On Discovering Deterministic Relationships in Multi-Label Learning via Linked Open Data

Eirini Papagiannopoulou, Grigorios Tsoumakas, and Nick Bassiliades

Department of Informatics, Aristotle University of Thessaloniki,
54124 Thessaloniki, Greece
{epapagia,greg,nbassili}@csd.auth.gr
http://www.csd.auth.gr/

Abstract. In multi-label learning, each instance can be related with one or more binary target variables. Multi-label learning problems are commonly found in many applications, e.g. in text classification where a news article is possible to be both on politics and finance. The main motivation of multi-label learning algorithms is the exploitation of label dependencies in order to improve prediction accuracy. In this paper, we present ongoing work on a method that uses the linked open data cloud to detect relationships between labels, enriches the set of labels with new concepts which are super classes of two or more labels, trains a model on the enhanced training set and finally, makes predictions on the enhanced test set in order to improve the prediction accuracy of the initial labels.

Keywords: multi-label learning, linked open data, semantics, WordNet

1 Introduction

In multi-label data, instances are related to one ore more binary target variables, commonly called labels. Learning from such data has received a lot of attention from the machine learning and data mining communities in recent years. This is due to the multitude of practical applications it arises in, and the interesting research challenges it presents, such as exploiting label dependencies, learning from rare labels and scaling up to large number of labels [1].

In several multi-label learning problems, the labels are organized as a tree or a directed acyclic graph, and there exist approaches that exploit such structure [2]. However, in most multi-label learning problems, flat labels are only provided without any accompanying structure. Yet, it is often the case that implicit deterministic relationships exist among the labels. For example, in the ImageCLEF 2011 photo annotation task [3], the learning problem involved 99 labels without any accompanying semantic meta-data, among which certain deterministic relationships did exist. In particular, there were several groups of mutually exclusive labels, such as the four seasons *autumn, winter, spring, summer* and the person-related labels *single person, small group, big group, no persons*. There were also several positive entailment (consequence) relationships, such as *river → water* and *car → vehicle*.

These observations motivated us to consider the automated discovery of such deterministic relationships as potentially interesting and useful knowledge, and the exploitation of this explicit knowledge for improving the accuracy of multi-label learning algorithms. In our previous work in this direction [4], we discovered such relationships from the data using techniques based on association rule mining and exploited them via a deterministic Bayesian network representation. Here, we investigate the possibility of exploiting Linked Open Data (LOD) for discovering such relationships. In particular we aim at finding common ancestor concepts among the labels. We then introduce these ancestors as additional labels, expanding thus the label space, inspired by approaches that similarly expand the feature space for data mining problems based on LOD [5]. Finally, we apply traditional multi-label learning algorithms that indirectly exploit relationships and notice that accuracy gains are often achieved this way.

The rest of this paper is organized as follows. Section 2 discusses related work on discovering and exploiting deterministic relationships in multi-label learning. In Section 3, we present our method in detail, giving examples to be comprehensible. In Section 4, we present the evaluation approach we follow and describe the results of our experiments. Finally, in the last Section, we set open research issues for discussion and some future extensions for this work.

2   Related Work

The idea of discovering and exploiting deterministic label relationships from multi-label data was first discussed in [6], where relationships were referred to as *constraints*. An interesting general point of [6] was that label constraints can be exploited either at the learning phase or at a post-processing phase. In addition, it presented four basic types of constraints and noted that more complex types of constraints can be represented by combining these basic constraints with logical connectors. For discovering constraints, it proposed association rule mining, followed by removal of redundant rules that were more general than others. For exploiting constraints, it proposed two post-processing approaches for the label ranking task in multi-label learning. These approaches correct a predicted ranking when it violates the constraints by searching for the nearest ranking that is consistent with the constraints. They only differ in the function used to evaluate the distance between the invalid and a valid ranking. Results on synthetic data with known constraints showed that constraint exploitation can be helpful, but results on real-world data and automatically discovered constraints did not lead to predictive performance improvements.

A rule-learning approach for discovering and exploiting deterministic label relationships in multi-label learning was recently proposed in [7]. In particular, for each label a separate rule model is constructed, but this model uses the rest of the labels as additional features. This can lead to rules that include labels with/without ordinary input features in the preconditions of a rule deriving a particular label. Such rules are a natural representation of entailment relationships among labels (and input features).

3   Our Approach

Our general goal is to discover deterministic relationships among labels via the LOD cloud. In this first step towards our goal, we focused specifically on discovering common ancestors of existing labels, with which we extend the existing label space of a learning problem. We expect standard multi-label learning algorithms to benefit from this explicit representation of latent relationships among labels. We also focused on WordNet[1] as the LOD resource to search for these relationships, and leave as future work the exploitation of DBPedia and other resources.

