Resampling Multilabel Datasets by Decoupling Highly Imbalanced Labels

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

[1] Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain
{fcharte,herrera}@ugr.es
http://sci2s.ugr.es

[2] Department of Computer Science, University of Jaén, Jaén, Spain
{arivera,mjjesus}@ujaen.es
http://simidat.ujaen.es

Abstract. Multilabel classification is a task that has been broadly studied in late years. However, how to face learning from imbalanced multilabel datasets (MLDs) has only been addressed latterly. In this regard, a few proposals can be found in the literature, most of them based on resampling techniques adapted from the traditional classification field. The success of these methods varies extraordinarily depending on the traits of the chosen MLDs.

One of the characteristics which significantly influences the behavior of multilabel resampling algorithms is the joint appearance of minority and majority labels in the same instances. It was demonstrated that MLDs with a high level of concurrence among imbalanced labels could hardly benefit from resampling methods. This paper proposes an original resampling algorithm, called REMEDIAL, which is not based on removing majority instances nor creating minority ones, but on a procedure to decouple highly imbalanced labels. As will be experimentally demonstrated, this is an interesting approach for certain MLDs.

Keywords: Multilabel classification · Imbalanced learning · Resampling · Label concurrence

1 Introduction

While in traditional classification the models get a set of input attributes aiming to predict only one output, whether it is binary (binary classifiers) or not (multiclass classifiers), in multilabel classification (MLC) [1] the algorithms have to figure out several outputs from the same set of inputs. There are many real-world applications for MLC, including automatic email classification [2], semantic annotation of music [3], and object recognition in images [4].

The imbalance problem [5], which has been profoundly studied in non-MLC, is also present in MLC. Actually, almost all MLDs suffer from imbalance. Some

E. Onieva et al. (Eds.): HAIS 2015, LNAI 9121, pp. 489–501, 2015.
DOI: 10.1007/978-3-319-19644-2_41

labels are scarcely represented (minority labels), while others are very frequent (majority labels). The use of resampling techniques is a common approach in non-MLC [6]. Therefore, when it came to face this problem in MLC, a clear path was adapting existent resampling methods to work with MLDs.

Since 2012 several ways to deal with imbalance in MLC have been proposed, including oversampling [7–9] algorithms, undersampling [10,11] algorithms, and ensemble based solutions [12]. As was stated in [13], the success of some resampling techniques is highly influenced by the concurrence of minority and majority labels in the same instances. The level of concurrence in an MLD can be computed with a measure called *SCUMBLE* (*Score of ConcUrrence among iMBalanced LabEls*), also proposed in [13]. The higher the *SCUMBLE* the harder it will be for a resampling algorithm to balance the labels distribution.

MLC raises new challenges, since MLDs exhibit traits unseen in traditional datasets. Therefore, new solutions have to be considered, specific to these traits. This is the goal of REMEDIAL (*REsampling MultilabEl datasets by Decoupling highly ImbAlanced Labels*). REMEDIAL evaluates the concurrence among imbalance labels of each sample in an MLD, by means of the aforementioned *SCUMBLE* measure, and splits those with a high level, decoupling minority and majority labels. As will be experimentally demonstrated, this is by itself an interesting approach for certain MLDs. However, the goal of REMEDIAL is not so much to compete with oversampling or undersampling techniques, but to facilitate the work of those methods.

The remainder of this paper is divided into four sections. In Sect. 2 a brief introduction to MLC is provided, along with a description on how learning from imbalanced MLDs has been tackled until now. Section 3 defines the problem of concurrence among imbalanced labels in MLDs, introducing the assessment of this concurrence with the *SCUMBLE* measure and how REMEDIAL addresses it. In Sect. 4 the experimental framework is described, and the results obtained are discussed. Finally, in Sect. 5 some conclusions are yield.

2 Background

This section provides a concise presentation of MLC, describing as well how learning from imbalanced MLDs has been recently tackled.

