

A System for multi-label classification of learning objects

V. F. López, F. Prieta, A. B. Gil, S. Rodríguez, M. N. Moreno

Abstract The rapid evolution within the context of e-learning is closely linked to international efforts on the standardization of Learning Object (LO), which provides ubiquitous access to multiple and distributed educational resources in many repositories. This article presents a system that enables the recovery and classification of LO and provides individualized help with selecting learning materials to make the most suitable choice among many alternatives. For this classification, it is used a special multi-label data mining designed for the LO ranking tasks. According to each position, the system is responsible for presenting the results to the end user. The learning process is supervised, using two major tasks in supervised learning from multi-label data: multi-label classification and label ranking.

1 Introduction

The concept of LO has evolved into a central component within the current context of e-learning. Chiappe et al. recently described [3] a learning object (LO) as a digital, self-contained and reusable entity with a clearly instructional content, containing at least three internal and editable components: content, learning activities, and elements of context. Additionally, LO should have an external information structure, the metadata, which can facilitate its identification, storage and retrieval. Given this definition, it is possible to arrive at a certain consensus regarding LOs: they must be a minimal content unit (self-contained) that intends to teach something (instructional purpose) and can be reused (reusability) on different platforms without any compatibility problems. To achieve the LO classification, each object must be tagged with descriptive metadata or information about that resource in order to be easily located and later retrieved from repositories. The problem is that there are no easy or automated ways to the tagging these objects. LOs will be tagged according to personal

Departamento Informática y Automática, University of Salamanca, Plaza de la Merced S/N, 37008. Salamanca, {vivian;fer;abg;srg;mmg}@usal.es

criteria that have been considered most appropriate by the editor or teacher. By the way, repositories have a lack of basic characteristics that are expected of any general search engine, such as classification tasks, sorting results, the use of different filtering techniques (such as the collaborative technique), the automated management of repositories and the extraction of statistics that serve to improve the global query process.

The focus of this paper is on multi-label classification methods for searching LOs because every LO must be tagged with metadata, which descriptive information that allows the easy search of LO. LOs are frequently annotated with more than a single label, we would like to be able to retrieve LOs based on any of the associated tag, therefore the single-label classification cannot model this multiplicity. This paper describes an approach that uses multi-label classification methods for searching LOs tagged by Learning Object Metadata (LOM) [7], specifically the model offers a methodology that illustrates the task of multi-label mapping of LOs into types queries through an emergent multi-label space, and that can improve the first choice of learners or teachers. The system provides individualized help in selecting learning materials establish a ranking system for the LOs.

This paper is structured as follows: section 2 explains the main concepts and characteristics that establish LOs as the fundamental base within the current context of web-based e-learning. Section 3 provides some background information on the problem of multi-label classification, the details of the dataset used in this paper and experimental results comparing the two multi-label classification algorithms. We conclude with section 4, which explains some of the more relevant aspects and future work.

2 The actual context of the e-learning

Existing standards and specifications LOs focus on facilitating the search, evaluation, acquisition, and reuse of LOs so that they can be shared and exchanged across different learning systems. The most notable standards used for LO with metadata are: DublinCore [6] and, most importantly, the IEEE-LOM [7]. Since 2002, the learning object metadata (LOM) has been the standard for specifying the syntaxes and semantics of LOM. It uses a hierarchical structure that is commonly coded in XML, and includes element names, definitions, data types, taxonomies, vocabularies, and field lengths. LOM is focused on the minimal set of attributes needed to allow these LOs to be managed. LOs are placed inside repositories, in an attempt to facilitate its reusability so that they can be more easily stored and retrieved. The LOR are highly heterogeneous, each one with a different storage system, query methods, etc. But the heterogeneity is not in and of itself a problem, because there are currently different systems that are interoperable [4]. The basic functioning of an interoperability interface is trivial; it is based on web services through which a client queries a LOR. This simple concept gave way to the birth of new types of applications dedicated to a federated search for LOs in repositories. This software is used to

perform simultaneous queries in different LOR, allowing a better reusability of the resources. As noted in [1], the searching of LOs is a challenging task of e-Learning, there are many approaches and techniques developed within e-Learning initiatives that facilitate the search and delivery of appropriate LO, for example: concept-based search, context-sensitive delivery and personalization, ontology-based course assembly and learning content development, adaptive learning and adaptive media, etc. Due to the continual research in search systems, the ability to create standardized and interoperability processes that can be applied to recovering LO has made it possible to formalize search and retrieval processes for LO in different repositories.

