Computer Engineering and Applications
Computer Engineering and Applications Vol. 4, No. 1, February 2015

Learning from Imbalanced Multi-label Data Sets by Using Ensemble Strategies

Mohammad masoud Javidi and Fatemeh Shamsezat

Department of Computer Science ,Faculty of Mathematics and Computer,Shahid Bahonar University of Kerman,Kerman,Iran.

ABSTRACT

Multi-label classification is an extension of conventional classification in which a single instance can be associated with multiple labels. Problems of this type are ubiquitous in everyday life. Such as, a movie can be categorized as action, crime, and thriller. Most algorithms on multi-label classification learning are designed for balanced data and don't work well on imbalanced data. On the other hand, in real applications, most datasets are imbalanced. Therefore, we focused to improve multi-label classification performance on imbalanced datasets. In this paper, a state-of-the-art multi-label classification algorithm, which called IBLR_ML, is employed. This algorithm is produced from combination of k-nearest neighbor and logistic regression algorithms. Logistic regression part of this algorithm is combined with two ensemble learning algorithms, Bagging and Boosting. My approach is called IB-ELR. In this paper, for the first time, the ensemble bagging method whit stable learning as the base learner and imbalanced data sets as the training data is examined. Finally, to evaluate the proposed methods; they are implemented in JAVA language. Experimental results show the effectiveness of proposed methods.

Keywords: Multi-label classification, Imbalanced data set, Ensemble learning, Stable algorithm, Logistic regression, Bagging, Boosting

1. INTRODUCTION

Conventional classification is concerned with learning from a set of instances that are associated with a single label ℓ from a set of finite labels $\mathcal{L}$, $|\mathcal{L}| > 1$. In multi-label classification the instances are associated with a set of labels L, $|L| \subseteq \mathcal{L}$. The goal in multi-label classification is to predict a set of relevant binary labels for a given input. Originally, multi-label learning came from the investigation of text categorization problem, where each document may belong to several predefined topics simultaneously [1, 2]. Nowadays, multi-label classification methods are increasingly required by modern applications: Such as, a scene can be categorized as bench and urban[3]; in biology, each gene may be associated with a set of functional classes, such as metabolism, transcription and protein synthesis [4].

One important challenge of multi-label data is the class imbalance problem, where sampled data for the classifier training is non-uniformly distributed over the data space. In other words, each label has usually more negative than positive examples, but still some labels have much more positive examples than others [5-8]. Many data sets in real-world applications, such as remote-sensing [9], risk management [10], pollution detection [11], especially medical diagnosis [12] and fraud detection [13] are imbalanced. There are only a limited number of approaches have been proposed to address this imbalanced problem for multi-label data, in contrast to single label data. In this paper, we focus on this problem of multi-label learning, and tackle highly imbalanced data distributions using ensemble of multi-label classifiers.

Ensemble techniques are becoming increasingly important as they have repeatedly demonstrated the generalization ability of an ensemble is usually much stronger than that of a single learner [14], especially with highly imbalanced data populations [15, 16]. It is well known that an ensemble of classifiers can provide higher accuracy than a single best classifier if the member classifiers are diverse and accurate [8, 17-19]. In this paper, we employ two data variation-based ensembles, which consist in the manipulation of the training examples in such a way that each classifier is trained with a different training set. These are AdaBoost [20, 21] and Bagging [22] that are the most common and successful ensemble learning algorithms [8].

The aim of this paper is to use homogeneous ensembles of learners to improve the performance of multi-label classifier for imbalanced data set. This is different from the existing work in the sense that we are proposing to apply ensemble technique to particular situation of a state-of-the-art multi-label learner, moreover this sub-algorithm as the base learner for ensemble method is stable. In this paper, a multi-label classification algorithm, which called IBLR_ML [23], is employed. This algorithm is produced from combination of k-nearest neighbor and logistic regression algorithms. In fact, we are applying ensemble techniques within the logistic regression part of IBLR-ML. while the bagging fails with stable learning algorithms whose output is insensitive to small changes in the input and also logistic regression is stable method, we employ logistic regression as the base learner. Additionally, this presented approach takes correlation and interdependencies between labels into account, because the base classifier, IBLR-ML, is inherently considering correlation among labels.

The proposed ensemble multi-label learning approach (IB-ELR)1 is applied to seven publicly available multi-label data sets from different domains (Emotions, Genbase, Mediamill, Image, text, Yeast and Scene) and furthermore we create several imbalanced data sets from real balanced data sets. The performance evaluation of multi-label classifiers is evaluated by using five different important multi-label classification measures, to find how the learning algorithms behave under variety of imbalance degrees. Finally, to evaluate the proposed methods; they are implemented in JAVA language. Experimental results show the effectiveness of proposed methods.

The paper is organized as follows: The related work of multi-label classification and ensemble learning are reviewed in section 2. Section 3 describes the dilates of purposed approach. Section 4 is devoted to experimenters with several benchmark data sets and evaluation metrics of multi-label classification problem. Finally, the paper ends with some conclusions in section 5.

2. RELATED WORK

[1] A shorter version of this work.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Nowadays, multi-label data are becoming ubiquitous. They arise in an increasing number and diversity of applications. In the past several years, a variety of multi-label methods have been proposed in pattern recognition, machine learning and statistics. The existing methods for multi-label classification can be grouped into two main categories: a) problem transformation methods, and b) algorithm adaptation methods [24, 25]. Methods of the first group transform the learning task into one or more single-label classification or ranking tasks. Algorithm adaptation extends some specific multi-class classification algorithms to handle an entire multi-label training data set directly. The following paragraph describes a number of problem transformation methods from the literature.

Binary relevance (BR) [2, 3, 25, 26] is a popular problem transformation method that learns M binary classifiers, one for each different label in L. An obvious disadvantage of this approach is that it ignores correlations and interdependencies between labels. Another transformation method is LP [3, 24]. This method considers each possible label combination of more than one class in a multi-label training data set as a new single class, and then converts a multi-label problem into a standard multi-class one. LP typically works well if the original label set L is small but quickly deteriorates for larger label sets. Ranking by pairwise comparison (RPC) [27] transforms the multi-label data set into M binary label datasets, one for each pair of labels. Each dataset contains those examples of training set that are annotated by at least one of the two corresponding labels, but not both. Calibrated Label Ranking (CLR) [28] extends RPC. The key idea in this approach is to introduce an artificial calibration label that, in each example, separates the relevant label from the irrelevant labels.

In this and next paragraphs briefly report the significant algorithm adaptation methods. Through modifying the formula of entropy calculation and permitting multiple labels at the leaves of the tree, a C4.5-like multi-label classification [29] algorithm is proposed, in which it is possible to generate a large number of leaves for all combinations of different labels, just like the original LP method. Rank-SVM [4] is a support vector machine algorithm for multi-label classification. This method employs ranking loss as its empirical loss function. For finding a natural zero to determine the relevant labels in this algorithm, a virtual label is simply added in [30, 31]. Another adaptation method is MMAC [32], that follows the paradigm of associative classification, which deals with the construction of classification rule sets using association rule mining. Back-propagation for multi-label learning (BP-MLL) [33] is an adaptation of the back-propagation algorithm to multi-label learning problems by introducing a new error function. Every algorithm of adaptation methods that described, considers all classes and all instances simultaneously [31, 34].

