scenario in the Table where 80% of the data are used for training and 20% are used for testing with 5-fold cross-validation. As can be seen from the Table, SVM outperforms the other classifiers for the COREL1000, COREL2000, and Elephant datasets. On the other hand, all classifiers perform similarly for the Tiger dataset and DT gives the best accuracy for the Fox dataset. Active learning results with different selection strategies are given in the next subsection. Then we compare the classification performance of the proposed method with different base classifiers and active learning instance selection strategies. Note that, as SCCE-MIL does not perform active learning, the initial accuracy of the algorithms with 20% training data also corresponds to SCCE-MIL results in the following experiments.

Table. Classification accuracy of the SCCE-MIL algorithm in a supervised learning scenario.

Classifier/Dataset	COREL1000	COREL2000	Elephant	Fox	Tiger
DT	81.3	66.76	82.5	63.52	80.95
MLP	78.38	49.17	84.00	59.51	80.52
SVM	84.9	71.90	84.55	57.04	80.50

[1] http://www.cs.columbia.edu/∼andrews/mil/datasets.html

4.1. Comparisons with the kernel-based MI active learning method

In the first set of experimental results, we compared the DEMIAL algorithm with the kernel-based active learning method. As the kernel-based MI active learning algorithm uses the SVM classifier, in our first experiment we compared it with the proposed DEMIAL algorithm using SVM as a base classifier. We also compare the DEMIAL active learning query strategies using uncertainty and entropy measures. The results of the experiments are shown in Figure 3.

As shown in Figure 3, increasing the training data size with active learning increases the classification accuracies for most of the datasets. In Fox and Tiger, the improvement varies compared to the other datasets. As given in [15], increasing the number of labeled examples does not always increase the classification accuracy during active learning iterations. Note that none of the algorithms control the labels of the queried data samples. If all of the queried data samples are labeled with the same class label, this may affect the balance of the training dataset and may radically change the decision boundary of the classifier at the next iteration. If we compare the algorithms, we can see that the proposed DEMIAL algorithm performs better than the kernel-based MI active learning method. We note that the initial training size is 20% of the whole dataset and the active learning continues until 50% of the data are in the training set.

In addition, if we compare the query strategies in the DEMIAL method, most of the accuracy results are similar. In fact, the results of the Elephant, Fox, and Tiger datasets are exactly the same. This is because these three datasets are binary classification problems and both query strategies choose the same data in the query stage. However, in Corel 1000 and 2000, the results are similar to each other.

4.2. Active learning with different base classifiers

In this experiment, we compared the proposed active learning framework DEMIAL with different classifiers using entropy and uncertainty based selection strategies. Note that the informativeness selection strategy can be only used with kernel-based classifiers and the results for informativeness were given in the previous section. The classification accuracies with respect to different number of training data samples that are obtained using entropy and uncertainty based active learning are given in Figure 4, which shows that using active learning in the algorithm generally increases the classification accuracy in all of the datasets.

In terms of the base classifiers, we can see different performances for each dataset. For example, in the Corel 1000 and 2000 datasets, SVM is significantly better than DT and MLP. However, in Elephant and Tiger, it performs poorly. For the Fox dataset, neither different base classifiers nor different query strategies produce significant results. This is partly because of the dataset itself. In some previous works, the Fox dataset does not produce good results in terms of active learning [27] and classifier ensemble strategy [28].

In addition, if we compare the active learning query strategies for the Elephant, Fox, and Tiger datasets, uncertainty and entropy measures have the same results using DT and SVM because they query the same data samples. As MLP starts with random initial weights, it has slightly different results when we use different query strategies.

5. Discussion and future work

In this paper, DEMIAL, an MI active learning method based on a dictionary ensemble, is proposed. The proposed algorithm uses dictionary learning to obtain different sparse feature sets for classifier ensembles. Moreover, it has the ability to use different base classifiers and different query strategies in active learning. Experimental results are obtained on 5 MIL datasets and the DEMIAL algorithm is compared with the

