Exploiting Label Dependency and Feature Similarity for Multi-Label Classification

Prema Nedungadi, H. Haripriya
Amrita CREATE, Amrita University

***Abstract* — Multi-label classification is an emerging research area in which an object may belong to more than one class simultaneously. Existing methods either consider feature similarity or label similarity for label set prediction. We propose a strategy to combine both k-Nearest Neighbor (kNN) algorithm and multiple regression in an efficient way for multi-label classification. kNN works well in feature space and multiple regression works well for preserving label dependent information with generated models for labels. Our classifier incorporates feature similarity in the feature space and label dependency in the label space for prediction. It has a wide range of applications in various domains such as in information retrieval, query categorization, medical diagnosis and marketing.**

Keywords—multilabel;multiple rgression;kNN

I. INTRODUCTION

In today's world we are mostly dealing with multi-label data, data assigned to more than one class or label. A lot of ongoing research is focusing on multi-label data sets. Let us consider some real world examples for multi-labeled data. A text document about scientific contributions in medical science can belong to both science and health categories. An image that captures a field and fall colored trees can belong to both field and fall foliage categories. A newspaper report may belong to a person, a country, some local or national or international news. A movie can simultaneously belong to action, crime, thriller, and drama categories. An email message can be tagged as both work and research project, etc.

Thus, in single label classification, an object x_i is assigned to one class c_i but in multi-label classification an object x_i will be assigned to a set of classes $c_1, c_2 \ldots c_l \subseteq$ L simultaneously, where L is the total number of classes or labels. There are two main approaches for dealing with multi-label classification [11] problems, problem transformation method and algorithm adaptation method. Problem transformation method converts multi-label problems into a set of single label problems. It transforms multi-label datasets in such a way to fit for traditional algorithms. Here, the preprocessing part is the transformation of dataset, and the classification or prediction part is similar to already existing mechanisms. But in algorithm adaptation method, an extension of already existing single label learning algorithms is used for predicting the label set of multi-labeled data. In this approach a modification is made on label prediction or error computation to handle multi-label data.

In our work, we combined multiple regression with kNN for label set prediction for incorporating feature similarity in the feature space and label dependency in the label space for label set prediction. Multiple regression is used for generating models for label sets. The generated model is later used for label set prediction. The nearest neighbor of the test data is calculated using the traditional k Nearest Neighbor algorithm. A combination of label similarity and feature dependency method produces a more effective text classifier. Our multi-label classifier has a wide range of applications in various domains such as: information retrieval, query categorization, medical diagnosis, marketing, etc.

The rest of the paper is organized as follows. Section II explains the related works of multi-label classification. Section III, IV explains briefly the basic works related to our paper. Section V discusses our proposed approach in detail. Section VI presents experimentation and result analysis. Section VII is a brief conclusion of our work.

II. RELATED WORK

Most of the existing classifiers in machine learning[3] are dealing with single label classification. Methods to extend existing classifiers in order to deal with multi-labeled data is discussed in [1,2,10,14,16,17]. Some methods also under research [4,11,13] convert multi-label dataset into a set of single label dataset to fit with existing classifiers.

ML-kNN [1] traditional k-Nearest Neighbor algorithm and Maximum a posteriori (MAP) principle is used for label set prediction. A Ranking-based kNN is an approach for Multi-Label Classification[2,11], is a k-nearest-neighbor-based ranking approach to solve the multi-label classification problem. Ranking SVM and kNN concepts are used for multi-label prediction. An approach to analyze features for specific label is discussed in [17]. But these approaches have not considered interdependencies between the label sets and thus ignored the possibility of co-occurrence of labels.

A Naive Bayesian Multi-label Classification Algorithm With Application to Visualize Text Search Results [4] is a problem transformation approach. During dataset conversion phase there may be a loss in label dependent information. Naive Baye's ignores feature dependency and so in real world application it will result in a decrease in prediction accuracy.

978-1-4799-3080-7/14/$31.00 ©2014 IEEE

In[6],[7] and [8] Ranking SVM is used for document retrieval. Ranking SVM belongs to the category of multi-label in which a single query matches with multiple documents. In the case of nonlinear datasets, and when the number of samples is less than the number of features, the possibility of low accuracy will be greater.

A decision tree algorithm C4.5 [9] is used for the analysis of phenotype data is discussed in [10]. It is simple and easy to learn. More informative attributes are used for tree splitting. But attempts for generalization results in decrease in performance . BPMLL[11] is a neural network approach to deal with multi-label problems. It is an extension of traditional back-propagation algorithm by introducing a new global error function that captures the characteristics of multi-label learning. But the neural network complexity becomes high in training phase.

