Improving Multilabel Classification Performance by Using Ensemble of Multi-label Classifiers

Muhammad Atif Tahir, Josef Kittler, Krystian Mikolajczyk, and Fei Yan

Centre for Vision, Speech and Signal Processing
University of Surrey
Guildford, GU2 7XH, UK
{m.tahir,j.kittler,k.mikolajczyk,f.yan}@surrey.ac.uk

Abstract. Multilabel classification is a challenging research problem in which each instance is assigned to a subset of labels. Recently, a considerable amount of research has been concerned with the development of "good" multi-label learning methods. Despite the extensive research effort, many scientific challenges posed by e.g. highly imbalanced training sets and correlation among labels remain to be addressed. The aim of this paper is use heterogeneous ensemble of multi-label learners to simultaneously tackle both imbalance and correlation problems. This is different from the existing work in the sense that the later mainly focuses on ensemble techniques within a multi-label learner while we are proposing in this paper to combine these state-of-the-art multi-label methods by ensemble techniques. The proposed ensemble approach (EML) is applied to three publicly available multi-label data sets using several evaluation criteria. We validate the advocated approach experimentally and demonstrate that it yields significant performance gains when compared with state-of-the art multi-label methods.

1 Introduction

A traditional multi-class classification system assigns each instance x a single label l from a set of disjoint labels L. However, in many modern applications such as music categorisation [1], text classification [2,3], image/video categorisation [4,5] etc, each instance is to be assigned to a subset of labels $Y \subseteq L$. This problem is known as multi-label learning.

There is considerable amount of research concerned with the development of "good" multi-label learning methods. Despite the extensive research effort, there exist many scientific challenges. They include highly imbalanced training sets, as very limited data is available for some labels, and capturing correlation among classes. Interestingly, most state-of-the-art multi-label methods are designed to focus mainly on the second problem and very limited effort has been devoted to handling imbalanced data populations. In this paper, we focus on the first problem of multi-label learning, and tackle highly imbalanced data distributions using ensemble of multi-label classifiers.

N. El Gayar, J. Kittler, and F. Roli (Eds.): MCS 2010, LNCS 5997, pp. 11–21, 2010.

Ensemble techniques are becoming increasingly important as they have repeatedly demonstrated the ability to improve upon the accuracy of a single-classifiers [6]. Ensembles can be homogeneous, in which every base classifier is constructed using the same algorithm, or heterogeneous in which base classifiers are constructed using different algorithms. In fact, some state-of-the-art multi-label learners use homogeneous or heterogeneous ensemble techniques to improve the overall performance. For example, in [7] Logistic Regression and Nearest Neighbour classifiers are combined, in [8] random subsets of training data are used. The aim of this paper is to use heterogeneous ensemble of multi-label learners to improve the performance. This is different from the existing work in the sense that the latter mainly focuses on ensemble techniques within a multi-label learner while we are proposing to combine these state-of-the-art multi-label methods by ensemble techniques.

Interestingly another advantage of combining multi-label classifiers is that both imbalance and correlation problems can be tackled simultaneously. The imbalance problem can be handled by using ensemble of multi-label classifiers while the correlation problem can be solved by using state-of-the-art multi-label classifiers as base classifiers that inherently consider correlation among labels. The proposed ensemble approach (EML) is applied to three publicly available multi-label data sets from different domains (Scene, Yeast, and Enron) using 18 different multi-label classification measures. We validate the advocated approach experimentally and demonstrate that it yields significant performance gains when compared with individual state-of-the art multi-label methods.

The paper is organised as follows. In Section 2, we review state-of-the-art multi-label methods. Section 3 discusses the proposed ensemble of multi-label classifiers. Experiments are discussed in Section 4 followed by the results and discussion in Section 5. Section 6 concludes the paper.

2 Related Work

The sparse literature on multi-label classification driven by problems in text classification, bioinformatics, music categorisation, and image/video classification, has recently been summarised by Tsoumakas et al [9]. This research can be divided into two different groups: i) *problem transformation* methods, and ii) *algorithm adaptation* methods. The problem transformation methods aim to transform multilabel classification task into one or more single-label classification [10,11], or label ranking [12] tasks. The algorithm adaptation methods extend traditional classifiers to handle multi-label concepts directly [13,14,7]. In this section, we review the state-of-the-art multi-label learners that are used as base classifiers in our ensemble approach namely RaKEL [11], Calibrated Label Ranking (CLR) [12], Multi-label KNN (MLKNN) [13], Instance Based Logistic Regression (IBLR) [7] and Ensemble of Classifier Chains (ECC) [8].

