2nd International Conference on Science, Technology and Management
University of Delhi (DU), Conference Centre, New Delhi (India)
27 September 2015, www.conferenceworld.in, www.icstmdu.com
978-81-931039-6-8

TRANSDUCTIVE BASED COST-SENSITIVE MULTI-LABEL CLASSIFICATION

Ms.Pranali Dhongade[1], Ms. Sonali Bodkhe[2]

[1,2]Department of Computer Science and Engineering,

G. H. Raisoni Academy of Engineering and Technology, Nagpur, (India)

ABSTRACT

In last decade, the problem of multi-label classification gain huge importance. The Multi-label classification problem includes example which belongs to multiple labels. Current research on multi-label classification concentrated on supervised learning whose assumption is that huge amount of labeled training data is available. Unfortunately, in many applications labeling the training example is expensive and time consuming, mainly when it consist of more than one label. However, there are often large amount of unlabeled data is available. To solve the problem of labeling to unlabeled data transductive multi-label learning method will be used. In this paper, TRAM is used to assign multiple labels to each instance and then the result of transductive multi-label learning will be verified by Cost-sensitive Multi-label learning which will use to identify & minimize the misclassification cost of labels.

Keywords: Data Mining, Multi-Label Leaning, Unlabeled & labeled Data, Transductive.

I. INTRODUCTION

Traditional classification approaches considers each instance is associated with single label. On the other hand, various real world applications often involves multiple labels in which each instance is associated with a set of many labels [1] . For example, in an image annotation task an image may be applicable to three categories: mountains, sky and rivers [4][5]; a gene may have multiple functions, such as metabolism and energy [6]; In text categorization , a document may be associated with economics , sports and health [7]. In the literature, multi-label learning has been briefly studied [8]. Conventional approaches focuses on supervised setting where large amount of labeled data is necessary. On the other hand, in real world applications, the labeling process is very expensive and time consuming particularly with multi-label data. Creating a large training data set, where each example is labeled with a set of multiple labels within the candidate classes, is usually infeasible in practice [1]. For example, in image annotation task, for tagging set of possible labels to an image human specialist must go throughout the entire list of all candidate words so it consumes excessive time, efforts, and resources to manually tag each image with all its candidate labels, and as a result only a limited amount of labeled images can be obtained in practice. But there is often abundant large amount of unlabeled data is available and easy to obtain. To improve multi-label classification performance, it is much desired that the large amount of unlabeled data can be utilized together with the limited amount of labeled data.

One of the most popular approach is Semi-supervised learning [9]where unlabeled data is exploited to facilitate the learning process in addition to labeled data without human involvement. Transductive learning [11] is a type of approach which utilizes unlabeled data in classification processes. Testing data assumes by transductive learning by exploiting the unlabeled testing data in the classification process to achieve better performance. By utilizing the information from both labeled and unlabeled data Transductive Multi-label classification problem effectively assign a set of multiple labels to each instance. Under transductive setting, previously developed algorithms making predictions on existing unlabeled data although cannot generalize to new unseen data. we have also extended the TRAM [1] to handle cost-sensitive multi-label classification. In cost-sensitive multi-label classification aim is to reduce misclassification costs of instance pair labels.

II. RELATED WORK

2.1. Multilabel Classification

Algorithm adaptation and Problem transformation [1] are the two categories of Multi-label classification [12]. The 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 extend some specific learning algorithms for single-label classification to solve the multi-label classification problem [12]. Label power set (LP) [1] is multi-label classification algorithm which converts the multi-label classification problem to a single-label multi-class classification problem by considering different combination of labels in the training set as a separate class. The multi-class LP classifier predicts the most probable class, which can be transformed to a set of labels [12]. RAKEL is ensemble based multi-label classification methods which splits original label set into label subset and then arbitrarily selects the number of label subsets by training related multi-class classifier by using label power set [14].