A probabilistic kNN and logistic regression based approach for label set prediction is discussed in [12] . The distance between neighboring instances and their labels are used for prediction. Random k-Labelsets(Rakel) [13] divides the label set and considers the label correlation ship. But both these approaches consumes more time in applications with large number of training samples and labels.

AdaBoost[14,15] creates accurate hypothesis by utilizing a set of weak hypotheses. It uses information from misclassified data. This algorithm is extended to handle multi-labeled data, but it is sensitive to noisy data and outliers.

A method to reduce dimensionality reduction in label space with association rule is discussed in [16]. An association rule mining algorithm is applied in order to reduce label cardinality, as well as total number of labels in the dataset. But this approach does not guarantee reduction without information loss. Dimensionality reduction technique in feature space [5] is used as a preprocessing phase of kNN in order to cope with handling big data. By searching k neighbors along each principal component after reducing the number of dimensions. It's not meant for multi-label classification.

In our proposed approach we combined the advantages of kNN and multiple regression for multi-label classification. We considered both feature space and label space for efficient prediction.

III. MULTIPLE REGRESSION

A simple linear regression model predicts a single variable based on one dependent variable. The response measurement Y_{pred} is dependent on variable X for each observation. The model assumes that the function is linear.

$$Y_{pred} = \alpha + \beta X \quad \text{---> Eqn (1)}$$

where Y_{pred} is associated with X in a linear way β is the co-efficient of X and α is the intercept.

Multiple regression is an extension of simple linear regression for incorporating the dependencies of more than one variable. we considered the dependence property of label in a specific way multiple regression is used for label prediction.

The main objectives of multiple regression are:

1. To predict a variable from a combined knowledge of several other variables.
2. To determine which labels of a larger set are contributing for the prediction of some criterion label than others.
3. To understand the percentage of accuracy for predicting a variable if we add one or more predictor variables.

The equation for multiple regression is:

$$Y_{pred} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n \quad \text{---> Eqn (2)}$$

Y_{pred} is the variable to be predicted, the X's are the dependent variables of the response variable or the predictors, and the β's are the weights or coefficients associated with the predictors.

IV. TRADITIONAL K NEAREST NEIGHBOUR

k Nearest Neighbor algorithm is a supervised machine learning algorithm. A similarity computation need to be performed with all the training data to predict the class of a newly arrived data. The most similar *k* neighbors from the training data are used for testing. We used cosine similarity for similarity computation. Consider two feature vectors x and y, cosine similarity is computed by using the Eqn(3)

$$\cos(\Theta) = similarity(x, y) = \frac{x.y}{\|x\| * \|y\|} \quad \text{----> Eqn (3)}$$

The similarity values range from 0 to 1. When the value is 1 they are equal or most similar, and when it is 0 they are least similar. Cosine similarity will generate a metric that indicates the similarity of two vectors by looking at the angle instead of magnitude.

V. SYSTEM ARCHITECTURE

Our objective is to classify data into more than one class instead of a single class. Most of the real datasets belong to multiple classes. We combined traditional kNN and multiple regression to predict multiple labels of data. When test data is given to our multi-label classifier, it will predict the label set of the test data as shown in Fig 1.

Fig. 1. Overview of Multi-label Classifier

Algorithm 1 trains the multi-label classifier. Multiple regression is used to obtain the information from dependent labels, i.e., from label space. We are assuming that for the occurrence of one label, all other labels will contribute. Most of the real world problems demonstrate label co-occurrence. Multiple regression predicts one label by considering all other existing labels. A weight is associated with the labels that describe its dependency. Labels from the training set generate a linear model for each label. Thus, the training phase is actually a model generation phase for label sets.

A. Algorithm 1 - MULTILABEL_CLASSIFIER_TRAIN ( )

Input: {Labels of training dataset}
Output: {Model generation for each Label}

1: Read the Class labels datasets.
2: Generate models using multiple regression
3: *for i :1 to L*
  Use Eqn(2) to generate model for *i-th* label
4: *end*
5: Output the generated models of each label

Algorithm 2 is our multi-label classifier. We used traditional kNN for finding the most similar *k* neighbors from our training set. Step 2 identifies the nearest neighbors. Steps 3 and 4 predict the label value by using its *k* neighbors. Then, compute average of the predicted value along each label . If the average value is less than a threshold value, the test has no label named *i* and if its average value is greater than threshold value, it has the label *i*. The label set of neighbors from the feature space predicts the labels. Steps 5 to 7 is for label set prediction.