3 Multi-label classification

In this application it is intended to demonstrate that multi-label classification can be applied to the organization of LO to illustrate the idea of using the MAS in finding a LO between learning materials of different heterogeneous LOR. According to Tsoumakas et al. [16] the learning from multi-label data has attracted recently significant attention, motivated by an increasing number of new applications, to name a few typical like: social network [10] [23], text categorization [19] [9] [24], semantic annotation of images [26][2], music categorization into emotions [11] [22] and bioinformatic [5][13][25].

In [8] it is indicated, traditional *single-label* classification is concerned with learning from a set of examples that are associated with a single label λ from a set of disjoint labels L , $|L| \leq 1$. If $|L| = 2$, then the learning task is called *binary classification* (or *filtering* in the case of textual and web data), while if $|L| \geq 2$, then it is called *multi-class classification*. In multi-label classification, the examples are associated with a set of labels $Y \subseteq L$.

There exist two major tasks in supervised learning from multi-label data: *multi-label classification* (MLC) and *label ranking* (LR). We would like to implement methods that are able to mine both an ordering and a bipartition of the set of labels from multi-label data. Such a task has been recently called *multi-label ranking* (MLR) [1] and poses a very interesting and useful generalization of MLC and LR.

Multi-label classification methods can be categorized into two different groups: i) *problem transformation methods*, and ii) *algorithm adaptation methods* [14]. The first group of methods are algorithm independent. They transform the multi-label classification task into one or more single-label classification, regression or label ranking tasks. The second group of methods extend specific learning algorithms in order to handle multi-label data directly.

3.1 Multi-label classification algorithms

We taking account, the resulting view in [18], to compare classification algorithms: binary relevance (BR), label powerset (LP), random k-labelsets (RAKEL) [15] and multi-label k-nearest neighbor (MLkNN) [26] [12] [15].

We used a machine learning method for performing an empirical evaluation of both algorithms RAKEL and MLkNN based on one multi-label data sets of LOs. We also experimented with a machine learning, in building a multi-label model using a training data set of LOs and then applying it to a new (unlabeled) data set, in order to obtain predictions and for new instances classification.

Multilabel classifiers such as RAKEL could be used for the automated classification of LO collections in multiple types queries (classes). We have used search for locating LOs by keyword based upon the metadata' contents. Such querying capability would be useful for LO selection in various LOR in the implementation of LO retrieval systems.

The RAKEL algorithm was selected, as a recent method that has been shown to be more effective than the first two [18]. The RAKEL method constructs an ensemble of LP classifiers. This way RAKEL manages to take label correlations into account, while avoiding LP's problems. A ranking of the labels is produced by averaging the zero one predictions of each model per considered label. Thresholding is then used to produce a bipartition as well.

MLkNN was selected, as a recent high-performance representative of problem adaptation methods that is based on k Nearest Neighbors (kNN) lazy learning algorithm.

3.2 Experimental work

The LO dataset was taken after making 60 queries to different repositories because according to [20] the access LOs can take advantage of queries upon metadata for selecting the objects that are most suited to the needs of learners or teachers. In addition, many LOs include textual material that can be indexed, and such indexes can also be used to filter the objects by matching them against user-provided keywords.

We present the experimental results for a LO dataset contains 253 LO examples, annotated with one or more out of 38 labels corresponding to types queries identified by teachers and pupils as necessary to support their learning discovery activities, such as Programming languages, Algorithm construction, etc. Each LO is described with 1442 features extracted from the LOM. The data set format (number of labels, number of features, order of attributes, etc) must conform to the format of the training data set based on which the model we was built.

