Concurrence among Imbalanced Labels and Its Influence on Multilabel Resampling Algorithms

Francisco Charte[1], Antonio Rivera[2],
María José del Jesus[2], and Francisco Herrera[1]

[1] Dep. of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain
[2] Dep. of Computer Science, University of Jaén, Jaén, Spain
{fcharte,herrera}@ugr.es, {arivera,mjjesus}@ujaen.es
http://simidat.ujaen.es, http://sci2s.ugr.es

Abstract. In the context of multilabel classification, the learning from imbalanced data is getting considerable attention recently. Several algorithms to face this problem have been proposed in the late five years, as well as various measures to assess the imbalance level. Some of the proposed methods are based on resampling techniques, a very well-known approach whose utility in traditional classification has been proven.

This paper aims to describe how a specific characteristic of multilabel datasets (MLDs), the level of concurrence among imbalanced labels, could have a great impact in resampling algorithms behavior. Towards this goal, a measure named *SCUMBLE*, designed to evaluate this concurrence level, is proposed and its usefulness is experimentally tested. As a result, a straightforward guideline on the effectiveness of multilabel resampling algorithms depending on MLDs characteristics can be inferred.

Keywords: Multilabel Classification, Imbalanced Learning, Resampling, Measures.

1 Introduction

Multilabel classification (MLC) [1] models are designed to predict the subset of labels associated to each instance in an MLD, instead of only one class as traditional classifiers do. It is a task useful in fields such as automated tag suggestion [2], protein classification [3], and object recognition in images [4], among others. Many different methods have been proposed lately to accomplish this problem.

The number of instances in which each label appears is not homogeneous. In fact, most MLDs show big differences in label frequencies. This peculiarity is known as imbalance [5], and it has been profoundly studied in traditional classification. In the context of MLC, several proposals to deal with imbalanced MLDs [6–12] have been made lately. Despite these efforts, there are still some aspects regarding imbalanced learning in MLC that would need additional analysis.

M. Polycarpou et al. (Eds.): HAIS 2014, LNAI 8480, pp. 110–121, 2014.

Resampling techniques are commonly used in non-MLDs [13], hence they are an obvious choice to face the same problem with MLDs. Notwithstanding, the nature of MLDs can be a challenge for resampling algorithms. In this paper we will show how a specific characteristic of these datasets, the joint presence of labels with different frequencies in the same instance, could prevent the goal of these algorithms. We hypothesized that this symptom, the concurrence among imbalanced labels, would influence the resampling algorithms behavior. A new measure, named *SCUMBLE* (*Score of ConcUrrence among iMBalanced LabEls*) and designed explicitly to assess this causality, will be proposed. Its effectiveness will be experimentally demonstrated.

The *SCUMBLE* measure was conceived aiming to know how difficult would be to work with a certain MLD for resampling algorithms. Its goal is to appraise the concurrence among imbalanced labels, giving as result a score easily interpretable. This score will be in the range [0,1]. A low score would denote an MLD with not much concurrence among imbalanced labels, whereas a high one would evidence the opposite case. Our hypothesis was that the lower the score obtained, the better the resampling algorithms would work. In the future, some recently published ideas, such as the modularity-based label grouping introduced in [14], could be included in our framework as additional means to obtain label concurrence data.

The rest of this paper is structured as follows. Section 2 offers a brief introduction to MLC, as well as a description on how the learning from imbalanced MLDs has been faced. In Section 3 the problem of concurrence among imbalanced level in MLDs will be defined, and how to assess this concurrence using the proposed measured will be explained. Section 4 describes the experimental framework used, as well as the obtained results from experimentation. Finally, Section 5 will offer the conclusions.

2 Preliminaries

In this section a concise introduction to multilabel classification is offered, along with a description on how the learning from imbalanced MLDs has been faced until now.

2.1 Multilabel Classification

Currently, there are many domains [3, 4, 15–18] in which each instance is not associated to an exclusive class, but to a group of them. In this context the classes are named labels, and the set of labels that belongs to a data sample is called labelset. Let D be an MLD, D_i the *i-th* instance, and L the full set on labels on D. The goal of a multilabel classifier is to predict a set $Z_i \subseteq L$ with the labelset for D_i.