The other algorithm adaptation methods, which described in this paragraph, still deals with each class independently after using some problem transformation tricks [31, 34]. ML-RBF [35] is a recent approach for adapting radial basis function networks to multi-label data. During its clustering procedure, a q-label problem is divided into q sub-problems using the one-by-one method, and then each class instances are clustered independently. Nearest neighbor (kNN) or instance-based (IB) algorithm has been extended to construct slightly different multi-label methods: ML-kNN [36] and IBLR-ML [23]. At ML-KNN, for each unseen instance, its K nearest neighbors in the training set is firstly identified. After that, based on statistical information gained from the label sets of these neighboring instances, maximum a posteriori (MAP) principle is utilized to determine the label set for the unseen instance. While the basic idea in IBLR-ML is to consider the information that derives from examples similar to a query instance as a feature of that instance, thereby blurring the distinction between instance-based and model-based learning to some extent. Therefore IBLR-ML takes more correlations between labels as possible in account than ML-kNN does.

On the other hand, classifier learning with data-sets that suffer from imbalanced class distributions is a challenging problem in data mining community. Several well-established classification modeling systems for conventional classification have been extended to the imbalanced case, including decision trees [37-42] , support vector machines [39, 43-46] , neural networks [39] , Bayesian network[47] , nearest neighbor [38, 48] and the newly reported associative classification approaches [49-51]. There are existing reviews and categorizations on imbalanced data learning can also be found in [6, 38, 39, 52, 53]. In a recent work, taxonomy for ensemble-based methods is proposed to address the class imbalance problem [8]. In addition, it develops a thorough empirical comparison by the consideration of the most significant ensemble-based published approaches. This comparison has shown the good behavior of the simplest approaches which combine random under sampling techniques with bagging or boosting ensembles. Diversity also plays an important role in improving the performance of ensemble classifier. In general, two popular directions for diversification are the bagging and boosting methods. In bagging methods, diversification is maintained by creating individual classifiers on different subsets of the training data. While in the boosting algorithms, example distributions are updated iteratively by giving more weights for those previously misclassified examples, and thus diversity means building classifiers on training data with progressively updated distributions. There are several studies which explain why bagging improves the predictive performance by reduction of the variance of the mean squared error. The amount of improvement depends on the bias-variance decomposition for base learners, which suggests that unstable models with high variances such as decision trees are preferable as the base learner for bagging rather than stable ones logistic regression and K nearest neighbor methods. A learning algorithm is unstable if small changes in their training sets tend to induce significant differences in the models. On the other hand, it can slightly degrade the performance of stable procedures[18, 22]. Four main algorithms in Bagging-based Ensembles family, which deal with class imbalance problems, are OverBagging [54], UnderBagging [55], UnderOverBagging [54], and IIVotes [56]. In addition, several popular Boosting-based Ensembles strategies for imbalance learning include SMOTEBoost [57], MSMOTEBoost [58], RUSBoost [59], and DataBoost-IM [60] algorithms.

Afterwards, we would also like to note that although the current efforts in the community are focused on multiclass imbalanced problems, multi-label imbalanced learning problems exist and are of equal importance. A limited number of approaches have also been proposed to address the multi-label and imbalanced data problems. In [61], a min-max modular network was proposed to decompose a multi-label imbalanced learning problem into a series of small two-class sub problems. This paper also presents several decomposition strategies to improve the performance of min-max modular networks. Another approach, that addressed multi-label classification which imbalanced data problems, is presented in [62]. This method uses an enrichment process in neural net training. The enrichment process

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

can manage the imbalanced data and train the neural net with high classification accuracy. Also, in [63] concept drift and class imbalanced in multi-label data in a data stream context is studied. Outers introduce a sophisticated parameterized windowing mechanism for dealing with it, which they exemplify with an efficient instance-incremental multi-label kNN method. Last, in a recent work, a heterogeneous ensemble multi-label learners is proposed [64], by combining state-of-the-art multi-label methods. This method simultaneously tackles both the sample imbalance and label correlation problems.

It is noteworthy, in machine learning, the ensemble of classifiers are known to increase the accuracy of single classifiers by combining several of them, but neither of these learning techniques alone solve the class imbalance problem, to deal with this issue the ensemble multi-label learning algorithms have to be designed specifically. The data sparseness problem of the LP approach was addressed in [65]. The authors propose Pruned Sets (PS) and Ensemble of Pruned Sets (EPS) methods to concentrate on the most important correlations. The two random k-labelsets (RAkEL) methods proposed in [66, 67], that construct an ensemble of LP classifiers. Each LP classifier is trained using a different small random subset of the set of labels. triple-random ensemble learning method (TREMLC) [68] is presented to handling multi-label classification problems. This proposed method integrates and develops the concepts of random subspace; bagging and random k-label sets ensemble learning methods to form an approach to classify multi-label data. Another approach in [69] calls it Multi-label Boosting by the selection of heterogeneous features with structural Grouping Sparsity (MtBGS). MtBGS induces a (structural) sparse selection model to identify subgroups of homogenous features for predicting a certain label. Moreover, the correlations among multiple tags are utilized in MtBGS to boost the performance of multi-label annotation. In classifier chain methods [70, 71], q sub-classifiers are linked in a cascade way and further the outputs of previous sub-classifier are added to the inputs of current sub-classifier. To relieve the effect of classifier order, an ensemble ECC framework is used to create different random chain ordering.

3. PROPOSED APPROACH

Most algorithms on multi-label classification learning are designed for balanced data and don’t work well on imbalanced data. The aim of this paper is to develop a state-of-the-art multi-label classification algorithm to tackle imbalance problem. Therefore, a state-of-the-art multi-label learning algorithm must be chosen. [72] presented a comparison between different methods of multi-label classification for different domain application. Of the algorithm adaptation methods, ML-kNN has provided the best results in almost all analyzed cases. On the other hand, extensive empirical study, [23], has clearly shown that IBLR improves upon existing methods, in particular the MLKNN method that can be considered as the state-of-the-art in instance based multi-label classification. In conclude, IBLR-ML consistently outperforms all other methods, regardless of the evaluation metric, indicating that it is the strongest method overall. This method considers label information of neighbored examples as features of a query instance, the idea of IBLR is to reduce instance-based learning formally to logistic regression. Moreover, this approach allows capturing interdependencies between labels .Consequently, the chosen successful multi-label learning algorithm is IBLR-ML.

After choosing the multi-label learning method with the good behavior, two ensemble strategies are devoted to develop it for tackling imbalance problem. Bagging is one of the ensemble-based meta-learning algorithms which samples subsets with replacement from the training set, building multiple base learners and aggregating their predictions to make final predictions. Another most significant ensemble-based published approach is boosting where the performance of weak classifiers is improved by focusing on hard examples which are difficult to classify. Boosting produces a series of classifiers and the outputs of these classifiers are combined using weighted voting in the final prediction of the model. In each step of the series, the training examples are re-weighted and selected based on the performance of earlier classifiers in the training series [60]. Boosting methods are able to significantly improve classification performance in many applications. One of the most popular boosting methods is the AdaBoost introduced in [73]. Adaboost is for binary classification problems. One version of Adaboost is Adaboost. M1 that is for multiple classification problems.