2.1 Introducing MLC

Many real world classification problems have an intrinsic multilabel nature. Some of them have been mentioned above [2–4]. Additionally, tasks as protein classification [14], gene functional classification [15], and automatic code assignment to medical texts [16] have to predict not a class label for each instance, but a group of them. D being an MLD, L the full set of labels in D, D_i the *i-th* instance, and $Y_i \subseteq L$ the subset of labels relevant to D_i, a multilabel classifier aims to predict a subset $Z_i \subseteq L$ which is as closer to Y_i as possible.

There have been plenty of MLC algorithm proposals during the last decade. Most of them follow one of two main approaches:

- **Data Transformation:** Addresses the task by converting the MLD into one or more traditional datasets. Although there are several transformation algorithms documented in the literature [1], the best known ones are called BR (*Binary Relevance*) [17] and LP (*Label Powerset*) [18]. BR generates one binary dataset for each label in the MLD, while LP produces a multiclass dataset using each label combination as class. Both have been used as foundation for designing different ensemble based algorithms, such as CC [19] and HOMER [20].
- **Method Adaptation:** Attempts to adapt existent non-MLC algorithms to work with MLDs natively, without transforming them. There are proposals of MLC classifiers based on kNN [21], SVMs [15], trees [22], and ANNs [23], among others. A description of many of them can be found in [24], a recently published review.

In addition to new algorithms, MLC also needed new measures to assess both MLDs characteristics and classification results. Many of them are described in [1,24]. The best known characterization measures are $Card$ (*label cardinality*), calculated as the average number of labels in the instances of an MLD, and a dimensionless measure known as *label density*, computed as $Card/|L|$.

2.2 MLC and Imbalanced Datasets

The imbalance problem is well known in the context of non-MLC. Essentially, during training most classifiers tend to be biased to the most frequent class, since they are designed to maximize a global performance measure, such as precision or accuracy. This problem has been tackled mainly by means of resampling techniques [25,26] and algorithmic adaptations [27], as well as through a mix of both methods called *cost-sensitive classification* [28].

As stated in [9], where specific measures to assess the imbalance levels in MLDs were proposed, imbalance is usually present in most MLDs, and the imbalance levels tend to be higher than those in traditional datasets. Learning from imbalanced MLDs has been faced through algorithmic adaptations in [29–31], and there is also some ensemble-based studies [12], as well as several proposals which have chosen the resampling approach [8–10].

Resampling is a classifier independent approach, therefore it can be applied in a broader spectrum of cases than adapted classifiers. Moreover, undersampling and oversampling algorithms have proven to be effective in many scenarios [6]. However, the specificities of MLDs can be a serious obstacle for these algorithms. The most noteworthy are the huge differences in imbalance levels between labels, and the join appearance of minority and majority labels in the same instances.

The imbalance ratio for each label in an MLD, as well as the mean imbalance ratio, can be determined as proposed in [7]. The first measure suggested is *IRLbl*, shown in Eq. 1. Its goal is to assess the imbalance ratio for an individual label. The rarer a label is the higher its *IRLbl* will be. The value 1 will correspond to the most frequent label, whereas the least frequent one will have the largest *IRLbl*. The second measure, called *MeanIR* (see Eq. 2), provides a global estimate

of how imbalanced an MLD is. In both equations D stands for any MLD, L the full set of labels in D, l the label being assessed, and Y_i the labelset of the *i-th* sample in D.

$$IRLbl(y) = \frac{\underset{l'=L_1}{\overset{L_{|L|}}{\operatorname{argmax}}}(\sum_{i=1}^{|D|} h(l', Y_i))}{\sum_{i=1}^{|D|} h(l, Y_i)}, \quad h(l, Y_i) = \begin{cases} 1 & l \in Y_i \\ 0 & l \notin Y_i \end{cases}. \tag{1}$$

$$MeanIR = \frac{1}{|L|} \sum_{l=L_1}^{L_{|L|}} (IRLbl(l)). \tag{2}$$

In order to know the extent to which minority and majority labels jointly appear in the same instances in an MLD, a measure called *SCUMBLE* was presented in [13]. As can be seen in Eq. 3, *SCUMBLE* relies on the *IRLbl* measure previously mentioned. First the concurrence level for each instance ($SCUMBLE_{ins}$) is obtained, then the mean *SCUMBLE* for the whole MLD is computed in Eq. 4. The value of *SCUMBLE* is normalized in the range $[0, 1]$, denoting a higher value a larger concurrence of imbalanced labels.