RaKEL: Multilabel classification can be reduced to the conventional classification problem by considering each unique set of labels as one of the classes.

This approach is referred to as *label powerset* (LP) in the literature. However, this approach leads to a large number of label subsets with the majority of them with a very few examples and it is also computationally expensive. Many approaches have been proposed in the literature to deal with the aforementioned problem [11,15]. One state-of-the-art approach is RaKEL (Random k-Labelsets) [11] that constructs an ensemble of LP classifiers where each LP classifier is trained using a different small random subset of the set of labels. In order to get near-optimal performance, appropriate parameters (subset size, number of models, threshold etc) must be optimised using internal cross validation. However, these parameters are hard to optimised when the number of training samples is insufficient.

Ensemble of Classifier Chains (ECC): Multilabel classification can be reduced to the conventional binary classification problem. This approach is referred to as *binary relevance* (BR) learning in the literature. In BR learning, the original data set is divided into $|Y|$ data sets where $Y = \{1, 2, ..., N\}$ is the finite set of labels. BR learns one binary classifier $h_a : X \rightarrow \{\neg a, a\}$ for each concept $a \in Y$. BR learning is theoretically simple and has a linear complexity with respect to the number of labels. Its assumption of label independence makes it attractive to situations where new examples may not be relevant to any known subset of labels or where label relationships may change over the test data [8]. However, BR learning is criticised for not considering correlations among the labels [7,12]. The work in [8] is a state-of-the-art BR approach. Their classifier chain (CC) approach only requires a single training iteration like BR and uses labels directly from the training data without any internal classification. Classifiers are linked in a chain where each classifier deals with the BR problem associated with label $y_j \in Y$. However, the order of the chain can clearly have an effect on accuracy. An ensemble framework (ECC) is used to create different random chain orderings. This method was shown to perform well against BR and other multi-label classifiers.

Calibrated Label Ranking (CLR): Like *one-vs-all* approach in BR learning, the binary pairwise *one-vs-one* approach has also been employed for multi-label classification, therefore requiring $|Y|^2$ classifiers as opposed to $|Y|$. Calibrated label ranking (CLR) [12] is an efficient pairwise approach for multilabel classification. 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. This method was shown to perform well against other multi-label classifiers but mainly on ranking measures.

Multi-label KNN (MLKNN): Instance-based approach is also quite popular in multilabel classification. In [13], a lazy learning approach (MLKNN) is proposed. This method is derived from the popular k-Nearest Neighbour (kNN) algorithm. It consists of two main steps. In the first step, for each test instance, its k nearest neighbours in the training set are identified. Next, in the second step, the maximum a posteriori probability label set is identified for a test instance based on the statistical information gained from the label sets of these

neighbouring instances. This method was shown to perform well in some domains e.g. in predicting the functional classes of genes in the Yeast Saccharomyces cerevisiae [7].

Instance Based Logistic Regression (IBLR): IBLR is also a novel approach to instance-based learning with main idea to combine instance-based learning (ILR) and logistic regression [7]. The key idea is to consider the labels of neighboring instances as "features" of unseen samples and thus reduce ILR to logistic regression. This approach captures interdependencies between labels for multilabel classification.

3 Ensemble of Multi-label Classifiers (EML)

Let X denote a set of images (instances) and let $Y = \{1, 2, ..., N\}$ be a set of labels. Given a training set $S = \{(x_1, y_1), ..., (x_m, y_m)\}$ where $x_i \in X$ is a single instance and $y_i \subseteq Y$ is the label set associated with x_i, the goal is to design a multi-label classifier H that predicts a set of labels from an unseen example.

As discussed previously, ensembles methods are well-known for overcoming over-fitting problems especially in highly unbalanced data sets. Ensemble of multi-label classifiers train q multi-label classifiers H_1, H_2, ..., H_q. Thus, all q models are diverse and able to give different multi-label predictions. For an unseen instance x_j, each kth individual model (of q models) produces an N-dimensional vector $P_{jk} = [p_{1k}, p_{2k}, ..., p_{Nk}]$, where the value p_{bk} is the probability of the b^{th} class label assigned by classifier k being correct.