In this paper, we applied ensemble techniques to particular situation of IBLR-ML. This algorithm is produced from combination of k-nearest neighbor and logistic regression algorithms. In fact, we are applying ensemble techniques, Bagging and Adaboost.M1, within the logistic regression part of IBLR-ML.

The pseudo-code of the EB-ELR training process is described in Fig. 1. In this algorithm, for each of the class labels in the training data set, builds a new binary data set. In each of them, per one of the class labels in the original data set, a new feature is added. A class label as well as the class label for new data set is added. To determine the values of the features, at the beginning, k nearest neighbor of each example in training data are selected. Then the neighbors that have a label corresponding to the desired feature, is determined. Divide these values by the number of neighbors, if the result is greater than or equal to .5, the value one and the number zero otherwise as the attribute value is chosen. The values of examples for class label are loaded with the same value of initial training data set. Finally, for each class label, a classifier learning algorithm is created by using the corresponding new data set. This classifier learning method is the ensemble method, bagging or boosting, with logistic regression as the based algorithm.

4. EXPERIMENTAL RESULTS

In this section, we compare our IB-ELR with five existing multi-label classification approaches experimentally. Before presenting our experimental the results, we briefly introduce learning algorithms, benchmark data sets included in the study and five evaluation measures for multi-label classification.

4.1 LEARNING ALGORITHMS

In this paper, we choose six successful multi-label classification methods to compare with my proposed approach. The first algorithm is IBLR-ML. For the reasons mentioned earlier, our main interest is focused on IBLR-ML, which is disputably the state-of-the-art in instance-based multi-label classification. Since

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

IBLR-ML consistently outperforms the expended version of this, we use the pure instance-based version of this algorithm. Another state-of-the-art machine learning method is MLKNN that quite will in practice. Both of IBLR-ML and MLKNN are parameterized by the size of the neighborhood, for which we adopted the value k=10. This value is nominated in [35], where it was found to yield the best performance. As an additional baseline we used binary relevance learning (BR) with three different base learners: logistic regression, C4.5 (the Weka [74] implementation J48 in its default setting), and KNN (again K=10). Finally, we also included label powerset (LP) with C4.5 as a base learner. we used their implementations in the MULAN package [75].

```
Inputs: training multi-label data set that consists the following
items.
            NumInstance ← Number of instances
            NumAttributes ← Number of attributes
            NumLabels ← Number of labels
Steps: % Create a training data with label info as features for every
         label.
             T[] ← Create an array of new training data set with
                   NumLabel elements.
             For i=0 to NumInstance do
                   k ← Number of neighbours.
                   Knn ← Specify k neighbors by KNearestNeighbours
                         method.
                   Confidence [] ← The label confidence vector as the
                                   additional features.
                   For j=0 to NumLabels do
                         C[j] ← Compute sum of counts for jth label in
                                Knn.
                         Confidence[j] ← C[j]/k.
                   End for
                   NewIns ← Create new instance with "NumLabel+1"
                            attributes.
                   % The last attribute is added for class label.
                   Copy Confidence vector as added for features to
                   NewIns.
                   For j=0 to NumLabels do
                         Add the value of jth label of instance i in
                         the training data to NewIns as the class
                         label.
                         Add NewIns to T[j].
                   End for
             End for
             % For every label create a corresponding classifier.
             Classifier[] ← create an array of classifiers.
             For i=0 to NumLabels do
                   Classifier [i] ← train an ensemble Bagging
                                    (Boosting) classifier using
                                    Logistic
                                          Regression as he base
                              classifier and T[i] as training data
                              set.
             End for
```

FIGURE 1. Pseudo code of the EB-ELR training process.

4.2 DATA SETS

Benchmark imbalanced data for multi-label classification is not as abundant as for conventional classification, and indeed, experiments in this field are often restricted to a very few or even only a single data set. We empirically evaluated the proposed approach by measuring its performance on eleven benchmark multi-label

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

datasets from different domains, variable sizes and imbalance ration. All datasets along with their properties are listed in Table 1 for balanced data sets and in table 2 for imbalanced ones. We collect five of them: Emotions, Scene, Genbase, Mediamill and Yeast from http://mulan.sourceforge.net/datasets.html, and Image and Reuters from http://lambda.nju.edu.cn/data.htm.

The first criterion to consider for imbalance data is the imbalance ratio[76]. We also consider the imbalance ratio, defined as the number of negative class examples that are divided by the number of positive class examples, to organize the different data-sets. Although [77] defined that for (significantly) imbalanced data, the ratio should be no less than 19:1, in the actual experimental settings, some non-significant imbalanced data should also be tested, in order to find how the learning algorithms behave under variety of imbalance degrees. Therefore, we created imbalanced data sets from the balance ones by eliminating some instances with largest relevant label set. Emotions(v2) And Emotions(v3) are derived from Emotions and as well as scene(v2) and scene(v3) are derived from scene to create imbalanced data sets with different imbalance ratios from balanced multi-label data sets.

TABLE 1.
Characteristic of the balanced multi-label datasets used in the experiments.

Balanced Data set	Domain	#Instances	#Attributes	#Labels	Cardinality	#max/#min	Imb. Ration
Emotions	Music	593	72	6	1.87	264/148	1.78
Emotions(v2)	Music	500	72	6	1.75	192/56	3.43
Image	Vision	2000	135	5	1.24	533/364	1.46
Reuters	Text	7119	243	7	1.24	2256/589	3.83
Scene	Vision	2407	294	6	1.07	580/409	1.42

TABLE 2.
Characteristic of the imbalanced multi-label datasets used in the experiments.

Imbalanced Data set	Domain	#Instances	#Attributes	#Labels	Cardinality	#max/#min	Imb. Ration
Emotions(v3)	Music	434	72	6	1,59	173/14	12.36
Genbase	Biology	662	1186	27	1.25	171/1	171
Mediamill	Multimedia	5000	120	6	4.27	3828/1	3828
Scene(v2)	Vision	1940	294	6	1.001	405/38	10.66
Scene(v3)	Vision	1227	294	14	1.001	405/18	22.5
Yeast	Biology	2417	103	101	4.24	1799/34	52.91

The emotions data set consists of 100 songs from each of the following 7 different genres: Classical, Reggae, Rock, Pop, Hip-Hop, Techno and Jazz. The collection was created from 233 albums choosing three songs from each album. From each song a period of 30 seconds after the initial 30 second was extracted. The resulting sound clips were stored and converted into wave files of 22050 HZ sampling rate, 16-bit per sample and mono [78]. From each wave file, 72 features have been extracted. Then, in the emotions labeling process, 6 main emotional clusters are retained.

The Image data set consists of 2,000 natural scene images belonging to the classes *desert*, *mountains*, *sea*, *sunset*, and *trees*. Some images were from the COREL image collection while some were collected from the Internet. Over 22% images belong to multiple classes simultaneously [79].

The *scene* image dataset contains 2407 images annotated with up to 6 concepts such as beach, mountain and field. Each image is described with 294 visual numeric features and these features are represented with spatial color moments in Luv color space. Each instance in the train and test datasets is labeled with possible 6 object classes as mentioned above [3, 28].