Generalized k-Label sets Ensemble for multi-label classification method is presented by Lo, Lin and wang which is based on the concept of label powerset method. In this, LP classifier trained basis expansion model is used to reduce global error between predicted and ground truth label by learning the expansion coefficients efficiently and also extended this model for cost-sensitive multi-label classification to reduce misclassification cost [12]. Independent labels assumed by Decomposition based Binary Relevance learning method [16] where, it alters original dataset into subsets in which classifiers trains on each dataset. If label set is present in original dataset then it labeled as positive [15].In addition to the popular ensemble learning method ADABOOST Schapire and Singer presented BOOSTEXTER [17] which maintains a set of weights over both training examples and their labels, which will be incrementally enlarged if labels are difficult to be predicted correctly. Zhang and Zhou presented multi-label version Ml-KNN [18]. It is the extension of lazy learning algorithm kNN, It utilizes label prior probabilities achieved from each example's k nearest neighbors and use Maximum a Posteriori (MAP) principle to decide labels. For multi-label learning, BP-MLL is an adapted version of back-propagation algorithm uses novel error function to capture the characteristics of multi-label learning. In this, unrelated labels are ranked lower than related labels [19].

2.2. Transductive Learning

In this paper focused on transductive learning and cost-sensitive multi-label classification. Vapnik proposed Transductive learning [20] in which utilizes unlabeled data for labeling where testing data is exactly same as

unlabeled data. TRAM [1] is Transductive multi-label classification which effectively assigns set of multiple labels to each instance. On the other hand, supervised multi-label learning evaluates the set of labels of unlabeled data from the information of both labeled and unlabeled instances. On the basis of smoothness property first it evaluates label concept composition for multi-label instance after that make multi-label predictions based on those label concept compositions. In [1] Supervised version and Transductive version of label set prediction proposed by Kong Ng and Zhou. 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 [21]

On composite kernel, a graph-based Transductive multi-label classifier (TMC) is developed by Yu, Domeniconi, Rangwala, Zhang. In which they proposed data integration method by using ensemble framework i.e. transductive multi-label ensemble classifier (TMEC) [22]. TMEC trains a graph based multi-label classifier and then combine predicted output of the different models for each different kernel. To capture relationships between pairs of proteins and pairs of functions

Bi-relational directed graph is used.

2.3. Cost-Sensitive Multi-label Classification

In data mining, Cost-sensitive learning considers misclassification costs. The aim of Cost-Sensitive Learning is to reduce total cost. The goal of this type of learning is to follow a high accuracy of classifying examples into a set of known classes [23].GLE model is extended to handle the cost-sensitive multi-label classification problem. Apply it in social Tagging, assuming tag counts as the misclassification cost to handle the issue of the noisy training set [12]. Condensed filter tree (CFT) is novel algorithm for optimizing evolution criteria in CSMLC. CFT is derived from famous filter tree algorithm for cost-sensitive multi-class classification via constructing the label power set[24]. In this training and prediction is done by designing the tree structure and focusing on the key nodes for cost-sensitive multi-label classification.

III. TRANSDUCTIVE BASED COST-SENISITIVE MULTILABEL CLASSIFICATION

Consider entire dataset D = { x1, x2….xn } which consists of n instances (xi € IR^d). The data set is divided into both labeled and unlabeled instances.

we assume the first n1 (n1<<n) instances within D are labeled by { Y1, Y2…..Yn } where Yi is subset of C which denotes the set of multiple labels that are assigned to xi. Here, set of possible label concepts are C = { l1, l2,…..ln }.The multi-label classification task corresponds to finding an optimal label set Yi for each unlabeled instance xi in the space of label sets. On both labeled and unlabeled instances construct weighted neighborhood graph G (V, E) to find similar instances. Find k nearest neighbors for each instance by using kd-tree. Before Kd-tree MDDM i.e. multi-label dimensionality reduction approach is used to build KNN graphs which attempts to project the original data into a lower dimensional feature space maximizing the dependence between the original feature description and the associated class labels[1][26]. Define a sparse matrix W which denotes similarities between neighboring instances

$$W_{iz} = \begin{cases} \frac{1}{Z_i}\exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_z\|^2}{2\sigma^2}\right), & \text{if } z \in \mathcal{N}_i, \\ 0, & \text{otherwise.} \end{cases} \quad \text{…….. [1]}$$

Where,

Ni is the index set of ith instance's k nearest neighbors.