$$SCUMBLE_{ins}(i) = 1 - \frac{1}{\overline{IRLbl_i}} (\prod_{l=1}^{|L|} IRLbl_{il})^{(1/|L|)} \tag{3}$$

$$SCUMBLE(D) = \frac{1}{|D|} \sum_{i=1}^{|D|} SCUMBLE_{ins}(i) \tag{4}$$

3 Multilabel Resampling with REMEDIAL

In this section the context in which the proposed algorithm, REMEDIAL, has been developed is depicted. First, how multilabel resampling has been confronted until now is reviewed. Then, the REMEDIAL approach, as a specific method for MLDs with concurrence of highly imbalanced labels, is described.

3.1 Related Work

In general, resampling methods aimed to work with non-MLDs can be divided into two categories, oversampling algorithms and undersampling algorithms. The former technique produces new samples with the minority class, while the latter removes instances linked to the majority class. The way in which the samples to be removed or reproduced are chosen can also be grouped into two categories, random methods and heuristic methods. Since this kind of datasets use only one class per instance, the previous techniques effectively balance the distribution of classes. However, this is not always true when dealing with MLDs. Moreover, most MLDs have more than one minority and one majority label.

The preceding approaches have been migrated to the multilabel scenario at some extent, giving as result proposals such as the following:

- **Random Undersampling:** Two multilabel random undersampling algorithms are presented in [9], one of them based on the LP transformation (LP-RUS) and another one on the *IRLbl* measure (ML-RUS). The latter determines what labels are in minority, by means of their *IRLbl*, and avoids removing samples in which they appear.
- **Random Oversampling:** The same paper [9] also proposes two random oversampling algorithms, called LP-ROS and ML-ROS. The former is based on the LP transformation, while the latter relies on the *IRLbl* measure. Both take into account several minority labels, and generate new instances cloning the original labelsets.
- **Heuristic Undersampling:** In [10] a method to undersample MLDs following the ENN (*Edited Nearest Network*) rule was presented. The instances are not randomly chosen, as in LP-RUS or ML-RUS, but carefully selected after analyzing their *IRLbl* and the differences with their neighborhood.
- **Heuristic Oversampling:** The procedure proposed in [8] is based on the original SMOTE algorithm. First, instances of an MLD are chosen using different criteria, then the selected samples are given as input to SMOTE, producing new samples with the same labelsets.

A major disadvantage of these algorithms is that they always work over full labelsets, cloning the set of labels in existent samples or completely removing them. Although this approach can benefit some MLDs, in other cases the result can be counterproductive depending on the MLD traits.

3.2 The Label Concurrence Problem

Since each instance in an MLD has two or more labels, it is not rare that some of them are very common ones while others are minority labels. This fact can be depicted using an interaction plot as the shown in Fig. 1[1]. The top half of this plot corresponds to two MLDs with a high level of concurrence between imbalanced labels, denoted by a *SCUMBLE* above of 0.1. As can be seen, the minority labels (on the right side) are entirely linked with some majority labels.

In some MLDs the concurrence between majority and minority labels is low, as shown in the bottom half of Fig. 1. In these cases the level of *SCUMBLE* is below 0.1, and as can be seen there are many arcs between minority labels, denoting interactions between them but not with the majority ones.

The aforementioned multilabel resampling algorithms will not have an easy work while dealing with MLDs which have a high *SCUMBLE*. Undersampling

[1] Visualizing all label interactions in an MLD is, in some cases, almost impossible due to the large number of labels. For that reason, only the most frequent labels and the most rare ones for each MLD are represented in these plots. High resolution version of these plots can be found at http://simidat.ujaen.es/remedial and they can be generated using the `mldr` R package [32].

