Multi Label Classification based on Logistic Regression (MLC-LR)

Prathibhamol C P
Department of CSE
Amrita School of Engineering
Amritapuri,
Amrita Vishwa Vidyapeetham
Amrita University
India
prathibhamolcp@am.amrita.edu

Jyothy K V
Department of CSA
Amrita School of Engineering
Amritapuri,
Amrita Vishwa Vidyapeetham
Amrita University
India
jyothykvenu@gmail.com

Noora B
Department of CSA
Amrita School of Engineering
Amritapuri,
Amrita Vishwa Vidyapeetham
Amrita University
India
noorabasheer313@gmail.com

***Abstract*— Numerous class labels associated with each data instance is the main feature of any multi-label classification (MLC) problem. Correct prediction of class labels related to any test data is a big challenge in this domain. MLC can be applied in many fields such as personality prediction, cancer prognosis prediction, image annotation etc. In this paper(MLC-LR), we have employed problem transformation method for solving MLC. The proposed method uses initially clustering in the feature space. It is then followed by FP-growth algorithm for finding the relationship between labels. Once the desired clusters are obtained, then normalization of data associated with each cluster is performed. Also logistic regression is then applied over the normalized data for each particular cluster pertaining to all labels. When a new instance arrives in the testing phase, immediately the nearby cluster is identified by means of Euclidean distance metric as the measure. Rules related to label space for the nearby cluster is extracted to check for hypothesis of each antecedent label. If the calculated value is higher than a predefined threshold, it is assumed that both antecedent and consequent labels as the estimated labels for that test instance.**

Keywords-k-means, FP growth, logistic regression, multi-label data set.

I. INTRODUCTION

In traditional classification methods each instance is related with single label which makes it easy to classify. For example if we are classifying people based on their citizenship they will be having only one label since they cannot have citizenship of two countries. But today we are mostly dealing with multi label data, that is set of labels are associated with single instance. The set of label may be different for each instance. There are many situations in real life where we need to classify these multi-label data. An example shown in Fig.1, if we classify people based on their diseases, each person can comes under multiple categories such as heart patient, kidney patient etc. For solving these types of problems MLC is used.

The two methods commonly used for solving MLC, are algorithm adaptation method and problem transformation method. In the former multi-label problem is transformed into binary classification problems which can then be solved using single class classifiers. In the latter one algorithm is adapted directly to accomplish MLC

patient	Disease Type(Class Label)		
	Heart	Kidney	Brain
A		✓	✓
B	✓		✓
D		✓	✓
D	✓		✓

Fig. 1. Multi-label classification

There are two phases in our work, training phase and testing phase. In training phase k-means algorithm is used on the feature space of the multi label data set in order to make number of clusters with similar features. k-means is a simple, unsupervised learning approach which resolve the well-known clustering problem. Then FP growth algorithm is used in the label space for finding the association between the labels. FP-Growth stands for frequent pattern growth. It is a scalable technique for excavating frequent pattern in a record. Once the desired number of clusters with association rules are obtained, then logistic regression approach is applied on the clustered data for each label. Prior to logistic regression normalization will be performed on the data as a preprocessing step. Logistic regression is a technique used for studying a data set in which there are one or more independent variables that find out a result. There will be only two possible outcomes in logistic regression such as pass/fail, absent/present etc. It helps in predicting the probability of presence of a label. When the training phase is over, testing phase is started by introducing a test data instance. Using Euclidean distance calculation we can get the nearest

978-1-5090-2029-4/16/$31.00 @2016 IEEE

cluster of the test data and the test data belongs to that particular cluster. In Euclidean distance calculation the test instance is being compared with the cluster centroids. This reduces the execution time. That is the test instance is checked only for the corresponding rules of that particular cluster. In testing phase, for each antecedent label, corresponding theta values obtained from the training phase are taken. For the new data instance theta transpose x is measured, where x is the attribute values. If the calculated value is greater than or equal to zero both antecedent and consequents labels are considered as the class labels for the test instance. The process is repeated for the next rule.