From the text processing field, a text data is derived from the widely studied Reuters 21578 collection [80]. The seven most frequent categories are considered. After removing documents whose label sets or main texts are empty, 8866 documents are retained where only 3.37% of them are associated with more than one class label. After randomly removing documents with only one label, a text categorization data set containing 2,000 documents is obtained. Thereafter, each instance is represented as a 243- dimensional feature vector.

The *mediamill* dataset is based on the *mediamill* challenge data set [66, 81, 82]. It contains pre-computed low-level multimedia features from 85 hours of international broadcast news video of the TRECVID 2005/2006. This dataset contains Arabic, Chinese, and US news broadcasts that were recorded during November 2004, and the contents are annotated with multiple labels. Every instance of this data set has 120 numeric features including visual, textual, as well as fusion information. The trained classifier should be able to categorize an unseen instance to some of these 101 labels, e.g., face, car, male, soccer, and so on.

The *yeast* dataset contains 2417 gene examples, and each of which is related to a set of 14 functional gene classes from the comprehensive Yeast Genome Database of the Munich Information Center for protein Sequences. Each gene is expressed with 103 numeric features [4, 65, 66].

4.3 EVALUATION MEASURES

The performance evaluation of a multi-label classifier is different from that of a classical single-label classifier, which induces more than ten performance evaluation measures [25]. The five evaluation metrics for label ranking used in [23, 36] are used in this paper: Hamming loss, One-error, Ranking loss, Coverage and average precision.

For a classifier$h$, let $h(x) \subseteq \mathcal{L}$ denote its multi-label prediction for an instance x, and let l_x denote the true set of relevant labels. Moreover, in case a related scoring function f is also defined, let $f(x,\lambda)$ denote the score assigned to label λ for instance x. The most commonly used evaluation measures are defined as follows:

a) *Hamming loss* computes the percentage of labels whose relevance is predicted incorrectly:

$$Hamloss(h) = \frac{1}{|\mathcal{L}|} \cdot |h(x)\ \Delta\ \mathrm{L_x}| \qquad (1)$$

where is the symmetric difference between two sets.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

b) *One error* computes how many times the top-ranked label is not relevant:

$$One\ error(f) = \begin{cases} 1, & If arg\ max_{\lambda \epsilon \mathcal{L}} f(x,\lambda) \notin L_x \\ 0 & otherwise \end{cases} \quad (2)$$

c) *Coverage* determines how far one needs to go in the list of labels to cover all the relevant labels of an instance. This measure is loosely related to the precision at the level of perfect recall:

$$Coverage(f) = max_{\lambda \epsilon L_x}\, rank_f\ (x\ , \lambda) - 1 \quad (3)$$

Where $\mathrm{rank}_f(x\ ,\lambda)$ denotes the position of label x in the ordering induced by$f$.

d) *Rank loss* computes the average fraction of label pairs that are not correctly ordered:

$$Rankloss(f) = \frac{\#\{(\lambda,\lambda')|f(x,\lambda) \leq f(x,\lambda'),(\lambda,\lambda\prime)\epsilon L_x \times L'_x\}}{|L_x|\ |\ L'_x|} \quad (4)$$

Where, $\mathcal{L}/L_x = L'_x$ is the set of irrelevant labels.

e) *Average precision* determines for each relevant label $\lambda \in L_x$ the percentage of relevant labels among all labels that are ranked above it, and averages these percentages over all relevant labels:

$$Aveprec(f) =$$

$$\frac{1}{|L_x|} \sum_{\lambda \epsilon L_x} \frac{|\{\lambda' | rank_f(x,\lambda') \leq rank_f(x,\lambda), \lambda\prime \epsilon L_x\}|}{rank_f(x,\lambda)} \quad (5)$$

It should be noted that smaller values indicate better performance for all measures except average precision. Finally, except for coverage, all measures are normalized and assume values between 0 and 1.

4.4 RESULTS AND DISCUSSION

This section presents the results of the evaluation experiments that we conducted. The predictive performances of the examined IB-ELR algorithms are evaluated using the 10-fold cross-validation are summarized from Table 3 to table 12. Tables 3 to 7 are related to the results on balanced data sets and tables 8 to 12 are related to the result on imbalanced data sets. Multi-label classification evaluation measures including the example-based Hamming-loss, ranking-based one error, Ranking Loss and average precision are employed to present the evaluation results of the examined IB-ELR algorithms.

The average ranks of experimental results for balanced data sets shown that IBLR-ML outperforms all other methods. To comparing IBLR-ML with IB-ELR $_{\text{Bagging}}$, we are looked that IBLR-ML is stronger than IB-ELR for One Error and

Coverage measures, for Ranking Loss measure IB-ELR $_{\text{Bagging}}$ is powerful and for other measures they are similar. This result is not unexpected, because logistic regression is stable method and clearly bagging will have little benefit when used with stable base learning algorithms (i.e., most ensemble members will be very similar). Moreover, [83] had concluded that bagging is systematically detrimental to performance for logistic regression. It is noteworthy that these results have been obtained on balanced data sets.

This is surprising in attention to experimental results of imbalanced data sets that IB-ELR$_{\text{Bagging}}$ consistently outperforms all other methods, especially IBLR-ML, for all measures. Our experimental results show that the performance of IBLR-ML in the imbalanced data sets is undesirable.

Diversity also plays an important role in improving the performance of ensemble classifier. In bagging methods, diversification is maintained by creating individual classifiers on different subsets of the training data. Here though logistic regression does not suffered from variance in balanced data sets, the diversity of the base learner produced from imbalanced data set. Actually imbalanced data set can convert a stable learning algorithm into unstable one for ensemble bagging strategy.

According to IB-ELR$_{\text{Boosting}}$ results, it is perceived that IB-ElLR$_{\text{Boosting}}$ had disappointing experimental results on the both of balanced and imbalanced data sets. These results have two causes. One of them is that the base classifiers should be *weak learners*; a classifier learning algorithm is said to be weak when low changes in data produce big changes in the induced model; this is why the most commonly used base classifiers are tree induction algorithms. Another one is that Adaboost algorithm by itself can’t deal with the imbalance problem directly; it has to be changed or combined with another technique, since it focus their attention on difficult examples without differentiating their class. In an imbalanced dataset, majority class examples contribute more to the accuracy (they are more probably difficult examples); hence, rather than trying to improve the true positives, it is easier to improve the true negatives, also increasing the false negatives, which is not a desired characteristic [8].

TABLE 3.
Experimental results based on the Hamming Loss for the balanced data sets.

Hamming Loss ↓	IB-ELR $_{\text{Bagging}}$	IB-ELR $_{\text{Boosting}}$	IBLR-ML	MLKNN	BR-LR	BR-C4.5	BR-KNN	LP
Emotions	.1886 (2)	.1887 (3)	.1883 (1)	.1951 (5)	.2190 (6)	.2474 (7)	.1934 (4)	.2777 (8)
Emotions(v2)	.1923 (2)	.1914 (1)	.1943 (3)	.2040 (5)	.2333 (6)	.2553 (7)	.2020 (4)	.2810 (8)
Image	.1874 (3)	.1864 (1.5)	.1864(1.5)	.1913 (4)	.2013 (6)	.2406 (7)	.1914 (5)	.2571(8)
Reuters	.0820 (4)	.0821 (5.5)	.0821(1.5)	.0826 (7)	.0489 (1)	.0583 (2)	.0903 (8)	.0669 (3)
Scene	.0833 (1)	.0845 (3)	.0834 (2)	.0862 (4)	.1393 (7)	.1368 (6)	.0920 (5)	.1437 (8)
Ave. rank	**(2.4)**	(2.8)	**(2.4)**	(5)	(5.2)	(5.8)	(5.2)	(7)

TABLE 4.
Experimental results based on the One Error for the balanced data sets.