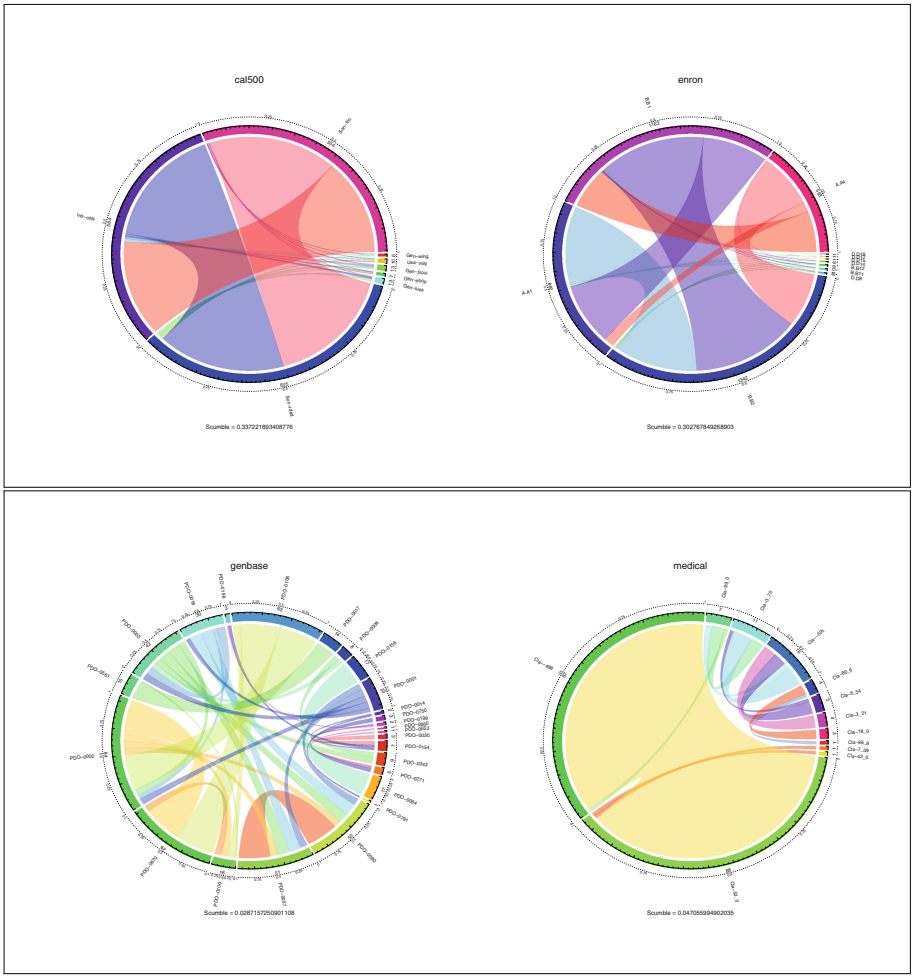

Fig. 1. Concurrence among minority and majority labels in four MLDs.

algorithms can produce a loss of essential information, as the samples selected for removal because majority labels appear in them can also contain minority labels. In the same way, oversampling algorithms limited to cloning the labelsets, such as the proposals in [8,9], can be also increasing the presence of majority labels. These facts were empirically demonstrated in [13].

3.3 Algorithm Description

As its name suggests, REMEDIAL (*REsampling MultilabEl datasets by Decoupling highly ImbAlanced Labels*) is a method specifically designed for MLDs which suffer from concurrence between imbalanced labels. In this context, *highly*

imbalanced labels has to be understood as labels with large differences in their *IRLbls*. This is a fact assessed with the *SCUMBLE* measure, thus REMEDIAL is directed to MLDs with a high *SCUMBLE* level.

