

Multi-Label Classification using an Ontology

Yaya TRAORE¹, Didier BASSOLE³

Mathematics and Informatics Laboratory (LAMI)
University Joseph KI-ZERBO (UJKZ)
Ouagadougou, BURKINA FASO

Sadouanou MALO², Abdoulaye SERE⁴

Mathematics and Informatics Laboratory (LAMI)
University Nazi Boni (UNB)
Bobo-Dioulasso, BURKINA FASO

Abstract—During these last few years, the problem of multi-label classification (ML) has been studied in several domains, such as text categorization. Multi-label classification is a main challenging task because each instance can be assigned to multiple classes simultaneously. This paper studies the problem of Multi-label classification in the context of web pages categorization. The categories are defined in an ontology. Among the weakness of the multi-label classification methods, exist the number of positive and negative examples used to build the training dataset of a specific label. So the challenge comes from the huge number of labels combinations that grow exponentially. In this paper, we present an ontology-based Multi-label classification which exploit dependence between the labels. In addition, our approach uses the ontology to take into account relationships between labels and to give the selection of positive and negative examples in the learning phase. In the prediction phase, if a label is not predicted, the ontology is used to prune the set of descendant labels. The results of experimental evaluation show the effectiveness of our approaches.

Keywords—Multi-label classification (ML); Binary Relevance (BR); ontology; categorization; prediction

I. INTRODUCTION

In recent years, Multi-label classification (ML) has been studied in several domains, such as text categorization with its many applications on the World Wide Web. Multi-label classification is a main challenging task because each instance can be assigned to multiple classes simultaneously. This paper studies the problem of Multi-label classification in the context of web pages categorization. The categories are defined in an ontology. So the challenge comes from the huge number of labels combinations that grow exponentially. In text categorization, Multi-label (ML) classification problem [1], [12] consist in predicting the categories of a new instance according to its features. In our context, categories are looked upon as text labels. Binary Relevance (BR) [14] is the most intuitive method for multi-label classification. It decomposes the multi-label classification task into a set of binary classification tasks. This approach is easy to understand and learn. However, this approach has a deficiency. The semantic relations between the labels are ignored.

In order to use dependence between the labels to build the training dataset, we associate an ontology with ML method. An ontology [2] is used to present the domain knowledge. In this paper, we propose a novel approach that uses a method of ML based on ontology to predict the categories of a new web page. Experiments are implemented to evaluate the performance of

the proposed approach on the datasets of the uniprot¹ web site. The results of experimental evaluation indicate that the approach has a better performance.

The remainder of the paper is organized as follows: Section 2 presents Multi-label classification task for web page. Section 3 presents an external Ontology used to annotate the page of semantic web platform. Section 4 presents related work. In Section 5, we describe and give details on the proposed approach. Section 6 specifies primary experiment results. Finally, Section 7 ends with a conclusion and perspectives.

II. MULTI-LABEL CLASSIFICATION (ML)

According to [9], traditional classification tasks deal with assigning instances to a single label. Multi-label classification is an extension of traditional classification, where the task is to find the set of labels that an instance can belong. In this case, each instance may belong to many classes simultaneously. For the rest of paper, we apply a ML for web page categorization to illustrate our proposed method. A multi-label classification for web page in Semantic web platform deals with a situation where web page [19] can belong to more than one category. Semantic web platform is basically defined by a set of categories and pages. Each page is assigned to one or more categories and includes a set of tags. Formally, semantic web platform web is defined by:

- P is the finite set of web pages, let n the number of pages,
- Let T the finite set of tags and $R_T \subseteq P \times T$ a binary relation between P and T , let m the number of tags. We denote : $\forall i \in [0, n], \forall j \in [0, m], \forall P_i \in P, \forall T_j \in T, R_T(P_i, T_j) = t_{ij}$ with $t_{ij} = 1$ if the page $P_i \in P$ is tagged by $T_j \in T$ and 0 otherwise,
- C the finite set of categories and $R_C \subseteq P \times C$ a binary relation between P and C , let k the number of categories. We denote : $\forall i \in [0, n], \forall j \in [0, k], \forall P_i \in P, \forall C_j \in C, R_C(P_i, C_j) = c_{ij}$ with $c_{ij} = 1$ if the page $P_i \in P$ is categorized by $C_j \in C$ and 0 otherwise. We define the function g which allows to obtain all the pages associated with a category as follows:
- $g: C \rightarrow P$ such that, $\forall c \in C, g(c) = \{p / p \in P \text{ and } R_C(p, c) = 1\}$.