Names of label attributes sometimes contain an auxiliary keyword that emphasizes their role as target variables. This typically arises in textual datasets, where the name of label attributes may coincide with names of input attributes (words). For example, all labels of the *bibtex* and *bookmarks* datasets [8], the *delicious* dataset [9] and the *EUR-Lex* datasets [10] start with "tag_". To deal with this issue automatically, we first tokenize label names with a standard set of delimiters, such as dash, underscore, space, comma and period. Tokens that appear in all label names are then removed and the rest of the tokens are concatenated again with a space separator between them.

We then look up each label name in WordNet. If a label name has multiple senses, we assume that the correct sense is the most frequent one according to WordNet, which is based on semantically tagged corpora. Next, we obtain recursively the hypernym-synsets of the determined sense of the label (i.e. the broader concepts) up to the root of WordNet. We then examine all pairs of labels that we managed to find in WordNet and their common ancestors are used for expanding the original label space of the learning problem that we are examining. In section 4, we describe two different approaches that ignore some very general senses typically arising as common parents of labels that seem to have no semantic relationship.

To exemplify our approach, let's consider we are examining the pair of labels (*Winter*, *Summer*) from the ImageCLEF 2011 photo annotation task [3]. No auxiliary keywords are used inside these label names, so the first pre-processing step leaves them unaltered. We then examine the senses that are related to these labels (the number inside parentheses at the end of the sense description denotes the sense's frequency):

- Winter
  1. the coldest season of the year; in the northern hemisphere it extends from the winter solstice to the vernal equinox (24)
  2. spend the winter (2)
- Summer
  1. the warmest season of the year; in the northern hemisphere it extends from the summer solstice to the autumnal equinox (58)
  2. the period of finest development, happiness, or beauty (0)
  3. spend the summer (0)

---

[1] http://wordnet.princeton.edu/

Based on their frequency, we assume that the 1st sense of each label is the correct one and continue to obtain the following branches of hypernym-synsets of these senses up to the root of WordNet:

- Winter:wintertime → season → period →measure:quantity:amount → abstraction → entity
- Summer:summertime → season → period → measure:quantity:amount → abstraction → entity

This will lead to the addition of the following labels: *season*, *period*, *measure:quantity:amount*, *abstraction* and *entity*. Recall that our method ignores some common parents that appear at the top of the WordNet hierarchy, because they will be added for all label pairs (or for most of them) and thus will not bring any new information. In this example we would not add as labels *abstraction* and *entity*.

After the label-addition process we should fill in the values of the new labels: for every new label of each instance if at least one of its children is true then the specific new label will be true. In all other cases the new label will be false for the specific instance. In our running example, if label *Winter* or label *Summer* is true, then *Season* will be true, too. Otherwise, *Season* will be false.

4 Experiments

We use the Calibrated Label Ranking (CLR) [11] problem transformation method for learning multi-label models, which we have mentioned above (Section 2). For the binary classification problems we employ linear Support Vector Machines. We use the implementations from Mulan [12] and Weka [13] respectively. We experiment on the 6 multi-label datasets that are shown in Table 1. Those datasets were selected because their labels don't have obscure names (e.g. label1, label2 or class1, class2, etc). We split each dataset in train set and test set with percentage 70% - 30%, respectively. We discuss the results in terms of Mean Average Precision (MAP) and Logarithmic Loss (LL).

Table 1 presents the above measures for CLR applied to: (i) the original data, and (ii) the expanded label space via two versions of our approach. In the 1st version, called LOD1, we take into consideration the hypernym-synsets of the determined sense of the label, for up to two layers, i.e. we obtain parent and grandparent senses of the determined sense and we ignore the following 32 very general senses that typically arise as common parents of labels that seem to have no semantic relationship: *substance, content, message, theme, topic, subject, whole, unit, object, entity, abstraction, domain, activity, individual, someone, somebody, mortal, soul, organism, being, cause, go, locomote, formation, alter, modify, change, alteration, modification, happening, occurence, occurent*. These were selected based on early experiments. In the 2nd version of our approach, called LOD2, we take into consideration the hypernym-synsets of the determined sense of the label for all layers up to the root of WordNet and we only ignore the following 6 general senses: *whole, unit, object, entity, abstraction, tag*.

In Table 1, we notice that in all datasets LOD1 leads to improvements in the Log-Loss measure (smaller is better). In addition, there are 4 datasets (bibtex, delicious, bookmarks, IMDB-F) where our method improves MAP but there are also 2 (IC2011, corel5k) where we observe reduction in this measure. Furthermore, we notice that in almost all datasets LOD2 leads also to improvements in the Log-Loss measure (except for ImageCLEF2011 where we have a reduction). Moreover, there are 3 datasets (bibtex, delicious, bookmarks) where our method improves MAP but in the other 3 (IC2011, corel5k, IMDB-F) we observe reduction in this measure.