II. LITERATURE REVIEW

In [1] many methods have been described for resolving problems related to multi-label classification, which we use in various application. They have used some common evaluation metrics such as subset loss and hamming loss. Some of them are concentrated on probabilistic models. New probabilistic methodologies has been cultivated for multi-label classification. One important way of doing this is centered on nearest neighbhour classifiers and logistic regression. In this method, the training phase is very similar to binary relevance method. In this method, real problem has to be converted as number of single classification problems and apply binary classification to each of these problems. Testing phase of this method includes three phases. Finding out the set of labels associated with test sample is the aim of the first phase. The process of applying binary logistic regression on the data set is elaborated in the second phase. This is for predicting the labels of every test samples. According to the outcome of the first two phases, group of labels is alloted to every test sample in the third phase. Another approach is grounded on labels which are mutually related which in turn form groups. One classifier is derived for every sets in this approach and the resultant label is named as group representative. By considering the labels which are predicted for the group representative gues the other labes. Main aim of this approach is to group the correlated labels together.

There exists many multilabel classification methods. In [2] the authors suggests an easy and effective context for multilabel classification. The method is known as group sensitive classifier chains. Here actual feature space is expanded with label space, then it is clustered into different groups. Then study the label dependency graph in each group and form the classifier chains on each group particular label dependency graph. This group specific classifier chains are constructed on the nearby group of the test sample are used for prediction.

Another approach is using a combination of apriori and ML-KNN. Apriori algorithm is used to search the relationship between every labels. When frequent item sets are being generated in an iterative manner, already existing single labels replace compound labels which has strong association. Then it uses ML-KNN algorithm to classify the multi label data set. While predicting the labels of new instance, compound labels are filled based on the relationship between labels [3].

Fuzzy rules are also used for solving multi label classification problem [4]. Multi-label classification problem is switched over into various single label classification probelms by using various approaches. These alterations make it possible to use single label classifiers for the classification. So this approach doesnt use any type of algorithms for the classification purpose. Fuzzy rules offer interpretability to the model between different labels. Paper results shows that classifier chain and binary relevance methods are good for multilabel problem transformation technique.

Authors put their focus on dependencies between labels in a probabilistic method in paper [5], and also they mine characteristic structures in a system of probabilistic distribution functions by using some data mining practices. They show some experimental outcomes, that is dependencies among labels to ensure the effectiveness of the method.

One of the enhanced MLC method is based on label constraints and label ranking strategy [6]. One Vs all deomposition technique can be used to split an MLC problem into numerous seperate binary classification sub-problems. Different class labels are trained for each label. Then based on training data, association rules between labels are mined. Then a label constraint based correction model is used to make up the output of this classifier.

Label compression coding method is a different way of solving multi label classification problem[12]. In this method, the dependence between labels and features is maximized using Hilbert-Schmidt independence criterion and thereby considering both label and feature information simultaneously. A small-scale coding matrix and a fast decoding operation are results of this method, obtained by working out an eigenvalue problem.

combination of multi-kernel fusion and multi-instance multi-label learning is adopted for multi label classification problem of scene data set[13]. By framing graphs, correlation between instances are put forward first. Then derivation of different kernel matrices is performed by kernel functions based on graphs in different scales. And to predict multi labels multi-kernel SVM classifier based

on multiple-kernel fusion is used.

Binary relevance method trains a single binary model for each different label ignoring other labels. Chaining and stacking techniques are used to derive the dependences among labels to improve accuracy for MLC in predicting labels. These two techniques are put together to get dependent binary relevance (DBR) method which is an advanced form of binary relevance method [14]. In this work a pruning method is adopted for DBR. This method uses a Phi coefficient function to calculate correlation degrees among labels for discarding irrelevant labels.

The proposed method for multi label classification in [15] is ML-PLS which is based on Partial Least Squares (PLS) regression. PLS can deal with the relations between the matrices of independent variables and dependent variables through a multivariate linear model. In this approach, the details of the label memberships are included in the matrix of dependent variables and thereby predicting the labels of dependent variables through the multivariate linear model.

