2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Applying the Self-training Semi-Supervised Learning in Hierarchical Multi-Label Methods

Araken M Santos
Exact, Technology and Human Sciences Department
Federal Rural University of Semi-Arido
Angicos, RN - BRAZIL 59.515-000
Email: araken@ufersa.edu.br

Anne M P Canuto
Informatics and Applied Mathematics Department
Federal University of Rio Grande do Norte
Natal, RN - BRAZIL 59.072-970
Email: anne@dimap.ufrn.br

***Abstract*—In classification problems with hierarchical structures of labels, the target function must assign several labels that are hierarchically organized. The hierarchical structures of labels can be used either for single-label (one label per instance) or multi-label classification problems (more than one label per instance). In general, classification tasks are usually trained using a standard supervised learning procedure. However, the majority of classification methods require a large number of training instances to be able to generalize the mapping function, making predictions with high accuracy. In order to smooth out this problem, the idea of semi-supervised learning has emerged. It combines labelled and unlabelled data during the training phase. Some semi-supervised methods have been proposed for single-label classification methods. However, very little effort has been done in the context of multi-label hierarchical classification. This paper proposes the use of a semi-supervised learning method for the multi-label hierarchical problems. In order to validate the feasibility of these methods, an empirical analysis will be conducted, comparing the proposed methods with their corresponding supervised versions. The main aim of this analysis is to observe whether the semi-supervised methods proposed in this paper have similar performance to the corresponding supervised versions.**

I. Introduction

In Machine Learning, there are problems in which instances are assigned to a single label l from a previously known finite set of disjoint labels L and these problems are known as single-label classification problems. In this context, the training instances are assumed to be drawn from some unknown distribution over tuples of the form (x_i,y_i), where x_i is an m-dimensional vector that represents a data point in the m-dimensional space and y_i represents one label to which x_i belongs. Nevertheless, there are some real classification problems in which an instance can belong to more than one class simultaneously and they are known as multi-label (ML) classification problems [1]. In this case, instances can be associated with a set of labels Y, where $Y \subseteq L$.

Independently of the number of assigned labels, the vast majority of classification methods are related to flat classification, where an algorithm generates a function that maps instances to one (or more) label of the available set without any hierarchical structure. In other words, the labels are arranged on one level. However, in several tasks the target function must assign not only a single label, but a series of hierarchically organized labels [2]. This hierarchical structure can be used in single-label (HSC) and multi-label (HML) classification problems. In the second case, we call it multi-label hierarchical problems.

One drawback of the classification algorithms is that they need to have a set of labelled instances with a reasonable size in order to work in an efficient way. However the labelling process is often hard, expensive, and slow to obtain, because it may require human expertise. For instance, in speech recognition, the accurate transcription of speech utterance at phonetic level is extremely time consuming and requires linguistic expertise. In text categorization, we need to take expensive actions to label the instances, such as: filtering out spam emails, categorizing user messages, recommending Internet articles, among others. In protein structure prediction, it may take several months of expensive lab work by expert to identify the 3D structure of a single protein. On the other hand, unlabelled instances are usually available in large quantity and costs little to collect. Utterances can be recorded from radio broadcast; Text documents can be crawled from the Internet and DNA sequences of proteins are readily available from gene databases. The problem with traditional classification methods is that they can not use unlabelled data to train classifiers.

The problem cited above is particularly important when the class labels follow a hierarchical structure, since it must have a good number of instances for all possible classes. Additionally, this problem becomes even more critical in the multi-label context, since the number of possible combinations in the label attributes increases considerably. Given that the cost of manually labelling instances is a very high and time-consuming process, researchers have been trying to smooth out this problem by using the information carried by the unlabelled instances in order to increase the performance of the classification models. This is called semi-supervised learning [3].

There is a wide range of studies on the use of semi-supervised learning in classification problems. However, in general, it involves either hierarchical single-label problems or flat multi-label problems. Thus, in this paper, we propose the use of semi-supervised learning in the context of the hierarchical multi-label classification tasks. In this case, a class of methods, known as semi-supervised multi-label hierarchical algorithms, is proposed. We propose the use of a specific

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

semi-supervised approach, called self-training, to increase the number of labelled training set using unlabelled instances, and then we build a final model using the new increased set of instances. Self-training is the most popular semi-supervised method because of its conceptual and algorithmic simplicity and these are the main reasons we selected this semi-supervised method in our work here. In addition, although it may be straightforward to incorporate the self-training idea into the existing HML algorithms, to the best of our knowledge, there is no study involving these three important subjects.

In order to analyse the performance of the proposed methods, an empirical analysis will be conducted. In this analysis, the proposed methods will be compared with the original hierarchical multi-label methods. These methods will be evaluated using different evaluation metrics. As a result of this analysis, we aim to investigate the effect of the use of semi-supervised learning in the hierarchical multi-label classification context, under different circumstances.

This paper is divided into seven sections and structured as follows. Classification approaches are described in Section 2, focusing on multi-label and hierarchical classification, while Section 3 presents a detailed explanation about semi-supervised learning. In Section 4, we describe the proposed methods of this paper. The experimental work is presented in Section 5, while Section 6 presents and discusses the experimental results. Finally, in Section 7, some final observations about this work are presented.

II. Classification Approaches

There are several approaches for classification methods, that can be categorized by the number of assigned label and/or structure of these labels. In this paper, as we work with multi-label and hierarchical classifications, in the next two subsection, we will describe these two approaches for classification methods.