When the few samples in which a minority label is present also contain one or more majority labels, whose frequency in the MLD is much higher, the power of the input features to predict the labels might be biased to the majority ones. Our hypothesis is that, in a certain way, majority labels are masking the minority ones when they appear together, a problem that could be solved to some extent by decoupling the labels in these instances.

REMEDIAL is a resampling algorithm. It could be seen as an oversampling method, since it produces new instances in some cases. At the same time it also modifies existent samples. In short, REMEDIAL is an editing plus oversampling algorithm, and it is an approach which has synergies with traditional resampling techniques. The method pseudo-code is shown in Algorithm 1.

Algorithm 1. REMEDIAL algorithm.

1: **function** REMEDIAL(MLD D, Labels L)
2: $IRLbl_l \leftarrow$ calculateIRLbl(l in L) ▷ Calculate imbalance levels
3: $IRMean \leftarrow \overline{IRLbl}$
4: $SCUMBLEIns_i \leftarrow$ calculateSCUMBLE(D_i in D) ▷ Calculate SCUMBLE
5: $SCUMBLE \leftarrow \overline{SCUMBLEIns}$
6: **for each** *instance* i **in** D **do**
7: **if** $SCUMBLEIns_i > SCUMBLE$ **then**
8: $D'_i \leftarrow D_i$ ▷ Clone the affected instance
9: $D_i[labels_{IRLbl<=IRMean}] \leftarrow 0$ ▷ Maintain minority labels
10: $D'_i[labels_{IRLbl>IRMean}] \leftarrow 0$ ▷ Maintain majority labels
11: $D \leftarrow D + D'_i$
12: **end if**
13: **end for**
14: **end function**

The *IRLbl*, *IRMean* and *SCUMBLE* measures are computed in lines 2–5. $SCUMBLE_{Ins_i}$ is the concurrence level of the D_i instance. The mean *SCUMBLE* for the MLD is obtained by averaging the individual *SCUMBLE* for each sample.

Taking the mean *SCUMBLE* as reference, only the samples with a *SCUMBLEIns* > *SCUMBLE* are processed. Those instances, which contain minority and majority labels, are decoupled into two instances, one containing only the majority labels and another one with the minority labels. In line 8 D_i, a sample affected by problem at glance, is cloned in D'_i. The formula in line 9 edits the original D_i instance by removing the majority labels from it. Majority labels are considered as those whose *IRLbl* is equal or below to *IRMean*. Line 10 does the opposite, removing from the cloned D'_i the minority labels. D_i belongs to the D MLD, but D'_i has to be added to it (line 11).

4 Experimental Analysis

This section describes the experimental framework used to test the proposed algorithm, then presents the obtained results, and finally analyzes them.

4.1 Framework

To check the influence of REMEDIAL in classification results six MLDs have been chosen (see Table 1). These are the MLDs used in [13] with *SCUMBLE* values above 0.1, which were the more problematic to process with classic resampling methods, and two more with low *SCUMBLE* levels. These MLDs are given as input, before and after preprocessing them with REMEDIAL, to three different MLC algorithms: BR [18], HOMER [20] and IBLR [33]. These are representatives of three main approaches to MLC classification, ensembles of binary classifiers, ensembles of label powerset classifiers, and instance based classifiers.

Table 1. Datasets used in experimentation.

Category	Dataset	SCUMBLE	max(MeanIR)	MeanIR	Ref
High SCUMBLE > 0.1	cal500	0.3369	133.1917	21.2736	[3]
	corel5k	0.3932	896.0000	168.7806	[4]
	enron	0.3023	657.0500	72.7730	[2]
	yeast	0.1044	53.6894	7.2180	[15]
Low SCUMBLE < 0.1	genbase	0.0283	136.8000	32.4130	[14]
	medical	0.0465	212.8000	72.1674	[16]

