ReliefF for Multi-label Feature Selection

Newton Spolaôr, Everton Alvares Cherman,
Maria Carolina Monard
Laboratory of Computational Intelligence
Institute of Mathematics and Computer Science
University of São Paulo
São Carlos, Brazil

Huei Diana Lee
Laboratory of Bioinformatics
Western Paraná State University
Foz do Iguaçu, Brazil
{*newtonspolaor,hueidianalee*}@*gmail.com,*
mcmonard@icmc.usp.br

***Abstract*—The feature selection process aims to select a subset of relevant features to be used in model construction, reducing data dimensionality by removing irrelevant and redundant features. Although effective feature selection methods to support single-label learning are abound, this is not the case for multi-label learning. Furthermore, most of the multi-label feature selection methods proposed initially transform the multi-label data to single-label in which a traditional feature selection method is then applied. However, the application of single-label feature selection methods after transforming the data can hinder exploring label dependence, an important issue in multi-label learning. This work proposes a new multi-label feature selection algorithm, *RF-ML*, by extending the single-label feature selection ReliefF algorithm. *RF-ML*, unlike strictly univariate measures for feature ranking, takes into account the effect of interacting attributes to directly deal with multi-label data without any data transformation. Using synthetic datasets, the proposed algorithm is experimentally compared to the ReliefF algorithm in which the multi-label data has been previously transformed to single-label data using two well-known data transformation approaches. Results show that the proposed algorithm stands out by ranking the relevant features as the best ones more often.**

***Keywords*-feature ranking; filter feature selection; Hamming distance; RReliefF; systematic review**

I. Introduction

Feature Selection (FS), usually applied as a data preprocessing step in machine learning and data mining, aims to find a small number of features that describes the dataset as well as or even better than the original set of features does [1]. FS provides support to tackle the "*curse of dimensionality*" problem by removing irrelevant and/or redundant features, speeding up learning algorithms and sometimes improving their performance [2].

Feature selection has been widely considered to support single-label learning, where each example (or instance) in the dataset is associated with only one class. However, this is not the case in multi-label learning, where each example is associated with a subset of labels, *i.e.*, each example can simultaneously belong to multiple classes. In fact, the main difference between multi-label and single-label learning is that classes in multi-label learning are often correlated, while the class values in single-label learning are mutually exclusive.

Multi-label learning is an emerging research topic due to the increasing number of applications where examples are annotated using more than one class, such as bioinformatics, emotion analysis, semantic annotation of media and text mining [3].

However, research on multi-label feature selection is scarce. The standard approach [4], which transforms multi-label data into one or more single-label data before selecting features is often used. Nevertheless, this transformation can hinder exploring label dependence, an important issue in multi-label learning [5].

In this work, we propose an extension of the single-label ReliefF algorithm [6] for multi-label feature selection, named *RF-ML*, which deals with multi-label data directly, *i.e.*, without any data transformation. In addition, it takes into account, as ReliefF does, the effect of interacting attributes [7]. Experimental results in synthetic datasets show that *RF-ML* ranks the relevant features as the best ones more often when compared to the single-label ReliefF algorithm in which the multi-label data has been transformed using two well-known data transformation approaches.

The rest of this paper is organized as follows: Section II briefly presents multi-label feature selection and Section III summarizes related work carried out by a systematic literature review. ReliefF and the proposed algorithm *RF-ML* are described in Section IV. Section V describes and discusses the experimental evaluation. Section VI concludes and highlights future work.

II. Background

This section presents basic concepts and terminology of multi-label learning and FS.

A. Multi-label learning

Let D be a dataset composed of N examples $E_i = (\mathbf{x}_i, Y_i)$, $i = 1..N$. Each example (instance) E_i is associated with a feature vector $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iM})$ described by M features (attributes) X_j, $j = 1..M$, and its multi-label Y_i, which consists of a subset of labels $Y_i \subseteq L$, where $L = \{y_1, y_2, \ldots y_q\}$ is the set of q labels. Table I shows this representation. In this scenario, the multi-label classification task consists in generating a classifier H which, given an unseen instance $E = (\mathbf{x}, ?)$, is capable of accurately predicting its multi-label Y, *i.e.*, $H(E) \rightarrow Y$.