One Error ↓	IB-ELR $_{\text{Bagging}}$	IB-ELR $_{\text{Boosting}}$	IBLR-ML	MLKNN	BR-LR	BR-C4.5	BR-KNN	LP
Emotions	.1886 (2)	.1887 (3)	.1883 (1)	.1835 (5)	.2869 (6)	.3913 (7)	.2565 (4)	.4672 (8)

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Emotions(v2)	.1923 (2)	.1914 (1)	.1943 (3)	.2980 (5)	.3440 (6)	.4260 (7)	.2840 (4)	.4880 (8)
Image	.3700 (3)	.4030 (6)	.3665 (2)	.3715(4)	.3660 (1)	.5000 (7)	.3830 (5)	.5100 (8)
Reuters	.2181 (5)	.2408 (8)	.2187 (6)	.2163 (4)	.0871 (1)	.1458 (2)	.2278 (7)	.1724 (3)
Scene	.2243(3.5)	.2418 (5)	.2235 (2)	.2243(3.5)	.3665 (6)	.4138 (8)	.0889 (1)	.3984 (7)
Ave. rank	(3.1)	(4.6)	**(2.8)**	(4.3)	(4)	(6.2)	(4.2)	(6.8)

TABLE 5.
Experimental results based on the Coverage for the balanced data sets.

Coverage ↓	IB-ELR Bagging	IB-ELR Boosting	IBLR-ML	MLKNN	BR-LR	BR-C4.5	BR-KNN	LP
Emotions	1.7137(2)	1.8590(6)	1.7087(1)	1.7884(3)	1.8493(5)	2.5507(7)	1.8018 (4)	2.6854(8)
Emotions(v2)	1.6360(2)	1.7885(5)	1.6340(1)	1.6960(4)	1.8540(6)	2.5840(7)	1.6860 (3)	2.6520(8)
Image	1.0720(4)	1.2200(6)	1.0665(3)	1.0600(2)	1.0560(1)	1.599(7.5)	1.0975 (5)	1.599(7.5)
Reuters	.7543 (2)	.9614 (7)	.7561 (3)	.7644 (4)	.4443 (1)	.8829 (6)	.8130 (5)	1.0393(8)
Scene	.4607 (1)	.6223 (5)	.4642 (2)	.4744 (3)	.8855 (6)	1.3345(8)	.5314 (4)	1.1570(7)
Ave. rank	(2.2)	(5.8)	**(2)**	(3.2)	(3.8)	(7.1)	(4.2)	(7.7)

TABLE 6.
Experimental results based on the Rank Loss for the balanced data sets.

Rank Loss ↓	IB-ELR Bagging	IB-ELR Boosting	IBLR-ML	MLKNN	BR-LR	BR-C4.5	BR-KNN	LP
Emotions	.1512 (2)	.1766 (6)	.1496 (1)	.1633 (4)	.1731 (5)	.2915 (6)	.1610 (3)	.3442 (7)
Emotions(v2)	.1583 (1)	.1846 (5)	.1587 (2)	.1660 (4)	.1951 (6)	.3170 (7)	.1652 (3)	.3584 (8)
Image	.2007 (4)	.2380 (6)	.1991 (3)	.1983 (2)	.1969 (1)	.3303 (7)	.2090 (5)	.3365 (8)
Reuters	.0818 (2)	.1134 (7)	.0821 (3)	.0828 (4)	.0305 (1)	.0918 (6)	.0905 (5)	.1203 (8)
Scene	.0753 (1)	.1069 (5)	.0760 (2)	.0774 (3)	.1585 (6)	.2465 (8)	.0889 (4)	.2125 (7)
Ave. rank	**(2)**	(5.8)	(2.2)	(3.4)	(3.8)	(6.8)	(4)	(7.6)

TABLE 7.
Experimental results based on the Average Precision for the balanced data sets.

Ave. Prec. ↑	IB-ELR Bagging	IB-ELR Boosting	IBLR-ML	MLKNN	BR-LR	BR-C4.5	BR-KNN	LP
Emotions	.8103 (2)	.7899 (6)	.1826 (1)	.7965 (4)	.7903 (5)	.7014 (7)	.8037 (3)	.6608 (8)
Emotions(v2)	.8016 (1)	.7773 (5)	.8013 (2)	.7842 (4)	.7545 (6)	.6726 (7)	.7928 (3)	.6410 (8)
Image	.7602 (4)	.7310 (6)	.7619 (2)	.7609 (3)	.7638 (1)	.6577 (7)	.7531 (5)	.6494 (8)
Reuters	.8606 (4)	.8351 (8)	.8601 (5)	.8600 (6)	.9439 (1)	.8866 (2)	.8504 (7)	.8624 (3)
Scene	.8674 (1)	.8449 (5)	.8673 (2)	.8662 (5)	.7672 (6)	.7109 (8)	.8496(4)	.7306 (7)
Ave. rank	**(2.4)**	(6)	**(2.4)**	(4)	(3.8)	(6.2)	(4.4)	(6.8)

TABLE 8.
Experimental results based on the Hamming Loss for the imbalanced data sets.

Hamming Loss ↓	IB-ELR Bagging	IB-ELR Boosting	IBLR-ML	MLKNN	BR-LR	BR-C4.5	BR-KNN	LP
Emotions(v3)	.1863 (1)	.1869 (2)	.1894 (3)	.1913 (5)	.2217 (6)	.2438 (8)	.1897 (4)	.2612 (7)
Genbase	.0022 (5)	.0018 (5)	.0029 (6)	.0048 (8)	.0019(3.5)	.0011 (1)	.0038 (7)	.0019 (3.5)
Mediamill	.0314 (3)	.0325 (6)	.0320 (4)	.0305 (2)	.0322 (5)	.0357 (7)	.0304 (1)	.0445 (8)
Scene(v2)	.0796(1.5)	.0780 (3)	.0796(1.5)	.0810 (4)	.1365 (8)	.1216 (6)	.0888 (5)	.1332 (7)
Scene(v3)	.0819(2.5)	.0796 (1)	.0819(2.5)	.0853 (4)	.1469 (8)	.1193 (6)	.0907 (5)	.1262 (7)
yeast	.1928 (1)	.1933 (2.5)	.1934 (4)	.1933(2.5)	.2050 (6)	.2454 (7)	.1952 (5)	.2779 (8)
Ave. rank	**(2.4)**	(2.75)	(3.5)	(4.25)	(6.08)	(5.83)	(4.5)	(6.75)

TABLE 9.
Experimental results based on the One Error for the imbalanced data sets.