Multilabel classification has been traditionally faced through two different approaches [1]. The first one, called data transformation, aims to produce binary or multiclass datasets from an MLD, allowing the use of non-MLC algorithms. The

second, known as algorithm adaptation, has the goal of adapting established algorithms to work natively with MLDs. The two most common transformation methods are Binary Relevance (BR) [19] and Label Powerset (LP) [20]. The former produces several binary datasets from an MLD, one for each label. The latter transforms the MLD in a multiclass dataset, taking each labelset as class identifier. Regarding adapted algorithms, the number of proposals is quite high. There are multilabel KNN classifiers such as ML-kNN [21], multilabel trees based on C4.5 [22], and multilabel SVMs such as [17], as well as a profusion of algorithms based on ensembles of BR and LP classifiers. A recent review on multilabel classification algorithms can be found in [23].

Thus far, most proposed multilabel measures are focused in assessing the number of labels and labelsets. The most common are the total number of labels $|L|$, label cardinality ($Card$), which is the average number of labels per instance, and label density, obtained as $Card/|L|$.

2.2 Learning from Imbalanced Data

Imbalanced learning is a well-known problem in traditional classification [5], having been faced through three main approaches [24]. First, through algorithmic adaptations [25] of existent classifiers, the imbalance is taken into account in the classification process. Second, the preprocessing approach aims to balance class distributions by way of data resampling, creating [26] (oversampling) or removing [27] (undersampling) data samples. Third, cost sensitive classification [28] is a combination of the two previous approaches. The data resampling approach has the advantage of being classifier independent, and its effectiveness has been proven in many scenarios.

In the MLC field, both the algorithmic adaptation and the data resampling approaches have been applied. The former is present in [6–8], while the latter appears in [10–12]. There are also proposals based on the use of ensemble of classifiers, such as [9].

When it comes to assess the imbalance level in MLDs, the measures in Equation 1 and Equation 2 are proposed in [11]. Let D be an MLD, Y the full set of labels in it, y the label being analyzed, and Y_i the labelset of *i-th* instance in D. *IRLbl* is a measure calculated individually for each label. The higher is the *IRLbl* the larger would be the imbalance, allowing to know what labels are in minority or majority. *MeanIR* is the average *IRLbl* for an MLD, useful to estimate the global imbalance level.

$$IRLbl(y) = \frac{\underset{y'=Y_1}{\overset{Y_{|Y|}}{\operatorname{argmax}}}\left(\sum_{i=1}^{|D|} h(y', Y_i)\right)}{\sum_{i=1}^{|D|} h(y, Y_i)}, \quad h(y, Y_i) = \begin{cases} 1 & y \in Y_i \\ 0 & y \notin Y_i \end{cases}. \tag{1}$$

$$MeanIR = \frac{1}{|Y|} \sum_{y=Y_1}^{Y_{|Y|}} (IRLbl(y)). \tag{2}$$

Even though the previously cited proposals for facing imbalanced learning in MLC achieve some good results, their behavior is heavily influenced by MLDs characteristics. In the following we will focus in this topic, specifically in regard to data resampling solutions.

3 MLDs and Resampling Algorithms Behavior

Most traditional resampling methods do their job by removing instances with the most frequent class, or creating new samples from instances associated to the least frequent one. Since each instance can belong to one class only, these actions would effectively balance the classes frequencies. However, this is not necessarily the case when working with MLDs.

3.1 Concurrence among Imbalanced Labels in MLDs

The instances in a MLD are usually associated simultaneously to two or more labels. It is entirely possible that one of those labels is the minority label, while other is the majority one. In the most extreme situation, all the appearances of the minority label could be jointly with the majority one, into the same instances. In practice the scenario would be more complicated, as commonly there are more than one minority/majority label in an MLD. Therefore, the potential existence of instances associated to minority and majority labels at once is very high. This fact is what we called concurrence among imbalanced labels.