Table I
MULTI-LABEL DATA.

	X_1	X_2	$\ldots$	X_M	Y
E_1	x_{11}	x_{12}	$\ldots$	x_{1M}	Y_1
E_2	x_{21}	x_{22}	$\ldots$	x_{2M}	Y_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
E_N	x_{N1}	x_{N2}	$\ldots$	x_{NM}	Y_N

Multi-label learning methods can be organized into two main categories: algorithm adaptation and problem transformation [3]. The first one consists of methods which extend specific learning algorithms to handle multi-label data directly, such as Multi-label Naive Bayes (MLNB) [8]. The second category is algorithm independent, allowing one to use any state of the art single-label learning method. Methods which transform the multi-label classification problem into either several binary classification problems, such as the Binary Relevance (*BR*) approach or one multi-class classification problem, such as the Label Powerset (*LP*) approach, fall within this category. Recall that single-label learning is called binary whenever the class value can take two values, and it is called multi-class whenever the class value can take more than two values.

B. Multi-label feature selection

Most of the multi-label feature selection methods proposed [4] use the problem transformation approach to previously transform multi-label data into single-label data using *BR* or *LP*, for example. As this approach is algorithm independent, it allows one to use any single-label FS algorithm. On the other hand, few multi-label feature selection methods that directly deal with multi-label data have been proposed [3], [8].

Regardless of the multi-label learning approach, any feature selection method addresses two relevant issues: interacting with the learning algorithm and evaluating feature importance.

Considering the interaction with the learning algorithm, there are three feature selection approaches: wrapper, embedded and filter [1]. The first two approaches strongly interact with the learning algorithm. Wrappers use a specific learning algorithm as a “black box” to evaluate feature importance. Conversely, the embedded approach incorporates feature selection during training of the learning algorithm, as decision trees do.

Unlike the wrapper and embedded approaches, filters remove irrelevant and/or redundant features regardless of the learning algorithm. They only use general properties of the dataset to perform feature selection. Thus, the features chosen may not be the best ones for specific learning algorithms. The FS algorithms used in this work fall within this approach.

Another important issue is how the evaluation of the features is tackled by the feature selection algorithm, *i.e.*, whether the quality of the features is estimated individually or in subsets. The filter feature ranking methods used in this work evaluate one feature at a time.

III. RELATED WORK

Feature selection has been an active research topic in supervised learning, with several related publications and comprehensive surveys [2], [3], [1]. However, most of the research related to supervised feature selection has been mainly proposed to support single-label classification, and few results on multi-label classification have been reported.

A systematic review process, a method to perform a wide, replicable and rigorous bibliographic review, was carried out in [4] and was recently updated to search for multi-label feature selection publications. Results gathered from 60 selected papers suggest a growing interest in the subject. Table II summarizes related multi-label filter FS publications which consider label dependence.

Table II
MULTI-LABEL FILTER FEATURE SELECTION PUBLICATIONS.

Reference	Feature importance measure
[9]	chi-squared
[10]	information gain
[11]	mutual information
[12]	correlation-based feature selection
[13]	symmetrical uncertainty
[14]	relieff/f-statistic
[15], [16]	information gain/relieff
[17]	mutual information

Unlike strictly univariate measures, single-label ReliefF takes into account the important issue of interacting attributes during feature ranking [7], which makes it one of the most well-known single-label FS algorithms. In what follows, we focus on publications using the ReliefF algorithm for multi-label feature selection.

Two multi-label FS methods, *RF-BR* and *RF-LP*, in which ReliefF is used in conjunction with data transformation approaches, are presented in [16], [15], [18]. *RF-BR* transforms a multi-label dataset into q single-label ones, applies the conventional ReliefF in each binary dataset and selects the features with average weight greater than or equal to a threshold. On the other hand, *RF-LP* transforms the multi-label dataset into a multi-class dataset and applies ReliefF once to select the features with weight greater than or equal to a threshold. Results in ten benchmark datasets for *RF-BR* and *RF-LP* show that both methods select subsets of features, which do not diminish the quality of the classifiers constructed using these features.