¹ <https://www.uniprot.org/>

TABLE. I. TRAINING DATA OF MULTI-LABEL CLASSIFICATION FOR WEB PAGE CLASSIFICATION

	T_1	...	T_m	C_1	...	C_k
P_1	t_{11}	...	t_{1m}	c_{11}	...	c_{1k}
...
P_n	t_{n1}	...	t_{nm}	c_{n1}	...	c_{nk}
P_{n+1}	$t_{(n+1)1}$...	$t_{(n+1)m}$?	?	?

Table I shows the training data for multi-label classification for web page classification.

There are multiple approaches to deal the Multi-Label classification problem. We distinguish two methods families proposed in [13]: transformation methods and adaptation methods. We focus on the transformation methods. The popular approach is Binary Relevance (BR) method. In this approach we can use k binary classifiers. Let h the classifier of ML. BR [14] decomposes the learning of h into a set of binary classification tasks where each single model h_l (using only the information of that particular label) is learned independently, and ignoring the information of all other labels. Hence, $h_l(x) = 1$, if the label l is predicted for the instance x . $h(x)$ gives the set of relevant labels predicted by h for the instance x . Thus, for the categorization of a new example x , BR [14] outputs the union of the labels predicted by the k classifiers.

Table II shows BR method of ML problem with training dataset for $n=3, m=3$ and $k=2$.

TABLE. II. TRANSFORMED DATASETS PRODUCED BY BINARY RELEVANCE (BR) METHOD

	T_1	T_2	T_3	C_1		T_1	T_2	T_3	C_2	
P_1	t_{11}	t_{12}	t_{13}	c_{11}		P_1	t_{11}	t_{12}	t_{13}	c_{12}
P_2	t_{21}	t_{22}	t_{23}	c_{21}		P_2	t_{21}	t_{22}	t_{23}	c_{22}
P_3	t_{31}	t_{32}	t_{33}	c_{31}		P_3	t_{31}	t_{32}	t_{33}	c_{32}

III. ONTOLOGY

Ontology [3], [2] represents the relevant concepts (classes) of a domain. Each concept is defined by a set of consensual terms that is not specific to an individual but accepted by a community of users. Specifically, all the defining terms are organized in hierarchy. In semantic web platform, the classes of ontology are used by the experts to annotate the new page created by the users. Uniprot.org web site is an example of semantic web platform where web pages are annotated by keywords (tags) and classes (categories) of gene ontology (GO).

Let Co the set of classes of an external ontology and HC the hierarchy of classes of the ontology. $\forall c \in Co$, we denote: $ch(c)$: the set of children classes of c in HC ; $desc(c)$: the set of descendant classes of c in HC ; $sib(c)$: the set of sibling classes of c in HC .

IV. RELATED WORK

The problem of multi-label classification has been studied in the work of [1], [4], [5], [12], [13], [14], [15], [17], [25], [26]. In our context, we apply the ML to predict the categories of a new web page created. In the literature, Multi-label classification approaches [13] can be divided into transformation methods and adaptation methods. An overview on multi-label classification algorithms is given in [13]. We focus on transformation method in this paper. Transformation methods [20] decompose the multilabel problem into a set of binary classification problems. The most popular method is called Binary (BR), which trains a independent binary classifier for each class (against the others). In this paper, we apply a Binary relevance method to categorize web page.

Many similar studies have been realized about web page categorization. The studies differed with the methods used and the use of different algorithms of machine learning [6], [7], [8], [10], [11]. In this paper, we propose a novel method that uses Multi-label Classification and an ontology to categorize a new web page in a web platform used to share knowledge from different communities.

BR [14] is a naturally multi-label classification approach. BR [22] builds one binary data set for each label that contains all examples of the original data set. Examples assigned to label are considered positive examples and the rest are considered to be negative examples. While BR has been used in many practical applications, according to [21], it has been widely criticized for its implicit assumption of labels independence which might be completely incorrect in the data.