A. Multi-label Classification

As mentioned previously, in a multi-label classification problem, an instance can be assigned to more than one label from the set of possible labels. This approach has attracted significant attention from a lot of researchers in several applications such as semantic annotation of images [4], [5] and video [6], [7], functional genomics [1], [8], [9], [10], music categorization into emotions [11], [12], directed marketing [13], among others. In the literature, different methods have been proposed to be applied to multi-label classification problems, as in [14], [15], which can be broadly classified as problem transformation and algorithm adaptation methods. The majority of the methods is based on problem transformation approach. In this approach, the main idea is to transform the original multi-label problem into a set of single-label classification problems. It is an algorithm independent approach, since its functioning does not depend directly on the classification method used. In this paper, we have chosen to use the two most widely applied methods in the literature, which are:

- Label Powerset (LP): This method considers each unique set of labels that exists in a multi-label training set as one of the labels of a new single-label classification task [14].
- Random k-labelsets (RAkEL): This method an ensemble of LP classifiers. Each LP classifier is trained using a different small random subset of the set of labels [14].

B. Hierarchical Classification

The vast majority of classification tasks addressed in the literature involve flat classification, where each example is assigned to a class out of a finite (and usually small) set of classes. On the other hand, in hierarchical classification problems, the classes are disposed in a hierarchical structure. In these classification problems, one or more classes can be divided into sub-classes or grouped into super-classes. In these structures, the nodes represent classes [16].

There are two main types of hierarchical structures: a tree or a Directed Acyclic Graph (DAG). The main difference between them is that, in the tree structure, each node has just one parent node, while, in the DAG structure, each node may have more than one parent. For both structures, the root node corresponds to any class, denoting a total absence of knowledge about the class of an object [17].

Most of the solutions proposed to hierarchical problems are based on top-down approach [16]. The top-down approach consists of constructing a tree classifier by training one or more classifiers at each level of hierarchy. In this approach, a classifier is trained with all training instances at the root node of the hierarchy. At the next level, for each class, a classifier is trained using only a subset of instances of the classes predicted by the classifier. The training process of classifiers proceeds in a top-down scheme until the classes belonging to the leaf nodes are predicted by a classifier. After the training phase, a set of hierarchically organized classifiers is obtained. Thus, the examples belonging to the test set are classified iteratively in a top-down approach throughout the hierarchy, starting at the root node. When an instance is associated with a class, the instance is then passed to a classifier to predict which subclasses it belongs to. In the literature, there are some works that use the top-down approach, such as in [16], [18].

C. Hierarchical Multi-Label Classification

The hierarchical multi-label classification (HMC) has emerged as a new category of classification problems, using features of both hierarchical and multi-label classification problems. In HMC problems, an instance may belong to several labels simultaneously and, in addition, the labels are organized in a hierarchical structure. Thus, an instance belonging to a label belongs automatically to all of its predecessor labels in the hierarchical structure. Problems of this type are reasonably common, especially in text categorization and prediction of protein functions problems. Moreover, hierarchical multi-label classification problems are more complex than the other classification problems.

Several methods can be used for hierarchical multi-label tasks in different domains, such as text classification [19], functional prediction of proteins [20] and object recognition [21]. In [22], two methods based on local and global approaches are described and evaluated: the first method is called HMC-LP (Hierarchical Multi-label Classification with Label Powerset) and is based on the local approach, however, it uses LP classifiers. Therefore, HMC-LP is an hierarchical adaptation of LP (Label Powerset). For each example, the method combines all the classes marked as an example, at a specific level of the hierarchical structure, in a single new class (meta-class). After the combination of classes, the original HMC problem is transformed into a hierarchical single-label problem and the top-down approach is used in training and testing phases of the method. The second method, called HMC4.5 was proposed by [22], based on the overall approach in the algorithm C4.5.

III. Semi-Supervised Learning

The main aim to combine semi-supervised learning with multi-label classification is the definition of efficient techniques to deal with problems where previously-classified data is very scarce. It is well known that the labeling process in some classification problems is expensive and a time consuming task. In this case, it is hard to build a good classification model since the available information representing efficiently the data space is not representative enough. In order to cope with this problem, we use the semi-supervised learning [2]. In this case, it is possible to use partially-supervised information to guide to the learning process and increase the amount of evidence regarding the target problem. In general, semi-supervised learning uses both types of data, but mainly in situations where the number of labeled instances is small and the number of unlabeled instances is large [23], [24].

Early studies in semi-supervised learning used methods with conceptual and algorithmic simplicity, such as expectation-maximization (EM)-based algorithms and self-training. As a consequence of self-training, the Co-training semi-supervised learning was proposed. Recently, graph-based semi-supervised learning methods have attracted great attention. Graph-based methods start with a graph where the nodes are the labeled and unlabeled data points, and (weighted) edges reflect the similarity of nodes. As mentioned previously, as our work is the first attempt to incorporate semi-supervised learning in HMC problems, we chose the self-training method as the semi-supervised method in our work.

The Self-Training is one the most commonly used technique for semi-supervised learning [24]. In this technique, the classifier uses its own predictions to teach itself. In this approach, an underlying classifier is first trained with a small number of labelled instances which is called the initial training set. Next, the underlying classifier is used to classify the unlabelled data and label a proportion of the unlabelled instances (the most confident ones). The underlying classifier is then retrained and this procedure is repeated until all the unlabelled instances have been moved to the labelled training set. For more details about this method, see [24].

IV. The Proposed Methods

In this paper, as mentioned previously, we propose the use of semi-supervised learning in hierarchical multi-label classification problems. Figure 1 presents the general methodology of the proposed methods. The functioning of the proposed framework can be describe as follows: a dataset containing labelled and unlabelled instances are presented to the semi-supervised learning module. It is important to emphasize that this module is adapted to label hierarchical multi-label instances. This module labels all unlabelled instance and passes the complete labelled dataset to the hierarchical multi-label module, which is responsible for training and classifying the instances in the dataset.