In [14], ReliefF is extended by analyzing pairs of labels and ignoring instances in which both labels co-occur. The performance of the classifiers constructed using the features selected in three image annotation datasets are usually better than the ones obtained by *RF-BR* and *RF-LP*.

IV. The proposed algorithm: *RF-ML*

As the proposed algorithm is built upon ReliefF and RReliefF, we first briefly review the earlier Relief algorithms. The main idea of Relief for binary data is to reward a feature for having different values on a pair of the nearest examples from different classes, and penalize it for having different values on examples from the same class [7]. ReliefF extends Relief to deal with multi-class, missing and noisy data by using $k > 1$ nearest neighbors. Finally, RReliefF extends ReliefF to deal with regression problems, in which the predicted values (class) are continuous. To avoid searching for examples from the same class in those problems, the probability that the predicted values of two examples are different is introduced. This probability is modeled as a dissimilarity function between the predicted values of the examples [6].

As Kononenko and Robnik-Šikonja showed [19], ReliefF and RReliefF have similar weighting schemes. Indeed, RReliefF uses the Bayes' rule to approximate the ReliefF weighting, not needing probabilities of examples from the same class.

Based on ReliefF and RReliefF, we propose the *RF-ML* algorithm (Algorithm IV.1) for multi-label feature selection. *RF-ML* is similar to RReliefF, but it searches for k nearest multi-label instances and uses a dissimilarity function $mld(.,.)$, which deals with multi-labels instead of single-labels (Lines 9 and 13). It should be emphasized that *RF-ML* takes into account the effect of interacting attributes by analysing dissimilarity between instances, as ReliefF and RReliefF do.

Note that $mld(E_a, E_b)$ can be any dissimilarity function in the multi-labels Y_a and Y_b, such as the Hamming Distance (HD). The HD between two sets (multi-labels) is defined as $|Y_a \cup Y_b| - |Y_a \cap Y_b|$. Thus, it counts the number of labels which are different in Y_a and Y_b. Observe that the HD considers the presence and absence of labels equally. For example, the HD between $Y_a = \{y_3, y_6\}$ and $Y_b = \{y_1, y_4\}$ is 4, the same as the HD between $Y_a = \{y_2, y_3, y_5, y_6\}$ and $Y_b = \{y_1, y_2, y_4, y_5\}$. In this work, we used the normalized HD, given by Equation 1.

$$HD(Y_a, Y_b) = \frac{|Y_a \cup Y_b| - |Y_a \cap Y_b|}{q}. \quad (1)$$

Similar to RReliefF, *RF-ML* specifies the partial weights W_{dY}, W_{dX} and W_{dYX} according to dissimilarities between labels, feature values and labels $\wedge$ feature values, respectively. Furthermore, both algorithms support distance weighting by the term $d(E_i, EK_z)$ and estimate the output W by combining the partial weights in an equation given by Bayes' rule (Line 19). The $diff(X_j, E_i, EK_z)$ metric calculates the difference between the values of a feature X_j in two instances E_i and EK_z.

Algorithm IV.1 ReliefF for Multi-label Feature Selection: *RF-ML*

Input: $Dataset\ D$
$Number\ of\ iterations\ c$
$Number\ of\ nearest\ neighbors\ k$
Output: $Vector\ of\ feature\ importance\ values\ W$
1: $W_{dY} \leftarrow \emptyset$ ▷ Label dissimilarities
2: $W_{dX} \leftarrow \emptyset$ ▷ Feature dissimilarities
3: $W_{dYX} \leftarrow \emptyset$ ▷ Label and feature dissimilarities
4: $W \leftarrow \emptyset$
5: **for** $i = 1 \rightarrow c$ **do**
6: $E_i \leftarrow randomInstance(D)$
7: $EK \leftarrow kNearestNeighbors(k, E_i, D)$
8: **for** $z = 1 \rightarrow k$ **do**
9: $W_{dY} \leftarrow W_{dY} + mld(E_i, EK_z) \times d(E_i, EK_z)$
10: **for** $j = 1 \rightarrow M$ **do**
11: $W_{dX}(X_j) \leftarrow W_{dX}(X_j) + diff(X_j, E_i, EK_z) \times$
12: $d(E_i, EK_z)$
13: $W_{dYX}(X_j) \leftarrow W_{dYX}(X_j) + mld(E_i, EK_z) \times$
14: $diff(X_j, E_i, EK_z) \times d(E_i, EK_z)$
15: **end for**
16: **end for**
17: **end for**
18: **for** $j = 1 \rightarrow M$ **do**
19: $W(X_j) \leftarrow \frac{W_{dYX}(X_j)}{W_{dY}} - \frac{W_{dX}(X_j) - W_{dYX}(X_j)}{c - W_{dY}}$
20: **end for**