A multilabel oversampling algorithm that clones minority instances, such as the proposed in [11], or that generates new samples from existing ones preserving the labelsets, as is the case in [12], could be also increasing the number of instances associated to majority labels. Thus, the imbalance level would be hardly reduced if there is a high level of concurrence among imbalanced labels. In the same way, a multilabel undersampling algorithm designed to remove instances from the majority labels, such as the proposed in [11], could inadvertently cause also a loss of samples associated to the minority ones.

The ineffectiveness of these resampling methods, when they are used with certain MLDs, would be noticed once the preprocessing is applied and the classification results are evaluated. This process will need computing power and time. For that reason, it would be desirable to know in advance the level of concurrence among imbalanced labels that each MLD suffers, saving these valuable resources.

3.2 The SCUMBLE Measure

The concurrence of labels in an MLD can be visually explored in some cases, as shown in Figure 1. Each arc represents a label, being the arc's length proportional to the number of instances in which this label is present. The top diagram corresponds to the genbase dataset. At the position of twelve o'clock appears a label called *P750* which is clearly a minority label. All the samples associated to

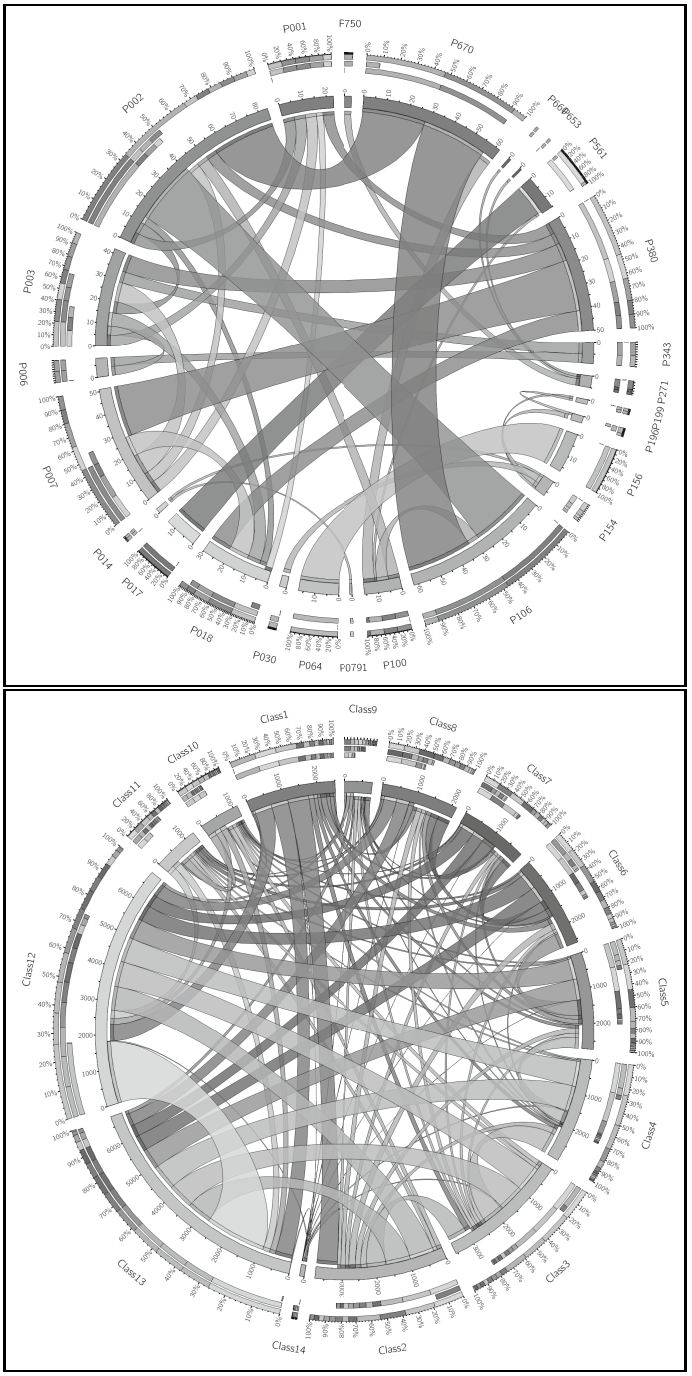

Fig. 1. Label concurrence in genbase (top) and yeast MLDs

this label also contains *P271*, another minority label. The same situation can be seen with label *P154*. By contrast, in the yeast MLD (bottom diagram) is easy to see that the samples associated to minority labels, such as *Class14* and *Class9*, appear always together with one or more majority labels. At first sight, that the concurrence between imbalanced labels is higher in yeast than in genbase could be concluded. However, this visual exploratory technique is not useful with MLDs having more than a few dozens labels.

The *SCUMBLE* measure aims to quantify the imbalance variance among the labels present in each data sample. This measure (Equation 3) is based on the Atkinson index [29] and the *IRLbl* measure (Equation 1) proposed in [11]. The former is an econometric measure directed to assess social inequalities among individuals in a population. The latter is the measure that lets us know the imbalance ratio of each label in an MLD. The Atkinson index is used to know the diversity among people's earnings, while our objective is to assess the extend to which labels with different imbalance levels appear jointly. Our hypothesis is that the higher is the concurrence level the harder would be the work for resampling algorithms, and therefore the worse they would perform.