BR+ algorithm [16], [23] is an extension of the BR algorithm where the relationship between labels are considered. The differences are its descriptor attributes, which merge all original attributes as well as all labels, except the label to be predicted.

Classifier Chains (CC) [5], [22] generates k binary classifiers. Each classifier incorporates the labels predicted by the previous classifiers as additional attributes. The label correlations are considered randomly.

In this paper, we propose to improve the transformation method of multilabel classification based on BR. So, we propose to use the relationship between labels to select the set of positive examples and negative examples used to build the training dataset of a label. In training dataset of a class cl [24], example assigned to cl are chosen as the positive examples, and other examples are selected as negative examples. An external ontology is used to select the set of positive and negative examples for each label (existing in an ontology) in the learning phase. Similarly, we improve the prediction phase by adding a pruning step if a label is not predicted. Our contribution is described in Section 5.

V. PROPOSED METHODOLOGY

In this section, we describe the methodology used to predict the categories of a new page. The proposed methodology contains two steps. Its main processes are given as follows. The first step is the learning phase. In this step, training data is divided into $|cl|$ subsets, where $|cl|$ means the total number of

labels. Then, $|cl|$ binary classifiers are built. The second step is the prediction phase. For a new page created and tagged, this step uses for each label cl , his binary classifier to predict if cl is affected.

A. Step 1: Learning Phase

The first phase is selecting sibling classes, descendant classes for each label by using the hierarchy of ontology and built the training set of cl . For each label cl , we use an ontology to select these siblings ($sib(cl)$) and descendant ($desc(cl)$) classes. For the training set of the category cl , examples assigned to cl or the descendants ($desc(cl)$) of cl are chosen as the positive examples, and other assigned to the siblings of cl are selected as negative examples. In the case that the class cl has no correspondent class in an ontology, positive examples of cl are calculated by $g(cl)$ and negative examples are the instances which are not selected by $g(cl)$. The set of positive examples $Tr^+(cl)$ and negative examples $Tr^-(cl)$ are given as follows:

If cl exist in HC then

- $Tr^+(cl) = g(cl) \cup g(desc(cl))$
- $Tr^-(cl) = g(sib(cl))$

Else

- $Tr^+(cl) = g(cl)$
- $Tr^-(cl) = P \setminus \{g(cl)\}$

The last phase of this step is the choice of the base classifier. For each label cl a base classifier [18] (for example: SVM, NaiveBayes, J48,...) is trained by using the training dataset generate for this label.

B. Step 2: prediction phase

This step uses the built classifiers to predict the label of a new example. For improving the prediction phase, if a label is not predicted, then these descendants are pruned in the list of candidates labels.

C. Proposed Algorithm

The proposed algorithm (Algorithm 1) takes as input, P, C, T, HC the hierarchy of classes of ontology and the base classifier, the new example. Ontology is used to select the descendant ($desc(cl)$), sibling ($sib(cl)$) labels of cl . The Algorithm generates all the set of positive examples and negative examples for the label cl (line 10).

For each example $p \in Tr(cl)$, it generates the rows (line 14) of the instance $p \in Tr(cl)$ of the training set. The algorithm traverses the set of labels and for each label, the corresponding classifier h_{cl} is invoked (line 26). For improving the prediction phase, if a class is not predicted, then these descendants (line 30) are pruned in the list of candidate labels. Finally, the algorithm returns the set of labels predicted (line 33) for a new example x .

Algorithm 1: ML using an ontology

Input:

P : set of examples, C : set of label,
 T : set of attributes,
 HC : Hierarchy classes of ontology,
classifier: base classifier (example :SVM)
 x : new example created and tagged

Output:

Y : classes predicted

Begin

```
1. For each  $cl$  in  $C$  do
2. // set of positive and negative examples
3.  $Tr^+(cl) = \emptyset$  If( $cl$  exist in  $HC$ )
4.  $Tr^+(cl) = g(cl) \cup g(desc(cl))$ 
5.  $Tr^-(cl) = g(sib(cl))$ 
6. Else
7.  $Tr^+(cl) = g(cl)$ 
8. //all instance without  $g(cl)$ 
9.  $Tr^-(cl) = P \setminus \{g(cl)\}$ 
10. End if
11.  $Tr^-(cl) = Tr^+(cl) \cup Tr^-(cl)$ 
12. For each example  $p \in Tr^-(cl)$  do
13. //set of attributes of a instance  $p$ 
14.  $x_p = \emptyset$ 
15. For each tag  $t \in T$  in  $BK$ 
16.  $x_p = x_p \cup \{R_T(p, t)\}$ 
17. End for
18. //build the training set of  $cl$ 
19.  $D(p, cl) = x_p \cup \{R_C(p, cl)\}$ 
20.  $InstancesTrain = D$ 
21. //build the binary classifier  $h_{cl}$  of  $cl$ 
22.  $h_{cl} = classifier.buildClassifier(InstancesTrain)$ 
23.  $h = h \cup \{h_{cl}\}$  // add  $h_{cl}$  to  $h$ 
24. End for
24.  $Y = \emptyset$ 
25. for each class  $cl \in C$  do
26. //  $v=1$  if the label  $cl$  is predict
27.  $v = h_{cl}(x)$ 
28. if( $v = 1$ ) then
29.  $Y = Y \cup \{cl\}$ 
30. Else
31. //delete the descendants of  $cl$ 
32.  $C = C \setminus desc(cl)$ 
33. end if
34. End for
35. // set of classes predicted of  $x$ 
36. return  $Y$ 
```

End

VI. EXPERIMENTS

A. Dataset and Experimental Setup

1) Platform: The experiments are implemented under macOS 10.13.4 (17E199), with Processor: Intel Core i5 @ 2,5 GHz and Memory RAM: 8 Go. The code is implemented in Java and used the weka library.

2) Dataset: As an application area we have chosen bioinformatics² in view of the fact that it is an important platform that has many pages annotated by some keywords and classes of GO (Gene Ontology). First we extract the pageID, Keywords and Gene ontology IDs to create the pages of web platform. Fig. 1 illustrates the whole process of collecting training data initial (a) and a fraction of Gene ontology (b) called hierarchy of categories *HC*.

The characteristics of the experimental dataset are summarized as following:

- pages number used for training data : 5199
- pages number used for Testing data: 3985
- Categories number : 630
- Tag number : 2815

The characteristics of ontology used: the number of classes used in hierarchy of categories *HC* is 630.

B. Evaluation Metrics

The metrics precision (P), recall (R) and Fmeasure (F1) are proposed to evaluate our method. The precision, recall and F1 for the example *i* are defined as:

$$P_i = \frac{|Y_i \cap Z_i|}{|Y_i|}, R_i = \frac{|Y_i \cap Z_i|}{|Z_i|}, F1_i = \frac{2 * P_i * R_i}{P_i + R_i}$$

where, for an example *i*, Y_i is the set containing all of the predicted classes, and Z_i the set including all of its true classes. To combine the performance of all instances to evaluate the results measured on a dataset with *n* instances labeled, we use the macro-averaging of the precision P, recall R and F1:

$$P = \frac{\sum_{i=1}^n P_i}{n}, R = \frac{\sum_{i=1}^n R_i}{n}, F1 = \frac{\sum_{i=1}^n F1_i}{n}$$

C. Experiments Results

In experimental evaluation, we use some base classifiers (SVM, NaiveBayes, J48) to observe the performance of the method proposed to place the pages in the good categories. We compare the proposed method with BR method.

Table III gives the experimental results of proposed method vs BR with some base classifiers.

Table IV gives the experimental results of the proposed method vs BR method from five categories. The base classifiers SVM is used.

D. Analysis and Discussion

The results (Table III) show: (i) with the classifier SVM, The macro-averaging *F1* for proposed method is 76.96% against 26.29% for BR; (ii) With the classifier NaiveBayes, the macro-averaging *F1* for proposed method is 65.08% against 3.88% for BR; (iii) With the classifier J48, the macro-averaging *F1* for Proposed method is 74.25% against 24.22% for BR. These results show that the proposed method gives the good performance than BR with use these base classifiers. For these three base classifiers, proposed method +SVM gives the better performance. These good performances are linked to the ontology used and by exploiting the dependencies between the categories.