We have used the Rakel classifier from the Java Library for Multi-Label Learning (MULAN) [17], for obtaining the predictions of a trained model for a data set with unlabeled instances. To the experiments we followed the directive that is available on-line in open-source MULAN system, which consists of three parts:

1. We load the multi-label data set that will be used for training the model. The training data were provided for the LO, in two text files required by Mulan for the specification of a multi-label data set: an XML file specifying the names of the labels (LO.xml), and an ARFF file specifying the actual data (LO.arff).
2. We create an instance of the learning algorithm that we want to train, in order to build a model and obtain predictions. We will create an instance of the RAKEL algorithm.

RAKEL is actually a meta algorithm and can accept any multi-label learner as a parameter, but is typically used in conjunction with the LP algorithm. It accepts a single-label classifier as a parameter. We will use the C4.5 algorithm from Weka for this purpose (J48).

RAKEL has three parameters that need to be selected prior to training the algorithm: a) the subset size, b) the number of models and c) the threshold for the final output. We used an internal 10-fold cross-validation on the training set, in order to automatically select these parameters. The subset size was varied from 2 to 5, the number of models from 1 to 100 and the threshold from 0.1 to 0.9 with a 0.1 step. 10 different 10-fold cross-validation experiments were run for evaluation. The results that follow are averages over these 100 runs of the different algorithms. The number of neighbors in MLkNN was set to 10.

3. We train the classifier using the LO dataset that we loaded of different LOR. Two different 10-fold cross-validation experiments were run for evaluation. The results that follow are averages over these 100 runs of the both algorithms.

For the empirical evaluation of both algorithms we use the `crossValidate` method of the `Evaluator` class of Mulan. This returns a `MultipleEvaluation` object, which we was printed to file to see the results in terms of all applicable evaluation measures available in Mulan. To obtain predictions we loaded the unlabeled data instances. The learner returns an instance of `MultiLabelOutput` class as a result from prediction. The results of the learner output contains bipartition of labels, label confidences and rankings as predicted for given instance.

3.2.1 Result

According to [22] the evaluation of methods that learn from multi-label data requires different measures than those used in the case of single-label data. There are various measures that have been proposed for the evaluation of bipartitions and rankings with respect to the multi-label data.

Here, we are interested in evaluating the quality of bipartitions and rankings. We then perform experiments using a variety of measures to compare both algorithms. Results are displayed in Table 1 shows the predictive performance of the both competing multilabel classification algorithms using a variety of measures. In the case of RAKEL all metrics significantly outperform the MLkNN algorithm in almost all measures, especially when taking into account the measure subset accuracy, which is equal to the zero-one loss for the single-label classification task of predicting the exact label subset.

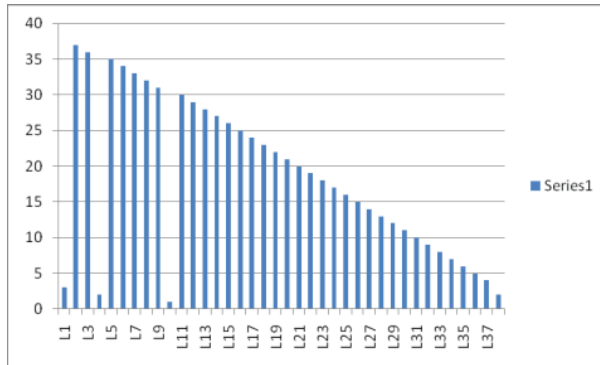
Table 1 Performance results

Class	RAKEL	MLkNN
Hamming Loss	0,078	0,0864
Subset Accuracy	0,1862	0,0948
Subset Recall	0,2195	0,1567
Example-Based Accuracy	0,2131	0,1538
Micro-averaged Precision	0,8677	0,7550
Micro-averaged Recall	0,2190	0,0717
Micro-averaged F-Measure	0,3397	0,2630
One-Error	0,1094	0,0888
Average Precision	0,3550	0,5581
Is-Error	0,7785	0,6962
Error Set Size	7,5411	4,3164
Coverage	22,0614	10,4075
Ranking Loss	0,4187	0,1766
Micro-averaged AUC	0,6127	0,8603

In relation to the test time, we notice that RAKEL is a fast algorithm, the time-consuming algorithm during testing is little nevertheless MLkNN is the most time-consuming algorithm during the experiments.

