Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 12, December 2014, pg.189 – 196

REVIEW ARTICLE

A Review on Classification of Multi-label Data in Data Mining

[1]Ms.Pranali Dhongade, [2]Mr.Rushi Longadge, [3]Mr.Deepak Kapgate

[1] Department of Computer Science and Engineering
G. H. Raisoni Academy of Engineering and Technology, Nagpur, India
pranali.dhongade@gmail.com

[2]Department of Computer Science and Engineering
G. H. Raisoni Academy of Engineering and Technology, Nagpur, India
rushi.longadge@gmail.com

[3]Department of Computer Science and Engineering
G. H. Raisoni Academy of Engineering and Technology, Nagpur, India
deepak.kapgate@raisoni.net

Abstract - Modern applications required multi-label classification such as protein function classification, music categorization, gene function analysis and semantic scene classification. Multi-label learning studies the problem where each instance associated with a set of labels simultaneously. This paper studies the problem of multi-label learning and its various methods for multi-label classification.

Keywords: Data mining, multi-label leaning, Unlabeled & labeled data

---

I. INTRODUCTION

In single label classification, each instance is associated with only one label from a set of different labels. But applications such as text categorization, semantic scene classification, music categorization may belong to more than one class [14]. This type of applications call for multi-label classification. In the previous years, multi-label classification was mostly demanded in the text categorization and medical diagnosis task. For example one news article can cover multiple aspects of an event, thus being related with a set of multiple topics, such as economics, sports and politics etc. [7], [22]. Similarly in medical diagnosis, a patient may be suffering for example from diabetes and prostate cancer at the same time [2]. Many real world applications necessitate multiple labels for each instance.

II. RELATED WORK

Problem transformation and Algorithm adaptation are the two methods of multi-label classification. In problem transformation methods, convert multi-label problem into set of binary classification problem which can then be handled using single-class classifiers and algorithm adaptation methods are those methods that extend specific learning algorithms in order to handle multi-label data directly [1].

A. Problem transformation methods

This method comprises Binary relevance and Label power set approach.

Examples	Attribute	Labelset
1	A_1	(L_1, L_3)
2	A_2	(L_2, L_4)
3	A_3	(L_1)
4	A_4	(L_1, L_2, L_3)

Table 1. Multi-label dataset

Binary Relevance learning method is based on decompositions assuming labels are independent in which it transforms original dataset into subsets where classifiers trains on each dataset. It labeled positively if the label set of original dataset In this approach binary learning method considered as base learner and it has linear complexity depends on number of labels. The main drawback of BR is that it does not take into account any label dependency and may fail to predict some label combinations if such dependence is present[14]. Figure 2 shows the four data sets that are constructed by BR when applied to the data set of Figure 1.

Examples	Label
1	L_1
2	$-L_1$
3	L_1
4	L_1

Examples	Label
1	$-L_2$
2	L_2
3	$-L_2$
4	L_2

Examples	Label
1	L_3
2	$-L_3$
3	$-L_3$
4	L_3

Examples	Label
1	$-L_4$
2	L_4
3	$-L_4$
4	$-L_4$

Table 2. data subset using the Binary relevance method

Label power set is another method of problem transformation. Label dependency considers in label power set. This approach changes multi-label problem into single-label classification which considers all labels as atomic label i.e. each set of label of multi-class classification problem as a one class of single label classification problem. One limitation of label power set is to predict the set of labels in the observed training dataset only. So it produces a high computational complexity.

Examples	Labelset
1	$(L_{1,3})$
2	$(L_{2,4})$
3	(L_1)
4	$(L_{1,2,3})$

Table 3. Transformed data set using the label powerset method

RAKEL [5] is the extension of label power set approach which eliminates the limitation of label power set method. It partitioned the original label set into label subset and from this original label set randomly selects the number of label subsets by this RAKEL method and then trains related multi-class classifier by using label power set. In RAKEL, ensemble the LP classifiers to predict the set of labels. RAKEL is also known as ensemble-based multi-label classification.