$\|\boldsymbol{x}_i - \boldsymbol{x}_z\|$ indicates Euclidean distance.

σ is the average distance between instances.

$Z_i = \sum_{z \in \mathcal{N}_i} \exp(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_z\|^2}{2\sigma^2})$

thus, $\sum_z W_{iz} = 1$ is for all instances.

For all unlabeled data, determine optimal alpha values by using following equation

$A_{\mathcal{UU}} \alpha_{\mathcal{U}j} = -A_{\mathcal{UL}} \bar{\alpha}_{\mathcal{L}j},$

Compute a sorted list of all potential labels for xi can be finding by ranking all candidate labels using their alpha values in descending order. Larger alpha value indicates the more likely xi will have the related label. The larger the alpha value is the more likely xi will have the corresponding label. For example, consider there are three class labels l1; l2; l3, and (l1 = 0.31 , l2 = 0.23 , l3 = 0.42) are the optimal alpha values. Then its sorted list is (l3, l1, l2).

$$\min_{\theta_1,\ldots,\theta_n} \sum_{i \in \mathcal{U}} \left(\theta_i - \sum_{z \in \mathcal{N}_i} W_{iz} \theta_z \right)^2$$

$$s.t.\ \theta_i = |Y_i|\ (\forall i \in \mathcal{L}).$$

Then predict optimal number of labels by using following equation

$$A_{\mathcal{UU}} \boldsymbol{\theta}_{\mathcal{U}} = -A_{\mathcal{UL}} \boldsymbol{\theta}_{\mathcal{L}},$$

Where , $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^\top = \left[\begin{smallmatrix} \boldsymbol{\theta}_{\mathcal{L}} \\ \boldsymbol{\theta}_{\mathcal{U}} \end{smallmatrix}\right]$.

After this operation we get predicted labels for each instance. We extend this TRAM for cost-sensitive multi-label classification. For cost-sensitive classification weak labels are taken into account. Basically untagged labels as well as higher misclassification cost labels are the weak labels. After applying TRAM weak labels are separated from dataset. After this operation we get predicted labels for each instance. We extend this TRAM for cost-sensitive multi-label classification. For cost-sensitive classification weak labels are taken into account. Basically untagged labels as well as higher misclassification cost labels are the weak labels. After applying TRAM weak labels are separated from dataset. Weak labels mark as 1 if label as assign is proper i.e predicted element% is same class otherwise assign 0 labels. We need to learn threshold to separates weak labels. Then build a positive semi-definite matrix W and partial label matrix to get cost-sensitive predicted label matrix. Decompose orthonormal eigenvector of W this belongs to the concept of multi-instance multi-label learning with defining C the parameter controls the similarity between a new learned kernel and the original one then [27] .

Table. I Summary of Experimental Datasets

Task Studied	DataSet	Instances	Attributes	Labels
Gene Function Analysis	Yeast	300	103	14
Natural Scene Classification	Scene	2407	294	6

IV. EXPERIMENTS

In this section, we show the performance of Cost-sensitive TRAM on several real-world multi-label classification tasks. Table I summarizes the data sets used. We conduct experiment on yeast and scene dataset for studying the task of the yeast gene functional analysis and natural scene classification (e.g., [6] and [25]). In [6], predicting the functional classes in the gene of yeast Saccharomyces cerevisiae. The data set is divided randomly into labeled/unlabeled data sets according to certain ratios. The whole data set of yeast consists of 2,417 instances of genes, 103 attributes and 14 possible class labels [1] and scene dataset consist of 2407 instances , 294 attributes and 6 labels.. In this paper TRAM i.e. transductive multi-label classification algorithm (implementation in MATLAB) is used for prediction of labels after then this algorithm is extended for cost-sensitive multi-label classification to reduce misclassification cost of instance label pairs. For this weak labels are separated and Cost sensitive TRAM is applied for prediction of labels of weak labeled instance pair. Table II summarizes the experimental results on yeast dataset. Table II shows the comparison of evaluation matrices of transductive based classification and transductive based cost-sensitive classification. Hamming loss is smaller in cost-sensitive TRAM as compared to TRAM. Micro F1 is larger in cost-sensitive TRAM rather than TRAM. No changes in Ranking loss and Average precision.