III. METHODOLOGY

In training phase as depicted in Fig.2 k means is applied on the feature space of the data set. k is the count of clusters to be formed as well as the number of centroids to be generated. Once the desired clusters are obtained the corresponding labels of that cluster is also extracted. Clustering is the grouping of items into various sets, so that the data in each cluster have some general characteristics or features often agreeing to some well-defined distance measures. k-means is an algorithm for assigning n objects into exactly one of the k groups represented by centroids and k is set before the execution of the algorithm. The grouping is done by calculating the Euclidean distances between data and the corresponding cluster centroid.

FP growth algorithm is applied on label space to obtain the association rules between labels. For each cluster there will be different set of association rules. As mentioned above[3], in this approach, Apriori algorithm is used for mining the association rules between the labels. In terms of efficiency FP-Growth is better than Apriori algorithm.

The end of the training phase is marked by normalization of data followed by logistic regression on it. So after logistic regression , hypothesis for each cluster related with all labels are obtained.

Finally we get the weights of θ. For this, initialize weight or theta value to be zero or it can also be selected randomly. Hypothesis function is defined as

$$g_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-z}}, where z = \theta^T x \qquad (1)$$

Using hypothesis function (1), calculate the θ value. θ can be calculated using any method, where as the correct value is returned. Using gradient descent θ value can be calculated as,

$$\theta_{present} = \theta_{past} - \alpha \sum_{i=1}^{m} (h_\theta(x^i) - y^i) x_j^i \qquad (2)$$

where α is the learning rate, Which can be set a default value 0.1, for the best results it can be varied with cross validation.

For predicting the class of a new instance, compute the sigmoid or logistic function. θ^Tx is greater than or equal to zero, when the logistic function is higher than or equal to 0.5, then the class is predicted as one. When θ^Tx is less than zero, then the logistic function is less than 0.5 and class is predicted as zero[7].

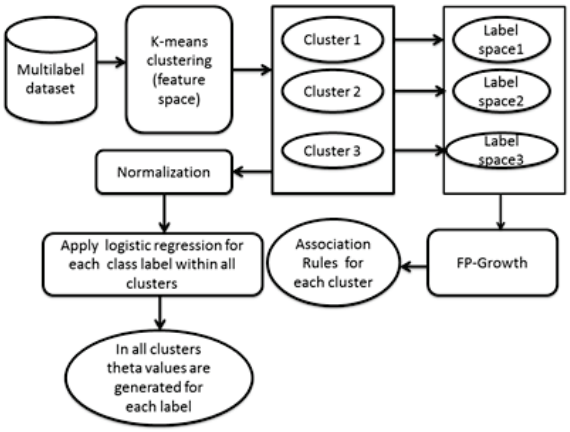

Fig. 2. training phase

In testing phase as in Fig.3 when a new instance is given, first, we have to find the nearby cluster by comparing distance between the new instance and the centroid of each clusters. The new instance belongs to that cluster, which has the minimum distance between the centroid and the new instance. Each cluster will be having a set of association rules, which are obtained from the training phase. Then we need to predict the labels of the new instance. Suppose one of the association rule between the labels is

L1=>L2,L3

it means that if the class label L1 (antecedent label)

is present, then we can ensure L2 and L3(consequent label) are also present for the instance. Presence of the antecedent label is found by taking the theta values generated in the training phase. Finally we can take θ^Tx. If θ^Tx is greater than or equal to zero, we can say hθ(X) is greater than or equal to 0.5. That means label L1 along with L2 and L3 are present for the instance.

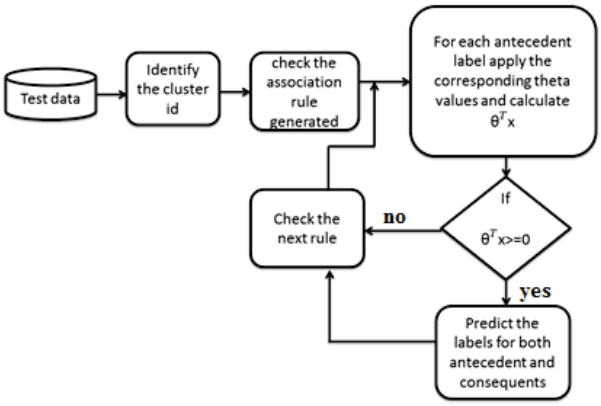

Fig. 3. testing phase

IV. EXPERIMENT AND EVALUATIONS

For testing the accuracy of our prospective method, yeast dataset from http://mulan.sourceforge.net/datasets-mlc.html is considered.