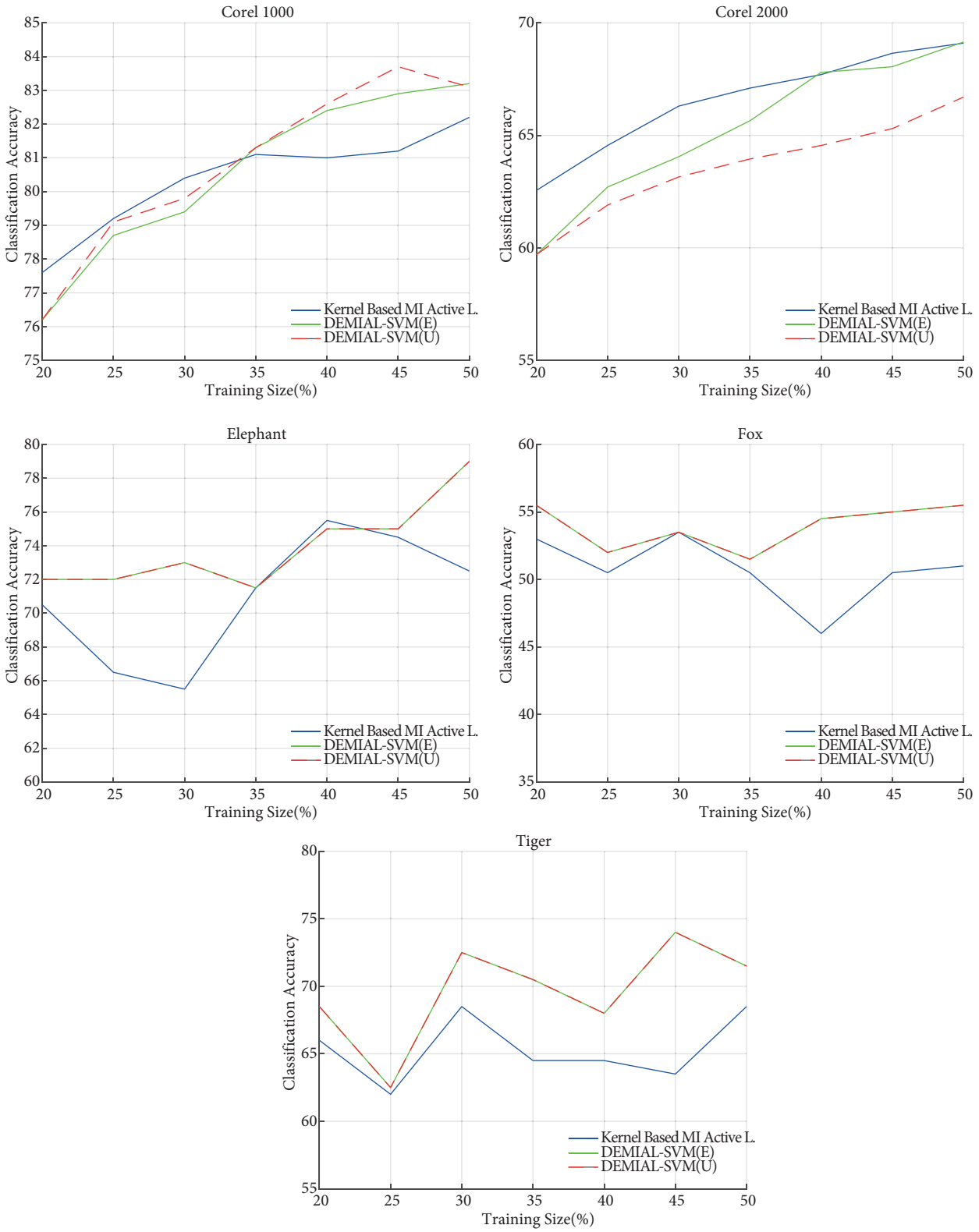

Figure 3. Performances of the active learning selection strategies on different datasets.