The Atkinson index is calculated using incomes, we used the imbalance level of each label instead, taking each instance D_i in the MLD D as a population, and the active labels in D_i as the individuals. If the label l is present in the instance i then $IRLbl_{il} = IRLbl(l)$. On the contrary, $IRLbl_{il} = 0$. $\overline{IRLbl_i}$ stands for the average imbalance level of the labels appearing in instance i. The scores for every sample are averaged, obtaining the final *SCUMBLE* value.

$$SCUMBLE(D) = \frac{1}{|D|}\sum_{i=1}^{|D|}[1 - \frac{1}{\overline{IRLbl_i}}(\prod_{l=1}^{|L|} IRLbl_{il})^{(1/|L|)}] \tag{3}$$

Whether our initial hypothesis was correct or wrong, and therefore this measure is able to predict the difficulty that an MLD implies for resampling algorithms or not, is something to be proven experimentally.

4 Experimentation and Analysis

This section starts describing the experimental framework used to assess the usefulness of the *SCUMBLE* measure, and follows giving all the details about the obtained results and their analysis.

4.1 Experimental Framework

To determine the usefulness of the *SCUMBLE* measure the six MLDs shown in Table 1 were used. The rightmost column indicates each dataset's origin. All of them are imbalanced, so theoretically they could benefit from the application of a resampling algorithm. Aside from the *SCUMBLE* measure, the *MaxIR* and *MeanIR* values are also shown. These will be taken as reference point to the

Table 1. Measures about imbalance on datasets before preprocessing

Dataset	SCUMBLE	MaxIR	MeanIR	Ref.
corel5k	0.3932	896.0000	168.7806	[4]
cal500	0.3369	133.1917	21.2736	[15]
enron	0.3023	657.0500	72.7730	[16]
yeast	0.1044	53.6894	7.2180	[17]
medical	0.0465	212.8000	72.1674	[18]
genbase	0.0283	136.8000	32.4130	[3]

posterior analysis. All the measures are average values from training partitions[1] using a 2x5 folds scheme. The datasets appear in Table 1 sorted by *SCUMBLE* value, from higher to lower. According to this measure, corel5k and cal500 would be the most difficult MLDs, since they have a high level of concurrence among labels with different imbalance levels. On the other hand, medical and genbase would be the most benefited from resampling.

Regarding the resampling algorithms, the two proposed in [11] were applied. Both are based on the LP transformation. LP-ROS does oversampling by cloning instances with minority labelsets, whereas LP-RUS performs undersampling removing samples associated to majority labelsets. All the dataset partitions were preprocessed, and the imbalance measures were calculated for each algorithm.