There are many ways of combining the outputs of these q classifiers. Among them, nontrainable combiners such as MEAN, MAX, MIN are the simplest and most popular way to combine the scores of classifiers with probabilistic outputs [16]. These combiners have no extra parameters to be trained. In this paper, nontrainable combiners are used to combine the scores from multi-label classifiers. The only exception is the adjustment of the thresholds using the method described below and proposed by [8]. It is reported in [17] that properly adjusting the decision thresholds (instead of the traditional value of 0.5) can improve the performance for a multi-label classifier. Let the sum of probabilities from q models be stored in a vector $W = (\theta_1, ..., \theta_N) \in \mathbb{R}^{\mathbb{N}}$ such that $\theta_b = \sum_{k=1}^{q} p_{bk}$. W is then normalised to W^{norm}, which represents a distribution of scores for each label in [0,1]. Let X_T be the training set and X_S the test set. A threshold t is then selected using Equation 1 to choose the final predicted multi-label set Z.

$$t = \arg \min_{\{t \in 0.00, 0.001, ..., 1.00\}} |LCard(X_T) - LCard(H_t(X_s))| \tag{1}$$

where LCard (Label Cardinality) is the standard measure of "multi-labelledness" [9]. It is the average number of labels relevant to each instance and is defined as $LCard(X) = \frac{\sum_{i=1}^{|X|} |E_i|}{|X|}$ where E_i is the actual set of labels for the training set and a predicted set of labels under threshold t for the test set. Equation 1 measures the difference between the label cardinality of the training

set and the predictions made on the test set. It avoids intensive internal cross-validation. Hence, the relevant labels in Z under threshold t represent the final predicted set of labels. It should be clear that the actual test labels are never seen by the presented threshold selection method. The threshold t is calculated using the predicted set of labels only.

4 Experiments

Datasets: We experimented with 3 multi-label datasets from a variety of domains. Table 1 shows certain standard statistics of these datasets. The publicly available feature vectors are used in this paper for all datasets [1]. The image dataset "scene" is concerned with semantic indexing of images of still scenes [10]. The "yeast" data set contains functional classes of genes in the Yeast Saccharomyces cerevisiae [11,7]. The "enron" is a subset of the Enron email corpus [15].

Table 1. Standard and multilabel statistics for the data sets used in the experiments

Datasets	Domain	Train	Test	Features	Labels	LCard
Enron	Text	1123	579	1001	53	3.38
Scene	Vision	1211	1196	294	6	1.07
Yeast	Biology	1500	917	103	14	4.24

Evaluation Measures: Multi-label classification requires different evaluation measures than traditional single-label classification. The details can be found in [9]. They are not shown here due to the lack of space. These measures can be categorised into three groups: example based, label-based and ranking-based. In this paper, 18 different evaluation measures are used to compare the proposed approach. These measures include Hamming Loss, Accuracy, Precision, Recall, F_1, and Classification Accuracy from the example-based category, and Micro Precision/Recall/F_1/AUC, Macro Precision/Recall/F_1/AUC from the label-based group. Additionally, we use One-error, Coverage, Ranking Loss and Average Precision from the ranking-based group.

Benchmark Methods: The proposed EML method is compared with the state-of-the-art multi-label classifiers discussed in Section 2: RaKEL [11], ECC [8], CLR [12] , MLKNN [13], and, IBLR [7]. Since all these multi-label classifiers are quite diverse, they are selected as base classifiers in the proposed EML method. MLKNN and IBLR are from the algorithm adaptation group while ECC, RaKEL and CLR are from the problem transformation group. Further, C4.5 is used as a base classifier in RaKEL while Linear SVM is used as a base classifier in ECC and CLR. For the training of MLKNN, IBLR, CLR and Rakel, the Mulan[1] open-source library in Java for multi-label classification is used. For the training of ECC, the MEKA [2] open-source library is used with the default parameters. Both

[1] http://mlkd.csd.auth.fr/multilabel.html
[2] http://www.cs.waikato.ac.nz/ jmr30/software

libraries are an extension of WEKA [18]. All multi-label classifiers are trained using default parameters which are also the best reported parameters e.g. the number of neighbours is 10 for IBLR and MLKNN; the number of iterations is 10 for ECC. The multi-label classifiers are compared with the following variants of the proposed method.