The complexity of traditional Relief-based algorithms [6], as well as the proposed algorithm *RF-ML*, is bound to the search for the k nearest instances, *i.e.*, $O(N^2.M)$, where N is the number of instances and M is the number of features. In fact, the N^2 term in the upper bound is reached when $c = N$ (default ReliefF configuration), which is important to obtain reliable results from the algorithm. On the other hand, *RF-BR* clearly has a higher complexity, *i.e.*, $O(N^2.M.q)$, as it has to estimate the ReliefF weights q times, one per label.

V. Experimental evaluation

This section presents an experimental comparison in synthetic datasets among the three multi-label FS algorithms: *RF-BR*, *RF-LP* and *RF-ML*. All methods were implemented using Mulan[1], a Java package for multi-label learning based on Weka[2]. Appart from the dissimilarity function in *RF-ML*, set as Hamming distance, all parameters were executed with default values.

A. *Synthetic datasets*

Synthetic (or artificial) datasets are useful to evaluate algorithms, as they provide researchers with a controlled

[1] http://mulan.sourceforge.net
[2] http://www.cs.waikato.ac.nz/ml/weka

environment based on known properties of the datasets [20]. In fact, showing the improvements of an algorithm on synthetic data sets can be more convincing that doing so in typical real world scenarios where the true solution is completely unknown [21].

However, few publicly available strategies to generate synthetic multi-label datasets have been provided. To this end, we have developed a publicly available framework *Mldatagen*[3] [22]. Up to now, *Mldatagen* implements two different strategies (*HyperCubes* and *HyperSpheres*) which are based on the one proposed by Zhang *et al.* [8]. Moreover, *Mldatagen* outputs the datasets in the well-known Mulan framework format[4]. After choosing the strategy, some mandatory parameters must be set, such as: the number of relevant (M_{rel}), irrelevant (M_{irr}) and redundant (M_{red}) features; number of labels (q) and number of instances (N) of the dataset. It is also possible to set several other optional parameters which have default values, such as the noise level μ.

In this work, we use the *HyperCubes* strategy to generate 45 synthetic datasets $(9\, settings \times 5\, noise\, levels)$, with $M = M_{rel} + M_{irr}$ features, for different values of M_{rel}, M_{irr}, N, q and five noise levels (μ): 0%, 5%, 10%, 20% and 40%. Table III describes the nine datasets created for each noise level.

Table III
NINE SYNTHETIC DATASETS WITH SAME NOISE LEVEL.

	N = 10000 q = 100	N = 5000 q = 30	N = 2000 q = 10
$M_{rel} = 10, M_{irr} = 10$	$dataset_1$	$dataset_4$	$dataset_7$
$M_{rel} = 5, M_{irr} = 15$	$dataset_2$	$dataset_5$	$dataset_8$
$M_{rel} = 1, M_{irr} = 19$	$dataset_3$	$dataset_6$	$dataset_9$

By inserting noise in a multi-label dataset, new combinations of labels can arise. In fact, the higher the noise level, the higher the number of distinct combinations of labels is.