Experimental results indicate that not only is Raket more efficient in training and testing than MLkNN algorithm, but that it also improves predictive accuracy.

The results demonstrate that Raket algorithm can be used to better the classification of LOs in types queries based upon the metadata' contents. Fig.1 shows how Raket separate the LOs according to their keywords for labeling the LO in types queries. In addition Fig.1 shows the number of examples annotated with each label. We notice that based on the ease of predictions we can rank the labels in the following descending order: L2, L3, L5, L6, L7,L8, L9, L11, ..., L38.

**Fig. 1** Number of examples annotated with each label

Using an instance of the RAKEL algorithm it has been possible to make both an ordering and a bipartition of the set of labels from multi-label data so the MLR task. Fig.2 shows an example of the ranking for Grammar LO.

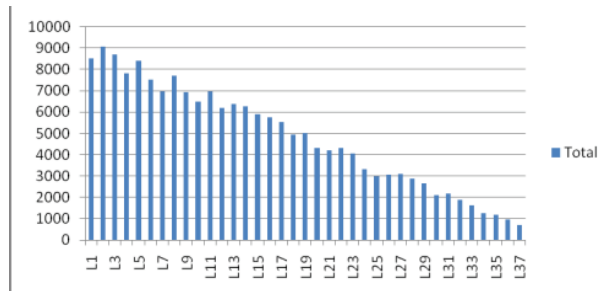


Fig. 2 The ranking for Grammar LO

4 Conclusions

The search and location services for educational content, and specifically LOs, presented in this paper constitute the core of the development of distributed, open computer-based educational systems. For this reason the research in this area has been so active in recent years. We have tried to utilize a multi-label classification algorithm in order to build a model to classifying and cataloguing the LOs in types queries.

The sorting system proposed is also very convenient, given that the LOM standard does not define a minimal set of fields that a LO must have; this makes it difficult to evaluate if a LO has a sufficient quality. Using the feedback provided by the users, from the daily use of the application, the multi-classifier goes through a learning process, which allows it to continually improve its results.

The RAKEL algorithm used for the classifier was very effective and was proposed for LO categorization. Its algorithm used for the classification was very effective and was too proposed for LOs ranking. Multi-label classifiers such as RAKEL could be used for the automated annotation of large LOR collections with multiple LO. This in turn would support the implementation of LO information retrieval systems that query LO collections by tags. Such a querying capability would be useful for LO selection in various applications. These LOs will be processed according to certain classification criteria that have been personalized and are considered most appropriate for the user.

Therefore this model offers a methodology that illustrates the task of multi-label mapping of LOs into types queries through an emergent multi-label space, and that can learning objects ranking tasks to selecting learning materials establishing a ranking system for the LOs.

Future work we will researcher with high data sets with other different feature sets for LO representation taking into account the global statistics of the LO.