B. Algorithm 2 - MULTILABEL_CLASSIFIER _TEST( )

Input: {Test data, Generated models}
Output: {Label set Prediction}

1: Read the test data
2: Find its *k* nearest neighbors using kNN algorithm
3: *for i :1 to k*
  Each Label value prediction using generated model
4: *end*
5: Average predicted value along each label
6: *for i :1 to L*
  if (average < threshold) then
    Test data has no label named *i*
  else
    Test data has label named *i*
7: *end*
8: Output the predicted label set

Fig 2. depicts the architecture of the training phase multi-label classifier. A literature survey on multi-label classifiers indicates that a little research has been done on multi-label classifiers with class correlation and feature similarity. So our work combined multiple regression as well as kNN approach to classify a multi-labeled data by considering the above mentioned aspects.
Multiple regression considers the class dependence for creating models. A model is generated for each class by incorporating the knowledge obtained from the class labels of training data.

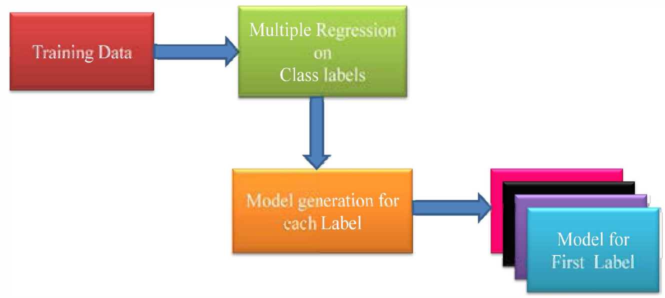

Fig. 2. Training phase of Multi-label Classifier

Fig 3. shows the architecture of the testing phase of multi-label classifier. When test data arrives, a search over the entire feature space is performed by kNN. Here, we have incorporated feature-dependent neighbor selection. The neighbors in the training set have both feature set and label set. The model obtained as a result of multiple regression is used for label set prediction from the label space.

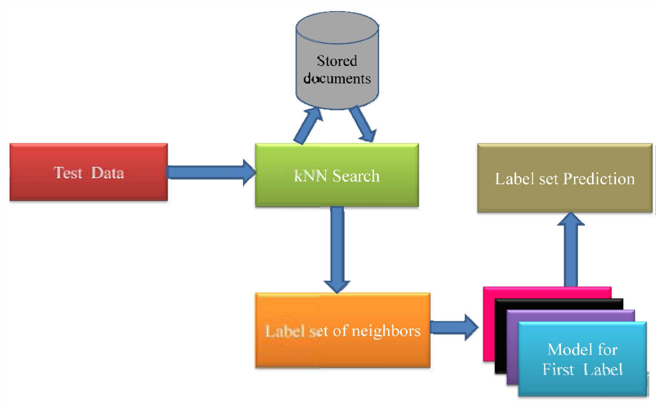

Fig. 3. Testing phase of Multi-label Classifier

VI. Experimental Evaluation

A. Evaluation Measure

The performance evaluation of a multi-label learning system is different from that of classic single-label learning system. Existing evaluation metrics for a single-label system include accuracy, precision, recall and F-measure. Whereas, in multi-label, the evaluation has to be made in a different way because it deals with multiple targets. We used hamming loss for evaluating the obtained clusters.

Hamming loss: It evaluates how many times an instance–label pair is misclassified [1,2], which means a label not belonging to the instance is predicted or a label belonging to the instance is not predicted. The performance is perfect when hamming loss is zero. The smaller the value of hamming loss, the better the performance. The value of hamming loss ranges between 0 and 1.

$$\text{hloss}_S(h) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{Q}|h(x_i)\Delta Y_i|$$

where Δ is the symmetric difference between two sets.

We have used a multi-label dataset from Knowledge Extraction based on the Evolutionary Learning (KEEL Dataset) data repository and from the UCI repository. We have evaluated the result with Yeast, Scene and Emotions datasets.

We have used 70 percent of our dataset for training and 30 percent for testing. The hamming loss computed after using our proposed method and MAP on the Yeast dataset is shown in Table I. The Yeast dataset has 103 attributes and with 14 labels. Table 1 shows even if MAP has a low value for hamming loss compared to Regression, it is almost equal at k =6.

TABLE I. HAMMING LOSS ON YEAST DATASET USING REGRESSION AND MAP

Yeast	Hamming Loss					
	k=5	**k=6**	k=7	k=8	k=9	k=10
Regression	0.2154	**0.2140**	0.2137	0.2151	0.2120	0.2123
MAP	0.2042	**0.2126**	0.2063	0.2074	0.2085	0.2051

Table II shows the hamming loss computed for the Scene dataset. The Scene dataset has 294 attributes and 6 labels. At k=6, hamming loss computed after using our proposed method is less than MAP. Our method outperforms MAP when tested with Scene dataset.