4.2 Results and Analysis

Once the LP-ROS and LP-RUS resampling algorithm were applied, the imbalance levels on the preprocessed MLDs were reevaluated. Table 2 shows the new *MaxIR* and *MeanIR* values for each dataset. Comparing these values with the ones shown in Table 1, it can be verified that a general improvement in the imbalance levels has been achieved. Although there are some exceptions, in most cases both *MaxIR* and *MeanIR* are lower after applying the resampling algorithms.

Table 2. Imbalance levels after applying resampling algorithms

	LP-ROS		LP-RUS	
Dataset	MaxIR	MeanIR	MaxIR	MeanIR
corel5k	969.4000	140.7429	817.1000	155.0324
cal500	179.35838	25.4685	620.0500	68.6716
enron	710.9667	53.2547	133.1917	21.2736
yeast	15.4180	2.6116	83.8000	19.8844
medical	39.9633	10.5558	46.5698	6.3706
genbase	13.7030	4.5004	150.8000	51.1567

[1] The dataset partitions used in this experimentation, as well as color version of all figures, are available to download at `http://simidat.ujaen.es/SCUMBLE`.

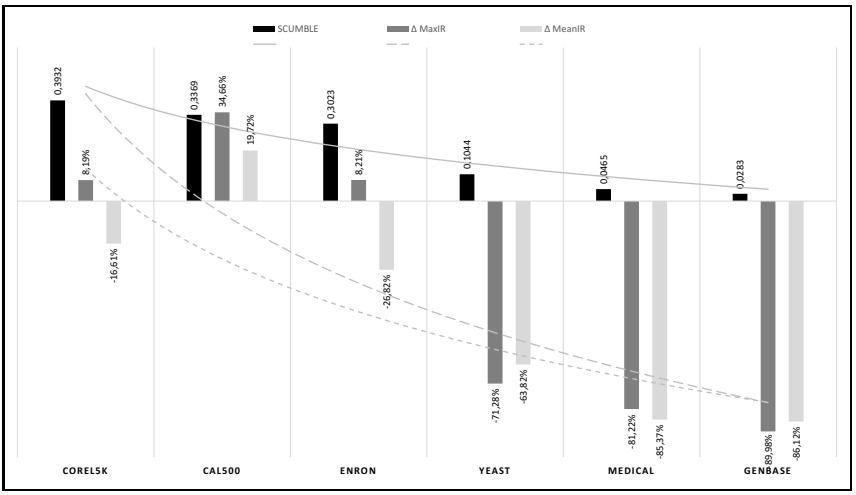

Fig. 2. SCUMBLE vs changes in imbalance level after applying LP-ROS

It would be interesting to know if the imbalance reduction is proportionally coherent with the values obtained from the *SCUMBLE* measure. The graphs in Figure 2 and Figure 3 are aimed to visually illustrate the connection between *SCUMBLE* values and the relative variations in imbalance levels. For each MLD, the *SCUMBLE* value from Table 1 is represented along with the percentage change in *MaxIR* and *MeanIR* after applying the LP-ROS/LP-RUS resampling methods. The tendency for the three values among all the MLDs is depicted by three logarithmic lines. As can be seen, a clear parallelism exists between the continuous line, which corresponds to *SCUMBLE*, and the dashed lines. This affinity is specially remarkable with the LP-RUS algorithm (Figure 3).

Although the previous figures allow to infer that an important correlation between the *SCUMBLE* measure and the success of the resampling algorithms exists, this relationship must be formally analyzed. To this end, a Pearson correlation test was applied over the *SCUMBLE* values and the relative changes in imbalance levels for each resampling algorithm. The resulting correlation coefficients and *p-values* are shown in Table 3. It can be seen that all the coefficients are above 80%, and all the *p-values* are under 0.05. Therefore, a statistical correlation between the *SCUMBLE* measure and the behavior of the tested resampling algorithms can be concluded.