References

1. K., Brinker, J., Fürkranz, E., Hüllermeier, A unified model for multi-label classification and ranking. In: Proceedings of the 17th European Conference on Artificial Intelligence (ECAI 06), Riva del Garda, Italy (2006) 489-493.
2. M., Boutell, J., Luo, X., Shen, C., Brown, Learning multi-label scene classification. *Pattern Recognition* 37 (2004) 1757-1771.
3. A., Chiappe, Y., Segovia, & H. Y. Rincon, H. Y., Toward an instructional design model based on learning objects. *Educational Technology Research and Development*, 55 (2007) 671-681.
4. D., Dagger, A., O'Connor, S., Lawless, E., Walsh, V. P., Wade, Service-Oriented E-Learning Platforms: From Monolithic Systems to Flexible Services, *IEEE Internet Computing*, vol. 11, no. 3, May/June (2007) 28-35.
5. S., Diplaris, G., Tsoumakas, P., Mitkas, I., Vlahavas, I., Protein classification with multiple algorithms. In: Proceedings of the 10th Panhellenic Conference on Informatics (PCI 2005), Volos, Greece (November 2005) 448-456.
6. DublinCore Metadata Initiative (DCMI), <http://dublincore.org> (2007).
7. IEEE Learning Technology Standard Committee (IEEE-LTSC). WG12 Learning Object Metadata. 2002. <http://ieeeltsc.org/wg12LOM/>
8. I., Katakis, G., Tsoumakas, I., Vlahavas, Multilabel Text Classification for Automated Tag Suggestion. Proceedings of the ECML/PKDD 2008 Discovery Challenge, Antwerp, Belgium, (2008).
9. A., McCallum, Multi-label text classification with a mixture model trained by em. In: Proceedings of the AAAI 99 Workshop on Text Learning. (1999).
10. P., Mika, Ontologies are us: A unified model of social networks and semantics. In Proc. of ISWC05, (2005).
11. L., Li, M., Ogihara, M., Detecting emotion in music. In: Proceedings of the International Symposium on Music Information Retrieval, Washington D.C., USA (2003) 239-240.
12. <http://lamda.nju.edu.cn/datacode/MLkNN.htm>
13. V., Roth, B., Fischer, Improved functional prediction of proteins by learning kernel combinations in multilabel settings. In: Proceeding of 2006 Workshop on Probabilistic Modeling and Machine Learning in Structural and Systems Biology (PMSB 2006), Tuusula, Finland (2006).
14. G., Tsoumakas, I., Katakis, I., Multi-label classification: An overview. *International Journal of Data Warehousing and Mining* 3 (2007) 1-13.
15. G., Tsoumakas, 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 (2007) 406-417.
16. G., Tsoumakas, I., Katakis, and I., Vlahavas, Mining multi-label data. In O. Maimon and L. Rokach, editors, *Data Mining and Knowledge Discovery Handbook*. Springer, 2nd edition, (2010).
17. G., Tsoumakas, J., Vilcek, E., Spyromitros, I., Vlahavas, Mulan: A Java Library for Multi-Label Learning, *Journal of Machine Learning Research* (accepted for publication conditioned on minor revisions), (2010).

18. K., Trohidis, G., Tsoumakas, G., Kalliris, I., Vlahavas, I., Multilabel classification of music into emotions. In: Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR), (2008).
19. R. E., Schapire and Y., Singer, Boostexter, A boosting-based system for text categorization. *Machine Learning*, 39(2/3):135 - 168, (2000).
20. S., Ternier, D., Massart, A., Campi, S. Guinea, S., Ceri, E., Duval, Interoperability for Searching Learning Object Repositories. The ProLearn Query Language. *D-Lib Magazine*, Volume 14 Number , ISSN 1082-9873, January/February (2008).
21. S., Ternier, K., Verbert, G., Parra, B., Vandeputte, J., Klerkx, E., Duval, V., Ordoñez, X., Ochoa, The Ariadne Infrastructure for Managing and Storing Metadata, *IEEE Internet Computing*, vol. 13, no. 4, pp. 18-25, July/Aug. (2009).
22. K., Trohidis, G., Tsoumakas, G., Kalliris, I., Vlahavas, Multilabel classification of music into emotions. In: Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR). (2008).
23. X. Wu, L. Zhang, and Y. Yu, Exploring Social Annotations for the Semantic Web. In Proc. of WWW06, (2006).
24. Y., Yang, An evaluation of statistical approaches to text categorization. *Journal of Information Retrieval* 1 (1999) 67-88.
25. M.L., Zhang, Z.H., Zhou, Multi-label neural networks with applications to functional genomics and text categorization. *IEEE Transactions on Knowledge and Data Engineering* 18(10) (2006) 1338-1351.
26. M.L., Zhang, Z.H., Zhou, M1-knn: A lazy learning approach to multi-label learning. *Pattern Recognition* 40 (2007) 2038-2048