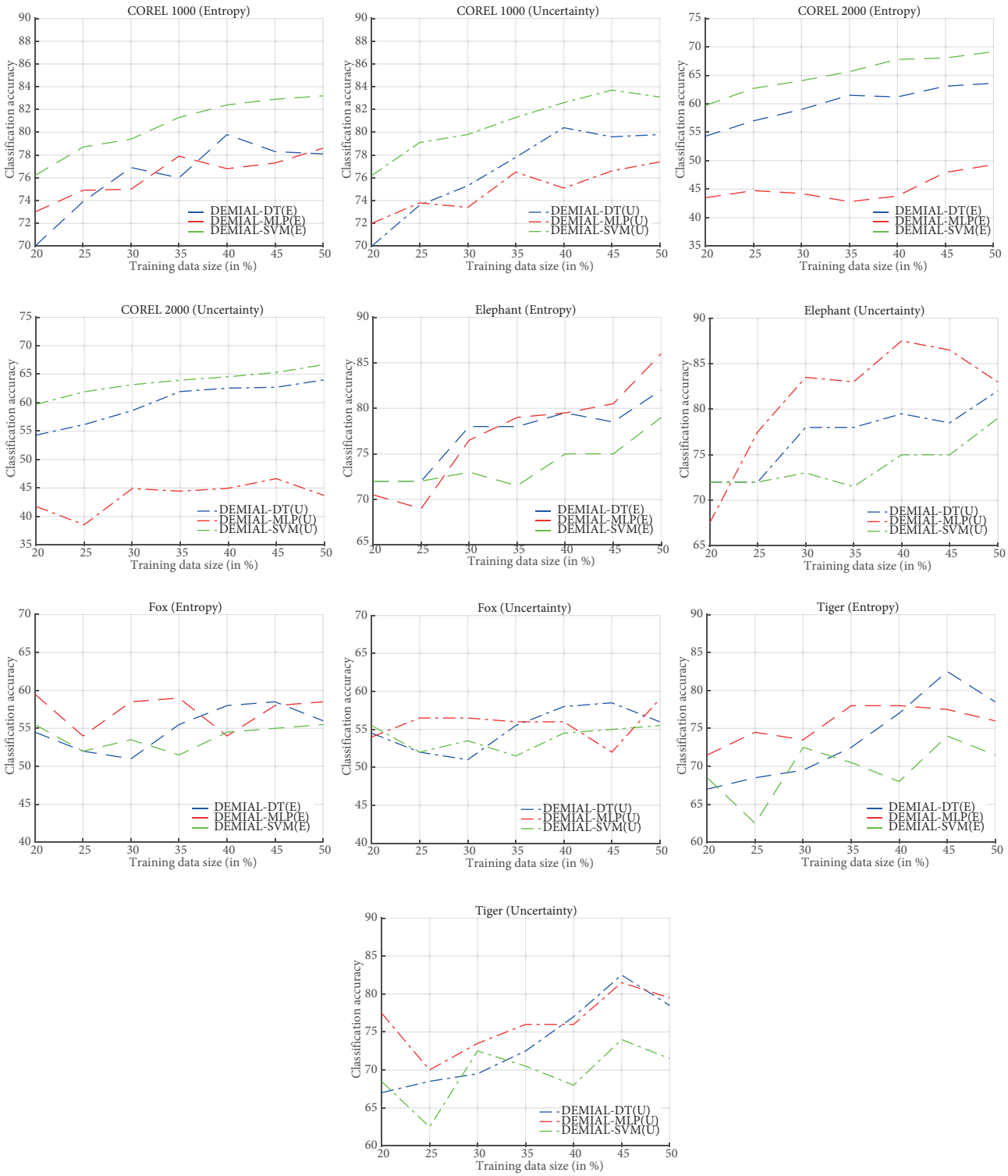

Figure 4. Results of the proposed DEMIAL algorithm with query strategy entropy and uncertainty.

kernel-based active learning method using different active learning query strategies. We can conclude from the experiments that the proposed method achieves better classification accuracy than the kernel-based method.

The experiments also showed that the performance of the base classifiers is heavily dependent on the datasets used. This variation is not seen in the multiclass Corel 1000 and 2000 datasets, and for these datasets the DEMIAL algorithm with SVM has the best classification accuracy. However, it is an issue for the Elephant, Fox, and Tiger datasets, which have binary class labels. This can be caused by class inequality after the active learning iteration. In future work, we will investigate the issues behind the variance of the results by comparing with more datasets and more query strategies. We are also planning to investigate the relationship between sparse feature representation and diversity of the classifiers. Classifier diversity plays an important role for classifier ensembles. Strategies that can lead to more diverse classifiers for DEMIAL is another future research direction.

Acknowledgment

We express our thanks to Prof Dr Zehra Çataltepe for her valuable comments.

References

[1] Dietterich TG, Lathrop RH, Lozano-Perez T. Solving the multiple instance problem with axis-parallel rectangles. Artif Intel 1997; 89: 31-71.

[2] Chen YX, Bi JB, Wang JZ. MILES: Multiple instance learning via embedded instance selection. IEEE T Pattern Anal 2006; 28: 1931-1947.

[3] Chen Y, Wang JZ. Image categorization by learning and reasoning with regions. J Mach Learn Res 2004; 5: 913-939.

[4] Andrews S, Tsochantaridis I, Hofmann T. Support vector machines for multiple instance learning. In: Advances in Neural Information Processing Systems; 9–14 December 2002; Vancouver, BC, Canada. Cambridge, MA, USA: MIT Press. pp. 561-568.

[5] Amores J. Multiple instance classification: review, taxonomy and comparative study. Artif Intel 2013; 201: 81-105.

