Middle-East Journal of Scientific Research 23 (7): 1341-1345, 2015
ISSN 1990-9233
© IDOSI Publications, 2015
DOI: 10.5829/idosi.mejsr.2015.23.07.102

An Efficient Multi-Label Classification Based on Random K-Label Sets Ensemble

A. Ramajayam, A. Subashini, S. Velu, M. Aishwarya and K. Vijayakumar

Computer Science and Engineering, SKP Engineering College, Tiruvannamalai, Tamil Nadu, India

Abstract: In this paper, we propose the label power set method. This method is one class of multi label concept. A fundamental development model for multi-label classification where a basic purpose is an LP classifier trained on a random k-label set is to be determined. The extension coefficients learned to reduce the global error between the estimate and the ground truth. And we can further extend this model to solve the multi label problem. We can apply this methodology in social tagging to handle the issues of the noisy instruction set by treating the tag counts as the misclassification cost. Our proposed work is based on the number of likes received from the user based on which the rank of the product will be increased. A number of similar products related to the product searched by the user will be displayed in order to minimize the search time.

Key words: Multi-label classification · Ensemble method · Label set · Tag count · Business application

INTRODUCTION

Multi-label classification is has involved a great deal of consideration in recent years. In a predictable single label classification task, specified set of K feasible transfer classes, each instance is coupled with solitary and just one class. In multi-label classification, an occasion can be connected with a set of labels jointly. Label power set (LP) [1] method is a kind of multi-label learning algorithm, which reduced to the multi-label classification trouble into single-label multi-class classification trouble by treating each distinct arrangement of labels in the instruction set as a dissimilar class. Given an experiment occurrence, the multi-class LP classifier predicts the for the most part of achievable class, which can be transformed to a group of labels. Table 1 display the example of multi-label dataset with altered multi-class label based on the model of LP using mobile phone parts. In compare to be the binary relevance approach, which loses to the label craving information while education to the binary classifier for each label separately, the LP method exploit provisional label dependence information by instruction to the joint label distribution [2].

Table 1: Multi-Label Dataset with Transformed

	Example	
	Label Set	Transformed Class
1	Battery	1
2	Memory card, battery	2
3	Touch screen, battery	3
4	Camera	4
5	Memory card, battery	2
6	Ram, Rom	5
7	Touch screen, battery	3
8	Os, Ram	6

Multi-Class Labels: However, one major concern for this model is that, when the amount of labels increases, the amount of likely classes increases proportionally and each class will be connected with very small amount of training instances. Furthermore, LP can simply guess label sets experimental in the training information. In [3], a technique called Random k- Label sets (RAkEL) is planned to conquer to the drawback of the traditional LP method. RAkEL arbitrarily selects an amount of label subsets from the unique set of labels and uses the LP method to guide the equivalent multi-class classifiers. The final calculation of RAkEL is made by appointment of the LP classifiers in the ensemble. This method can not only reduce the number of classes, but also allow each class to have more training instances. We select the multi-label classification and proposed a ranking method. Business products are important application of ranking method which can be used in multi-label classification.

Corresponding Author: A. Ramajayam, Computer Science and Engineering, SKP Engineering College, Tiruvannamalai, Tamil Nadu, India. E-mail: ramajayamr33@gmail.com.

Related Works: Multi-label classification methods can be grouped into two categories: algorithm adaptation and problem transformation [1]. The algorithm adaptation methods extend some detailed learning algorithms for single-label classification to solve the multi-label classification problem. Zhang and Zhouextended the famous back-propagation algorithm for multi-label learning (BPMLL). Some algorithms are extended from the instance-based learning, such as multi-label K-nearest neighbor (MLKNN) [4, 5] and instance based learning by logistic regression (IBLR) [6]. The problem transformation methods convert the multi-label classification problem to one or many single-label classification tasks. Binary relevance and label power set are two trendy problem transformation approaches. The binary relevance method trains a binary classifier for each label independently. Sun et al. [7] proposed two stage learning method to improve the binary relevance method based on hyper graph spectral learning. The label power set method treats each distinct combination of labels as a different class. Our proposed method belongs to the LP-based method. Rokach and Itach design methods to select the k- Label sets in RAkEL using a greedy algorithm [8] which solves a set coverage problem. Their method assumes that every base classifier in the ensemble is equally important. Zhang [9] proposed to use different feature sets for each binary relevance classifier. We note that these algorithms can hardly be applied to cost-sensitive multi-label classification without significant modification. General ensemble learning methods consist of two steps: learning the base classifiers and assigning weights for the base classifiers. Some methods, such as Random Forest [10] and RAkEL [3], only focus on the first step and assign equal weights for every base classifier. Ensemble selection [7] is another ensemble learning method which focuses on the latter step and its success has been proved by winning ACM KDD Cup 20091. Social tag prediction is one major application of multi-label classification. Previous researches have empirically studied different multi-label learning methods for tag annotation on the music [11-18] and social bookmark [13]. In [9] introduce the concept of multi-label classification which is a -dimensional input space and, which is a finite set of possible labels. To facilitate the discussion, hereafter, is represented by a vector. Given a training set that contains samples, the goal of multi label classification is to learn a classifier such that predicts which labels ought to assign to an unseen sample. Cost-sensitive multi-label classification [14] extends multi-label classification by combination of cost vector to each teaching sample. The component denote the cost to be paid when the label is misclassified. In this work, the false harmful cost is set as the tag count while the false activist cost is uniformly set to one. We extend two existing multi-label learning algorithms, namely, stacking [17] and [14], to solve the CSML problem.