Following this analysis, it seems reasonable to avoid resampling algorithms when the *SCUMBLE* measure for an MLD is well above 0.1, such as is the case with corel5k, cal500 and enron. In this situation the benefits obtained from resampling, if any, are very small. The result can even be a worsening of the imbalance level. In average, the *MeanIR* for the three MLDs with *SCUMBLE* > 0.3 has been reduced only a 6%, while the *MaxIR* is actually increasing in the same percentage. By contrast, the average *MeanIR* reduction for the other three MLDs, with *SCUMBLE* $\lesssim 0.1$, reaches 52% and the *MaxIR* reduction 54%.

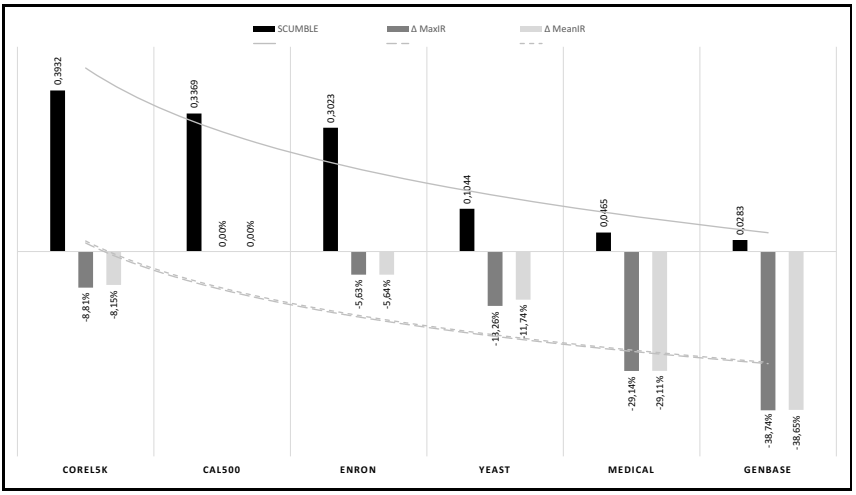

Fig. 3. SCUMBLE vs changes in imbalance level after applying LP-RUS

Table 3. Results from the Pearson correlation tests

	SCUMBLE vs ΔMaxIR		SCUMBLE vs ΔMeanIR	
Algorithm	Cor	p-value	Cor	p-value
LP-ROS	0.8120	0.0497	0.9189	0.0096
LP-RUS	0.8607	0.0278	0.8517	0.0314

Aiming to know how these changes in the imbalance levels would influence classification results, and if a correlation with *SCUMBLE* values exists, the HOMER [30] algorithm was used, following a 2x5 folds cross-validation scheme. It must be highlighted that the interest here is not in the raw performance values, but in how they change after a resampling algorithm has been applied and how this change correlates with *SCUMBLE* values. Therefore, the HOMER algorithm is used only as a tool to obtain classification results before and after applying the resampling. Any other MLC algorithm could be used for this task. Additionally, the proposed *SCUMBLE* measure is not used in the experimentation to influence the behavior of LP-ROS, LP-RUS or HOMER by any means. The goal is to theoretically explore the correlation between changes in classification results and *SCUMBLE* values.

Table 4 shows these results assessed with the F-measure, the harmonic mean of precision and recall measures. It can be seen that with the three MLDs which show high *SCUMBLE* values, the preprocessing has produced a remarkable deterioration in classification results. Among the other three MLDs the resampling has improved them in some cases, while producing a slight worsening (less than 1%) in others. Therefore, even though the MLC algorithm behavior would be also affected by other dataset characteristics, that the *SCUMBLE* measure would offer valuable information to determine the convenience of applying a resampling method can be concluded.

Table 4. F-Measure values obtained by HOMER MLC algorithm

Dataset	Base	LP-RUS	LP-ROS	ΔRUS	ΔROS
corel5k	0.3857	0.2828	0.2920	-26.6788	-24.2935
cal500	0.3944	0.3127	0.3134	-20.7150	-20.5375
enron	0.5992	0.5761	0.5874	-3.8551	-1.9693
yeast	0.6071	0.6950	0.6966	14.4787	14.7422
medical	0.9238	0.9158	0.9162	-0.8660	-0.8227
genbase	0.9896	0.9818	0.9912	-0.7882	0.1617

5 Conclusions

Multilabel classification has many applications nowadays, but usually MLDs are imbalanced. This is a fact that challenges most MLC algorithms, and several approaches to face it have been proposed lately. Some of them rely on resampling techniques, through adaptations of algorithms that have proven their usefulness in traditional classification. However, the specific nature of MLDs has to be taken into account, since some of their characteristics could influence these algorithms behavior.