B. Evaluation measure

To evaluate the features selected, the Area Under the Curve for Feature Ranking evaluation (*FR-AUC*) is used. Considering M_{rel} and M_{irr} as the lengths of the *y* and *x*-axis respectively, the feature ranking is analyzed in descending order, such that the best ranked feature is taken into account first. For each feature X_j, the curve increases one unit in the *y*-axis if X_j is relevant (X_R); otherwise, it increases one unit in the *x*-axis (X_I). *FR-AUC* is the area under this curve, normalized by $M_{rel} \times M_{irr}$, *i.e.*, the maximum possible *FR-AUC* value. Thus, the *FR-AUC* value ranges from 0 to 1. The higher the *FR-AUC* value, the better the ranking is.

To illustrate this, consider a dataset with $M_{rel} = 5$, $M_{irr} = 15$ and feature ranking $(X_R, X_R, X_I, X_R, X_R, X_I, X_I, X_R, X_I, \ldots, X_I)$. Figure 1 shows the curve and the area under the curve obtained from this ranking.

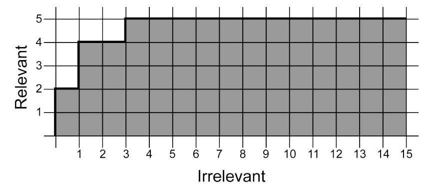


Figure 1. Area under curve for feature ranking: *FR-AUC*.

C. Results and discussion

As already mentioned, *RF-BR*, *RF-LP* and the proposed algorithm *RF-ML* were experimentally compared in 45 synthetic datasets according to the *FR-AUC* evaluation measure. Due to lack of space, in what follows, the main results are presented. Detailed results showing the *FR-AUC* and the feature rankings for all datasets can be found at http://www.labic.icmc.usp.br/pub/mcmonard/ExperimentalResults/BRACIS2013.pdf.

Table IV shows the number of times in which *FR-AUC* did not achieve its maximum value for each noise level (9 datasets) and all the 45 datasets.

Table IV
NUMBER OF CASES IN WHICH *FR-AUC* DID NOT ACHIEVE ITS MAXIMUM VALUE.

	RF-BR	*RF-LP*	*RF-ML*
$\mu = 0\%$	1	2	0
$\mu = 5\%$	1	5	0
$\mu = 10\%$	1	6	0
$\mu = 20\%$	3	7	0
$\mu = 40\%$	6	8	5
all 45 datasets	12	28	5
%	27%	62%	11%

As can be observed, *RF-ML* obtained better results than *RF-BR* and *RF-LP* in all cases, failing to achieve the best *FR-AUC* value in only 11% of the cases. In fact, up to a noise level of 20%, *RF-ML* always found the best features, *i.e.*, *FR-AUC* = 1. It only failed when the noise level reaches 40%, suggesting robustness to noise. *RF-BR* obtained the second best results, although it failed more than twice as much as *RF-ML* (12 against 5) in achieving the maximum *FR-AUC* value.

The *FR-AUC* value was used to rank the results obtained for each noise level averaged across the nine datasets with the same noise level. Table V presents these results, where the best ones are highlighted in bold.

Note that *RF-ML* stands out in all these results, leading to the lowest average rankings, followed by *RF-BR*. However, as mentioned in Section IV, the complexity of *RF-BR* is higher than the one of *RF-ML*, as it has to apply ReliefF q

[3] http://sites.labic.icmc.usp.br/mldatagen
[4] http://mulan.sourceforge.net/format.html

Table V
AVERAGE RANKING (AND STANDARD DEVIATION) OF *FR-AUC*.

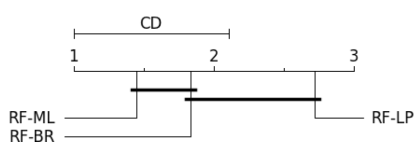
	RF-BR	*RF-LP*	*RF-ML*
$\mu = 0\%$	2 (0.4)	2.2 (0.5)	**1.8** (0.3)
$\mu = 5\%$	1.8 (0.3)	2.6 (0.5)	**1.7** (0.4)
$\mu = 10\%$	1.7 (0.3)	2.7 (0.5)	**1.6** (0.3)
$\mu = 20\%$	1.8 (0.5)	2.7 (0.6)	**1.4** (0.3)
$\mu = 40\%$	1.7 (0.4)	2.9 (0.3)	**1.4** (0.5)
all datasets	1.8 (0.4)	2.6 (0.5)	**1.6** (0.4)

times. As in Table IV, Table V shows that *RF-LP* obtained the worst results.