Table 1. Mean Average Precision (MAP) and Log Loss of CLR: without using our method i.e. standard CLR (std CLR), following the 1st approach (LOD1) and following the 2nd approach (LOD2). The next three columns show the number of additional labels for the LOD1 and LOD2 approach and the proportion of labels that are found in WordNet per dataset. The last columns show the number of examples of each dataset.

dataset	MAP			Log Loss			extra labels		% found	samples
	CLR	LOD1	LOD2	CLR	LOD1	LOD2	LOD1	LOD2		
IC2011	.3275	.3263	.3245	.7221	.7003	.7315	17	69	75/99	8000
bibtex	.3757	.3838	.3833	.9288	.8840	.7881	8	63	119/159	7395
delicious	.1596	.1646	.1661	.9273	.8305	.7901	190	319	778/983	16105
bookmarks	.2353	.2435	.2358	.9401	.8842	.7945	22	79	148/208	87856
corel5k	.0612	.0584	.0580	.9795	.8503	.7481	84	214	367/374	5000
IMDB-F	.1161	.1163	.1160	.8256	.7109	.6411	5	14	23/28	120919

In Table 1, we also give the number of additional labels for the LOD1 and LOD2 approach per dataset in order to have an impression of the increase in the number of labels. We also give the proportion of labels that are found in WordNet for each dataset.

5 Conclusions and Future Work

This work has introduced a method that detects relationships among labels within multi-label datasets using WordNet (as LOD resource) and exploits them, enriching the set of labels with the common parents of them in order to make the learning process more accurate. It is in our immediate plans to conduct experiments where we will add the lowest common subsumer (LCS) instead of all common parents for each pair of labels. It would be useful to insert a criterion that determines whether or not to reject the LCS depending on how far it is from the root or the initial label of the dataset (its child). We believe that our approach can be further extended and improved by exploiting additional resources of the LOD cloud, such as DBPedia, LinkedGeoData, Geospecies knowledge base and Bio2RDF. It is also in our immediate plans to include a general first check of all labels in order to detect the domain that they are referred to. In this way, we will

select the appropriate sense of a label based on the domain of the dataset's labels and not on the assumption that the correct sense is the most common among all the senses. Another important direction is the generalization of our approach so as to be able to discover additional types of relationships among the existing labels (e.g. mutual exclusion, spouse). In this work, we simply extended the original label space with the ancestor labels and relied on standard multi-label learning algorithms to exploit the discovered knowledge. In the future we can also investigate direct exploitation of this knowledge via techniques such as [4]. Finally, we intend to apply our approach to additional datasets and to employ additional evaluation measures using cross-validation as well as including some statistical tests for the significance of our method's improvements.

References

1. Tsoumakas, G., Zhang, M.L., Zhou, Z.H.: Introduction to the special issue on learning from multi-label data. Machine Learning **88**(1-2) (2012) 1–4
2. Vens, C., Struyf, J., Schietgat, L., Džeroski, S., Blockeel, H.: Decision trees for hierarchical multi-label classification. Machine Learning **73**(2) (2008) 185–214
3. Nowak, S., Nagel, K., Liebetrau, J.: The clef 2011 photo annotation and concept-based retrieval tasks. In: CLEF (Notebook Papers/Labs/Workshop). (2011)
4. Papagianopoulou, C., Tsoumakas, G., Tsamardinos, I.: Discovering and exploiting entailment relationships in multi-label learning. arXiv preprint arXiv:1404.4038 [cs.LG] (2014)
5. Paulheim, H., Fürnkranz, J.: Unsupervised generation of data mining features from linked open data. In: 2nd International Conference on Web Intelligence, Mining and Semantics, WIMS '12, Craiova, Romania, June 6-8, 2012. (2012) 31
6. Park, S.H., Fürnkranz, J.: Multi-label classification with label constraints. In: ECML PKDD 2008 Workshop on Preference Learning. (2008)
7. Loza Mencía, E., Janssen, F.: Stacking label features for learning multilabel rules. In: Discovery Science - 17th International Conference, DS 2014, Bled, Slovenia, October 8-10, 2014, Proceedings. Volume 8777 of Lecture Notes in Computer Science. Springer (2014) 192–203
8. Katakis, I., Tsoumakas, G., Vlahavas, I.: Multilabel text classification for automated tag suggestion. In: Proceedings of the ECML/PKDD 2008 Discovery Challenge, Antwerp, Belgium (2008)
9. Tsoumakas, G., Katakis, I., Vlahavas, I.: Effective and efficient multilabel classification in domains with large number of labels. In: Proc. ECML/PKDD 2008 Workshop on Mining Multidimensional Data (MMD'08). (2008) 30–44
10. Loza Mencia, E., Fürnkranz, J.: Efficient pairwise multilabel classification for large scale problems in the legal domain. In: 12th European Conference on Principles and Practice of Knowledge Discovery in Databases, PKDD 2008, Antwerp, Belgium (2008) 50–65
11. Fürnkranz, J., Hüllermeier, E., Mencia, E.L., Brinker, K.: Multilabel classification via calibrated label ranking. Machine Learning **73**(2) (2008) 133–153
12. Tsoumakas, G., Spyromitros-Xioufis, E., Vilcek, J., Vlahavas, I.: Mulan: A java library for multi-label learning. Journal of Machine Learning Research (JMLR) **12** (July 12 2011) 2411–2414
13. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: An update. SIGKDD Explorations **11** (2009)