Furthermore, we conducted comparisons with BR using the base classifier SVM on five categories. The results (Table IV) show that our approach has the best performance than BR for the five categories used. The Precision of the category GO0008270 (95.15%) is better with BR than the Precision of proposed method (92.36%). This shows that our selection method presents limits in the choice of positive and negative examples. The macro-averaging *F1* for each category used is better for proposed method than macro-averaging *F1* of BR. This best performance is due firstly to the use of the ontology and the relation between category. Also, our method uses a method of selecting positive and negative instances in the building of the training set for each category while the BR method does not use it in the construction of the model. Thus, we can come to a conclusion that our method proposed improves the performance of the method BR of Multi-label classification.

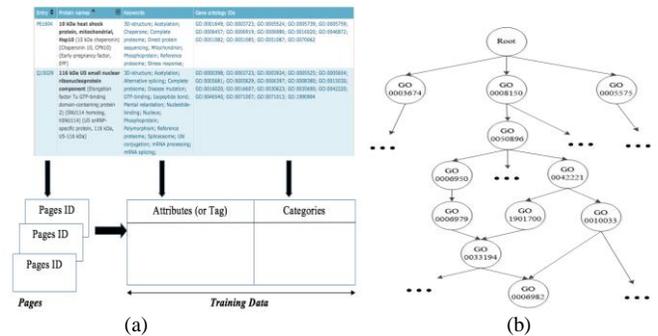


Fig. 1. (a) The Whole Process of Collecting Training Data (b) a Fraction of the Gene Ontology.

TABLE. III. THE EXPERIMENTAL RESULTS OF PROPOSED METHOD VS BR WITH SOME BASE CLASSIFIERS

Base classifiers	Proposed method			BR		
	P	R	F1	P	R	F1
SVM	76.85%	77.67%	76.96%	26.67%	25.92%	26.29%
NaiveBayes	61.25%	69.94%	65.08%	1.99%	74.20%	3.88%
J48	74.81%	73.76%	74.25%	24.83%	23.63%	24.22%

TABLE. IV. BR METHOD VS PROPOSED METHOD FOR FIVE CATEGORIES

Categories	BR			Proposed method		
	P	R	F1	P	R	F1
GO0003700	72.09%	38.59%	50.27%	97.54%	98.76%	98.15%

² https://www.uniprot.org/

GO0016818	81.58%	72.66%	76.86%	98.45%	99.22%	98.83%
GO0008270	95.15%	63.64%	76.27%	92.36%	94.12%	93.23%
GO0003735	75.00%	2.80%	5.398%	99.03%	95.33%	97.14%
GO0003773	66.67%	38.46%	48.78%	96.15%	96.15%	96.15%

VII. CONCLUSION

In this paper, we propose a novel method of Multi-label Classification (ML) using an ontology. In the learning phase, the ontology is used to select the set of positive and negative instances for learning and building the training set. In the prediction phase, ontology is used to delete the descendants of the category not predicted. Experiment results on a data downloaded from the biological database Uniprot (www.uniprot.org) show that proposed method improve BR method of ML. The good performance of proposed method demonstrates that our proposal is expected to be a potential approach to solve the automatic web page categorization in a semantic web platform.

In the future work, first we compare our approach with BR+ and CC methods of transformation methods of multi-label classification. Next we will test and compare our approach in more methods of Multi-label classification and more datasets. Also, we will use our approach with the selection methods of features and study their impact in web page categorization.

REFERENCES

[1] Zhang, M., Zhou, Z.: A review on multi-label learning algorithms. IEEE Transactions on Knowledge and Data Engineering (Volume: 26, Issue:8, Aug. 2014).

[2] Gomez Perez A. Ontological engineering : A state of the art. Expert update 2. Technique et Science Informatiques 28, page 81233–126,1999.

[3] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. in : International Journal of Human-Computer Studies - Special issue : the role of formal ontology in the information technology, V, pages 907 – 928, 1993.

[4] Madjarov G., Kocev D., Gjorgjevikj D., and Dzeroski S. An extensive experimental comparison of methods for multi-label learning. Pattern recognition, 45(9) :3084–3104., 2012.