4.1. Evaluation Metrics

Multi-label classification systems require much more complex evaluation criteria than single label systems. In this paper, briefly summarize the criteria used for performance evaluation from various perspectives [1].

- **MicroF1** –

With equal significance it evaluates micro average of Precision and micro average of Recall with equal significance

$$MicroF1 = \frac{2 \times \sum_{i \in \mathcal{U}} |h(\boldsymbol{x}) \cap Y_i|}{\sum_{i \in \mathcal{U}} |h(\boldsymbol{x})| + \sum_{i \in \mathcal{U}} |Y_i|}.$$

For better performance, MicroF1 value must be bigger.

- **Hamming loss**

It evaluates how many times aninstance-label pair is misclassified.

$$HammingLoss(h, \mathcal{D}_{\mathcal{U}}) = \frac{1}{|\mathcal{D}_{\mathcal{U}}|} \sum_{i \in \mathcal{U}} \frac{1}{m} |h(\boldsymbol{x}_i) \Delta Y_i|,$$

where Δ stands for the symmetric difference of two

sets. For the better the performance, hamming loss value must be smaller.

- **Ranking loss.**

It evaluates the average fraction of instance label pairs that incorrectly ordered.

$$RankLoss(f, \mathcal{D}_{\mathcal{U}}) = \frac{1}{|\mathcal{D}_{\mathcal{U}}|} \sum_{i \in \mathcal{U}} \frac{1}{|Y_i||\overline{Y_i}|} |\{(y_1, y_2) \in Y_i \times \overline{Y_i} | f(\boldsymbol{x}_i, y_1) \leq f(\boldsymbol{x}_i, y_2)\}|,$$

- **Average Precision**

It evaluates the average fraction of labels ranked above a particular label y € Yi which actually is in Yi.

$$AvePrec(f, \mathcal{D}_{\mathcal{U}}) = \frac{1}{|\mathcal{D}_{\mathcal{U}}|} \sum_{i \in \mathcal{U}} \frac{1}{|Y_i|} \sum_{y \in Y_i} \frac{|\{y' \in Y_i | r_f(\boldsymbol{x}_i, y') \leq r_f(\boldsymbol{x}_i, y)\}|}{r_f(\boldsymbol{x}_i, y)}.$$

For better the performance value must be bigger. Note that all the criteria evaluate the of multilabel classification systems performance from different aspects.

Table .II Experimental Results on Yeast Datasets

Evaluation Metrices			RL ↓	AP ↑	HL ↓	Micro F1 ↑
TRAM	Number of Neirest Neighbors	K=8	0.17974	0.75721	0.21119	0.65393
		K=9	0.17861	0.76041	0.20823	0.65943
		K=10	0.17608	0.76131	0.20942	0.65633
		K=11	0.17495	0.76221	0.21119	0.65343
Cost-TRAM		K=8	0.17974	0.75721	0.19422	0.65794
		K=9	0.17861	0.76041	0.19156	0.66207
		K=10	0.17608	0.76131	0.19186	0.6603
		K=11	0.17495	0.76221	0.19437	0.65795

