A Review on multi-label classification algorithms

Yong-xian Jin
(College of Mathematics and Computer Science, Zhejiang Normal University, Jinhua Zhejiang 321004, China)
jyx@zjnu.cn

ABSTRACT

Multi-label classification has become a research hotspot in data mining technology. Its research results are widely used in various fields, such as image and video semantic annotation, functional genome, music emotional classification and marketing guidance etc. So far, researchers have proposed a variety of multi-label classification algorithms and significant amount of progresses have been made towards this emerging machine learningfield. This paper aims to provide a timely review on this area with emphasis on state-of-the-art multi-label learning algorithms. Finally, it summarizes the problems and challenges in the current research and looks forward to the development trend in this field.

Keywords: multi-label classification;classificationalgorithm;problem conversion; algorithm adaptive;integration method

1.INTRODUCTION

Classification, as one of the important research topics in the field of data mining, is to analyze and study known category samples to predict unknown category samples. The traditional classification is mainly single-label classification, that is, each sample belongs to only one category. In practice, however, a sample can have multiple categories at the same time. For example, in document classification [1-4], each document may belong to multiple predefined topics.In image classification [5-8], each picture may have different semantics, and in bioinformatics [9-11], each gene may have multiple functions simultaneously. This led to the study of multi-label learning [12-14]. With the application of multi-label classification more and more widely, multi-label classification has gradually become a hot spot in the international machine learning field. After nearly ten years of development, multi-label classification technology has been widely applied in bioinformatics, medical diagnosis, scene classification, music emotion classification [15-18] and other fields. Therefore, the study of multi-label classification has important theoretical and practical significance.

2. THE CURRENT SITUATION OFMULTI-LABEL CLASSIFICATION ALGORITHM RESEARCH

With the development of multi-label classification in practical application, the researchers have proposed a variety of multi-label classification methods[14]. It is divided into problem conversion method, algorithm adaptation method and integration method.

2.1Problem conversion method

Themain idea of problem conversion methodis to convert the learning of multi-label data into one or more single-label learning. Representative algorithms are Binary Relevance(BR) algorithm, Classifier Chain[19] (CC) algorithm and Label powset (LP) algorithm. In which, BR method learns multiple binary classifiers, each of which classifies only one tag. CC still uses the BR method to learn binary classifiers and then concatenate these base classifiers into a chain. The main idea of LP algorithm is to treat each different tag set in the training set as a new atomic tag or single tag, and then convert the multi-label classification problem into multi-class single-label classification problem.

PPT (Pruned problem transformation) [20] is an improvement of LP algorithm, which introduces pruning technology. Firstly, delete all instances in training data set *D* where the occurrence frequency of label subset *Yi* is less than a certain threshold, and then select some label subsets whose occurrence frequency is more than a certain threshold instead of the deleted label combination to join the original training set, so as to solve the problem of too few individual label training samples in LP algorithm that makes learning difficult. However, due to the direct deletion of some tag information, the trained classifier can only predict the tag combination in the new training set.

A method called generalized k-labelset ensemble (GKL) [21] is proposed. Through the combination of AdaBoost and LP algorithm, the equation $H(x)=\sum_{m=1}^{M}\beta_m h_m(x)$ is constructed to control the weight influence of LP algorithm classification results, and improve the prediction accuracy of RAKEL algorithm[34].

In reference [22], EML algorithm is proposed to improve performance by using heterogeneous clustering multi label classifier. Clustering algorithm is good enough to overcome the problem of over fitting, especially the highly unbalanced data set. There are many ways to combine multiple base classifiers, among which the most common and simple are the maximum combination, the minimum combination and the average combination. There are no redundant parameters for training. In addition, the weight voting method is also a potential way for multi label classification prediction, and it can also make the classification algorithm have better robustness.

Ranking by pairwise comparison (RPC) [23] is to convert the multi label training data set into a $\frac{M(M-1)}{2}$ binary label data set, and the binary label in the data set is (λ_i,λ_j), $1 \le i \le j \le M$ .Then the binary label of the binary label dataset is modeled. For a new test sample, call all binary classifiers to predict, then vote the data on both sides, finally, sort the voting results. The result of voting is controlled by threshold, so that the final queue can get a better prediction accuracy. However, it is difficult to select reasonable threshold, which will affect the prediction results. Because the number of algorithm modeling is $\frac{M(M-1)}{2}$, when the number of tags is large, the complexity of the algorithm will be too large, which will affect the efficiency of the algorithm.