TABLE II. HAMMING LOSS ON SCENE DATASET USING REGRESSION AND MAP

Scene	Hamming Loss					
	k=5	**k=6**	k=7	k=8	k=9	k=10
Regression	0.1503	**0.1499**	0.1511	0.1556	0.1474	0.1560
MAP	0.1466	**0.1716**	0.1429	0.1491	0.1302	0.1503

Table III shows the hamming loss computed for the Emotions dataset. The Emotions dataset has 72 attributes and 6 labels. At k=7, hamming loss is lower than other k values. The hamming loss computed after regression is less compared to MAP. Here also our approach outperforms MAP.

TABLE III. HAMMING LOSS ON EMTOIONS DATASET USING REGRESSION AND MAP

Emotions	Hamming Loss					
	k=5	k=6	**k=7**	k=8	k=9	k=10
Regression	0.2912	0.2846	**0.2781**	0.2790	0.2790	0.2856
MAP	0.2940	0.3081	**0.2959**	0.2978	0.2828	0.2846

VII. CONCLUSION

Existing approaches for multi-label classification either consider feature space similarity or label space dependency for label set prediction. In our paper we proposed a novel method for multi-label classification. A model generation phase is incorporated to gather label dependent information from label space and kNN is used to compute most similar neighbors from feature space. A combined strategy with multiple regression and kNN is used for prediction. This combination is good enough to capture useful information from label, as well as from the feature space.

The results obtained with various multi-labeled dataset justify our method. Future work would involve experimentation with more datasets and would focus on increasing processing and calculation speed.

ACKNOWLEDGMENT

This work derives inspiration and direction from the Chancellor of Amrita University, Sri Mata Amritanandamayi Devi.

REFERENCES

[1] Min-Ling Zhang, Zhi-Hua Zhou, ML-KNN:Alazy learning approach to multi-label learning, 2007 Pattern Recognition Society. Published by Elsevier Ltd

[2] Tsung-Hsien Chiang, Hung-Yi Lo ,Shou-De Lin, A Ranking-based KNN Approach for Multi-Label Classification,2012 Asian Conference on Machine Learning

[3] Charu C. Aggarwal , ChengXiang Zhai ,A Survey of Text Classification Algorithms, Chapter 6: MINING TEXT DATA

[4] Zhihua Wei, Hongyun Zhang_, Zhifei Zhang, Wen Li, Duoqian Miao,A Naive Bayesian Multi-label Classification Algorithm With Application to Visualize Text Search Results, International Journal of Advanced Intelligence, July 2011

[5] Prema Nedungadi, Haripriya Harikumar,Maneesha Ramesh, A high performance hybrid Algorithm for Text Classification, ICADWIT 2014

[6] Hang LI,A Short Introduction to Learning to Rank , IEICE TRANS. INF. & SYST., VOL.E94{D, NO.10 OCTOBER 2011

[7] Yunbo CAO, Jun XU, Tie-Yan LIU, Hang LI, Yalou HUANG, Hsiao-Wuen HON,Adapting Ranking SVM to Document Retrieval , *SIGIR'06*, August 6–11, 2006, Seattle, Washington, U.S.A.

[8] Thorsten Joachims,Optimizing Search Engines using Clickthrough Data, *SIGKDD* 02 Edmonton, Alberta, Canada Copyright 2002 ACM

[9] J.R. Quinlan, Induction of Decision Trees, Kluwer Academic Publisher, 1986

[10] Amanda Clare, Ross D. King, Knowledge Discovery in Multi-Label Phenotype Data Springer-Verlag Berlin Heidelberg 2001

[11] Purvi Prajapati, Amit Thakkar, Amit Ganatra "A Survey and Current Research Challenges in Multi-Label Classification Methods", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012

[12] Weiwei Cheng, Eyke H¨ ullermeier, Combining Instance-Based Learning and Logistic Regression for Multilabel Classification, Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2009)

[13] Grigorios Tsoumakas, Affiliate, CS, Ioannis Katakis, and Ioannis Vlahavas,Random *k*-Labelsets for Multi-Label classification, IEEE Transactions on Knowledge and Data Engineering

[14] Robert E. Schapire, Yoram Singer,BoosTexter: A Boosting-based System for Text Categorization, Machine Learning, 39, 135–168, 2000.

[15] G. Ratsch, T. Onoda, K.R.Muller, Soft Margins for AdaBoost, Kluwer Academic Publishers, 2001

[16] Francisco Charte, Antonio Rivera, MarEto Josst del Jesus, and Francisco Herrera, Improving Multi-label Classifiers via Label Reduction with Association Rules, Springer-Verlag Berlin Heidelberg 2012

[17] Min-Ling Zhang,Lei Wu, LIFT: Multi-Label Learning with Label-Specific Features, IEEE Transactions on Pattern Analysis and Machine Intelligence,2014