International Journal of Mechanical Engineering

Multi-label Classification Methods and Challenges

Raghav Agrawal[1], Rohit Sah[2]

[1,2]*Department of Applied Mathematics, Delhi Technological University, Delhi, India*

Abstract

Conventional classification problems usually deal with classification of a single label from one or more (single-label) classes or two or more (multi-class) classes. But when an instance can be associated with multiple classes simultaneously, it is known as Multi-label Classification. With more and more use of machine learning algorithms in the real world applications, the need for multi-label classification has also arisen. This has caused in development of a range of multi-label classification algorithms.Keeping this in mind, with this study we aim to give an overview label classification algorithms. Our goal is to provide a comprehensive study of challenges faced during the training of classifier.

Keywords–Algorithm Adaptation,Dimensionality Reduction, Label Drifting,Multi-label, Problem Transformation

1. INTRODUCTION

Multi-label classification is the problem where each instance may be associated with multiple labels. The problem of single-label classification is the one where each instance is associated with a single label selected from a finite set of labels. When there are at least two labels in the label set for single-label classification, it is known as multi-class classification. It can be said that multi-label classification is a generalization of multi-class classification wherein there is no limitation to the number of labels which are selected for an instance and instead it produces a ranking for the labels.

Nowadays, with most of the people having access to the internet, everyone has a different profile on the internet which is used for showing relevant ads and targeting on social media. These user or customer profiles are made with the help of multi-label classification. For example, if an iPhone user is also a cricket fan, he can be targeted with ads of iPhone accessories with cricket players on it.

Multi-label classification can be divided mainly into two categories:

(a) Algorithm Adaptation Methods
(b) Problem Transformation Methods

The major challenges of classifier training discussed are:

a) Dimensionality
b) Label Drifting
c) Label Imbalance
d) Data cleaning
e) Label dependency
f) Label Uncertainty

The mounting number of labels and interdependencies among them make multi-label problems difficult to resolve. Therefore, there is a pressing need to comprehend and handle this issue. This paper talks about different algorithm adaptation and problem transformation methods. It further talks about usual challenges faced when doing multi-label classification and also tried to bring light to some related research.

2. MULTI-LABEL LEARNING METHODS

With so many technological developments in the recent years. Multi-label classification can be done using many methods. They can be divided mainly into two categories:

(a) Algorithm Adaptation Methods: An adaptation of existing single-label methods.

(b) Problem Transformation Methods: A conversion of multi-label problems to one or more single label problems.

2.1 Algorithm Adaptation Methods

Algorithm Adaptation methods are multi-label methods that tailor already known algorithms for the task of multi-label learning.

2.1.1 Decision Trees

Multi-label C4.5 (ML-C4.5) [1] is a variation of famous C4.5 algorithm. This was done by modifying formula for calculating entropy. This is done by calculating sum of entropy of the labels recursively. This uses the concept of hierarchical multi-label classification.

2.1.2 Boosting

This refers to tree-based boosting, which is based on ADABOOST. ADABOOST.MH and ADABOOST.MR [2] are extensions of ADABOOST for multi-label learning. ADABOOST.MR finds a hypothesis to rank the labels at top, while ADABOOST.MH minimizes hamming loss. Its goal is to improve loss functions, but it is prone to noisy data.

2.1.3 Lazy Learning

Many methods exist which are based on k-Nearest Neighbor (i.e. Lazy Learning). ML-kNN [3] is an extension of kNN for multi-label data. It is based on maximum a posteriori principle. This method follows the concept of error function in backpropagation.

2.1.4 Support Vector Machines

This a ranking approach [4] used for multi-label learning that is based on SVMs. Their cost function is average fraction of incorrectly ordered pairs of labels. It uses a maximum margin ranking strategy. MLTSVM [5] is a recently proposed algorithm which uses a nonparallel hyperplane.

2.2 Problem Transformation Methods

Problem transformation methods are multi-label methods that aim to convert multi-label problems to one or more single-label problems. Since, there exists a lot of machine learning algorithms for single-label problems. Finally, the results of single-label problems are again transformed to multi-label.

2.2.1 Binary Relevance

Binary Relevance (BR) [6] is a type of problem transformation algorithm. It considers prediction of each label as an independent binary classification task. It is a one-against-all strategy. Thus, BR builds binary classifiers, one for each different label. This method also has a shortcoming as it doesn't consider correlation amongst the labels.

2.2.2 Label Power-set

Label Power-set is to group label sets into single-labels to form single-label problems. The set of single-labels represent different label subsets from multi-label representation. This gets rid of the problem of label independency in binary relevance method. Calibrated Label Ranking (CLR) [7] is based on popular pair-wise method for multi-label classification.All classifiers are trained using first label instances as positive example and second label's as negative. Majority voting algorithm is instinctively used to combine the classifiers.