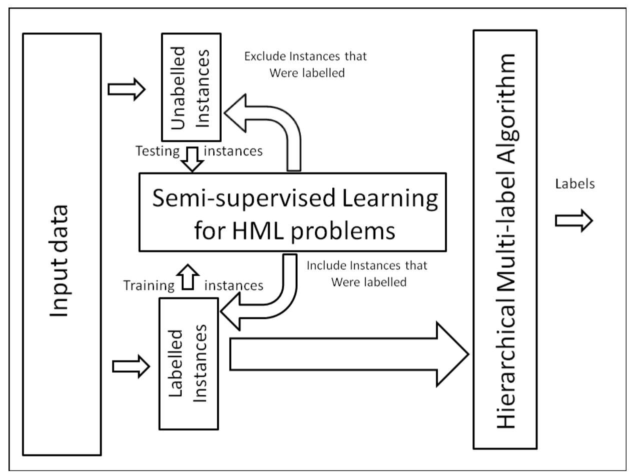

Fig. 1. The general structure of the proposed methodology

Still in Figure 1, during the labelling process (semi-supervised module) the choice of the instances to be labelled is made at random. It is important to emphasize that the proposed methods use a top-down approach as a strategy for hierarchical classification in conjunction with the semi-supervised multi-label classification methods proposed in [25].

Now, in order to better understand the proposed methods, consider a scenario in which the training set is $D = \{D_l \cup D_u\}$, where $D_l = \{E_{l1}, ..., E_{lN}\}$ is the set of labelled instances, $D_u = \{E_{u1}, ..., E_{uK}\}$ is the set of unlabelled instances, $L = \{\lambda_1, ..., \lambda_M\}$ is the set of labels, N is the number of labelled instances, K is the number of unlabelled examples and M is the number of possible labels. Furthermore, $\bigvee E_{li} \in D_l$, E_{li} receives r labels, with $1 \leq r \leq M$.

A. Semi-Supervised Multi-Label Methods

In a previous work of the authors, [25], two methods that apply the semi-supervised technique self-training in multi-label classification were proposed, which are: SSLP (Semi-Supervised Label Powerset) and SSRA*k*EL (Semi-Supervised Random *k*-labelsets). These methods are based on their corresponding supervised methods LP and RA*k*EL. The original and proposed methods differ only in the approach used during training, supervised or semi-supervised training. A brief description of these methods is shown below :

- SSLP: Like the original LP, the first step of SSLP is the transformation of the set of training data, considering the combinations of labels, in order to create a meta-label t for each combination, such that $1 \leq t \leq |L|$. After the transformation of the data, the semi-supervised process is similar to the one used in SSBR;

In this paper, we would like to extend the work done in [25], incorporating semi-supervised learning in HMC problems. In order to do this, we will use the multi-label methods used in [25], SSLP and SSRA*k*EL, in the hierarchical context.

B. Semi-Supervised Hierarchical Multi-Label Methods

As mentioned previously, the top-down approach is widely used for treating of HMC problems. In this approach a tree of classifiers is built, through the training of one or more classifiers at each hierarchy level. Thus, at the root node of the hierarchy, a classifier is trained with all training instances. In the following levels, for each class, a classifier is trained using only the subset of instances belonging to classes of the corresponding sub-tree.

The adaptation of the self-training method for multi-label problems is straightforward, allowing more than one label to be set to 1. In this case, a threshold is defined (usually 0.5) and label values higher than this threshold is set to 1 and values lower than this threshold is set to 0. In addition, its adaptation to hierarchical problems with the top-down approach is also straightforward, using in each node of the tree the semi-supervised approach to label the instances of the corresponding classes. Therefore, the adaptation of the self-training method for hierarchical multi-label problems uses the multi-label adaptation in each node of the tree used in the top-down approach of the hierarchical methods.

In [22], it is described and evaluated one hierarchical method based on the aforementioned top-down approach, called HMC-LP (Hierarchical Multi-label Classification with Label Powerset). The HMC-LP is an hierarchical adaptation of LP (Label Powerset). For each example, this method combines all the classes marked as an example, at a specific level of the hierarchical structure, in a single new class (meta-class). After the combination of classes, the original HMC problem is transformed into a hierarchical single-label problem, and top-down approach is used in training and testing phases of the method. In this paper, we will use an adaptation of this method, using the semi-supervised learning as a pre-processing phase.

In this work, we also propose and evaluate one adaptation of HMC-LP, called HMC-RA*k*EL (Hierarchical Multi-label Classification with Random *k*-Labelsets). The only difference between the original HMC-LP and HMC-RA*k*EL is that instead of using the original LP classifiers, as in HMC-LP, the HMC-RA*k*EL uses the multi-label classification method called RA*k*EL .

Additionally, this paper proposes two algorithms which combine the top-down approach with the semi-supervised learning strategy, which are:

- HMC-SSLP (Hierarchical Multi-label Classification using Semi-Supervised Label Powerset): it is a semi-supervised version of HMC-LP [22], in which the original LP is replaced by SSLP [25]. In this case, a semi-supervised learning is perform to label all unlabelled instances and, after that, the HMC procedure (HMC-LP) is performed;
- HMC-SSRA*k*EL (Hierarchical Multi-label Classification using Semi-Supervised Random *k*-Labelsets): It is a variation of HMC-RA*k*EL. The only difference between the original HMC-RA*k*EL and this proposed extension is that instead of using the original RA*k*EL, as in HMC-RA*k*EL, it is used in the semi-supervised hierarchical multi-label classification SSRA*k*EL in this proposed method. As the preivous item, in the proposed method, a semi-supervised learning is perform to label all unlabelled instances and, after that, the HMC procedure (HMC-RA*k*EL) is performed;

In all algorithms proposed in this paper, the classifiers are trained according to the hierarchical structure of the problem. In this case, each level of hierarchy is treated as a multi-label problem, which is similar to the idea described in [22].