A 5×2 fold cross validation has been used. Classification results are evaluated using three usual multilabel measures: HammingLoss (HL), Macro-FMeasure (MacroFM) and Micro-FMeasure (MicroFM). HL (see Eq. 5) is a global sample based measure. It assesses differences between Z_i, the predicted labelset, and Y_i, the real one, without distinction among labels. The lower the HL the better the predictions are. MacroFM and MicroFM are label based measures. As can be seen in Eqs. 6 and 7, in MacroFM the *F-measure* is evaluated independently for each label and then is averaged, while in MicroFM the counters for all labels are aggregated and then used for calculating the *F-measure*. The former approach is more sensitive to performance classifying minority labels.

$$HammingLoss = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i \Delta Z_i|}{|L|}. \tag{5}$$

$$MacroFM = \frac{1}{|L|}\sum_{i=1}^{|L|} F\text{-}measure(TP_i, FP_i, TN_i, FN_i) \tag{6}$$

$$MicroFM = F\text{-}measure(\sum_{i=1}^{|L|} TP_i, \sum_{i=1}^{|L|} FP_i, \sum_{i=1}^{|L|} TN_i, \sum_{i=1}^{|L|} FN_i) \quad (7)$$

The results obtained from each classifier over the datasets, before and after preprocessing, are the shown in Table 2. Best results are highlighted in bold.

Table 2. Results before and after applying REMEDIAL

		BR		HOMER		IBLR	
	Dataset	Before	After	Before	After	Before	After
HL	cal500	0.1630	**0.1496**	0.1888	**0.1794**	0.2340	**0.2125**
	corel5k	0.0098	**0.0094**	0.0132	**0.0118**	0.0242	**0.0148**
	enron	**0.0522**	0.0524	0.0583	**0.0560**	**0.0572**	0.0573
	yeast	0.2505	**0.2240**	0.2632	**0.2433**	**0.1942**	0.2139
	genbase	**0.0012**	0.0084	**0.0016**	0.0062	**0.0022**	0.0092
	medical	**0.0107**	0.0131	**0.0108**	0.0125	0.0198	0.0198
MacroFM	cal500	**0.2934**	0.2516	0.3316	**0.3358**	**0.2772**	0.2597
	corel5k	0.1774	**0.1826**	0.1916	**0.1924**	0.1059	**0.1432**
	enron	0.4029	**0.4190**	0.3790	**0.3793**	**0.3458**	0.3114
	yeast	0.4341	**0.5204**	0.4334	**0.4626**	**0.4944**	0.4156
	genbase	0.9890	**0.9924**	**0.9780**	0.9697	**0.9655**	0.8450
	medical	**0.8166**	0.8013	**0.7942**	0.7780	**0.6404**	0.6216
MicroFM	cal500	**0.3488**	0.2506	0.3978	**0.4008**	**0.3184**	0.2934
	corel5k	**0.1096**	0.0782	**0.1744**	0.1627	**0.0542**	0.0530
	enron	**0.5334**	0.4745	**0.5265**	0.5036	**0.4561**	0.3541
	yeast	0.5787	**0.5898**	0.5763	**0.5974**	**0.6502**	0.5546
	genbase	**0.9867**	0.9012	**0.9820**	0.9284	**0.9768**	0.8902
	medical	**0.8006**	0.7350	**0.7994**	0.7582	**0.6324**	0.5830

4.2 Analysis

Beginning with the two MLDs which have low *SCUMBLE* values, the results produced by REMEDIAL are not good almost in any case. Although some differences are quite small, in general the decoupling of labels has worsened classification performance. As a consequence a clear guideline follows from these results, REMEDIAL should not be used with MLDs with low *SCUMBLE* levels, since it is an algorithm specifically designed to face the opposite casuistic. The analysis of results from the other four MLDs can be divided into two parts, depending on where the focus is.

Looking at the results by evaluation measure, it is clear that REMEDIAL is benefiting minority labels, with better MacroFM values, and has a good overall

behavior, denoted by the HL values after resampling. There are mixed results when MicroFM is used, as for some MLDs the results are improved while for others there is a worsening.