The yeast dataset includes the information about various kinds of genes of one specific organism. It includes 1500 objects having 14 labels and 103 numerical valued attributes.

This scene dataset includes various kinds of scene environmental information such as beach, mountain and sunset. It includes 2407 objects having 6 labels and 294 numerical valued attributes.

A. Evaluaiton Metrics

In this paper four commonly used evaluation metrics have been used for comparing the proposed method with existing multi label classification methods[8]. That are Hamming loss, precision, recall, accuracy.

Hamming Loss: It is the fraction of labels that are incorrectly predicted. Here it will check the predicted label with the original label. It will predict one if the labels are wrong and it will predict zero if the labels are correct.

$$HammingLoss(x,y) = \frac{1}{|m|}\sum_{i=1}^{|m|}\frac{|x_i \oplus y_i|}{|L|} \quad (3)$$

Here x_i indicates the actual label and y_i indicates the predicted label of the i^{th} instance. Here L and m stand for total count of labels and count of objects tested respectively.

Precision: It is the probability that the retrieved instances that are relevant. That means it evaluates the number of exact matches gained between the count of predicted and actual labels.

$$Precision = \frac{1}{|m|}\sum_{i=1}^{|m|}\frac{|x_i \cap y_i|}{|y_i|} \quad (4)$$

Recall: It is the probability that the relevant instances that are retrieved. It evaluates the number of correct matches gained between number of the predicted label and the count of actual labels.

$$Recall = \frac{1}{|m|}\sum_{i=1}^{|m|}\frac{|x_i \cap y_i|}{|x_i|} \quad (5)$$

Accuracy:It is an extent to which a predictive model is correct when applied to data[9].

$$Accuracy = \frac{1}{|m|}\sum_{i=1}^{|m|}\frac{|x_i \cap y_i|}{|x_i \cup y_i|} \quad (6)$$

V. RESULTS AND DISCUSSIONS

The above mentioned evaluation matrices are used for comparing various algorithms of multi label classification with our method.

From the Table I and II, the results shows that the methods used in this paper is more effective than existing multilabel classification algorithms. In K-means algorithm number of clusters is taken as 4. Minimum support in FP growth algorithm is set as 2 and minimum confidence as 75 percentage. The evaluation results of algorithms ML-KNN, CLR, RAKEL, I-BLR on scene and yeast data were referred from [10] and the evaluation results of C4.5, Binary-SVM and Nave Bayes were taken from [6]. Evaluation results of SMO and kNN are taken from [11].

VI. CONCLUSIONS

The proposed approach in this paper is logistic regression based multilabel classification. K-means algorithm is applied for clustering the dataset and FP-growth algorithm for generating the association rules between the labels. Finally the logistic regression algorithm is

TABLE I
EXPERIMENTAL RESULT (YEAST DATASET)

Algorithms	Hamming Loss	Accuracy	Precision	Recall
ML-KNN	0.198	0.492	0.732	0.549
C4.5	0.259	0.423	0.561	0.593
Nave Bayes	0.301	0.421	0.610	0.531
Binary-SVM	0.202	0.530	0.586	0.633
CLR	0.210	0.497	0.674	0.596
RAKEL	0.244	0.465	0.601	0.618
I-BLR	0.199	0.506	0.712	0.581
kNN	0.208	0.514	0.623	0.665
SMO	0.233	0.337	0.337	0.748
MLC-LR	0.119	0.751	0.845	0.814

TABLE II
EXPERIMENTAL RESULT (SCENE DATASET)

Algorithms	Hamming Loss	Accuracy	Precision	Recall
ML-KNN	0.099	0.629	0.661	0.655
C4.5	0.148	0.576	0.579	0.588
Nave Bayes	0.139	0.605	0.615	0.624
Binary-SVM	0.103	0.702	0.715	0.720
CLR	0.122	0.577	0.600	0.669
RAKEL	0.112	0.571	0.598	0.612
I-BLR	0.091	0.647	0.676	0.655
kNN	0.125	0.637	0.651	0.669
SMO	0.114	0.571	0.628	0.596
MLC-LR	0.090	0.774	0.783	0.934

hypothesized for predicting the class labels of the instances. The result confirm that logistic regression based multilabel classification is more effective. In future, it has been planed to perform the proposed method on various datasets with advanced clustering methods. The result confirms that logistic regression based multi-label classification is more effective than any other existing approaches.