V. Experimental Methodology

In order to investigate the feasibility of the proposed methods, an empirical analysis is performed, comparing the proposed methods with their corresponding supervised versions. For the supervised versions, the whole labelled datasets are considered, while the datasets are divided into labelled and unlabelled instances in the semi-supervised methods, always maintaining the class distribution of the original datasets. The main aim of this analysis is to observe whether the semi-supervised methods proposed in this paper have similar performance of the corresponding supervised versions. In this case, we could use the semi-supervised learning methods without deteriorating the performance of the hierarchical multi-label methods.

For the use of semi-supervised classification methods, we need to define two additional parameters, which are: the initial percentage of labelled instances and the percentage of instances to be labelled in each iteration. We performed an initial empirical analysis and we used a percentage value of the initial labelled instances of 50%. At each iteration, the percentage of the unlabelled instances to be labelled is 17%. In this case, we will have six iterations to label all the unlabelled data. An initial analysis has shown that this is an interesting value of instances to be labelled at each step. Finally, the choice of the instances to be labelled is made at random.

All hierarchical multi-label classification methods and supervised learning algorithms used in this work are implementations of the Weka-based package of Java classes for multi-label classification, called Mulan [26]. This package includes implementations of multi-label classification methods such as LP and RAKEL and hierarchical multi-label classification method HMC. The implementations of semi-supervised methods (SSLP and SSRAkEL) were obtained from adjustments made in Mulan, changing the training phase to be used in a semi-supervised strategy.

The empirical analysis was conducted using the 10-fold cross-validation methodology. Thus, all results presented in this paper refer to the mean over 10 different test sets. An initial investigation was conducted in order to define the parameter values used in supervised learning algorithms.

In order to compare the obtained results of the different learning methods, a statistical test is applied. The t-test was used for pairwise comparisons. In other words, it compares two samples (set of results). In this paper, we will compare the proposed methods with their corresponding supervised versions. For this test, the confidence level is 95% ($\alpha = 0.05$).

A. Evaluation Measures

Unlike the single-label problems, which the classification of an instance is correct or incorrect, in a multi-label problem, a classification of an instance may be partially correct and partially incorrect. This can happen when a classifier correctly assigns an instance to one of the labels it belongs to, but it does not assign it to all labels it belongs to. In addition, a classifier could assign an instance to one or more labels it does not belong [1]. In this sense, the evaluation of multi-label classifiers requires the use of different evaluation measures from those used in single-label problems.

Several measures have been proposed in the literature for the evaluation of multi-label classifiers. According to [14], these measures can be broadly categorized in two groups: bipartition-based and ranking-based. Some of the bipartition-based measures, called example-based-measures, evaluate bipartition over all examples of the evaluation dataset. Furthermore, the ranking-based measures evaluate rankings with respect to the ground truth of multi-label dataset. In this paper six multi-label measures are used: three bipartition-based (Hamming Loss (HL), F-Measure (FM) and Accuracy (Acc)) and three ranking-based (One-Error (1-Err), Average Precision (AvPre) and Ranking Loss (RL)).

The evaluation of hierarchical multi-label classifiers is still a widely debated issue. Although some works assesses the performance of hierarchical multi-label classification, we are still missing the proposal of well defined measures. In this paper, one of the most widely used measures is applied, called *hierarchical Loss* [27]. The hierarchical loss function is used to measure the discrepancy between the set of predicted labels and the set true label. The leading idea underlying our hierarchical loss function is: if a parent class has been predicted wrongly, then errors in the children should not be taken into account [27]. Regarding that, this measure is calculated only from the second hierarchical level on.

B. Datasets

The datasets used in this paper are related to gene functions of the *Saccharomyces cerevisiae* fungus, often used in the fermentation of sugar for the production of ethanol, and also in the fermentation of wheat and barley for the production of alcoholic beverages. It is one of Biology′s classic model organisms, and it has been subject of intensive study for years [20]. We selected these datasets, that are from the same domain, because they are, to the best of our knowledge, the only public datasets available for hierarchical multi-label classification. In addition, these datasets have been widely used in several studies involving HMC methods.

The datasets are structured as a tree, similar to the scheme proposed in FunCat, that was developed by MIPS [28]. They are public datasets [1]. The Funcat annotation scheme consists of 28 main categories that cover fields such as cellular transport, metabolism and cellular communication. Its hierarchy is structured as a tree, with up to six levels deep and a total of 1362 functional classes.

In [25], we performed an extensive investigation with these datasets, involving all six levels of hierarchy. However, we noticed that the there is a high computation cost involved and the pattern of behaviour of the second level of hierarchy is very similar to the following ones. Therefore, for simplicity reasons, the investigation performed in this work uses only the first two levels of the hierarchy.

Table I shows the main characteristics of the used datasets. This table describes the number of instance (NumIn) and the amount of numeric attributes (NUM) and categorical one (CAT) of the datasets. In addition, L1 represents the number of labels in the first level of hierarchy and L2 presents the number of labels in the second level of hierarchy. These datasets can be considered as small, but we decided to use them because the use of semi-supervised learning is more difficult when we have a small number of instance to train and label the unlabelled ones.