CLR (calibrated label ranking) algorithm [24] is to extend RPC algorithm by adding a virtual label. This virtual tag divides tags into related tags and unrelated tags. In the process of training, for any related label, its relevance is greater than the virtual label, and any uncorrelated label is less than the virtual label. The related labels are the prediction results corresponding to the samples.

2.2Algorithm adaptation method

This method is mainly to improve the existing single label classification algorithm, so that it can be directly used in multi label data classification, or some new algorithms are proposed for multi label classification.

ML-KNN [25] algorithm combines Bayesian theory and nearest neighbor idea to process multi label data. This algorithm is a method which combines the traditional KNN method and Bayesian method to deal with the problem of multi label classification. The basic idea of the algorithm is to use KNN method to count the tag information of the nearest neighbor samples, and then to predict the tag of the unknown samples by maximizing the posterior probability.

ML-C4.5 [26], a multi label learning method of extended decision tree, is an improved algorithm of C4.5 decision tree. Clare et al. Modify the definition formula of entropy in the original algorithm to adapt to multi label classification in order to allow multiple labels in the leaf nodes of the tree.

Predictive clustering trees (PCTs) [27] is a hierarchical clustering of decision trees. The top node corresponds to the cluster containing all the data, and the lower tree node is to divide the data into smaller clusters. The PCTs algorithm uses the standard top-down decision tree algorithm. PCTs are instantiated by task variance and prototype functions. Therefore, PCTs can deal with a variety of structured output results: continuous meta output or discrete variables, and time series constitute clustering. For the prediction of discrete tags, the variance function is obtained by calculating the Gini index of the target variable. For a given target variable, its prototype function returns a possibility vector.

In reference [28], a probability generation model is proposed, which considers that each tag generates different words. Based on this model, a multi tag document is generated by mixing the word distribution of its tags. In reference [29], a deconvolution method is proposed to estimate the individual contribution of each label to a given term.

Based on the conditional random field method [30], two graph model algorithms are proposed to represent tag co-occurrence with parameters. First, collective multi label captures the pattern of tag co-occurrence, and second, collective multi label with features, attempts to capture the impact of a single feature on the probability of a pair of tags co-occurrence.

Multi-class multi-label perception (MMAC) [31] is a perceptron based online algorithm for label sorting. MMAC maintains a perceptron for each tag. By updating the weight of each perceptron, it can get a better ranking of all tags.

Extended support vector machine multi label learning method [32]. Elisseeff and Weston proposed a Rank-SVM algorithm based on SVM to adapt to multi label classification. A set of linear classifiers is defined with the goal of minimizing rank loss evaluation index by using the idea of maximum interval, and “kernel technique” is introduced to make the learning model have the ability to deal with nonlinear classification problems.

The ML-KNN[25] method mentioned above aims at the independent statistics of relevant information for each tag, and does not take into account the correlation between tags. Based on this, the author proposes an improved IMLLA algorithm based on label correlation [33]. The algorithm first finds out the neighbor samples of unknown samples, then constructs a tag count vector based on the multi tag information of the neighbor samples, and then submits it to the trained classifier for prediction.

2.3 The integrated method

The integrated method is a multi-label classification method based on problem transformation or algorithm adaptive.

Common integration methods based on problem transformation include RAKEL[34] and ECC[35]. RAKEL method is to randomly select appropriate samples from LP transformed data to train the corresponding classifier, and then integrate all classification results as multi-label classification results. The ECC method first integrates the classification results of all classifiers, and then uses threshold method to select relevant labels as the labels of unknown data.

The ensemble method based on the algorithm adaptive method is usually based on the algorithm adaptive acquired multi-label classifier. For example, RF-PCT and RF-MLC4.5 are integrated methods that use PCTs[27] and MLC4.5[26] as base classifiers, respectively. Because of the advantages of integration in reducing over-fitting and dealing with unbalanced data problems, it is also well applied to multi-label problems, which improves the overall performance.

RAKEL algorithm obtains the final result through ensemble learning, and the effectiveness of ensemble learning lies in the difference and accuracy of classifiers. Because this algorithm randomly selects a subset of labels from the label set L, when L is small, the number of pre-set sub classifiers can better reflect the correlation of labels, and also ensure the difference and accuracy of sub classifiers. However, for large label data sets, the sub model composed of a certain number of randomly selected label subsets can not fully reflect the correlation, which has a great impact on the accuracy of integrated prediction. For this reason, we propose an integrated LC-RAKEL [36]. In the process of label subset selection, this algorithm pays attention to finding the label combination with few times in training set, which makes the constructed sub model more representative.