REFERENCES

[1] Gangadhara Rao Kommu, M.Trupthi, Suresh Pabboju, "A novel approach for multilabel classification using probabilistic classifiers", 2014 International Conference on Advances in Engineering and Technology Research (ICAETR), pp. 1-8, Aug. 2014 .

[2] Jun Huang,Guorong Li, Shuhui Wang, Weigang Zhang,Qingming Huang,"Group Sensitive Classifier Chains For Multilabel Classification",2015 IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6, 2015.

[3] Feng Q in,Xian Juan,Tang Ze-Kai Cheng"Applications of Apriori Algorithm in Multi label Classification using probabilistic classifiers",13 Proceedings of the 2013 International Conference on Computational and Information Sciences(ICCIS), pp.717-720, June, 2013.

[4] Ronaldo C.Prati,"Fuzzy rule classifiers for multi-label Classification",2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp.1-8,Aug.2015.

[5] Yuichiro Kase, Takao Miura,"Multi-Label Classification Using Labelled Association",2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), pp.90-95, Aug.2015.

[6] Benhui Chen, Xuefen Hong, Lihu Duan and Jinglu HU," Improving Multi-Label Classification Performance by Label Constraints",2013 International Joint Conference on Neural Networks, IJCNN 2013 - Dallas, TX, Aug.2013.

[7] Ameena Ismail, Unmesha Sreeveni U.B, Shiju Sathyadevan, Jayasree Narayanan,"ABDF Integratable Machine Learning Algorithms-MapReduce Implementation", Second International Symposium on Computer Vision and the Internet(VisionNet15).

[8] Jiawei Han, Micheline Kamber, Jian Pei ,"Data Mining Concepts and Techniques",Third edition

[9] $https://en.wikipedia.org/wiki/Multi-label_classification$

[10] Muhammad Atif Tahir, Josef Kittler, Krystian Mikolajczyk, and Fei Yan, Improving Multilabel Classification Performance by Using Ensemble of Multi-label Classifiers, 9th International Workshop, MCS 2010, Cairo, Egypt,pp.11-21, April 2010.

[11] Grigorios Tsoumakas, Ioannis Katakkis," Multi-label Classification:An Overview", Int J Data Warehousing and Mining,2007.

[12] Lei Cao, Jianhua Xu,"A label compression coding approach through maximizing dependence between features and labels for multi-label classification",2015 International Joint Conference on Neural Networks (IJCNN),pp.1-8,July 2015.

[13] Chen Tong-tong , Liu Chan-juan , Zou Hai-lin , Zhou Shusen, Liu Ying , Ding Xin-miao ," A multi-instance multi-label scene classification method based on multi-kernel fusion ",SAI Intelligent Systems Conference (IntelliSys),pp.782 - 787,Nov. 2015.

[14] Yahong Zhang, Yujian Li , Zhi Cai,"Correlation-based pruning of dependent binary relevance models for Multi-label classification",2015 IEEE 14th International Conference on Cognitive Informatics and Cognitive Computing (ICCI*CC),pp.399 - 404,July 2015.

[15] Qiande Ren , Farong Zhong ," A multi-label classification algorithm based on Partial Least Squares regression", 2012 3rd International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM),vol:2,pp.21-24,Oct 2012.

[16] Prathibhamol C. P,Asha Ashok, "Solving Multi Label Problems with Clustering and Nearest Neighbor by Consideration of Labels", Advances in Signal Processing and Intelligent Recognition Systems. Springer International Publishing, pp. 511-520, 2016.