[5] W. Bi and J. T. Kwok. Multi-label classification on tree-and DAG-structured hierarchies in Proc. 28th Int. Conf. Mach. Learn., Bellevue, WA, USA, 2011, pp. 17–24.

[6] O. W. Kwon and J. H. Lee, Web page classification based on knearest neighbor approach, IRAL, 2003.

[7] O. W. Kwon and J. H. Lee, Web page classification based on knearest neighbor approach, Proceedings of the Fifth International Workshop on Information Retrieval with Asian Languages, 2000, Hong Kong, China, September 30 - October 01, 2000.

[8] Ajay S. Patil, B.V. Pawar. Automated Classification of Web Sites using Naive Bayesian Algorithm, Proceedings of the International Multi Conference of Engineers and Computer Scientists 2012 IMECS 2012, 14-16 March, 2012, Hong Kong.

[9] Noor, A., Chandan, K. R., Farshad, F. Exploiting Label Dependency for Hierarchical Multi-label Classification. In PAKDD (1), 294-305 (2012)

[10] S. Shibu, A. Vishwakarma, and N. Bhargava, A combination approach for web page classification using page rank and feature selection technique, International Journal of Computer Theory and Engineering, 2010.

[11] A. P. Asirvatham and K. K. Ravi, Web Page Categorization Based on Document Structure, 2002.

[12] F. HERRERA, F. CHARTE, A.J. RIVERA, M.J. DEL JESUS, Multi-label Classification Problem Analysis. Metrics and Techniques. Springer International Publishing, ISBN 78-3-319-41111-8, 2016.

[13] Tsoumakas G. and Katakis I. Multi-label classification: An overview. International Journal of Data Warehousing and Mining (IJDWM), 3(3) :1–13, 2007.

[14] Tsoumakas G., Katakis I., and Vlahavas I. Mining multi-label data. In Data mining and knowledge discovery handbook, pages 667–685. Springer, 2010.

[15] Tsoumakas G. and Vlahavas I. Random k-labelsets : An ensemble method for multilabel classification. In Proceedings of the 18th European conference on Machine Learning, 2007.

[16] Cherman EA, Metz J, Monard MC. Métodos multirrotulo independientes de algoritmo: um estudo de caso. In: Anais da XXXVI Conferencia Latinoamericana de Informática (CLEI). Asuncion, Paraguay; 2010. p. 1–14.

[17] Clare A, King RD. Knowledge discovery in multi-label phenotype data. Lect Notes Comp Sci 2001:42–53.

[18] Sun, A.; Lim, E.P.; Liu, Y. On strategies for imbalanced text classification using SVM: A comparative study. Decis. Support Syst. 2009, 48, 191–201.

[19] B. Choi. "Web Page Classification*", Studies in Fuzziness and Soft Computing, 2005.

[20] André Melo, Heiko Paulheim. "Local and global feature selection for multilabel classification with binary relevance", Artificial Intelligence Review, 2017.

[21] Alazaidah, Raed, Fadi Thabtah, and Qasem Al-Radaideh. "A Multi-Label Classification Approach Based on Correlations Among Labels", International Journal of Advanced Computer Science and Applications, 2015.

[22] Gibaja, Eva, and Sebastián Ventura. "Multi-label learning: a review of the state of the art and ongoing research : A review on multi-label learning", Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery, 2014.

[23] Tanaka, Erica Akemi, Sérgio Ricardo Nozawa, Alessandra Alaniz Macedo, and José Augusto Baranauskas. "A multi-label approach using binary relevance and decision trees applied to functional genomics", Journal of Biomedical Informatics, 2015.

[24] Carlos N. Silla, Alex A. Freitas. "A survey of hierarchical classification across different application domains", Data Mining and Knowledge Discovery, 2010.

[25] Chang Xu, Dacheng Tao, Chao Xu. "Robust Extreme Multi-label Learning", Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, 2016.

[26] Amirreza Mahdavi-Shahri, Mahboobeh Houshmand, Mahdi Yaghoobi, Mehrdad Jalali. "Applying an ensemble learning method for improving multi-label classification performance", 2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS), 2016.