3.EXISTING PROBLEMS AND CHALLENGES

Although multi-label data mining has made great progress in the past years, multi-label classification algorithm is faced with the following problems.

3.1Algorithm complexity and prediction accuracy

As the number of models or tags will increase after the problem transformation of the multi-label classification algorithm, the algorithm complexity will increase significantly when processing the data volume and multi-label data with large tag set size, and meanwhile, the prediction accuracy will decrease. Therefore, it is still necessary to find more universally applicable algorithms or methods to reduce the computational complexity and improve the prediction accuracy.

3.2Multi-label data set labeling bottleneck and algorithm generalization ability

A large number of annotated samples are needed to construct the model, but the information provided by annotated samples is limited. On the other hand, compared with the labeled samples, the unlabeled samples are more easily obtained and closer to the data distribution in the whole sample space. Providing as many labeled samples as possible requires a lot of time-consuming manual labeling, which leads to the bottleneck problem of labeling. At the same time, the learning system trained with only a small number of labeled samples often has the phenomenon of overfitting, which makes it difficult to make it have a strong generalization ability.Therefore, the research on how to train a more generalized model with only a small number of samples still needs to be further deepened.

3.3 Data set skew

Through many studies in the field of machine learning, it is found that the distribution of data sets about categories is often biased or unbalanced, that is, there may be an order of magnitude difference in the number of samples between categories, which is an important factor leading to the unsatisfactory classification effect. It is easy to generate skewed data sets through problem transformation with BR and LP methods. In this case, the samples cannot accurately reflect the data distribution of the whole space, and the classifier is easy to be overwhelmed by the large class and ignore the small class.

In order to improve the prediction accuracy of multi-label classification algorithm and enhance the generalization ability of the algorithm, LC-RAKEL algorithm was proposed by us. During the selection of tag subset, mutual exclusion relation between labels was introduced to reconstruct the training data set, so as to make the classification accuracy of the trained submodel higher.

4. REFERENCES

[1] Zhang Jing, Li Deyu, Wang sugE, et al. Multi marker text classification based on robust fuzzy rough set model [J]. Computer science, 2015, 42 (7): 270-275.

[2] LV Xiaoyong. Research on multi label text classification algorithm [D]. Shanxi University of Finance and economics, 2010.

[3] Schapire R E, Singer Y. BoosTexter: A boosting-based system for text categorization[J]. Machine learning, 2000, 39(2): 135-168.

[4] McDonald R, Crammer K, Pereira F. Flexible text segmentation with structured multilabel classification[C]//Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2005: 987-994.

[5] Liu Peng, ye Zhipeng, Zhao Wei, et al. An image classification method of multi-level abstract semantic decision [J]. Acta automaticaSinica, 2015, 41 (5): 960-969.

[6] Tengzhou, Guo Yuefei. EM based multi label region calibration algorithm for unsupervised image [J]. Computer application and software, 2012, 29 (2): 5-8.

[7] Vezhnevets V, Konouchine V. GrowCut: Interactive multi-label ND image segmentation by cellular automata[C]//proc. of Graphicon. 2005: 150-156.

[8] Grady L, Funka-Lea G. Multi-label image segmentation for medical applications based on graph-theoretic electrical potentials[M].Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis. Springer Berlin Heidelberg, 2004: 230-245.

[9] Cerri R, de Carvalho A C. Hierarchical multilabel protein function prediction using local neural networks[M].Advances in Bioinformatics and Computational Biology. Springer Berlin Heidelberg, 2011: 10-17.

[10] Berry A F H, Heal W P, Tarafder A K, et al. Rapid multilabel detection of geranylated proteins by using bioorthogonal ligation chemistry[J]. ChemBioChem, 2010, 11(6): 771-773.

[11] Otero F E B, Freitas A A, Johnson C G. A hierarchical multi-label classification ant colony algorithm for protein function prediction[J]. Memetic Computing, 2010, 2(3): 165-181.

[12] Li Sinan, Li Ning, Li zhanhuai. Multi label data mining technology: research review [J]. Computer science, 2013, 40(4): 14-21.