Lo, Lin and wang presents basis expansion model i.e. Generalized *k*-Labelsets Ensemble for multi-label classification based on the concept of label powerset method. This expansion model is LP classifier trained

To reduce global error between predicted and ground truth label learn the expansion coefficients efficiently. This model is also extended for cost-sensitive multi-label classification and used in social tagging by considering tag counts as the misclassification costs [7].

B. Algorithm Adaptation methods

This method simply adopts the algorithms to directly perform multi-label classification to address problem in its full form.

Following are some algorithm adaptation methods of multi-label classification are as follows.

Boosting approach is developed for high accurate prediction to machine learning by combining weak and inaccurate learners. Adaboost extended methods for multi-label classification are Adaboost.MH, Adaboost.MR and BoosTexter (Schapire & Singer, 2000). In AdaBoost.MH, Consider the examples are labeled with l and it predicts new examples x with set of labels lin case of positive sign weak classifier only otherwise it is not labeled with l. In AdaBoost.MR the ranking of each of the labels is based on the output of the weak classifiers.

BoosTexter [7] is ensemble learning method extended from Ad-aBoost [23] which is proposed by Schapire and Singer for text categorization. In the training phase, BoosTexter maintains a set of weights over both training examples and their labels, where training examples and their corresponding labels that are hard (easy) to predict correctly get incrementally higher (lower) weights [20].

ML-KNN is multi-label lazy learning approach [20].ML-*k*NN (Zhang & Zhou, 2005) is an extension of the *k*NN for multi-label data. In training set, it first identifies the k-nearest neighbors for each unobserved instance. Based on statistical information received from the label sets of these neighboring instances, i.e. the number of neighboring instances belonging to each possible class, maximum a posteriori (MAP) principle is utilized to determine the label set for the test instance [6]. It is capable of producing ranked labels.

BP-MLL is an adaptation of the popular back-propagation algorithm for multi-label learning. In this novel error function is derived to capture the characteristics of multi-label learning, i.e. the labels belonging to an instance should be ranked higher than those not belonging to that instance [24]. In the training phase, BP-MLL has high computational complexity while based on training model time cost of making prediction is quite less.

A SVM ranking algorithm for multi-label classification invented by Elisseeff and Weston (2002) [10] to minimize ranking loss. (Godbole & Sarawagi, 2004) presents three improvements with SVM in collaboration with BR method for multi-label classification. In first improvement, extends original dataset with q additional features containing the predictions of each binary classifier then trained q binary classifier with extended dataset. In the second improvement, based on confusion matrix which is estimated using fast and moderately accurate classifier removes negative training instances of complete label which is similar to positive label. In the third improvement, discard similar negative instances within threshold distance from learned hyperplane [19].

Another hybrid approach is presented in which, decision tree is integrated with SVM for multi-label classification. In this built decision tree architecture for multi-label classification that utilizes local SVMs where, binary SVM classifier is used in each leaf for making multi-label predictions [26].

Transductive learning

Transductive learning was proposed by vapnik [11] which automatically exploits unlabeled data where testing data is exactly same as unlabeled data. Transductive inference is reasoning from observed specific (training) cases to specific (test) cases. TRAM [1] is the transductive Multi-label classification which effectively assigns set of multiple labels to each instance. With contrast from supervised multi-label learning it evaluates the label sets of the unlabeled instance from the information of both labeled and unlabeled instances. It first defines label concept composition for multi-label instance based on smoothness property after that make multi-label predictions based on label concept compositions. Kong Ng and Zhou presents Supervised version of label set prediction and Transductive version of label set prediction method. In Supervised version, predicts label set directly based upon estimated alpha values by using labeled data. In Transductive version, estimate the cardinality of label set by using both labeled and unlabeled instances.

A graph-based Transductive multi-label classifier (TMC) is developed that is evaluated on a composite kernel in [29] by Yu, Domeniconi, Rangwala, Zhang, and Yu which presents, a method for data integration using the ensemble framework, called transductive multi-label ensemble classifier (TMEC). For each distinct kernel, TMEC trains a graph-based multi-label classifier and then combines predicted output of the distinct models. In this uses a bi-relational directed graph that captures relationships between pairs of proteins, between pairs of functions, and between proteins and functions.