TABLE I
DESCRIPTION OF THE USED DATASETS

Dataset	NumIn	Attribute		Labels	
		NUM	CAT	L1	L2
Cellcycle	848	77	0	18	80
Church	844	27	0	18	80
Derisi	842	63	0	18	80
Eisen	529	79	0	18	76
Pheno	353	69	0	18	74

In the empirical analysis, we divided the hierarchical multi-label classification methods into two groups: supervised methods (HMC-LP and HMC-RA*k*EL) and semi-supervised methods (HMC-SSLP and HMC-SSRA*k*EL). For each method of each group, we applied five base classifiers, which are: *k*-Nearest Neighbour (*k*-NN), Decision Tree (DT), Support Vector Machine (SVM), Nave Bayesian (NB) and Repeated Incremental Pruning to Produce Error Reduction (RIPPER). These specific classifiers were chosen for being very distinct in their classification criteria, in this way, performing a broader and wider search in the databases. In this paper, we represent the average of all these five base classifiers.

[1]available at http://www.cs.kuleuven.be/dtai/clus/hmcdatasets.html

VI. Experimental Results

Tables II and III illustrate the results of HMC-LP-based and HMC-RAk*E*L-based methods, respectively. Each table presents the results obtained by hierarchical multi-label supervised method and hierarchical multi-label semi-supervised method in the first and second hierarchy levels. As already mentioned, the results are presented using seven different measures evaluation, where three are bipartition-based multi-label measures, three are ranking-based multi-label measures and the remaining one is a hierarchical measure. In these tables, the symbols ↓ and ↑ define the expected behavior of the evaluation metrics, with ↓ meaning that the lowest values are the best ones and ↑ meaning that the highest values are the best ones. The best value at each level is bold (the best of the first and third columns and the better of the second and fourth columns), for each measure. The statistical test compared the results of the two hierarchical multi-label classification methods, in a two-by-two basis. The results that showed no statistical superiority of supervised approach have shaded cells in these tables. In the following sections, the experimental results obtained by each HMC method will be described in more details.

A. HMC-LP-based Methods

Table II presents the experimental results obtained by the HMC-LP-based methods using the supervised learning approach (HMC-LP) and the semi-supervised learning approach (HMC-SSLP). In analysing Table II, on the first hierarchical level (Level 1 columns), it is possible to observe that the HMC-LP obtained superior results (bold numbers in Table II) compared to HMC-SSLP in 80% (24 out of 30) of the cases, while the HMC-SSLP obtained superior results compared to HMC-LP in 13.33% (4 out of 30) of the cases. When applying the t-test (shaded cells in Table II), it was verified that the HMC-LP and HMC-SSLP have similar performance in almost all cases (26 out of 30 cases), and these results are represented by shades cells. Therefore, on the first hierarchical level, it is possible to observe that the semi-supervised strategy presented results similar to its corresponding supervised version, from the statistical point of view, in almost 86.67% (26 out of 30) of the cases.

On the second hierarchical level (Level 2 columns), it is possible to observe that that the HMC-LP obtained superior results compared to HMC-SSLP in 56.67% (17 out of 30) of the cases, while the HMC-SSLP obtained superior results compared to HMC-LP in 43.33% (13 out of 30) of the cases. When applying the t-test (shaded cells in Table II), MC-LP and HMC-SSLP have similar performance in most of cases (18 out of 30 cases). On the other hand, it was verified that HMC-LP showed itself statistically superior compared to HMC-SSLP in 30% (9 out of 30) of the cases, while the HMC-SSLP showed itself statistically superior compared to HMC-LP in 10% (3 out of 30) of the cases.

In analysing the used hierarchical measure, the HMC-LP presented superior results compared to HMC-SSLP in 100% (5 out of 5) of the cases. However, in none of the cases it was verified statistical relevance in the differences presented between the HMC-LP and the HMC-SSLP.

TABLE II
Results of HMC-LP-based Methods.

	HMC-LP		HMC-SSLP	
	Level 1	Level 2	Level 1	Level 2
Cellcycle				
HL ↓	0.15±0.01	**0.06±0.00**	**0.15±0.01**	0.07±0.01
FM ↑	**0.35±0.40**	**0.24±0.30**	0.32±0.23	0.19±0.04
Acc ↑	**0.24±0.03**	**0.16±0.03**	0.21±0.03	0.12±0.01
1-Err ↓	**0.61±0.03**	**0.75±0.05**	0.62±0.03	0.82±0.04
AvPrec ↑	**0.38±0.03**	**0.19±0.03**	0.35±0.02	0.16±0.01
RL ↓	**0.49±0.03**	**0.51±0.02**	0.52±0.02	0.52±0.02
HiLoss ↓	-	**3.29±0.24**	-	3.31±0.21
Church				
HL ↓	0.15±0.01	**0.06±0.00**	**0.15±0.01**	0.06±0.00
FM ↑	**0.33±0.35**	**0.23±0.34**	0.31±0.39	0.18±0.04
Acc ↑	**0.19±0.03**	0.10±0.02	0.18±0.04	**0.10±0.01**
1-Err ↓	**0.65±0.05**	0.91±0.02	0.66±0.05	**0.87±0.01**
AvPrec ↑	0.33±0.03	0.12±0.02	0.33±0.03	**0.13±0.02**
RL ↓	0.54±0.03	0.59±0.01	0.54±0.03	**0.57±0.01**
HiLoss ↓	-	**3.28±0.13**	-	3.32±0.15
Derisi				
HL ↓	0.15±0.01	**0.06±0.00**	**0.15±0.01**	0.06±0.01
FM ↑	0.22±0.28	**0.12±0.15**	**0.24±0.24**	0.11±0.10
Acc ↑	**0.20±0.03**	0.11±0.02	0.19±0.04	**0.11±0.00**
1-Err ↓	**0.63±0.06**	0.85±0.04	0.64±0.07	**0.85±0.02**
AvPrec ↑	**0.34±0.03**	0.15±0.03	0.34±0.04	**0.15±0.01**
RL ↓	**0.53±0.03**	0.56±0.02	0.53±0.04	**0.56±0.01**
HiLoss ↓	-	**3.25±0.12**	-	3.30±0.12
Eisen				
HL ↓	**0.15±0.01**	0.07±0.01	0.16±0.02	**0.07±0.00**
FM ↑	**0.35±0.33**	**0.30±0.27**	0.30±0.34	0.24±0.08
Acc ↑	**0.25±0.04**	**0.19±0.03**	0.22±0.02	0.16±0.02
1-Err ↓	**0.60±0.05**	**0.69±0.05**	0.64±0.06	0.74±0.03
AvPrec ↑	**0.38±0.03**	**0.23±0.02**	0.35±0.03	0.20±0.02
RL ↓	**0.50±0.04**	**0.54±0.02**	0.53±0.03	0.56±0.01
HiLoss ↓	-	**3.25±0.20**	-	3.34±0.28
Pheno				
HL ↓	**0.15±0.01**	0.07±0.00	0.16±0.02	**0.07±0.00**
FM ↑	**0.45±0.35**	**0.26±0.20**	0.34±0.31	0.18±0.05
Acc ↑	**0.20±0.04**	**0.12±0.03**	0.18±0.03	0.10±0.02
1-Err ↓	**0.63±0.06**	**0.70±0.06**	0.68±0.11	0.85±0.04
AvPrec ↑	**0.34±0.05**	0.14±0.02	0.32±0.03	**0.15±0.01**
RL ↓	**0.53±0.05**	0.59±0.05	0.55±0.04	**0.53±0.02**
HiLoss ↓	-	**3.25±0.20**	-	3.44±0.21