Going through the results by classifier, that REMEDIAL works better with BR and HOMER than with IBLR can be observed. Binary relevance based algorithms train a classifier for each label, taking as positive the instances containing it and as negative the remainder samples. When a majority label is being processed, all the instances in which it appears jointly with a minority label are processed as positive, disregarding the fact that they contain other labels. The decoupling of these labels tends to balance the bias of each classifier. LP based algorithms, such as HOMER, surely are favored by REMEDIAL, since the decoupling produces simpler labelsets. Moreover, the number of distinct labelsets is reduced after the resampling. The influence of REMEDIAL on instance based classifiers, such as IBLR, is easy to devise. The attributes of the decoupled samples do not change, so they will occupy exactly the same position with respect to the instance which is taken as reference for searching nearest neighbors. Therefore, the classifier will get two samples at the same distance but with disjoint labelsets, something that can be confusing depending on how the algorithm predicts the labelset of the reference sample.

Overall, REMEDIAL would be a recommended resampling for MLDs with high *SCUMBLE* levels and when BR or LP based classifiers are going to be used. In these cases the prediction of minority labels would be improved, and the global performance of the classifiers would be better. These are the benefits brought by itself, but REMEDIAL could be used as a first step aimed to ease the work of traditional resampling techniques.

5 Conclusions

In this paper REMEDIAL, a new resampling algorithm aimed to boost multilabel imbalanced learning, has been presented. This algorithm is specifically devised for MLDs with a high concurrence between minority and majority labels, a trait that can be assessed with a measure called *SCUMBLE*. REMEDIAL looks for instances with a high *SCUMBLE* level and decouples minority and majority labels, producing new instances.

The conducted experimentation has proven that REMEDIAL is able to improve classification results when applied to MLDs with a high *SCUMBLE*, although the chosen classifier also influences the obtained outputs.

Those results could be improved joining REMEDIAL with some of the existent resampling methods. Once the labels have been decoupled, traditional oversampling and undersampling algorithms would find less obstacles to do their work. Thus, this a potential path for future research into the imbalanced treatment for MLDs.

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-7765.

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, Boston (2010). doi:10.1007/978-0-387-09823-4_34
2. 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). doi:10.1007/978-3-540-30115-8_22
3. Turnbull, D., Barrington, L., Torres, D., Lanckriet, G.: Semantic annotation and retrieval of music and sound effects. IEEE Audio Speech Lang. Process. **16**(2), 467–476 (2008). doi:10.1109/TASL.2007.913750
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). doi:10.1007/3-540-47979-1_7
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). doi:10.1145/1007730.1007733
6. García, V., Sánchez, J., Mollineda, R.: On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. Knowl. Based Syst. **25**(1), 13–21 (2012). http://dx.doi.org/10.1016/j.knosys.2011.06.013
7. 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). doi:10.1007/978-3-642-40846-5_16
8. 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). doi:10.1007/978-3-642-41822-8_42
9. Charte, F., Rivera, A.J., del Jesus, M.J., Herrera, F.: Addressing imbalance in multilabel classification: Measures and random resampling algorithms, Neurocomputing to be published
10. Charte, F., Rivera, A.J., del Jesus, M.J., Herrera, F.: MLeNN: a first approach to heuristic multilabel undersampling. In: Corchado, E., Lozano, J.A., Quintián, H., Yin, H. (eds.) IDEAL 2014. LNCS, vol. 8669, pp. 1–9. Springer, Heidelberg (2014). doi:10.1007/978-3-319-10840-7_1
11. Tahir, M.A., Kittler, J., Yan, F.: Inverse random under sampling for class imbalance problem and its application to multi-label classification. Pattern Recogn. **45**(10), 3738–3750 (2012). doi:10.1016/j.patcog.2012.03.014
12. Tahir, M.A., Kittler, J., Bouridane, A.: Multilabel classification using heterogeneous ensemble of multi-label classifiers. Pattern Recogn. Lett. **33**(5), 513–523 (2012). doi:10.1016/j.patrec.2011.10.019