2.2.3 PT Methods

To demonstrate we will use Table 1 as dataset. It consists of four instances and four labels.

TABLE 1 Multi-label Dataset

Instance	L1	L2	L3	L4
1	X			X
2			X	X
3	X			
4		X	X	

Two simple problem transformations methods convert the learning problem into single-label classification problems [8]. The first one is PT1 in which one of the labels is randomly selected in a multiple label instance and the rest are discarded (shown in Table 2) and the second one is PT2 in which it drops every instance of the dataset which is multi-label (shown in Table 3). Both of these reject a lot of information and hence, affects the accuracy of the result.

TABLE 2 Transformed Dataset after PT1

Instance	L1	L2	L3	L4
1	X			X
2				X
3	X			
4		X		

TABLE 3 Transformed Dataset after PT2

Instance	L1	L2	L3	L4
3	X			

The third method, PT3, takes the relationship of the labels into consideration. It uses each set of label as a single label as shown in Table 4. It has been used by [8] and [9] in the past.

TABLE 4 Transformed Dataset after PT3

Instance	L1	L1 ∧ L4	L3 ∧ L4	L2 ∧ L3
1		X		
2			X	
3	X			
4				X

The fourth method, PT4, considers binary classifier for each label. This transforms the original dataset into multiple datasets that have all the instances of the original dataset.

TABLE 5 Transformed Dataset after PT4

Instance	L1	¬L1
1	X	
2		X
3	X	
4		X

Instance	L2	¬L2
1		X
2		X
3		X
4	X	

Instance	L3	¬L3
1		X
2	X	
3		X
4	X	

Instance	L4	¬L4
1	X	
2	X	
3		X
4		X

Another problem transformation method, PT5, converts an instance to multiple instances with same instance and different labels[10]. This uses a distribution classifier that outputs a distribution for all labels.

TABLE 6 Transformed Dataset after PT5

Instance	Label
1	L1
1	L4
2	L3
2	L4
3	L1
4	L2
4	L3

3. CHALLENGES

With continuous development in the field of multi-label methods, various problems are arising. Numerous solutions have been proposed which are based on assumptions. The nature of the dataset also contributes to the accuracy of the result. The major challenges faced are listed below.

3.1 Dimensionality

Dimensionality represents the number of features in a dataset.Dimensionality Reduction [11] aims to remove insignificant features for a clearer and better understanding of the data. This allows us to accentuate more relevant features of the data. There are many unsupervised methods for reducing dimensionality of high-dimensional data. These include Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Singular Value Decomposition (SVD), Independent Component Analysis (ICA), etc. A drawback of unsupervised methods is that they ignore information in labels.

3.2 Data Cleaning

There are a number of concerns that need to be addressed during the data cleaning. For example, data of job searchers, who are requested to mention their talents in a job recruitment application. Regardless of which domain their skills pertain to, every seeker enters them. Text analytics, tokenization, stemming, and stop word removal are required for this type of multi-label data. This can lead to the loss of crucial information and a reduction in data quality.

3.3 Label Dependency

When the data has a lot of labels, investigating correlations between labels may help to minimize complexity. Mainly, two types of multi-label data dependence have been recognized [12]. Conditional label dependency represents the likeliness of labels to occur together given the features of an instance. Unconditional (or marginal) label dependency represents the likeliness of labels to occur together independent of an instance. Since, multi-label classifier process multiple labels simultaneously, it affects the performance of classifier and introduces loss. A way to minimize loss by linking label dependencies [12] has been proposed.

3.4 Label Uncertainty

There is a lot of uncertainty in labels in real-world applications. For example, in a database which stores different skills of users, every user can have a different set of skills which can result into creation of multiple new labels. The major challenge is the preprocessing and categorization for each new label.

3.5 Drifting

Learning algorithms are usually used in dynamic environments in real world where the data is continuously growing. In such a case, the target concept changes over time, this is known as drifting. According to [13], drifting is one of the biggest challenges for a classifier.A method for drift detection in multi-label data streams using label grouping andentropy has been proposed by [14].

3.6 Data Imbalance

Data Imbalance is a common problem in which some labels with lesser frequency have higher importance than the labels which have lower frequency. This has a great effect on the performance of the classifier. An approach was proposed by [15] to solve this, in which they were able to reduce hamming loss but not completely eliminate it. In real world, most of such data is gathered by sensors, on which substantial research is going on to tackle these challenges.

4. CURRENT TRENDS

4.1 Active Learning

Labelling multi-label data manually is a time-consuming, impractical and laborious task. Active learning is a type of learning in which training dataset is kept to minimum and only valuable data points are selected. By this, it intends to reduce label cost and efforts.Mainly, active learning is used single label problems. Such a system consists of a teacher (or oracle), which is the source of data, and a learner. There are mainly three active learning methods:

a) Membership query synthesis
b) Pool-based sampling
c) Stream-based selective sampling.