Consequently, on the second hierarchical level, it is possible to observe that the semi-supervised strategy presented results similar to its corresponding supervised version, from the statistical point of view, in almost 74.29% (26 out of 35) of the cases.

From a more general perspective, we can observe from Table II that the semi-supervised strategy presented results similar to its corresponding supervised version, from the statistical point of view, in almost 80% (52 out of 65) of the cases. Based on these empirical results, we can state that an automatic labelling process can be used in the HMC-LP methods without deteriorating its performance, from a statistical point of view. It is important to emphasize that the LP method takes into consideration the correlation among labels in a multi-label problem.

B. *HMC-RAkEL-based Methods*

Table III illustrates the experimental results obtained by the HMC-RA*k*EL-based methods using the supervised learning approach (HMC-RA*k*EL) and semi-supervised learning approach (HMC-SSRA*k*EL). In analysing Table III, on the first hierarchical level, it is possible to observe that the HMC-RA*k*EL obtained superior results compared to HMC-SSRA*k*EL in 53.33% (16 out of 30) of the cases, while the HMC-SSRA*k*EL obtained superior results compared to HMC-RA*k*EL in 46.67% (13 out of 30) of the cases. By applying the t-test (shaded cells in Table III), it is possible to observe that the semi-supervised strategy presented results similar to its corresponding supervised version, from the statistical point of view, in 90% (29 out of 30) of the cases. In contrast, it was verified that the HMC-RA*k*EL showed itself statistically superior in only 3.33% (1 out of 30) of the cases, while the HMC-SSRA*k*EL showed statistically superior in 6.67% (2 out of 30) of the cases.

On the second hierarchical level, it is possible to observe that HMC-RA*k*EL obtained superior results compared to HMC-SSRA*k*EL in 43.33% (13 out of 30) of the cases, while the HMC-SSRA*k*EL obtained superior results compared to HMC-RA*k*EL in 56.67% (17 out of 30) of the cases. As a result of the t-test (shaded cells), it was verified that both methods (HMC-RA*k*EL and HMC-SSRA*k*EL) are similar in 56.67% (17 out of 30) of the cases. In contrast, the HMC-RA*k*EL showed to be statistically superior in 30% (9 out of 30) of the cases, while the HMC-SSRA*k*EL showed itself statistically superior in 13.33% (4 out of 30) of the cases.

Concerning the hierarchical measure, the HMC-RA*k*EL presented superior results compared to HMC-SSRA*k*EL in only 40% (2 out of 5) of the cases, while the HMC-SSRA*k*EL obtained superior results compared to HMC-RA*k*EL in 60% (3 out of 5) of the cases. When applying the t-test, it was verified that the HMC-RA*k*EL showed to be statistically superior in 20% (1 out of 5) of the cases, while the HMC-SSRA*k*EL did not present statistical superiority in any of the cases. Thereby, on the second hierarchical level, it is possible to observe that the semi-supervised strategy presented results similar to its corresponding supervised version, from the statistical point of view, in almost 71.43% (25 out of 35) of the cases.

From a more general perspective, as in the previous section, in the majority of cases, the performance of the supervised and semi-supervised versions of HMC-SSRA*k*EL are similar, from a statistical point of view. Once again, we can state that the an automatic labelling process can be used in the HMC-RA*k*EL methods without deteriorating its performance, from a statistical point of view. It is important to emphasize that the best results were obtained by RA*k*EL-based, when compared with LP-based methods. As mentioned in Section II-A, RA*k*EL is an extension of LP and the use of this more elaborated method had a positive impact in the functioning of the semi-supervised approach.

TABLE III
RESULTS OF HMC-RA*k*EL-BASED METHODS.