In this paper the concurrence among imbalanced labels has been explained and *SCUMBLE*, a new measure designed to assess this characteristic, has been proposed. The suitability of this measure has been experimentally demonstrated against six MLDs and two resampling algorithms. The conducted correlation analysis, summarized in Table 3, has shown that the *SCUMBLE* measure can be used to know in advance if resampling would be positive for a certain MLD or not. This assumption has been corroborated by classification results, shown in Table 4, which experiment a remarkable worsening when used with MLDs with the highest *SCUMBLE* values.

Given this reality, a further and deeper analysis should be directed, involving additional MLDs and other resampling algorithms. Notwithstanding it could be concluded that basic resampling algorithms, which clone the labelsets in new instances or remove samples, are not a general solution in the multilabel field. More sophisticated approaches, which take into account the concurrence among imbalanced labels, would be needed.

Acknowledgments. F. Charte is supported by the Spanish Ministry of Education under the FPU National Program (Ref. AP2010-0068). This work was partially supported by the Spanish Ministry of Science and Technology under projects TIN2011-28488 and TIN2012-33856, and the Andalusian regional projects P10-TIC-06858 and P11-TIC-9704.

References

1. Tsoumakas, G., Katakis, I., Vlahavas, I.: Mining Multi-label Data. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, ch. 34, pp. 667–685. Springer US, Boston (2010)

2. Katakis, I., Tsoumakas, G., Vlahavas, I.: Multilabel Text Classification for Automated Tag Suggestion. In: Proc. ECML PKDD 2008 Discovery Challenge, Antwerp, Belgium, pp. 75–83 (2008)
3. Diplaris, S., Tsoumakas, G., Mitkas, P.A., Vlahavas, I.: Protein Classification with Multiple Algorithms. In: Bozanis, P., Houstis, E.N. (eds.) PCI 2005. LNCS, vol. 3746, pp. 448–456. Springer, Heidelberg (2005)
4. Duygulu, P., Barnard, K., de Freitas, J.F.G., Forsyth, D.: Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part IV. LNCS, vol. 2353, pp. 97–112. Springer, Heidelberg (2002)
5. Chawla, N.V., Japkowicz, N., Kotcz, A.: Editorial: special issue on learning from imbalanced data sets. SIGKDD Explor. Newsl. 6(1), 1–6 (2004)
6. He, J., Gu, H., Liu, W.: Imbalanced multi-modal multi-label learning for subcellular localization prediction of human proteins with both single and multiple sites. PloS One 7(6), 7155 (2012)
7. Li, C., Shi, G.: Improvement of learning algorithm for the multi-instance multi-label rbf neural networks trained with imbalanced samples. J. Inf. Sci. Eng. 29(4), 765–776 (2013)
8. Tepvorachai, G., Papachristou, C.: Multi-label imbalanced data enrichment process in neural net classifier training. In: IEEE Int. Joint Conf. on Neural Networks, IJCNN, 2008, pp. 1301–1307 (2008)
9. Tahir, M.A., Kittler, J., Bouridane, A.: Multilabel classification using heterogeneous ensemble of multi-label classifiers. Pattern Recognit. Lett. 33(5), 513–523 (2012)
10. Tahir, M.A., Kittler, J., Yan, F.: Inverse random under sampling for class imbalance problem and its application to multi-label classification. Pattern Recognit. 45(10), 3738–3750 (2012)
11. Charte, F., Rivera, A., del Jesus, M.J., Herrera, F.: A first approach to deal with imbalance in multi-label datasets. In: Pan, J.-S., Polycarpou, M.M., Woźniak, M., de Carvalho, A.C.P.L.F., Quintián, H., Corchado, E. (eds.) HAIS 2013. LNCS, vol. 8073, pp. 150–160. Springer, Heidelberg (2013)
12. Giraldo-Forero, A.F., Jaramillo-Garzón, J.A., Ruiz-Muñoz, J.F., Castellanos-Domínguez, C.G.: Managing imbalanced data sets in multi-label problems: A case study with the smote algorithm. In: Ruiz-Shulcloper, J., Sanniti di Baja, G. (eds.) CIARP 2013, Part I. LNCS, vol. 8258, pp. 334–342. Springer, Heidelberg (2013)
13. García, V., Sánchez, J., Mollineda, R.: On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. Knowl. Based Systems 25(1), 13–21 (2012)
14. Szymański, P., Kajdanowicz, T.: MLG: Enchancing multi-label classification with modularity-based label grouping. In: Pan, J.-S., Polycarpou, M.M., Woźniak, M., de Carvalho, A.C.P.L.F., Quintián, H., Corchado, E. (eds.) HAIS 2013. LNCS, vol. 8073, pp. 431–440. Springer, Heidelberg (2013)
15. Turnbull, D., Barrington, L., Torres, D., Lanckriet, G.: Semantic Annotation and Retrieval of Music and Sound Effects. IEEE Audio, Speech, Language Process. 16(2), 467–476 (2008)
16. Klimt, B., Yang, Y.: The Enron Corpus: A New Dataset for Email Classification Research. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) ECML 2004. LNCS (LNAI), vol. 3201, pp. 217–226. Springer, Heidelberg (2004)
17. Elisseeff, A., Weston, J.: A Kernel Method for Multi-Labelled Classification. In: Advances in Neural Information Processing Systems 14, vol. 14, pp. 681–687. MIT Press (2001)