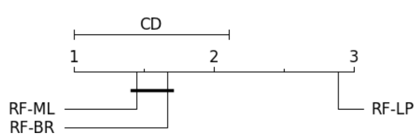
By analyzing the results according to μ in Table V, the higher the noise, the better the *RF-ML* average ranking is, strengthening its robustness to noise. On the other hand, *RF-LP* shows opposite behavior, which can be associated to the relation between μ and the number of distinct combinations of labels (Section V-A). In fact, *RF-LP* is sensitive to this number, as it transforms each different multi-label into a new single-label class value.

As mentioned, *RF-ML* leads to the highest percentage of maximum *FR-AUC*. The few datasets in which *FR-AUC* was not maximum have the highest noise ($\mu = 40\%$). These datasets are indeed hard to analyze, since there is 40% probability of changing a label in the multi-label of each example.

To verify if the difference among the methods is significant, we ran the Friedman's test with 95% of confidence level followed by the Nemenyi's post hoc test. Significant differences were found for $\mu = 20\%$ and 40%, in which the average rankings differ by at least the Critical Difference (CD) [23]. Figure 2 shows the graphical representation of these results.



(a) Datasets with $\mu = 20\%$.



(b) Datasets with $\mu = 40\%$.

Figure 2. Diagram of the Nemenyi's post hoc test. Methods which are not significantly different at $p < 0.05$ are connected.

For $\mu = 20\%$, Figure 2a shows that there is a significant difference among *RF-ML* and *RF-LP*. For $\mu = 40\%$, Figure 2b shows that both, *RF-ML* and *RF-BR*, are significantly better than *RF-LP*. Although there is no significant difference between *RF-ML* and *RF-BR*, in both cases *RF-ML* is ranked first. Nevertheless, in all the five cases analyzed, *RF-ML* is always ranked first, which strengthens *RF-ML* as a multi-label feature selection method.

VI. CONCLUSION

This work presents *RF-ML*, a new multi-label feature selection algorithm based on ReliefF and RReliefF, which uses a dissimilarity function in the multi-labels to find the nearest instances as a partial weight to estimate feature importance. Thus, no data transformation is required by *RF-ML*. The experimental evaluation in controlled environments shows that *RF-ML*, using the Hamming distance as the dissimilarity function, ranks first all the relevant features in 89% of all the 45 cases considered. Moreover, it outperforms two other methods, *RF-BR* and *RF-LP*, which use two well-known data transformation approaches and single-label ReliefF to rank the features.

As future work, we plan to evaluate other similarity functions which do not consider equally the presence and absence of labels in the multi-labels, as the Hamming distance does. In addition, we plan to compare *RF-ML* in benchmark and real multi-label datasets with other feature selection algorithms which also consider feature interactions, such as CFS [12]. Furthermore, to deal with large datasets we plan to parallelize the RF-ML implementation to speed up its execution, as suggested in [19].

ACKNOWLEDGMENT

This research was supported by the São Paulo Research Foundation (FAPESP), grants 2011/02393-4 and 2010/15992-0. The authors would like to thank the anonymous reviewers for their helpful comments.

REFERENCES

[1] H. Liu and H. Motoda, *Computational Methods of Feature Selection*. Chapman & Hall/CRC, 2008.

[2] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu, “Advancing feature selection research - ASU feature selection repository,” Technical Report, 2011, Arizona State University.

[3] G. Tsoumakas, I. Katakis, and I. P. Vlahavas, “Mining multi-label data,” in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Springer, 2010, pp. 667–685.

[4] N. Spolaôr, M. C. Monard, and H. D. Lee, “A systematic review to identify feature selection publications in multi-labeled data,” ICMC Technical Report No 374. 31 pg., 2012, University of São Paulo.