	HMC-RA*k*EL		HMC-SSRA*k*EL	
	Level 1	Level 2	Level 1	Level 2
		Cellcycle		
HL ↓	**0.13±0.01**	0.05±0.00	0.13±0.01	**0.05±0.00**
FM ↑	**0.29±0.31**	**0.03±0.01**	0.26±0.06	0.01±0.01
Acc ↑	**0.20±0.03**	**0.01±0.00**	0.19±0.01	0.01±0.00
1-Err ↓	**0.73±0.05**	**0.94±0.03**	0.74±0.02	0.95±0.01
AvPrec ↑	0.33±0.03	0.08±0.02	**0.33±0.01**	**0.08±0.00**
RL ↓	0.51±0.03	0.61±0.03	**0.50±0.02**	**0.61±0.00**
HiLoss ↓	-	2.39±0.18	-	**2.39±0.00**
		Church		
HL ↓	0.13±0.01	0.05±0.00	**0.13±0.00**	**0.05±0.00**
FM ↑	0.16±0.20	**0.09±0.21**	**0.18±0.03**	0.05±0.02
Acc ↑	0.16±0.02	0.02±0.01	**0.16±0.01**	**0.03±0.00**
1-Err ↓	0.85±0.02	0.95±0.02	**0.83±0.01**	**0.91±0.01**
AvPrec ↑	0.26±0.01	0.08±0.01	**0.27±0.00**	**0.09±0.00**
RL ↓	0.52±0.03	0.6±0.02	**0.51±0.01**	**0.60±0.00**
HiLoss ↓	-	2.39±0.12	-	**2.39±0.00**
		Derisi		
HL ↓	**0.13±0.01**	0.05±0.00	0.13±0.01	**0.05±0.00**
FM ↑	**0.22±0.28**	**0.14±0.15**	0.19±0.06	0.01±0.02
Acc ↑	0.17±0.01	**0.11±0.01**	**0.17±0.01**	0.03±0.00
1-Err ↓	0.78±0.06	**0.81±0.03**	**0.77±0.03**	0.90±0.01
AvPrec ↑	0.30±0.04	**0.16±0.02**	**0.31±0.01**	0.09±0.00
RL ↓	**0.47±0.02**	**0.46±0.01**	0.48±0.01	0.60±0.00
HiLoss ↓	-	**2.39±0.12**	-	2.40±0.00
		Eisen		
HL ↓	**0.13±0.01**	0.06±0.00	0.13±0.01	**0.06±0.00**
FM ↑	0.28±0.24	0.02±0.07	**0.31±0.03**	**0.02±0.03**
Acc ↑	**0.24±0.03**	**0.02±0.01**	0.22±0.02	0.01±0.01
1-Err ↓	**0.66±0.09**	0.94±0.03	0.69±0.03	**0.93±0.01**
AvPrec ↑	**0.39±0.04**	**0.09±0.01**	0.36±0.01	0.08±0.00
RL ↓	**0.46±0.04**	0.61±0.03	0.48±0.02	**0.61±0.00**
HiLoss ↓	-	2.48±0.19	-	**2.48±0.00**
		Pheno		
HL ↓	0.14±0.01	**0.06±0.01**	**0.14±0.00**	0.07±0.01
FM ↑	0.22±0.35	0.16±0.22	**0.23±0.11**	**0.20±0.04**
Acc ↑	**0.12±0.02**	0.08±0.02	0.11±0.01	**0.11±0.01**
1-Err ↓	**0.78±0.04**	0.84±0.05	0.79±0.01	**0.79±0.07**
AvPrec ↑	**0.28±0.03**	0.13±0.02	0.27±0.00	**0.13±0.01**
RL ↓	**0.58±0.05**	**0.59±0.02**	0.59±0.00	0.60±0.01
HiLoss ↓	-	**2.53±0.20**	-	3.45±0.05

VII. FINAL REMARKS

This paper proposed and evaluated the use of semi-supervised learning in two hierarchical multi-label classification methods. Once these methods are variations of existing supervised methods, a comparative analysis was performed, comparing both supervised and semi-supervised versions of the hierarchical multi-label classification methods. Finally, these methods were analysed using six different multi-label evaluation metrics and one hierarchical evaluation metrics.

We performed an exhaustive set of experiments, using five datasets applied to two groups of methods (supervised and semi-supervised) and each method used five existing classification methods. For simplicity reasons, we presented the results in the first and second level of hierarchy. As a result of the empirical analysis, from a more general perspective, it is possible to observe that the semi-supervised approach presented results statistically equal or superior compared to the supervised approach in 81.54% (106 out of 130) of the cases. In addition, the statistical test showed that the use of semi-supervised learning had a positive effect in almost 91.67% of the analysed cases (55 out of 60) in the first hierarchical level and 78.46% of the analysed cases (51 out of 65) in the second hierarchical level, either having better or similar performance than the corresponding supervised versions. In some cases they even have a superior performance, from a statistical point of view. Moreover, when analysing separately the hierarchical loss measure, it is possible to observe that the semi-supervised approach presented results statistically equal or superior compared to the supervised approach in 90% (9 out of 10) of the cases.

One interesting aspect of this analysis was that the semi-supervised methods had their best results for HMC-SSRA*k*EL, followed by HMC-SSLP. We believe that the use of an ensemble, used in HMC-SSRA*k*EL, had a positive effect in the labelling process.

In summarizing, it is possible to state that, in a general way, the process of automatic attribution of labels, used on the semi-supervised approach, did not deteriorate significantly the performance of the classifiers, showing that its use is favourable.

REFERENCES

[1] R. Cerri, R. R. O. da Silva, and A. C. P. L. F. de Carvalho, "Comparing methods for multilabel classification of proteins using machine learning techniques," in *BSB*, 2009, pp. 109–120.

[2] J. Metz and A. Freitas, "Extending hierarchical classification with semi-supervised learning," in *UK Workshop on Computational Intelligence*. United Kindom: UK, 2009, pp. 1–6.