Feature Selection (FS)

Feature Selection (FS) plays an important role in machine learning and data mining. Spolaor, Cherman, Monard and Lee presents comparative study of four multi-label feature selection methods like *RF-BR*, *RF-LP*, *IG-BR* and *IG-LP* which use the filter approach to select features and transforms the multi-label data to single-label data using problem transformation methods like BR and LP approach. In this, ReliefF (RF) and Information Gain (IG) are used to measure the goodness of each label [21].

The wrapper approach to feature selection [27] is directly applicable to multi-label data. Using multi-label algorithm, Explorer the subset of features to optimize a multi-label loss function on an evaluation dataset

A famous approach in text categorization uses the BR transformation in order to estimate the discriminative power of each feature with each of the labels independently of the rest of the remaining labels. Then the obtained values are aggregated to obtain an overall ranking. Most of common aggregation strategies contain maximum or a weighted average of the obtained values [28].

III. DISCUSSION

The analysis from the comparative study of the different multi-labeling algorithms is as follows.

BR has the linear complexity according to the number of labels and it is also easily parallelized. For prediction of labels BR requires label independence. The Label power set approach is method of problem transformation and removes the drawback of BR and considers label dependency in case of prediction. Label power set approach has the limitation is that predicts the set of labels in the observed training dataset only. Random k-label set approach overcomes the limitation of label power set but it cannot directly optimize the learning objective.GLE can be used for both multi-label classification and cost-sensitive multi-label classification. Based on the combine output of weak and inaccurate learners boosting approach yields high and accurate prediction of labels. Ml-kNN is better than some well-established multi-label Learning algorithms. Based on training model time cost of making prediction is quite less and the computational complexity of BP-MLL is high. Transductive learning make the use of labeled and unlabeled data for predict label set of unlabeled data in which test data is same as the unlabeled data. By using labeled data TRAM effectively boost the performance of multi-label classification for labeling unlabeled data.

IV. CONCLUSION

There are many ways to solve the problem of multi-label classification. The basic approach to solve the problem is to label the data, but it has been found that process of labeling to multi-label data is expensive and time consuming. Transductive based multi-label classification is an effective way of assigning multi-label to each instance. TRAM algorithm used label set method which utilize the information of label and unlabeled data which helps to optimize the problem of composite labeling.

REFERENCES

[1] Xiangnan Kong, Michael K. Ng, and Zhi-Hua Zhou, “Transductive Multi-label Learning via Label Set Propagation” IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 3, March 2013.

[2] GrigoriosTsoumakas, IoannisKatakis,“Multi-Label Classification: An Overview” International Journal Data Warehousing and Mining , 2007.

[3] Lei Wu , Min-Ling Zhang “Multi-Label Classification with Unlabeled Data: An Inductive Approach” JMLR: Workshop and Conference proceedings 29:197-212, 2013

[4] Charles X. Ling, Victor S. Sheng “Cost-Sensitive Learning and the Class Imbalance Problem” Encyclopedia of Machine Learning. C. Sammut (Ed.). Springer.

[5] Hung-Yi Lo, Shou-De Lin, and Hsin-Min Wang, “Generalized *k*-Label sets Ensemble for Multi-Label and Cost-Sensitive Classification” IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 7, July 2014 1679 Chun-Liang Li, Hsuan-Tien Lin “Condensed Filter for Cost-sensitive Multi-label Classification” International conference on machine learning , China JMLR :W/P volume 32.

[6] M.-L. Zhang and Z.-H. Zhou, “ML-kNN: A Lazy Learning Approach to Multi-Label Learning,” Pattern Recognition, vol. 40,no. 7, pp. 2038-2048, 2007.

[7] R.E. Schapire and Y. Singer, "BoostTexter: A Boosting-Based System for Text Categorization," Machine Learning, vol. 39, nos. 2/3, pp. 135-168, 2000

[8] Y. Freund and R.E. Schapire, "A Decision-Theoretic Generalization of On- Line Learning and an Application to Boosting," J. Computer and System Sciences, vol. 55, no.1, pp. 119-139, 1997.