Genralized K-Label Sets Ensamble: In this section, we start from introducing the concept of multi-label classification and plus cost-sensitive multi-label classification. Then, we review the concept of exploit the hyper graph representation for multi-label classification and describe the proposed Generalized k-Label Sets Ensemble for multi-label classification. Finally, we extend GLE for cost-sensitive multi-label classification.

Multi-Label Classification Methods: Let **a** Є P^d, which is a d-dimensional input space and Y ⊆ L = {λ1, λ2..., λ_k}, which is a finite set of K possible labels. To facilitate to the discussion, hereafter, B is represent by a vector.

b = (b1,b2,...,bK) {1, –1}k, in which bj = –
1 ⇔ λk B, bj = – 1 ⇔ λj / ∊B.

We denote the labels of the whole instances by **B** Є$P^{N\times K}$, where the i-th row of **B** is **B**i. Given a training set $(\mathbf{a}i, \mathbf{b}i)^{N}_{i=1}$ that contains N samples, the goal of multi-label classification is to learn a classifier H: Pd → 2K such that H(**x**) predicts which labels should be assigned to an unseen sample **a**.

Generalized K-Label Sets Ensemble for Multi-Label Classification: As introduced in [3], a k-label set is a label set P ⊆ L with |P| = k. GLE first trains M LP-based classifiers using randomly selected k-label sets from the original set of labels. Then, GLE uses the base classifiers as dictionary functions and learns a linear combination of these functions. Algorithms 1 and 2 describe the instruction and classification processes, respectively [19]. In Algorithm 2, the prediction of a multi-class LP classifier, gm, for a sample **a** is denoted by gm(**a**) ∊{1, 2,. . . , Z}, where Z is the number of classes, that is, the number of different label sets used for training the LP classifier. Note that Z is upper bounded by min (N, 2k) and is usually much smaller in practice [3]. The i, j^{th} element in H_m is calculated by t (gm (**a**i), j), which is defined as:

$$H(gm(xi),j) = \begin{cases} 1, \text{if } j \in Pm \text{ and } j \text{ is positive in } gm(ai), \\ -1, \text{if } j \in Pm \text{ and } j \text{ is negative in } gm(ai), \\ 0, \text{if } j \notin Pm. \end{cases}$$

For example, when k = 2, the classes 1, 2, 3 and 4 communicate to (1, 1), (1, -1), (-1, 1) and (-1, -1), correspondingly. If label j is not included in Pm, t (gm (**a**i), j) is 0. If label j corresponds to the first label of Pm, t(1, j), t(2, j), t(3, j) and t(4, j) will be 1, 1,-1 and -1, correspondingly. We note that the task q (gm (**a**), j) is used to create the qm (**a**) in the final classifier (1) by assembly to the predictions on all labels j. The weight coefficients β for the base classifiers are learned by solving a minimization problem formulated as follows with respect to [19],

$$\min_{\beta} \frac{1}{2} \|Y - \sum_{m=1}^{M} \beta_m H_m \|^{2}_F + \frac{y}{2} \|\beta\| \frac{2}{2} + \frac{y}{2} \text{ trace } \left(\sum_{m=1}^{M} \beta_m H_m\right)^T L \left(\sum_{m=1}^{M} \beta_m H_m\right)$$