Table. III Experimental Results on Scene Datasets

Evaluation Metrices			RL ↓	AP ↑	HL ↓	Micro F1 ↑
TRAM	Number of Neirest Neighbors	K=8	0.80757	0.85847	0.094053	0.73281
		K=9	0.076452	0.85871	0.096819	0.72495
		K=10	0.074015	0.86155	0.094398	0.73156
		K=11	0.075778	0.86113	0.095781	0.72709
Cost-TRAM		K=8	0.080757	0.85847	0.080548	0.73166
		K=9	0.076452	0.85871	0.096819	0.72495
		K=10	0.074015	0.86155	0.07951	0.73129
		K=11	0.075778	0.86113	0.080548	0.72745

("↓ " Indicates "the Smaller the Better," and " ↑ " Indicates "the Larger the Better")

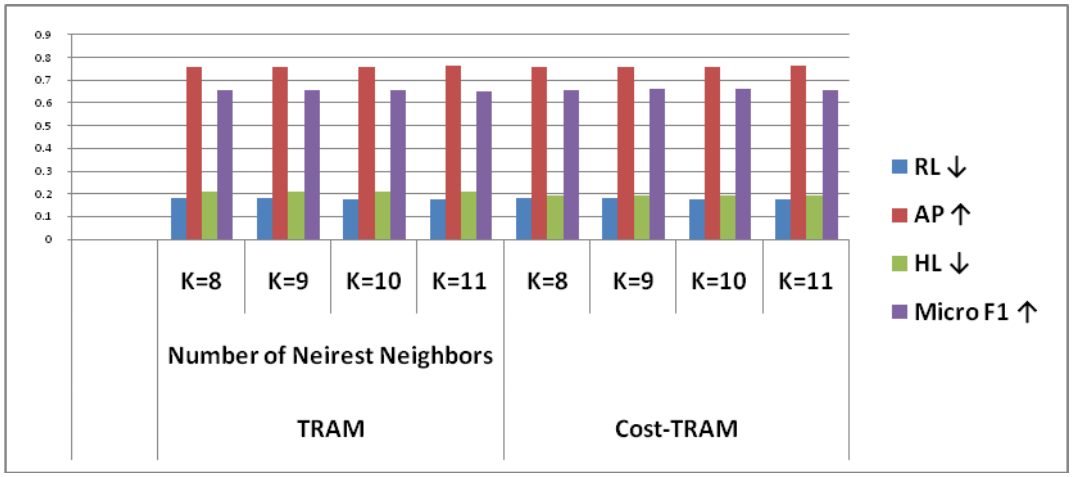

Fig. 1 Graph Showing Result Analysis on Yeast Dataset

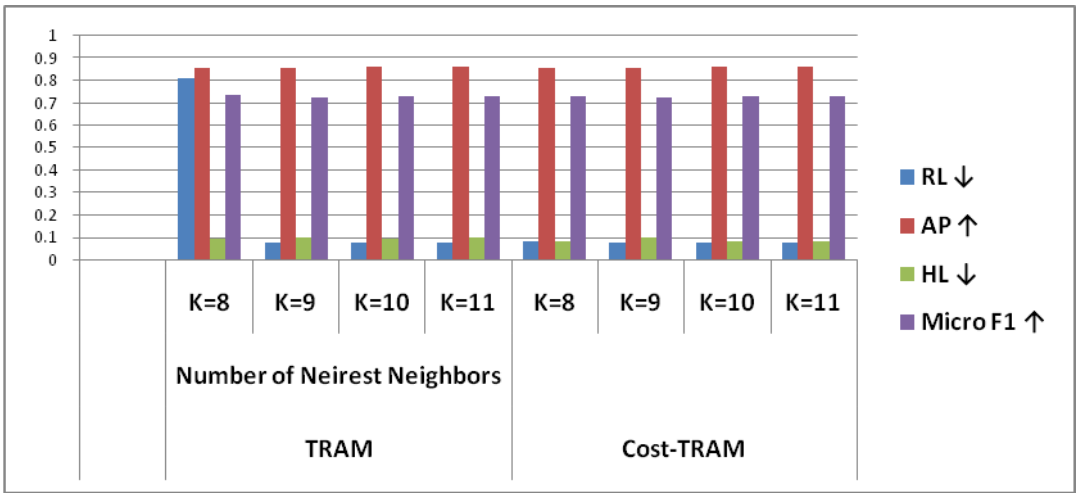

Fig. 2 Graph Showing Result Analysis on Scene Dataset

V. CONCLUSION

In this paper, our basic approach is to solve the problem of multilabeling. Transductive based multi-label classification is an effective way of assigning multi-label to each instance. In this TRAM algorithm is used to predict the labels of unlabeled data which utilize the information of label and unlabeled data which helps to optimize the problem of composite labeling. We extend this TRAM to cost-sensitive transductive multi-label classification method to reduce misclassification cost of instances which have weak labels. Experiential studies focuses on real-world tasks of yeast gene function analysis and natural scene classification and demonstrate that our Transductive based Cost sensitive method can effectively enhance the performance of multi-label classification by using unlabeled data in addition to labeled data.

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] G. Carneiro, A. Chan, P. Moreno, and N. Vasconcelos, "Supervised Learning of Semantic Classes for Image Annotation and Retrieval," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 3, pp. 394-410, Mar. 2007.

[3] G.J. Qi, X.S. Hua, Y. Rui, J. Tang, T. Mei, and H.J. Zhang, "Correlative Multi-Label Video Annotation," Proc. 15th Int'l Conf. Multimedia, pp. 17-26, 2007.

[4] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," Pattern Recognition., vol 37, 2004, pp. 1757-1771.

[5] S. J. Huang, and Z. H. Zhou, "Multi-label learning by exploiting label correlations locally," in Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI'12), Toronto, Canada, pp. 949-955, 2012.

[6] A. Elisseeff, and J. Weston, "A kernel method for multi-labelledclassificaiton," in Advances in Neural Information Processing Systems 14, 2001, pp. 681-687.