One Error ↓	IB-ELR Bagging	IB-ELR Boosting	IBLR-ML	MLKNN	BR-LR	BR-C4.5	BR-KNN	LP
Emotions(v3)	.3226 (4)	.3041 (1)	.3273 (5)	.3089 (2)	.3712 (6)	.4424 (7)	.3225 (3)	.4798 (8)
Genbase	.0030(2.3)	.0030(2.3)	.0135 (6)	.0136 (7)	.0121 (5)	.0030(2.3)	.0166 (8)	.0106 (4)
Mediamill	.1968 (3)	.3720 (6)	.2532 (4)	.1924 (2)	.3326 (5)	.4440(7)	.1880 (1)	.6936 (8)
Scene(v2)	.2407 (1)	.2567 (4)	.2423 (2)	.2459 (3)	.4052 (7)	.4211 (8)	.2866 (5)	.4000 (6)
Scene(v3)	.2428 (1)	.2576 (4)	.2461 (2)	.2534(3)	.4237 (8)	.4010 (7)	.2966 (5)	.3790 (6)
yeast	.2242 (1)	.0327 (6)	.2263 (2)	.2292 (3)	.2400 (5)	.3993 (7)	.2309 (4)	.5139(8)
Ave. rank	**(2.05)**	(3.88)	(3.5)	(4)	(3.33)	(6.38)	(4.33)	(6.67)

TABLE 10.
Experimental results based on the Coverage for the imbalanced data sets.

Coverage ↓	IB-ELR Bagging	IB-ELR Boosting	IBLR-ML	MLKNN	BR-LR	BR-C4.5	BR-KNN	LP
Emotions(v3)	1.5236(1)	1.6241 (5)	1.5261 (2)	1.5627(3)	1.6892(6)	2.2896(7)	1.6126(4)	2.5250(8)
Genbase	.3926 (2)	.4281 (4)	.4936 (6)	.5559 (8)	.3986 (3)	.3564 (1)	.4367 (5)	.4997 (7)
Mediamill	17.6362(5)	17.5282(4)	17.4482(3)	14.7694(1)	15.6048(2)	48.8870(7)	22.8582(6)	60.1690(8)
Scene(v2)	.3742 (1)	.4949 (5)	.3794 (2)	.3822 (3)	.9789 (6)	1.2015(8)	.4789 (4)	1.0309(7)
Scene(v3)	.3715 (1)	.4507 (4)	.3755(2)	.3826 (3)	.9419(7)	1.1940(8)	.4710 (5)	.9292(6)
yeast	6.1906 (1)	6.3419 (4)	6.1927 (2)	6.2324 (3)	6.4857(5)	9.2398(7)	6.5245(6)	9.3647(8)
Ave. rank	**(1.83)**	(4.33)	(2.83)	(3.5)	(4.83)	(6.33)	(5)	(7.33)

TABLE 11.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Experimental results based on the Ranking Loss for the imbalanced data sets.

Rank Loss ↓	IB-ELR Bagging	IB-ELR Boosting	IBLR-ML	MLKNN	BR-LR	BR-C4.5	BR-KNN	LP
Emotions(v3)	.1682 (1)	.1814 (5)	.1694 (2)	.1702 (3)	.1982 (9)	.3068 (7)	.1796 (4)	.3640 (8)
Genbase	.0029 (2)	.0031 (3)	.0053 (6)	.0062 (7)	.0046 (4)	.0028 (1)	.0052 (5)	.0076 (8)
Mediamill	.0506 (4)	.0533 (5)	.0497 (3)	.0415 (1)	.0421 (2)	.1706 (7)	.0667 (6)	.3459 (8)
Scene(v2)	.0746 (1)	.0987 (5)	.0757 (2)	.0776 (3)	.1956 (6)	.2401 (8)	.0955 (4)	.2060 (7)
Scene(v3)	.0739 (1)	.0896 (4)	.0747 (2)	.0761 (3)	.1882 (7)	.2384 (8)	.0937 (5)	.1857 (6)
yeast	.1632 (1)	.1782 (6)	.1635 (2)	.1652 (3)	.1763 (4)	.3097 (7)	.1778 (5)	.3993 (8)
Ave. rank	**(1.67)**	(4.67)	(2.83)	(3.33)	(5.33)	(6.33)	(4.83)	(7.5)

TABLE 12.
Experimental results based on the Average precision for the imbalanced data sets.

Ave. Prec. ↑	IB-ELR Bagging	IB-ELR Boosting	IBLR-ML	MLKNN	BR-LR	BR-C4.5	BR-KNN	LP
Emotions(v3)	.7723 (2)	.7702 (4)	.7712 (3)	.7736 (1)	.7421 (6)	.6702 (7)	.7645 (5)	.6382 (8)
Genbase	.9917 (3)	.9921 (2)	.9860 (7)	.9864 (6)	.9891 (4)	.9927 (1)	.9820 (8)	.9871 (5)
Mediamill	.7105 (3)	.6421 (6)	.7001 (5)	.7262 (1)	.7010 (4)	.5549 (7)	.7152 (2)	.3055 (8)
Scene(v2)	.8620 (1)	.8423 (4)	.8606 (2)	.8582 (3)	.7372 (7)	.7118 (8)	.8336(5)	.7374 (6)
Scene(v3)	.8619 (1)	.8464 (4)	.8602 (2)	.8559 (3)	.7351 (7)	.7227 (8)	.8295 (5)	.7559 (6)
yeast	.7698 (1)	.7470 (6)	.7687 (2)	.7658 (3)	.7549 (5)	.6216 (7)	.7599 (4)	.5723 (8)
Ave. rank	**(1.83)**	(4.33)	(3.5)	(2.83)	(5.5)	(6.33)	(4.83)	(6.83)

5. CONCLUSION

In this paper, a multi-label classification algorithm, which called IBLR_ML [23], is employed. This algorithm is produced from combination of k-nearest neighbor and logistic regression algorithms. We use ensemble techniques within the logistic regression part of IBLR-ML. while the bagging fails with stable learning algorithms whose output is insensitive to small changes in the input and also logistic regression is stable method, we employ logistic regression as the base learner.

We empirically evaluated the proposed approach by measuring its performance on eleven benchmark multi-label datasets from different domains, variable sizes and imbalance ration. In addition to, six successful multi-label classification methods are chosen to compare with my proposed approach.

The average ranks of experimental results for balanced data sets shown that IBLR-ML outperforms all other methods. This is because of the base classifiers should be unstable learner; a classifier learning algorithm is said to be unstable when low changes in data produce big changes in the induced model. This is surprising in attention to experimental results of imbalanced data sets that IB-ELR$_{\text{Bagging}}$ consistently outperforms all other methods, especially IBLR-ML, for all measures. Its cause is that, actually imbalanced data set can convert a stable learning algorithm into unstable one for ensemble bagging strategy. According to IB-ELR$_{Boosting}$ results in imbalanced data sets, it is perceived that IB-ElLR$_{Boosting}$ had disappointing experimental results since Adaboost algorithm by itself can't deal with the imbalance problem directly. In an imbalanced dataset, majority class examples contribute more to the accuracy (they are more probably difficult examples); hence, rather than trying to improve the true positives, it is easier to improve the true negatives, also increasing the false negatives, which is not a desired characteristic.

REFERENCES

[1] R. E. Schapire and Y. Singer, "BoosTexter: A boosting-based system for text categorization," *Machine Learning*, vol. 39, pp. 135-168, 2000.

[2] T. Joachims, C. Nedellec, and C. Rouveirol, "Text categorization with Support Vector Machines: Learning with many relevant features," in *Machine Learning: ECML-98*. vol. 1398: Springer Berlin / Heidelberg, pp. 137-142, 1998.