Algorithm 1 The Training Process of GKL

Input: number of models M, size of label set k, learning Parameters C and E, set of labels L and the training set $D = (a_i; b_i)_{i=1}^{N}$

Output: an ensemble of LP classifiers gm, the equivalent k-label sets Rm and coefficients βm

1. Initialize S ← L^k
2. for m ← 1 to min(M,$|L^k|$) do
   - Pm ←x k-label set arbitrarily selected from S
   - train the LP classifier g_m based on D and P_m
   - calculate a altered prediction of g_m
   - S ← S \Pm
3. end
4. Learn β

where || • ||F is the Frobenius norm of a matrix, **L** is the normalize hyper graph Laplacian and Qm × PN×K is a transformed calculation of gm which will be describe in more detail later. The first term in the point of function aims to minimize the global error between the prediction of H(**x**) and the multi-label ground truth **Y**. The next term is a two-norm regularization term of the coefficients **β**. The third term is a hyper graph regularization term.

Algorithm 2 The classification process of GKL

Input: integer of models M, a test sample a, an ensemble of LP classifiers g_m and the equivalent k-label sets P_m and coefficients β_m

Output: the multi-label classification vector v = $(v_1,v_2,\ldots,v_k)$

1. for j ← 1 to K do
   (a) $v_j = 0$
   (b) for each g_m, if j ∈ P_m do
   $v_j = v_j + \beta_m . h(g_m(x), j)$
   (c) end
2. end

Table 2: Data sets

Data set	Attributes [M]
Type of Camera	mb
Type of Os	6
Display type	2
Battery range	mAh
Memory range	RAM
Types of Platform	8

In this GLE Multi-label classification method is comparing the performance of GLE with that of six state-of the- art multi-label learning algorithms: RAkEL, BR (Binary Relevance) [1], CC (Classifier Chains) [17], MLKNN [5], IBLR [6] and BPMLL [4]. In this algorithm using evaluate the five different metrics and evaluate the different data set. Calculate average rank into the multi-label learning algorithm then GLE method is better than other algorithm method. The data set can be downloaded from the Mulan library websites. And also analyze the effect and importance of tag counts in multi-label classification for social tag prediction. Further extend the ranking the business products using multi-label classification algorithm.

Generalized K-Label Sets Ensemble For Multi-Label Classification Using Ranking Method: In introduced Multi-label classification [2] algorithm using ranking the business application. This application using good product search to the user.

Multi-Label Classification Using Ranking Method: Multi-label classification technique is attracted great deal of recent year. Multi-label is main concept of this paper and another concept is cost-sensitive classification method. Multi-label is derived to the products of ranking method. Some products using for mobile phones business application used. Multi-label is much tag to be count. Then similar products can be display to the main products of the page. Multi-label is best classification of this method. Mobile phone is the one of multi-label is using many types of mobile types of parts.

Data Sets: Some voluble data is mining to this paper. Business product using mobile phones, These product data sets is specification of products such as micromax, Samsung, I phone features is calculated camera, os, display, battery, memory, processor and more then use.

Then many types data set can used for business products given using for types of camera range and resolution then next for types of os its new version can be

Fig. 1: Home Page for Searching Business Products

Fig. 2: Displaying Business Products

Fig. 3: Ranking the Business Products

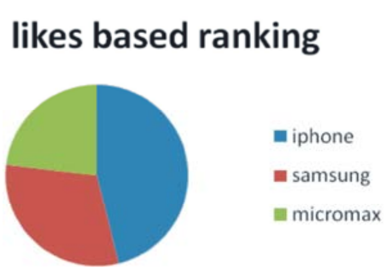

Fig. 4: Ranking Based on Popularity among the Users

updated and next for Battery power range display to the product next for memory range its using internal and external to be classified last one is platform is processor feature to be verified its all of for data set of features.

Generalized K-Label Sets Ensemble for Ranking Method: GLE concept is using for multi-label classification concept. This paper main concept is ranking for the business application using user likes and additional extra new version of attributes its likely to the ranking the application.

Experimental Result: The experimental results of multi-label classification are summarized in figures. First display to the home page of business products of ranking method its using user, admin page can display. In this concept is best solution of business products ranking method. Multi-label classification is using for one of the best method of ranking method. Ranking is based on the number of likes received from the user based on which the rank of the product will be increased. Comment is also received from the user. Comment is display for positive or negative comments its do not calculate for ranking.