[7] A. McCallum, "Multi-label text classification with a mixture model trained by EM," in Working Notes of the AAAI'99 Workshop on Text Learning, Orlando, FL, 1999.

[8] G. Tsoumakas and I. Katakis, "Multi-Label Classification: An Overview," Int'l J. Data Warehousing and Mining, vol. 3, no. 3, pp. 1-13, 2007.

[9] T. Joachims, "Transductive Inference for Text Classification Using Support Vector Machines," Proc. 16th Int'l Conf. Machine Learning, pp. 200-209, 1999.

[10] X.-J. Zhu and A.-B. Goldberg. Introduction to semi-supervised learning. In R. Brachman and T. Dietterich, editors, Synthesis Lectures on Arti_cial Intelligence and Machine Learning, pages 1{130. Maogen and Claypool, 2009.

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

[12] 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

[13] 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.

[14] GrigoriosTsoumakas , loannisKatakis and loannisVlahavas, " Random K-labelsets for Multi-label Classification" IEEE Transactions On Knowledge And Data Engineering,

[15] 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

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

[17] 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

[18] 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.

[19] 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.

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

[21] Ms. Pranali Dhongade, Mr. Rushi Longadge, Mr. Deepak Kapgate, "A Review on Classification of Multi-label Data in Data Mining", IJCMLC Vol.3 Issue.12, December- 2014, pg. 189-196

[22] 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.

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

[24] 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.

[25] P. Pavlidis, J. Weston, J. Cai, and W.N. Grundy, "Combining Microarray Expression Data and Phylogenetic Profiles to Learn Functional Categories Using Support Vector Machines," Proc. FifthInt'l Conf. Computational Biology, pp. 242-248, 2001.

[26] Yin zhang ,zhi-hua Zhou , "multilabeldimensinality reduction via dependence Maximization"AAAI conference on AI 2008.

[27] Shu-jun Yang , Yuan Jiang , Zhi-Hua Zhou , "Multi-Instance Multi-label Learning with weak label" , NFSFC

[28] Yu-Yin Sun , Yin Zhang , Zhi-Hua Zhou , "Multi-label Learning with weak label" , NFRPOC 2010