- EML_M: An ensemble of multilabel classifiers using the MEAN rule. It should be noted that we have also tried several others rules such as MAX, MIN and only the best is reported here due to lack of space.
- EML_T: An ensemble of multi-label classifiers using the threshold selection method discussed in Section 3.

5 Results and Discussion

In this section, we discuss the results obtained using EML for the Enron, Medical, Scene and Yeast datasets using Example, Label and Rank based measures.

5.1 Enron Dataset

Table 2 shows the comparison of EML with the state-of-the-art multilabel classifiers for the Enron dataset. First, when the individual multi-label classifiers are compared with each other using various performance measures, it is hard to pick between CLR and RakEL for this dataset. RakEL delivers excellent performance in the majority of Example-based measures while CLR gives the best performance in the majority of Label and Rank-based measures. Overall, CLR and RakEL show the best performance in eight and six evaluation measures respectively. MLKNN and ECC are also the winners in 2 measures. However, by using an ensemble of multi-label classifiers, significant performance gains have been observed in almost all measures. It is also observed that the presented

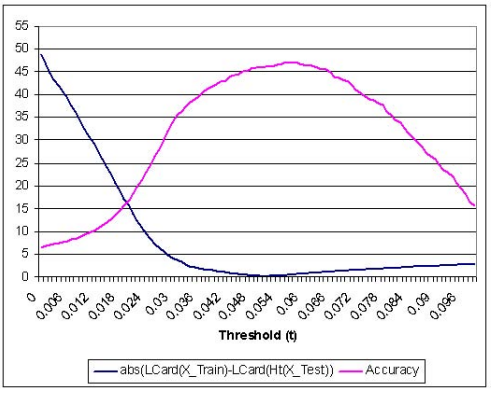

Fig. 1. Threshold t vs $\{|LCard(X_T) - LCard(H_t(X_s))|$,Accuracy$\}$ for Enron

Table 2. Comparison of proposed ensemble method (EML) with the state-of-the-art multi-label classifiers for *Enron* dataset. For each evaluation criterion, ↓ indicates "the smaller the better" while ↑ indicates "the higher the better".

	CLR	RakEL	MLKNN	IBLR	ECC	EML_M	EML_T
HammingLoss ↓	0.062	0.050	0.052	0.057	0.063	**0.047**	0.051
Accuracy ↑	0.401	0.429	0.352	0.337	0.405	0.411	**0.461**
Precision ↑	0.539	0.640	0.630	0.557	0.523	**0.698**	0.608
Recall ↑	0.577	0.509	0.390	0.396	0.589	0.447	**0.626**
Fmeasure ↑	0.557	0.567	0.482	0.463	0.554	0.545	**0.617**
SubsetAccuracy ↑	0.083	0.131	0.093	0.086	0.071	**0.133**	0.071
Micro Precision ↑	0.516	0.651	0.680	0.594	0.507	**0.746**	0.603
Micro Recall ↑	0.553	0.485	0.363	0.379	0.558	0.416	**0.603**
Micro F_1 ↑	0.534	0.556	0.473	0.463	0.531	0.534	**0.603**
Macro Precision ↑	**0.297**	0.213	0.163	0.220	0.217	0.212	0.256
Macro Recall ↑	**0.269**	0.137	0.078	0.130	0.229	0.099	0.169
Macro F_1 ↑	**0.257**	0.149	0.088	0.144	0.210	0.118	0.175
Micro AUC ↑	0.904	0.816	0.900	0.880	0.856	**0.918**	0.916
Macro AUC ↑	0.723	0.592	0.632	0.619	0.662	**0.724**	0.712
One-error ↓	0.309	0.287	0.269	0.378	0.307	**0.238**	**0.238**
Coverage ↓	11.957	24.073	12.751	14.534	19.411	**11.218**	**11.218**
Ranking Loss ↓	0.085	0.195	0.092	0.109	0.148	**0.075**	**0.075**
AvgPrecision ↑	0.649	0.612	0.641	0.612	0.624	**0.699**	**0.699**
# Wins (Ind. Classi.)	8/18	6/18	2/18	0/18	2/18	-	-
# Wins (All)	3/18	0/18	0/18	0/18	0/18	10/18	9/18

threshold selection technique (EML_T) has a significant effect on some evaluation measures. For example, there is a 12% increase in Accuracy while 48% increase in Macro F_1. In summary, a significant improvement is observed by using both variants of the proposed ensembles of multi-label classifier techniques (EML_M and EML_T).