User is search the new products and the user is satisfied with the search result then the user give likes to the products based on the likes the rank will be increase Fig 2. And admin work is adding to the products details. User is first register to the site and login to the site from user and display to the rank wise business products.

Next for ranking the product is display to order wise Fig 3.

Chart of the Likes Based Ranking: It's based on the likes to be calculated three mobiles average value to be display in pie chart.

There are three mobiles can be used iphone, Samsung, micromax. Blue color is choosing the iphone and red color is Samsung last one for green it chose for micromax. High value likes is increase to the chart of the color.

CONCLUSION

In this paper we have proposed a Generalized k-Label sets Ensembles.GLE based on the model of Label Power set method in multi-label classification. We have proposed a ranking method. Business products are important application of ranking method which can be used in multi-label classification. We have selected the mobile of business products and also display the similar products of the main products. Social tagging is one of the valuable applications in business products. GLE is the better performance of experimental results.

REFERENCES

1. Tsoumakas, G., I. Katakis and I. Vlahavas, 2010. Mining multilabel data,in Data Mining and Knowledge Discovery Handbook, O. Maimon and L. Rokach, Eds. New York, NY, USA: Springer.
2. Dembczy´ nski, K., W. Waegeman, W. Cheng and E. Hüllermeier, 2010. On label dependence in multi-label classification, in Proc. Int. Workshop Learn. Multi-Label Data, Haifa, Israel.
3. Tsoumakas, G., I. Katakis and I. Vlahavas, 2011. Random k-labelsets for multilabel classification, IEEE Trans. Knowl. Data Eng., 23(7): 1079-1089.
4. Zhang, M.L. and Z.H. Zhou, 2006. Multilabel neural networks with applications to functional genomics and text categorization, IEEE Trans. Knowl. Data Eng., 18(10): 1338-1351.
5. Zhang, M.L. and Z.H. Zhou, 2007. Ml-knn: A lazy learning approach to multi-label learning, Pattern Recognit., 40(7): 2038-2048.
6. Cheng, W. and E. Hüllermeier, 2009. Combining instance-based learning and logistic regression for multilabel classification, Mach. Learn., 76(2-3): 211-225.
7. Sun, L., S. Ji and J. Ye, 2008. Hypergraph spectral learning for multilabel classification, in Proc. ACM Int. Conf. KDD, Las Vegas, NV, USA.
8. Rokach, L. and E. Itach, 2010. An ensemble method for multi-label classification using an approximation algorithm for the set covering problem, in Proc. Int. Workshop Learn. Multi- Label Data.
9. Zhang, M.L., 2011. Lift: Multi-label learning with label-specific features, in Proc. Int. Joint Conf. Artif. Intell.
10. Breiman, L., 2001. Random forests, Mach. Learn., 45: 5-32.
11. Caruana, A. and A. Niculescu-Mizil, 2006. Getting the most out of ensemble selection, in Proc. 6th ICDM, Hong Kong, China.
12. Tsoumaka, G., I. Katakis and I. Vlahavas, 2008. Effective and efficient multilabel classification in domains with large number of labels,in Proc. ECML/PKDD Workshop Mining Multidimensional Data, Antwerp, Belgium.
13. Sanden, C. and J.Z. Zhang, 2011. An empirical study of multi-label classifiers for music tag annotation,in Proc. 12th Int. Soc. Music Inform. Retrieval, Miami, FL, USA.
14. Lo, H.Y, J.C. Wang, H.M. Wang and S.D. Lin, 2011. Cost-sensitive multi-label learning for audio tag annotation and retrieval,IEEE Trans. Multimedia, 13(3): 518-529.
15. Wolpert, D.H., 1992. Stacked generalization,Neural Network., 5(2): 241-259.
16. Tsoumakas, G. and I. Vlahavas, 2007. Random k-labelsets: An ensemble method for multilabel classification,in Proc. Eur. Conf. Machine Learning.
17. Read, J., B. Pfahringer, G. Holmes and E. Frank, 2009. Classifier chains for multi-label classification, in Proc. ECML PKDD, Bled, Slovenia.
18. Zhang, M.L. and Z.H. Zhou, 2006. Multilabel neural networks with application to functional genomics and text categorization, IEEE Trans.Known. Data Eng., 18(10): 1338-1351.
19. Lo Hung-Yi, Shou-De Lin and Hsin-Min Wang, 2014. Generalized K-Labelsets Ensemble for Multi-Label and Cost-Sensitive Classification, IEEE Transactions on Knowledge and Data Engineering, 26(7).