[3] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognition*, vol. 37, pp. 1757-1771, 2004.

[4] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification," in *In Advances in Neural Information Processing Systems 14*. vol. 14, pp. 681-687, 2001.

[5] Y. Sun, A. K. C. Wong, and M. S. Kamel, "Classification of imbalanced data: A review," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, pp. 687-719, 2009.

[6] H. Haibo and E. A. Garcia, "Learning from Imbalanced Data," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 21, pp. 1263-1284, 2009.

[7] O. Maimon, L. Rokach, and N. Chawla, "Data Mining for Imbalanced Datasets: An Overview," in *Data Mining and Knowledge Discovery Handbook*: Springer US, pp. 853-867, 2005.

[8] M. Galar, A. FernÃ¡ndez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 42, pp. 463-484, 2012.

[9] D. P. Williams, V. Myers, and M. S. Silvious, "Mine classification with imbalanced data," *Geoscience and Remote Sensing Letters, IEEE*, vol. 6, pp. 528-532, 2009.

[10] Y. M. Huang, C. M. Hung, and H. C. Jiau, "Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem," *Nonlinear Analysis: Real World Applications*, vol. 7, pp. 720-747, 2006.

[11] W.-Z. Lu and D. Wang, "Ground-level ozone prediction by support vector machine approach with a cost-sensitive classification scheme," *Science of The Total Environment*, vol. 395, pp. 109-116, 2008.

[12] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss, "A methodological approach to the classification of dermoscopy images," *Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society*, vol. 31, pp. 362, 2007.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

[13] D. A. Cieslak, N. V. Chawla, and A. Striegel, "Combating imbalance in network intrusion datasets," *network*, vol. 11, pp. 15, 2006.
[14] T. Dietterich, "Machine-learning research: four current directions," *AI Magazine*, vol. 18, pp. 97-136, 1997.
[15] V. C. Nitesh and S. Jared, "Exploiting diversity in ensembles: improving the performance on unbalanced datasets," in *Proceedings of the 7th international conference on Multiple classifier systems* Prague, Czech Republic: Springer-Verlag, 2007.
[16] T. Theeramunkong, C. Nattee, P. L. Adeodato, N. Chawla, P. Christen, P. Lenca, J. Poon, G. Williams, K.-W. Hsu, and J. Srivastava, "An Empirical Study of Applying Ensembles of Heterogeneous Classifiers on Imperfect Data," in *New Frontiers in Applied Data Mining*. vol. 5669: Springer Berlin Heidelberg, pp. 28-39, 2009.
[17] R. Polikar, "Ensemble based systems in decision making," *Circuits and Systems Magazine, IEEE*, vol. 6, pp. 21-45, 2006.
[18] Zhou Z., "Ensemble learning," in *Encyclopedia of biometrics*, 2009.
[19] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, pp. 1-39, 2010.
[20] R. E. Schapire, "The strength of weak learnability," *Machine Learning*, vol. 5, pp. 197-227, 1990.
[21] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol. 55, pp. 119-139, 1997.
[22] L. Breiman, "Bagging Predictors," *Machine Learning*, vol. 24, pp. 123-140, 1996.
[23] W. Cheng and E. Hüllermeier, "Combining instance-based learning and logistic regression for multilabel classification," *Machine Learning*, vol. 76, pp. 211-225, 2009.
[24] G. Tsoumakas and I. katakis, "Multi-label classification: an overview," *international journal of data warehousing and mining 3*, vol. 3, pp. 1-13, 2007.
[25] G. Tsoumakas, I. Katakis, I. Vlahavas, O. Maimon, and L. Rokach, "Mining multi-label data," in *Data Mining and Knowledge Discovery Handbook*, 2nd ed, O. Maimon and L. Rokach, Eds.: Springer US, pp. 667-685, 2010.
[26] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," *Machine learning: ECML-98*, pp. 137-142, 1998.
[27] E. Hullermeier, J. Furnkranz, W. Cheng, and K. Brinker, "Label ranking by learning pairwise preferences," *Artificial Intelligence*, vol. 172, pp. 1897-1916, 2008.
[28] J. Furnkranz, E. Hullermeier, E. Loza Mencia, and K. Brinker, "Multilabel classification via calibrated label ranking," *Machine Learning*, vol. 73, pp. 133-153, 2008.
[29] C. Amanda and D. K. Ross, "Knowledge Discovery in Multi-label Phenotype Data," in *Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2001)*, pp. 42-53, 2001
[30] A. Jiang, C. Wang, and Y. Zhu, "Calibrated rank-svm for multi-label image categorization," in *Neural Networks, 2008. IJCNN 2008.(IEEE World*

Congress on Computational Intelligence). IEEE International Joint Conference on, pp. 1450-1455, 2008.

[31] Xu J., "An efficient multi-label support vector machine with a zero label," *Expert Systems with Applications,* vol. 39, pp. 4796-4804, 2011.

[32] F. A. Thabtah, P. Cowling, Y. Peng, R. Rastogi, K. Morik, M. Bramer, and X. Wu, "MMAC: A New Multi-Class, Multi-Label Associative Classification Approach," in *Proceedings of Fourth IEEE International Conference on Data Mining, ICDM 2004*, pp. 217-224, 2004.

[33] M. L. Zhang and Z. H. Zhou, "Multilabel neural networks with applications to functional genomics and text categorization," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 18, pp. 1338-1351, 2006.

[34] Xu J., "An extended one-versus-rest support vector machine for multi-label classification," *Neurocomputing,* vol. 74, pp. 3114-3124, 2011.

[35] M.-L. Zhang, "ML-RBF: RBFneural networks for multi-label learning," *Neural Processing Letters,* vol. 29, pp. 61-74, 2009.

[36] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognition,* vol. 40, pp. 2038-2048, 2007.

[37] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Editorial: special issue on learning from imbalanced data sets," *ACM SIGKDD Explorations Newsletter,* vol. 6, pp. 1-6, 2004.

[38] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explorations Newsletter,* vol. 6, pp. 20-29, 2004.

[39] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intelligent data analysis,* vol. 6, pp. 429-449, 2002.

[40] G. M. Weiss, "Mining with rarity: a unifying framework," *Sigkdd Explorations,* vol. 6, pp. 7-19, 2004.

[41] J. R. Quinlan, "Improved estimates for the accuracy of small disjuncts," *Machine Learning,* vol. 6, pp. 93-98, 1991.

[42] B. Zadrozny and C. Elkan, "Learning and making decisions when costs and probabilities are both unknown," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 204-213, 2001.

[43] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," *Machine Learning: ECML 2004,* pp. 39-50, 2004.

[44] B. Raskutti and A. Kowalczyk, "Extreme re-balancing for SVMs: a case study," *ACM SIGKDD Explorations Newsletter,* vol. 6, pp. 60-69, 2004.

[45] G. Wu and E. Y. Chang, "Class-boundary alignment for imbalanced dataset learning," in *ICML 2003 workshop on learning from imbalanced data sets II, Washington, DC*, pp. 49-56, 2003.

[46] Y. Lin, Y. Lee, and G. Wahba, "Support vector machines for classification in nonstandard situations," *Machine Learning,* vol. 46, pp. 191-202, 2002.