[13] Spyromitros E, Tsoumakas G, Vlahavas I. An empirical study of lazy multilabel classification algorithms[M].Artificial Intelligence: Theories, Models and Applications. Springer Berlin Heidelberg, 2008: 401-406.

[14] Madjarov G., Kocev D, Gjorgjevkjj D, Dzerosji S. An extensive experimental comparison of methods for multi-label learning [J]. Pattern Recognition, 2012, 45(9):3084-3104.

[15] Trohidis K, Tsoumakas G, Kalliris G, et al. Multi-Label Classification of Music into Emotions[C]//ISMIR. 2008, 8: 325-330.

[16] Zhen Chao, Zheng Tao, Xu Jieping. Research on music genre classification based on music semantic information [C] / / Proceedings of the fifth national information retrieval academic conference, 2009.

[17] Zhang Danpu, Wang Lili, Fu Zhongliang, et al. Integrated learning algorithm of label matching based on double label set [J]. Computer application, 2014, 34(9): 2577-2580.

[18] Ness S R, Theocharis A, Tzanetakis G, et al. Improving automatic music tag annotation using stacked generalization of probabilistic svm

outputs[C]//Proceedings of the 17th ACM international conference on Multimedia. ACM, 2009: 705-708.

[19] Read J, Pfahringer B, Holmes G, et al. Classifier chains for multi-label classification[J]. Machine learning, 2011, 85(3):254-269.

[20] Read J. A pruned problem transformation method for multi-label classification[C]//Proc. 2008 New Zealand Computer Science Research Student Conference (NZCSRS 2008). 2008, 143-150.

[21] Lo H Y, Lin S D, Wang H M. Generalized k-labelsets ensemble for multi-label and cost-sensitive classification[J]. Knowledge and Data Engineering, IEEE Transactions on, 2014, 26(7): 1679-1691.

[22] Tahir M A, Kittler J, Bouridane A. Multilabel classification using heterogeneous ensemble of multi-label classifiers[J]. Pattern Recognition Letters, 2012, 33(5): 513-523.

[23] Hüllermeier E, Fürnkranz J, Cheng W, et al. Label ranking by learning pairwise preferences[J]. Artificial Intelligence, 2008, 172(16): 1897-1916.

[24] Fürnkranz J, Hüllermeier E, Mencía E L, et al. Multilabel classification via calibrated label ranking[J]. Machine learning, 2008, 73(2): 133-153.

[25] Zhang M L, Zhou Z H. ML-KNN: A lazy learning approach to multi-label learning[J]. Pattern recognition, 2007, 40(7): 2038-2048.

[26] Clare A, King R D. Knowledge discovery in multi-label phenotype data[M].Principles of data mining and knowledge discovery. Springer Berlin Heidelberg, 2001: 42-53.

[27] Todorovski L, Blockeel H, Dzeroski S. Ranking with predictive clustering trees[M]. Springer Berlin Heidelberg, 2002.

[28] McCallum A. Multi-label text classification with a mixture model trained by EM[C]//AAAI'99 workshop on text learning. 1999: 1-7.

[29] Streich A P, Buhmann J M. Classification of multi-labeled data: A generative approach[M].Machine learning and knowledge discovery in databases. Springer Berlin Heidelberg, 2008: 390-405.

[30] Ghamrawi N, McCallum A. Collective multi-label classification[C]//Proceedings of the 14th ACM international conference on Information and knowledge management. ACM, 2005: 195-200.

[31] Crammer K, Singer Y. A family of additive online algorithms for category ranking[J]. The Journal of Machine Learning Research, 2003(3): 1025-1058.

[32] Elisseeff A, Weston J. A kernel method for multi-labelled classification[C]//Advances in neural information processing systems. 2001: 681-687.

[33] Zhang MINLING. A new multi label lazy learning algorithm [J]. Computer research and development, 2012, 49(11): 2271-2282.

[34] Tsoumakas G, Katakis I, Vlahavas I. Random k-labelsets for multilabel classification [J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(7): 1079-1089.

[35] Read J, Pfahringer B, Holmes G, et al. "Classifierchainsformulti-labelclassification[J].MachineLearning, 2011,85(3):333–359.

[36] JIN Yongxian, ZHANG Weiwei, ZHOU Enbo.:An improved RAKEL method for multilabel classification[J]. Journal of Zhejiang Normal university,2016,39(4):386-391.