[3] O. Chapelle, *Introduction to Semi-supervised Learning*. MIT Press, 2006, ch. 1, pp. 1–12.

[4] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognition*, vol. 37, no. 9, pp. 1757–1771, 2004.

[5] M. Zhang and Z. Zhou, "Ml-knn: A lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, July 2007.

[6] G. J. Qi, X. S. Hua, Y. Rui, J. Tang, T. Mei, and H. J. Zhang, "Correlative multi-label video annotation," in *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*. New York, NY, USA: ACM, 2007, pp. 17–26.

[7] C. G. M. Snoek, M. Worring, J. C. van Gemert, J. M. Geusebroek, and A. W. M. Smeulders, "The challenge problem for automated detection of 101 semantic concepts in multimedia," in *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*. New York, NY, USA: ACM, 2006, pp. 421–430.

[8] A. Clare and R. D. King, "Knowledge discovery in multi-label phenotype data." in *PKDD*, ser. Lecture Notes in Computer Science, L. D. Raedt and A. Siebes, Eds., vol. 2168. Freiburg, Germany: Springer, 2001, pp. 42–53.

[9] H. Blockeel, L. Schietgat, J. Struyf, S. Dzeroski, and A. Clare, "Decision trees for hierarchical multilabel classification: A case study in functional genomics," *Machine Learning*, pp. 18–29, 2006.

[10] Z. Barutcuoglu, R. E. Schapire, and O. G. Troyanskaya, "Hierarchical multi-label prediction of gene function," *Bioinformatics*, vol. 22, no. 7, pp. 830–836, 2006.

[11] A. Wieczorkowska, P. Synak, and Z. Raś, "Multi-label classification of emotions in music." Berlin, Heidelberg: Springer Berlin / Heidelberg, 2006, pp. 307–315.

[12] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, "Multilabel classification of music into emotions," in *Proc. 9th International Conference on Music Information Retrieval (ISMIR 2008), Philadelphia, PA, USA, 2008*. Germany: Springer, 2008.

[13] Y. Zhang, S. Burer, and W. N. Street, "Ensemble pruning via semi-definite programming," *J. Mach. Learn. Res.*, vol. 7, pp. 1315–1338, December 2006.

[14] G. Tsoumakas and I. Vlahavas, "Random k-labelsets: An ensemble method for multilabel classification," in *Proceedings of the 18th European Conference on Machine Learning (ECML 2007)*. Warsaw, Poland: Springer, September 2007, pp. 406–417.

[15] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," 2009, pp. 254–269.

[16] A. Freitas and A. C. de Carvalho, *A Tutorial on Hierarchical Classification with Applications in Bioinformatics*. UK: Idea Group, 2007, ch. VII.

[17] E. P. Costa, A. C. Lorena, Carvalho, and A. A. Freitas, "A review of performance evaluation measures for hierarchical classifiers," in *2007 AAAI Workshop, Vancouver*. AAAI Press, 2007.

[18] A. Secker, M. N. Davies, A. A. Freitas, J. Timmis, M. Mendao, and D. R. Flower, "An experimental comparison of classification algorithms for the hierarchical prediction of protein function," in *3rd UK Data Mining and Knowledge Discovery Symposium (UKKDD 2007)*. UK: UKKDD, 2007, pp. 13–18.

[19] J. Rousu, C. Saunders, S. Szedmak, and J. Shawe-Taylor, "Kernel-based learning of hierarchical multilabel classification models," *Journal of Machine Learning Research*, vol. 7, pp. 1601–1626, 2006.

[20] C. Vens, J. Struyf, L. Schietgat, S. Dzeroski, and H. Blockeel, "Decision trees for hierarchical multi-label classification," *Mach. Learn.*, vol. 73, pp. 185–214, November 2008.

[21] B. Stenger, A. Thayananthan, P. H. S. Torr, and R. Cipolla, "Estimating 3d hand pose using hierarchical multi-label classification," *Image Vision Comput.*, vol. 25, pp. 1885–1894, December 2007.

[22] R. Cerri and A. C. P. L. F. de Carvalho, "Comparing local and global hierarchical multilabel classification methods using decision trees," in *Proceedings of the Brazilian Symposium on Bioinformatics, Lecture Notes on Bioinformatics*. Porto Alegre, Brazil: Springer-Verlag, 2009, pp. . 109–120.

[23] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training." New York, NY, USA: Morgan Kaufmann Publishers, 1998, pp. 92–100.

[24] C. Rosenberg, M. Hebert, and H. Schneiderman, "Semi-supervised self-training of object detection models," in *Seventh IEEE Workshop on Applications of Computer Vision*. Breckenridge, CO, USA: IEEE, 2005, pp. 29–36.

[25] A. M. Santos and A. M. P. Canuto, "Using semi-supervised learning in multi-label classification problems," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, june 2012, pp. 1 –8.

[26] E. S. G. Tsoumakas, J. Vilcek and I. Vlahavas, "Mulan: A java library for multi-label learning," *Journal of Machine Learning Research*, 2010.

[27] N. Cesa-Bianchi, C. Gentile, and L. Zaniboni, "Hierarchical classification: combining bayes with svm," in *Proceedings of the 23rd international conference on Machine learning*, ser. ICML 06. New York, NY, USA: ACM, 2006, pp. 177–184.

[28] H. W. Mewes, D. Frishman, U. Gildener, G. Mannhaupt, K. Mayer, M. Mokrejs, B. Morgenstern, M. Minsterktter, S. Rudd, and B. Weil, "Mips: a database for genomes and protein sequences," *Nucleic Acids Res*, vol. 30, pp. 31–34, 2002.