[6] Maron O, Lozano-Pérez T. A framework for multiple instance learning. In: Advances in Neural Information Processing Systems; 30 November–5 December 1998; Denver, CO, USA. Cambridge, MA, USA: MIT Press. pp. 570-576.

[7] Zhang Q, Goldman SA. EM-DD: An improved multiple instance learning technique. Advances in Neural Information Processing Systems; 3–8 December 2001; Vancouver, BC, Canada. Cambridge, MA, USA: MIT Press. pp. 1073-1080.

[8] Wang J, Zucker JD. Solving multiple instance problem: A lazy learning approach. In: International Conference on Machine Learning; 29 June–July 2 2000; Stanford, CA, USA. San Francisco, CA, USA: Morgan Kaufmann Publishers, Inc. pp. 1119-1125.

[9] Zhang J, Marszałek M, Lazebnik S, Schmid C. Local features and kernels for classification of texture and object categories: A comprehensive study. Int J Comput Vision 2007; 73: 213-238.

[10] Song X, Jiao L, Yang S, Zhang X, Shang F. Sparse coding and classifier ensemble based multi-instance learning for image categorization. Signal Process 2013; 93: 1-11.

[11] Mairal J, Bach F, Ponce J, Sapiro G. Online dictionary learning for sparse coding. In: International Conference on Machine Learning; 14–18 June 2009; Montreal, QC, Canada. New York, NY, USA: ACM. pp. 689-696.

[12] Lee H, Battle A, Raina R, Ng AY. Efficient sparse coding algorithms. In: Advances in Neural Information Processing Systems; 3–6 December 2007; Red Hook, NY, USA: Curran Associates, Inc. pp. 801.

[13] Settles B, Craven M, Ray S. Multiple instance active learning. In: Advances in Neural Information Processing Systems; 8–11 December 2008; Vancouver, BC, Canada. Red Hook, NY, USA: Curran Associates, Inc. pp. 1289-1296.

[14] Fu J, Yin J. Bag-level active multi-instance learning. In: IEEE 8th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD); 26–28 July 2011; Shanghai, China. Piscataway, NJ, USA: IEEE. pp. 1307-1311.

[15] Liu D, Hua XS, Yang L, Zhang HJ. Multiple instance active learning for image categorization. In: Advances in Multimedia Modeling, 15th International Multimedia Modeling Conference; 7–9 January 2009; Sophia-Antipolis, France. Berlin, Germany: Springer. pp. 239-249.

[16] Li M, Tang J, Zhao C. Active learning on sparse graph for image annotation. KSII T Internet Inf 2012; 6: 2650-2662.

[17] Melendez J, van Ginneken B, Maduskar P, Philipsen RH, Ayles H, Sánchez CI. On combining multiple instance learning and active learning for computer-aided detection of tuberculosis. IEEE T Med Imag 2016; 35: 1013-1024.

[18] Wang R, Kwong S. Active learning with multi-criteria decision making systems. Pattern Recogn 2016; 47: 3106-3119.

[19] Koçyiğit G, Yaslan Y. Dictionary ensemble based multi instance active learning method for image categorization. In: 24th Signal Processing and Communication Application Conference (SIU); 16–19 May 2016; Zonguldak Turkey. Piscataway, NJ, USA: IEEE. 1221-1224.

[20] Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. Ann Stat 2004; 32: 407-499.

[21] Sprechmann P, Sapiro G. Dictionary learning and sparse coding for unsupervised clustering. In: IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP); 14–19 March 2010; Dallas, TX, USA. Piscataway, NJ, USA: IEEE. pp. 2042-2045.

[22] Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE T Signal Proces 2006; 54: 4311-4322.

[23] Kuncheva LI. Combining Pattern Classifiers: Methods and Algorithms. 1st ed. Hoboken, NJ, USA: Wiley, 2004.

[24] Alpaydin E. Introduction to Machine Learning. 2nd ed. Cambridge, MA, USA: MIT Press, 2010.

[25] Vainsencher D, Mannor S, Bruckstein AM. The sample complexity of dictionary learning. J Mach Learn Res 2011; 12: 3259-3281.

[26] Chang CC, Lin CJ. A library for support vector machines. ACM T Intel Syst Tec 2011; 2: 1-27.

[27] Zhang D, Wang F, Shi Z, Zhang C. Interactive localized content based image retrieval with multiple instance active learning. Pattern Recogn 2010; 43: 478-484.

[28] Cheplygina V, Tax DMJ, Loog M. Dissimilarity-based ensembles for multiple instance learning. IEEE T Neur Net Lear 2016; 27: 1379-1391.