4.2 Multi-view Multi-label Learning

Multi-view multi-label dataset has instances with multiple views and multiple labels. Since there is lot of multi-label data easily accessible, it needs to be analyzed from using different views for best analysis. Multi-view learning methods [16] can be generally classified into three major categories:

a) Co-training
b) Subspace Learning
c) Multiple Kernel Learning

5. CONCLUSION

In this paper, we discussed several multi-label classification methods. The general challenges faced in a scenario of multi-label classification are also discussed in detail. We also highlighted the new methods which are becoming a need now for superior analysis in multi-label classification. The classification of data can become convoluted in real world application since the data is imbalanced and incomplete. Dimensionality reduction is also a very challenging task for large dataset.

REFERENCES

[1] A. Clare and R. King, "Knowledge Discovery in Multi-label Phenotype Data", Principles of Data Mining and Knowledge Discovery, pp. 42-53, 2001. Available: 10.1007/3-540-44794-6_4.

[2] R. Schapire and Y. Singer, Machine Learning, vol. 39, no. 23, pp. 135-168, 2000. Available: 10.1023/a:1007649029923.

[3] M. Zhang and Z. Zhou, "ML-KNN: A lazy learning approach to multi-label learning", Pattern Recognition, vol. 40, no. 7, pp. 2038-2048, 2007. Available: 10.1016/j.patcog.2006.12.019.

[4] "A kernel method for multi-labelled classification", Advances in Neural Information Processing Systems 14, 2002. Available: 10.7551/mitpress/1120.003.0092.

[5] W. Chen, Y. Shao, C. Li and N. Deng, "MLTSVM: A novel twin support vector machine to multi-label learning", Pattern Recognition, vol. 52, pp. 61-74, 2016. Available: 10.1016/j.patcog.2015.10.008.

[6] G. Tsoumakas and I. Katakis, "Multi-Label Classification", International Journal of Data Warehousing and Mining, vol. 3, no. 3, pp. 1-13, 2007. Available: 10.4018/jdwm.2007070101.

[7] S. Park and J. Fürnkranz, "Efficient Pairwise Classification", Machine Learning: ECML 2007, pp. 658-665, 2007. Available: 10.1007/978-3-540-74958-5_65.

[8] M. Boutell, J. Luo, X. Shen and C. Brown, "Learning multi-label scene classification", Pattern Recognition, vol. 37, no. 9, pp. 1757-1771, 2004. Available: 10.1016/j.patcog.2004.03.009.

[9] S. Diplaris, G. Tsoumakas, P. Mitkas and I. Vlahavas, "Protein Classification with Multiple Algorithms", Advances in Informatics, pp. 448-456, 2005. Available: 10.1007/11573036_42.

[10] G. Tsoumakas and I. Katakis, "Multi-Label Classification", International Journal of Data Warehousing and Mining, vol. 3, no. 3, pp. 1-13, 2007. Available: 10.4018/jdwm.2007070101.

[11] W. Siblini, P. Kuntz and F. Meyer, "A Review on Dimensionality Reduction for Multi-label Classification", IEEE Transactions on Knowledge and Data Engineering, pp. 1-1, 2019. Available: 10.1109/tkde.2019.2940014.

[12] K. Dembczyński, W. Waegeman, W. Cheng and E. Hüllermeier, "On label dependence and loss minimization in multi-label classification", Machine Learning, vol. 88, no. 1-2, pp. 5-45, 2012. Available: 10.1007/s10994-012-5285-8.

[13] E. Spyromitros-Xioufis, M. Spiliopoulou, G. Tsoumakas and I. Vlahavas, "Dealing with concept drift and class imbalance in multi-label stream classification", in IJCAI'11: Proceedings of the Twenty-Second international joint conference on Artificial Intelligence, 2011, pp. 1583-1588.

[14] Z. Shi, Y. Wen, C. Feng and H. Zhao, "Drift Detection for Multi-label Data Streams Based on Label Grouping and Entropy", 2014 IEEE International Conference on Data Mining Workshop, 2014. Available: 10.1109/icdmw.2014.92.

[15] L. Li and H. Wang, "Towards Label Imbalance in Multi-label Classification with Many Labels", 2016. Available: https://arxiv.org/abs/1604.01304. .

[16] P. Pant, A. SaiSabitha, T. Choudhury and P. Dhingra, "Multi-label Classification Trending Challenges and Approaches", Advances in Intelligent Systems and Computing, pp. 433-444, 2018. Available: 10.1007/978-981-13-2285-3_51.