Figure 1 shows the graph for different values of threshold t in the X-axis and $\{|LCard(X_T) - LCard(H_t(X_s))|$,Accuracy$\}$ in the Y-axis for the Enron data set. For clarity, only values with $t < 0.1$ are shown here as optimal value of the threshold selection measure is between 0 and 0.1 for this dataset. It is clear from this graph that the threshold selection method is able to deliver a near-optimal value of accuracy. The optimal value of accuracy is 47.2 under threshold $t = 0.056$ (accuracy curve is plotted using actual test labels for demonstration only). In contrast, the best value of accuracy obtained by Equation 1 is 46.1 under $t = 0.051$ which is very close to the optimal one. In summary, this graph clearly shows the merit of the presented threshold selection method as this simple approach attains near-optimal values without expensive internal cross validation.

5.2 Scene Dataset

Table 3 shows the comparison of EML with the state-of-the-art multilabel classifiers for the Scene dataset. It is interesting to observe that for this data set, IBLR and ECC achieve the highest performance in most of the measures when compared with the individual multi-label classifiers. Overall, IBLR, ECC and CLC

Table 3. Comparison of proposed ensemble method (EML) with the state-of-the-art multi-label classifiers for *Scene* dataset

	CLR	RakEL	MLKNN	IBLR	ECC	EML_M	EML_T
HammingLoss ↓	0.122	0.112	0.099	0.091	0.109	**0.084**	0.095
Accuracy ↑	0.577	0.571	0.629	0.647	0.683	**0.699**	0.694
Precision ↑	0.600	0.598	0.661	0.676	0.716	**0.730**	0.725
Recall ↑	0.669	0.612	0.655	0.655	0.727	0.716	**0.754**
Fmeasure ↑	0.632	0.605	0.658	0.665	0.722	0.723	**0.740**
SubsetAccuracy ↑	0.474	0.503	0.573	0.609	0.605	**0.651**	0.602
Micro Precision ↑	0.666	0.732	0.779	**0.824**	0.696	0.812	0.737
Micro Recall ↑	0.659	0.600	0.634	0.635	0.708	0.696	**0.737**
Micro F_1 ↑	0.663	0.660	0.699	0.717	0.702	**0.750**	0.737
Macro Precision ↑	0.680	0.729	0.784	**0.827**	0.713	0.817	0.764
Macro Recall ↑	0.662	0.609	0.647	0.642	0.715	0.701	**0.741**
Macro F_1 ↑	0.669	0.663	0.692	0.719	0.712	**0.754**	0.748
Micro AUC ↑	0.916	0.894	0.924	0.931	0.901	**0.949**	0.943
Macro AUC ↑	0.910	0.886	0.911	0.927	0.898	**0.943**	0.938
One-error ↓	0.261	0.293	0.242	0.234	0.273	**0.216**	**0.216**
Coverage ↓	0.543	0.691	0.569	0.551	0.639	**0.451**	**0.451**
Ranking Loss ↓	0.088	0.117	0.093	0.090	0.106	**0.070**	**0.070**
AvgPrecision ↑	0.845	0.817	0.851	0.856	0.831	**0.873**	**0.873**
# Wins (Ind. Classi.)	2/18	0/18	0/18	10/18	6/18	-	-
# Wins (All)	0/18	0/18	0/18	2/18	0/18	12/18	8/18

give the best performance in ten, six and two evaluation measures respectively. However, by using the proposed ensemble of multi-label classifiers (EML), significant performance gains have been observed in all measures except Micro/Macro Precision. In summary, in addition to performance gains, fusion of multi-label classifiers also has overcome some limitations of individual multi-label classifiers as the performance of these individual multi-label classifiers may vary in evaluation measures and from one data set to another.