[5] K. Dembczynski, W. Waegeman, W. Cheng, and E. Hüllermeier, “On label dependence and loss minimization in multi-label classification,” *Machine Learning*, vol. 88, pp. 5–45, 2012.

[6] M. Robnik-Šikonja and I. Kononenko, “Theoretical and empirical analysis of ReliefF and RReliefF,” *Machine Learning*, vol. 53, no. 1–2, pp. 23–69, 2003.

[7] J. Demšar, "Algorithms for subsetting attribute values with Relief," *Machine Learning*, vol. 78, pp. 421–428, 2010.

[8] M.-L. Zhang, J. M. Peña, and V. Robles, "Feature selection for multi-label Naive Bayes classification," *Information Sciences*, vol. 179, pp. 3218–3229, 2009.

[9] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, "Multi-label classification of music into emotions," in *International Conference on Music Information Retrieval*, 2008, pp. 1–6.

[10] S. Dendamrongvit, P. Vateekul, and M. Kubat, "Irrelevant attributes and imbalanced classes in multi-label text-categorization domains." *Intelligent Data Analysis*, vol. 15, no. 6, pp. 843–859, 2011.

[11] G. Doquire and M. Verleysen, "Feature selection for multi-label classification problems," in *Advances in Computational Intelligence*, J. Cabestany, I. Rojas, and G. Joya, Eds. Springer-Verlag/Heidelberg, 2011, ch. 2, pp. 9–16.

[12] S. Jungjit, A. A. Freitas, M. Michaelis, and J. Cinatl, "A multi-label correlation-based feature selection method for the classification of neuroblastoma microarray data," in *Industrial Conference on Data Mining - Workshops*, 2012, pp. 149–157.

[13] G. Lastra, O. Luaces, J. R. Quevedo, and A. Bahamonde, "Graphical feature selection for multilabel classification tasks," in *International Conference on Advances in Intelligent Data Analysis*, 2011, pp. 246–257.

[14] D. Kong, C. Ding, H. Huang, and H. Zhao, "Multi-label ReliefF and F-statistic feature selections for image annotation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2352–2359.

[15] N. Spolaôr, E. A. Cherman, M. C. Monard, and H. D. Lee, "Filter approach feature selection methods to support multi-label learning based on ReliefF and information gain," in *Advances in Artificial Intelligence - SBIA 2012*, ser. Lecture Notes in Artificial Intelligence, L. N. Barros, M. Finger, A. T. Pozo, G. A. Gimenénez-Lugo, and M. Castilho, Eds. Springer Berlin Heidelberg, 2012, vol. 7589, pp. 72–81.

[16] ——, "A comparison of multi-label feature selection methods using the problem transformation approach," *Electronic Notes in Theoretical Computer Science*, vol. 292, pp. 135–151, 2013.

[17] J. Lee and D.-W. Kim, "Feature selection for multi-label classification using multivariate mutual information," *Pattern Recognition Letters*, vol. 34, no. 3, pp. 349–357, 2013.

[18] N. Spolaôr, E. A. Cherman, and M. C. Monard, "Using ReliefF for multi-label feature selection (in Portuguese)," in *Conferencia Latinoamericana de Informática*, 2011, pp. 960–975.

[19] I. Kononenko and M. Robnik-Šikonja, "Weighting and local methods non-myopic feature quality evaluation with (R)ReliefF," in *Computational Methods of Feature Selection*, H. Liu and H. Motoda, Eds. Chapman & Hall/CRC, 2008, pp. 169–191.

[20] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowledge and Information Systems*, vol. 34, no. 3, pp. 483–519, 2013.

[21] L. A. B. Muñoz and F. F. González-Navarro, "Review and evaluation of feature selection algorithms in synthetic problems," *CoRR*, vol. abs/1101.2320, 2011.

[22] J. T. Tomás, N. Spolaôr, E. A. Cherman, and M. C. Monard, "A computational framework to generate synthetic datasets for multi-label learning (in Portuguese)," ICMC Technical Report No 391. 46 pg., 2013, University of São Paulo.

[23] J. Demšar, "Statistical comparison of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, no. 1, pp. 1–30, 2006.