18. Crammer, K., Dredze, M., Ganchev, K., Talukdar, P.P., Carroll, S.: Automatic Code Assignment to Medical Text. In: Proc. Workshop on Biological, Translational, and Clinical Language Processing, BioNLP 2007, Prague, Czech Republic, pp. 129–136 (2007)
19. 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)
20. Boutell, M., Luo, J., Shen, X., Brown, C.: Learning multi-label scene classification. Pattern Recognit. 37(9), 1757–1771 (2004)
21. Zhang, M., Zhou, Z.: ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognit. 40(7), 2038–2048 (2007)
22. Clare, A.J., King, R.D.: Knowledge discovery in multi-label phenotype data. In: Siebes, A., De Raedt, L. (eds.) PKDD 2001. LNCS (LNAI), vol. 2168, pp. 42–53. Springer, Heidelberg (2001)
23. Zhang, M., Zhou, Z.: A Review on Multi-Label Learning Algorithms. IEEE Trans. Knowl. Data Eng., doi:10.1109/TKDE.2013.39
24. López, V., Fernández, A., García, S., Palade, V., Herrera, F.: An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Inf. Sciences 250, 113–141 (2013)
25. Fernández, A., López, V., Galar, M., del Jesus, M.J., Herrera, F.: Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches. Knowl. Based Systems 42, 97–110 (2013)
26. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: Synthetic minority over-sampling technique. J. Artificial Intelligence Res. 16, 321–357 (2002)
27. Kotsiantis, S.B., Pintelas, P.E.: Mixture of expert agents for handling imbalanced data sets. Annals of Mathematics, Computing & Teleinformatics 1, 46–55 (2003)
28. Provost, F., Fawcett, T.: Robust classification for imprecise environments. Mach. Learn. 42, 203–231 (2001)
29. Atkinson, A.B.: On the measurement of inequality. Journal of Economic Theory 2(3), 244–263 (1970)
30. Tsoumakas, G., Katakis, I., Vlahavas, I.: Effective and Efficient Multilabel Classification in Domains with Large Number of Labels. In: Proc. ECML/PKDD Workshop on Mining Multidimensional Data, MMD 2008, Antwerp, Belgium, pp. 30–44 (2008)