[9] N. Ghamrawiand A. McCallum, "Collective Multi-Label,Classification"Proc. 14th Int'l Conf. Information and Knowledge Management, pp. 195-200, 2005

[10] A. Elisseeff and J. Weston, "A Kernel Method for Multi-Labelled Classification," Advances in Neural Information Processing Systems 14, T.G. Dietterich, S. Becker and Z. Ghahramani, eds., pp. 681-687, MIT Press, 2002.

[11] V.N. Vapnik, Statistical Learning Theory. Wiley, 1998.

[12] T. Joachims, "Transductive Inference for Text Classification Using Support Vector Machines," Proc. 16th International Conf. Machine Learning, pp. 200-209, 1999.

[13] Krzysztof Dembczy´, Weiwei Cheng ,Eyke H¨ullermeier1 " Bayes Optimal Multi-labelClassification via Probabilistic Classifier Chains "International Conference on Machine Learning, Haifa, Israel, 2010.

[14] HiteshriModi Mahesh Panchal, "Experimental Comparison of Different Problem Transformation Methods for Multi-Label Classification using MEKA" International Journal of Computer Applications Volume 59 No.15, December 2012

[15] Oscar Luaces, Jorge Díez, José Barranquero· Juan José del Coz · Antonio Baham "Binary relevance efficacy for multilabel classification" © Springer-Verlag Berlin Heidelberg 2012

[16] Erica Akemi Tanaka1 and Jos´e Augusto Baranauskas "An Adaptation of Binary Relevance for Multi-Label Classification applied to Functional Genomics" ISSN -2012.

[17] Cherman, E. A., J. Metz and M. C. Monard, "Incorporating label dependency into the binary relevanceframework for multi-label classification, Expert Systems with Applications" 39(2012), pp. 1647–1655.

[18] Newton Spolaor,EvertonAlvaresCherman, Maria Carolina Monard&Huei Diana Lee, "A Comparison of Multi-label Feature Selection Methods using the Problem Transformation Approach" ELSEVIER-Electronic Notes in Theoretical Computer Science 292 (2013) 135–151

[19] GrigoriosTsoumakas, IoannisKatakis, and IoannisVlahavas , "Mining Multi-label Data" Data Mining and Knowledge Discovery Handbook 2010, pp 667-685

[20] D. W. Aha, Lazy learning: Special issue editorial, Artificial Intelligence Review 11 (1-5) (1997) 7-10.

[21] Liu, H. and H. Motoda, "Computational Methods of Feature Selection," Chapman & Hall/CRC, 2008.

[22] A. McCallum, "Multi-Label Text Classification with a Mixture Model Trained by EM," Proc. Working Notes Am. Assoc. Artificial Intelligence Workshop Text Learning (AAAI '99), 1999.

[23] Freund, Y., &Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1),119–139.

[24] Zhang, M.L.; Zhou, Z.H. "Multi-label neural networks with applications to functional genomics and text categorization". IEEE Transactions on Knowledge and Data Engineering, 18 (10) (2006), pp. 1338–1351.

[25] 'Discriminative Methods for Multi-labeled Classification', paper presented to Proceedings of the 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2004).

[26] DejanGjorgjevikj,GjorgjiMadjarov&SasoDzeroski “Hybrid Decision Tree Architecture utilizing Local SVMs for Efficient Multi-Label Learning” International Journal of Pattern Recognition and Artificial Intelligence World Scientific Publishing Company ,October 17, 2013

[27] Kohavi, R., John, G.H “Wrappers for feature subset selection” Artificial Intelligence 97 (1997) 273–324

[28] Chen, W., Yan, J., Zhang, B., Chen, Z., Yang, “Document transformation for multi-label feature selection in text categorization” In: Proc. 7th IEEE International Conference on Data Mining, Los Alamitos, CA, USA, IEEE omputer Society (2007) 451–456

[29] G.Yu, C. Domeniconi, H. Rangwala, G. Zhang, Z.Yu, “Transductive Multi-label Ensemble Classification for Protein Function Prediction” *KDD’12,* August 12–16, 2012, Beijing, China.