[47] K. J. Ezawa, M. Singh, and S. W. Norton, "Learning goal oriented Bayesian networks for telecommunications risk management," in *MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-*, pp. 139-147, 1996.

[48] I. Mani, "knn approach to unbalanced data distributions: A case study involving information extraction," in *Proceedings of Workshop on Learning from Imbalanced Datasets*, 2003.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

[49] B. Liu, W. Hsu, and Y. Ma, "Mining association rules with multiple minimum supports," in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 337-341, 1999.
[50] A. K. C. Wong and Y. Wang, "High-order pattern discovery from discrete-valued data," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 9, pp. 877-893, 1997.
[51] B. Liu, Y. Ma, and C. Wong, "Improving an association rule based classifier," *Principles of Data Mining and Knowledge Discovery,* pp. 293-317, 2000.
[52] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Handling imbalanced datasets: A review," *GESTS International Transactions on Computer Science and Engineering,* vol. 30, pp. 25-36, 2006.
[53] S. L. Phung, A. Bouzerdoum, and G. H. Nguyen, "Learning pattern classification tasks with imbalanced data sets," 2009.
[54] S. Wang and X. Yao, "Diversity analysis on imbalanced data sets by using ensemble models," in *Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on*, pp. 324-331, 2009.
[55] R. Barandela, R. M. Valdovinos, and J. S. SÃ¡nchez, "New applications of ensembles of classifiers," *Pattern Analysis & Applications,* vol. 6, pp. 245-256, 2003.
[56] J. Blaszczynski, M. Deckert, J. Stefanowski, and S. Wilk, "Integrating selective pre-processing of imbalanced data with Ivotes ensemble," in *Rough Sets and Current Trends in Computing,* pp. 148-157, 2010.
[57] N. Chawla, A. Lazarevic, L. Hall, and K. Bowyer, "SMOTEBoost: Improving prediction of the minority class in boosting," *Knowledge Discovery in Databases: PKDD 2003,* pp. 107-119, 2003.
[58] S. Hu, Y. Liang, L. Ma, and Y. He, "MSMOTE: Improving classification performance when training data is imbalanced," in *Computer Science and Engineering, 2009. WCSE'09. Second International Workshop on*, pp. 13-17, 2009.
[59] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "Rusboost: A hybrid approach to alleviating class imbalance," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on,* vol. 40, pp. 185-197, 2010.
[60] G. Hongyu and L. V. Herna, "Learning from imbalanced data sets with boosting and data generation: the DataBoost-IM approach," *ACM SIGKDD Explorations Newsletter,* vol. 6, pp. 30-39, 2004.
[61] K. Chen, B. L. Lu, and J. T. Kwok, "Efficient classification of multi-label and imbalanced data using min-max modular classifiers," in *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, pp. 1770-1775, 2006.
[62] G. Tepvorachai and C. Papachristou, "Multi-label imbalanced data enrichment process in neural net classifier training," in *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on,* pp. 1301-1307, 2008.
[63] E. S. Xioufis, M. Spiliopoulou, G. Tsoumakas, and I. Vlahavas, "Dealing with concept drift and class imbalance in multi-label stream classification,"

in *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Two*, 2011, pp. 1583-1588.

[64] Tahir M., Kittler J., and Bouridane A., "Multilabel classification using heterogeneous ensemble of multi-label classifiers," *Pattern Recognition letters*, vol. 33, pp. 513-523, 2012.

[65] J. Read, B. Pfahringer, and G. Holmes, "Multi-label Classification Using Ensembles of Pruned Sets," in *International Conference on Data Mining. ICDM '08. Eighth IEEE*, pp. 995-1000, 2008.

[66] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Random k-Labelsets for Multilabel Classification," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 23, pp. 1079-1089, 2010.

[67] G. Tsoumakas, I. Vlahavas, J. Kok, J. Koronacki, R. Mantaras, S. Matwin, D. Mladenic, and A. Skowron, "Random k -Labelsets: An Ensemble Method for Multilabel Classification," *Machine Learning: ECML 2007*, vol. 4701, pp. 406-417, 2007.

[68] G. Nasierding, A. Z. Kouzani, and G. Tsoumakas, "A triple-random ensemble classification method for mining multi-label data," in *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pp. 49-56, 2010.

[69] W. Fei, H. Yahong, T. Qi, and Z. Yueting, "Multi-label Boosting for Image Annotation by Structural Grouping Sparsity," in *Proceedings of the international conference on Multimedia*, Firenze, Italy, pp. 15-24, 2010.

[70] K. Dembczynski, W. Cheng, and E. HÃ¼llermeier, "Bayes optimal multilabel classification via probabilistic classifier chains," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 279-286, 2010.

[71] J. Read, B. Pfahringer, G. Holmes, E. Frank, W. Buntine, M. Grobelnik, D. Mladenic, and J. Shawe-Taylor, "Classifier Chains for Multi-label Classification," *Machine Learning and Knowledge Discovery in Databases*, vol. 5782, pp. 254-269, 2009.

[72] A. Santos, A. Canuto, and A. Neto, "A comparative analysis of classification methods to multi-label tasks in different application domains," *International journal of computer Information systems and Industrial Management Applicationsâ€•. ISSN*, pp. 2150-7988, 2011.

[73] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-*, pp. 148-156, 1996.

[74] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, pp. 10-18, 2009.

[75] G. Tsoumakas, J. Vilcek, L. Spyromitros, and I. Vlahavas, "Mulan: A java library for multi-label learning," *Journal of Machine Learning Research*, vol. 1, pp. 1-48, 2010.

[76] A. Orriols-Puig and E. BernadÃ³-Mansilla, "Evolutionary rule-based systems for imbalanced data sets," *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, vol. 13, pp. 213-225, 2009.

[77] Z. Ding, "Diversified ensemble classifiers for highly imbalanced data learning and its application in bioinformatics," GEORGIA STATE UNIVERSITY, 2011.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

[78] K. Trohidis, G. Tsoumakes, G. Kalliris, and I. Vlahavas, "MultiLabel classification of music into emotions," in *9th International Conference on Music Information Retrieval (ISMIR 2008)*, Philadelphia, PA,USA, 2008.

[79] Z. H. Zhou and M. L. Zhang, "Multi-instance multi-label learning with application to scene classification," *Advances in Neural Information Processing Systems*, vol. 19, pp. 1609, 2007.

[80] F. Sebastiani, "Machine learning in automated text categorization," *ACM computing surveys (CSUR)*, vol. 34, pp. 1-47, 2002.

[81] C. G. M. Snoek, M. Worring, J. C. Van Gemert, J. M. Geusebroek, and A. W. M. Smeulders, "The challenge problem for automated detection of 101 semantic concepts in multimedia," in *Proceedings of the 14th annual ACM international conference on Multimedia*, pp. 421-430, 2006.

[82] G. Nasierding, G. Tsoumakas, and A. Z. Kouzani, "Clustering based multi-label classification for image annotation and retrieval," in *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, pp. 4514-4519, 2009.

[83] C. Perlich, F. Provost, and J. S. Simonoff, "Tree induction vs. logistic regression: a learning-curve analysis," *The Journal of Machine Learning Research*, vol. 4, pp. 211-255, 2003.