5.3 Yeast Dataset

Table 4 shows the comparison of EML with the state-of-the-art multilabel classifiers for the Yeast dataset. First, when the individual multi-label classifiers are compared with each other using various performance measures, ECC, MLKNN and IBLR demonstrate very good performance on this dataset. IBLR ranks first in the ranked-based measures while the performance vary for ECC and MLKNN in Example and Label based measures. Overall, ECC, IBLR and MLKNN report the best performance in seven, six and four evaluation measures respectively. As before, the fusion of multi-label classifiers has significantly improved the overall performance in all except classification accuracy. It is also observed that the presented threshold selection technique (EML_T) makes a significant impact on the performance especially on example-based measures e.g. 11% and 8.5% increase in Accuracy and Fmeasure respectively when compared with an ensemble using the MEAN rule (EML_M) for fusion.

Table 4. Comparison of proposed ensemble method (EML) with the state-of-the-art multi-label classifiers for *Yeast* dataset

	CLR	RakEL	MLKNN	IBLR	ECC	EML_M	EML_T
HammingLoss ↓	0.210	0.244	0.198	0.199	0.212	**0.193**	0.197
Accuracy ↑	0.497	0.465	0.492	0.506	0.535	0.500	**0.553**
Precision ↑	0.674	0.601	0.732	0.712	0.654	**0.738**	0.682
Recall ↑	0.596	0.618	0.549	0.581	0.669	0.553	**0.690**
Fmeasure ↑	0.633	0.609	0.628	0.640	0.661	0.633	**0.686**
SubsetAccuracy ↑	0.158	0.091	0.159	0.176	**0.197**	0.166	0.190
Micro Precision ↑	0.681	0.596	0.736	0.714	0.648	**0.750**	0.676
Micro Recall ↑	0.585	0.616	0.543	0.576	0.656	0.548	**0.677**
Micro F_1 ↑	0.629	0.605	0.625	0.637	0.652	0.633	**0.677**
Macro Precision ↑	0.447	0.430	0.600	0.560	0.460	**0.689**	0.504
Macro Recall ↑	0.364	0.420	0.308	0.342	0.423	0.315	**0.428**
Macro F_1 ↑	0.382	0.407	0.336	0.371	0.403	0.352	**0.420**
Micro AUC ↑	0.814	0.785	0.835	0.840	0.806	**0.844**	0.840
Macro AUC ↑	0.658	0.626	0.664	0.686	0.642	**0.708**	0.697
One-error ↓	0.251	0.338	0.234	**0.232**	0.278	0.240	0.240
Coverage ↓	6.589	7.834	6.414	6.350	7.067	**6.297**	**6.297**
Ranking Loss ↓	0.181	0.233	0.172	0.169	0.218	**0.163**	**0.163**
AvgPrecision ↑	0.749	0.693	0.758	0.760	0.730	**0.766**	**0.766**
# Wins (Ind. Classi.)	0/18	1/18	4/18	6/18	7/18	-	-
# Wins (All)	0/18	0/18	0/18	1/18	1/18	9/18	10/18

5.4 Discussion

The results presented in this paper show the merit of combining multi-label classifiers. In all three datasets, it is hard to pick a multi-label method that can perform consistently well. For example while CLC and RakEL performs quite well on Enron, they do not deliver similar superiority on Scene and Yeast when compared with ECC, MLKNN and IBLR. Furthermore, since multi-label data suffers from class imbalance problem, it is natural to apply ensemble techniques to overcome over-fitting and improve the accuracy of individual classifiers. Both EML_M and EML_T improve consistently when compared with the individual methods and across the majority of evaluation measures. However, this performance gain is at the expense of inherent computational complexity of ensemble techniques since several multi-label classifiers need to be trained separately. The easiest solution is to use parallel computing techniques to improve the efficiency since all base classifiers can be trained independently.

To the best of our knowledge, this is the first study that aims to combine the output of various multi-label classifiers. In this paper, we have investigated nontrainable ensemble techniques based on the MEAN rule and threshold selection. Since, multi-label classifiers inherently are computationally intensive and data is highly imbalanced, it opens new research challenges how to use other combination techniques efficiently such as trainable combiners (Weighted Average, Fuzzy Integral) or class indifferent combiners (Decision Templates and Dempster-Shafer Combination). The other interesting research issue that needs

to be investigated is how to select the base classifiers in EML since different combinations of base classifiers may perform differently for specific problem domains.