13. Charte, F., Rivera, A., del Jesus, M.J., Herrera, F.: Concurrence among imbalanced labels and its influence on multilabel resampling algorithms. In: Polycarpou, M., de Carvalho, A.C.P.L.F., Pan, J.-S., Woźniak, M., Quintian, H., Corchado, E. (eds.) HAIS 2014. LNCS, vol. 8480, pp. 110–121. Springer, Heidelberg (2014)
14. Diplaris, S., Tsoumakas, G., Mitkas, P.A., Vlahavas, I.P.: Protein classification with multiple algorithms. In: Bozanis, P., Houstis, E.N. (eds.) PCI 2005. LNCS, vol. 3746, pp. 448–456. Springer, Heidelberg (2005). doi:10.1007/11573036_42
15. Elisseeff, A., Weston, J.: A kernel method for multi-labelled classification. In: Dietterich, G., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems 14, vol. 14, pp. 681–687. MIT Press, Cambridge (2001)
16. Crammer, K., Dredze, M., Ganchev, K., Talukdar, P.P., Carroll, S.: Automatic code assignment to medical text. In: Proceedings of the Workshop on Biological, Translational, and Clinical Language Processing, BioNLP 2007. Prague, Czech Republic, pp. 129–136 (2007)
17. 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). doi:10.1007/978-3-540-24775-3_5
18. Boutell, M., Luo, J., Shen, X., Brown, C.: Learning multi-label scene classification. Pattern Recogn. **37**(9), 1757–1771 (2004). doi:10.1016/j.patcog.2004.03.009
19. Read, J., Pfahringer, B., Holmes, G., Frank, E.: Classifier chains for multi-label classification. Mach. Learn. **85**, 333–359 (2011). doi:10.1007/s10994-011-5256-5
20. Tsoumakas, G., Katakis, I., Vlahavas, I.: Effective and efficient multilabel classification in domains with large number of labels. In: Proceedings of the ECML/PKDD Workshop on Mining Multidimensional Data, MMD 2008. Antwerp, Belgium, pp. 30–44 (2008)
21. Zhang, M., Zhou, Z.: ML-KNN: a lazy learning approach to multi-label learning. Pattern Recogn. **40**(7), 2038–2048 (2007). doi:10.1016/j.patcog.2006.12.019
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, p. 42. Springer, Heidelberg (2001). doi:10.1007/3-540-44794-6_4
23. Zhang, M.-L.: Multilabel neural networks with applications to functional genomics and text categorization. IEEE Trans. Knowl. Data Eng. **18**(10), 1338–1351 (2006). doi:10.1109/TKDE.2006.162
24. Zhang, M., Zhou, Z.: A review on multi-label learning algorithms. IEEE Trans. Knowl. Data Eng. **26**(8), 1819–1837 (2014). doi:10.1109/TKDE.2013.39
25. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002). doi:10.1613/jair.953
26. Kotsiantis, S.B., Pintelas, P.E.: Mixture of expert agents for handling imbalanced data sets. Ann. Math. Comput. Teleinformatics **1**, 46–55 (2003)
27. 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. Sci. **250**, 113–141 (2013). doi:10.1016/j.ins.2013.07.007
28. Provost, F., Fawcett, T.: Robust classification for imprecise environments. Mach. Learn. **42**, 203–231 (2001). doi:10.1023/A:1007601015854
29. 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). doi:10.1371/journal.pone.0037155

30. 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)
31. Tepvorachai, G., Papachristou, C.: Multi-label imbalanced data enrichment process in neural net classifier training. In: IEEE International Joint Conference on Neural Networks, IJCNN 2008, pp. 1301–1307 (2008). doi:10.1109/IJCNN.2008.4633966
32. Charte, F., Charte, F.D.: How to work with multilabel datasets in R using the mldr package. doi:10.6084/m9.figshare.1356035
33. Cheng, W., Hüllermeier, E.: Combining instance-based learning and logistic regression for multilabel classification. Mach. Learn. **76**(2–3), 211–225 (2009). doi:10.1007/s10994-009-5127-5