6 Conclusion

In this paper, heterogeneous ensemble of multi-label learners is presented to simultaneously tackle both imbalance and correlation problems. For multi-label classification, this idea is especially appealing, as ensembles methods are well-known for overcoming over-fitting problems and improving the performance of individual classifiers. Two nontrainable ensemble techniques based on the MEAN rule and threshold selection are investigated and then applied to three publicly available multi-label data sets using several evaluation criteria. It has been shown that the presented approach provides a very accurate and efficient solution when compared with the state-of-the-art multi-label methods.

Acknowledgements. This work was supported by the EU VIDI-Video Project.

References

1. Li, T., Ogihara, M.: Toward intelligent music information retrieval. IEEE Trans. on Multimedia 8(3), 564–574 (2006)
2. Godbole, S., Sarawagi, S.: Discriminative methods for multi-labeled classification. In: Dai, H., Srikant, R., Zhang, C. (eds.) PAKDD 2004. LNCS (LNAI), vol. 3056, pp. 22–30. Springer, Heidelberg (2004)
3. Zhang, M.L., Zhou, Z.H.: Multilabel neural networks with applications to functional genomics and text categorization. IEEE Trans. on Knowledge and Data Engineering 18(10), 1338–1351 (2006)
4. Tahir, M.A., Kittler, J., Yan, F., Mikolajczyk, K.: Kernel discriminant analysis using triangular kernel for semantic scene classification. In: Proc. of the 7th International Workshop on CBMI, Crete, Greece. IEEE, Los Alamitos (2009)
5. Dimou, A., Tsoumakas, G., Mezaris, V., Kompatsiaris, I., Vlahavas, I.: An empirical study of multi-label learning methods for video annotation. In: Proc. of the 7th International Workshop on CBMI, Chania, Greece (2009)
6. Chawla, N.V., Sylvester, J.C.: Exploiting diversity in ensembles: Improving the performance on unbalanced datasets. In: Haindl, M., Kittler, J., Roli, F. (eds.) MCS 2007. LNCS, vol. 4472, pp. 397–406. Springer, Heidelberg (2007)
7. Cheng, W., Hullermeier, E.: Combining instance-based learning and logistic regression for multilabel classification. Machine Learning 76(2-3), 211–225 (2009)
8. Read, J., Pfahringer, B., Holmes, G., Frank, E.: Classifier chains for multi-label classification. In: Buntine, W., Grobelnik, M., Mladenić, D., Shawe-Taylor, J. (eds.) ECML PKDD 2009, Part II. LNCS (LNAI), vol. 5782, pp. 254–269. Springer, Heidelberg (2009)
9. Tsoumakas, G., Katakis, I., Vlahavas, I.: Mining Multi-label Data. In: Data Mining and Knowledge Discovery Handbook, 2nd edn. Springer, Heidelberg (2009)
10. Boutell, M.R., Luo, J., Shen, X., Brown, C.M.: Learning multi-label scene classification. Pattern Recognition 37(9), 1757–1771 (2004)

11. Tsoumakas, G., Vlahavas, I.: Random k-labelsets: An ensemble method for multilabel classification. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) ECML 2007. LNCS (LNAI), vol. 4701, pp. 406–417. Springer, Heidelberg (2007)
12. Furnkranz, J., Hullermeier, E., Menca, E.L., Brinker, K.: Multilabel classification via calibrated label ranking. Machine Learning 23(2), 133–153 (2008)
13. Zhang, M.L., Zhou, Z.H.: ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognition 40(7), 2038–2048 (2007)
14. Elisseeff, A., Weston, J.: A kernel method for multi-labelled classification. In: Advances in NIPS, vol. 14 (2002)
15. Read, J., Pfahringer, B., Holmes, G.: Multi-label classification using ensembles of pruned sets. In: Perner, P. (ed.) ICDM 2008. LNCS (LNAI), vol. 5077. Springer, Heidelberg (2008)
16. Kuncheva, L.I.: Combining Pattern Classifiers. Wiley, Chichester (2004)
17. Fan, R.E., Lin, C.J.: A study on threshold selection for multi-label classification. Technical report, National Taiwan University (2007)
18. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco (2005)