Feature and Search Space Reduction for Label-Dependent Multi-label Classification

Prema Nedungadi and H. Haripriya

Abstract The problem of high dimensionality in multi-label domain is an emerging research area to explore. A strategy is proposed to combine both multiple regression and hybrid k-Nearest Neighbor algorithm in an efficient way for high-dimensional multi-label classification. The hybrid kNN performs the dimensionality reduction in the feature space of multi-labeled data in order to reduce the search space as well as the feature space for kNN, and multiple regression is used to extract label-dependent information from the label space. Our multi-label classifier incorporates label dependency in the label space and feature similarity in the reduced feature space for prediction. It has various applications in different domains such as in information retrieval, query categorization, medical diagnosis, and marketing.

Keywords Multi-label · Multiple regression · Hybrid kNN · PCA

1 Introduction

In today's world, we most often deal with multi-label data, that is, data assigned to more than label at the same time. Defining this formally, in single-label classification, an object x_i is assigned to one class c_i, but in multi-label classification an object x_i is assigned to a set of classes, c_1, $c_2 ... c_l \subseteq C$ simultaneously, where C is the total number of labels.

P. Nedungadi (✉) · H. Haripriya
Amrita CREATE, Amrita University, Kollam, Kerala, India
e-mail: prema@amrita.edu

H. Haripriya
e-mail: haripriya@am.amrita.edu

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 380, DOI 10.1007/978-81-322-2523-2_57

There are many real world examples for multi-label data and we can quote a few. A movie can belong to the categories of action, crime, thriller, or drama simultaneously. A newspaper article may belong to a person, a country, some local, national, or international category.

High dimensionality is a curse for almost all machine-learning problems. Linear as well as non-linear dimensionality reduction methods are proposed for linear as well as non-linear datasets. Techniques for dimensionality reduction such as PCA [1], CCA [2], kernel PCA [3], Sammon's non-linear mapping [4], and SVD handle high dimensional data.

Like all existing classification techniques, the multi-label domain also has complexity problems when dealing data with high dimensions. Thus, our objective is to effectively reduce the dimensions of data residing in the high-dimensional space. A new version of PCA and kNN that is a hybrid kNN is proposed in one of our works [5] to reduce the search as well as the feature space of traditional kiln. A new approach for multi-label classification based on multiple regression is proposed in another work [6] of ours. In this paper, we propose a strategy for dimensionality reduction, and multi-label prediction by a hybrid approach using our works [5, 6].

We combined multiple regression with hybrid kNN for label set prediction of high-dimensional data. Multiple regression is used for generating models for label sets. Traditional kNN are computationally intensive since its search space is the entire training data. So applying kNN only for neighboring vectors in the principal components reduces the inputs for the kNN classifier as well as the feature space. Our proposed hybrid feature space reduced multi-label classifier has various applications in high-dimensional domains such as information retrieval, query categorization, medical diagnosis, marketing, and text categorization.

The remaining sections of the paper are the following; Sect. 2 explains the existing techniques for multi-label classification. Section 3 explains briefly the basics of our paper. Section 4 discusses our proposed approach in detail; Sect. 5 details the experimentations and result analysis and Sect. 6 concludes our work briefly.

2 Related Work

Most of the existing classifiers in machine-learning [3] deals with single-label classification. Methods to extend existing classifiers in order to deal with multi-labeled data are discussed in [7–12, 26]. Some methods that are also under research [4, 13, 14] convert multi-label dataset into a set of single-label dataset to fit with the existing classifiers. Dimensionality reduction in high-dimensional space for multi-label is discussed in [15–18].

ML-kNN [7] used the concept of traditional k-Nearest Neighbor algorithm and the maximum a posteriori principle for label set prediction. A Ranking-based kNN Multi-Label Classification [8, 13] also used k-nearest-neighbor-based ranking approach for the multi-label classification. Ranking SVM and kNN concepts are used for multi-label prediction. An approach to analyze features for specific label is discussed in [12]. But these approaches have not considered interdependencies between the label sets and thus ignored the possibility of co-occurrence of labels.

This problem can be eliminated by the mapping of each label in the feature space as a robust subspace, MSE [6], and formulating the prediction as finding the group sparse representation of a given instance on the subspace ensemble.

A Naive Bayesian Multi-label Classification Algorithm [19] is a problem transformation approach. Naive Baye's ignores feature dependency and so in real world application it will result in a decrease in prediction accuracy.

In [13, 14, 20], Ranking SVM is used for document retrieval. Ranking SVM belongs to the category of multi-label dataset in which a single query matches with multiple documents. For non-linear data sets when between the number of samples and the number of features is very low, the possibility of lower accuracy will be high.

A decision tree algorithm C4.5 [21] that is used for the analysis of phenotype data is discussed in [9]. 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 [22] is an extension of traditional back-propagation algorithm proposed an error function that adapt according to the characteristics of multi-label data and thus can be used for multi-label learning. But the neural network complexity becomes high in the training phase.

A probabilistic kNN and logistic regression-based approach for label set prediction is discussed in [23]. The distance between neighboring instances and their labels are used for prediction. Random k-Labelsets (Rakel) [24] divide the label set and considers the label correlation ship. But both these approaches consume more time in applications with large number of training samples and labels.

AdaBoost [10, 25] creates an 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 both noisy data and outliers. A method to reduce dimensionality reduction in the label space with association rule is discussed in [11]. But this approach does not guarantee reduction without information loss.

To solve this problem, a joint learning framework [18] is used, in which we perform dimensionality reduction and multi-label classification simultaneously. This is a novel joint learning framework [3] which performs reduction in dimensions and multi-label inference in semi-supervised setting. A multi-label dimensionality reduction method, MDDM [4], attempts to project the original data into a lower dimensional feature space by maximizing the dependence between the

original feature space and its class labels. But these methods cannot provide an explicit modeling of the label dependency and thus their performance improvements due to exploring label structure are of less significance.

In our proposed approach, we combined the advantages of kNN and multiple regression for multi-label classification. We not only combined feature as well as a label correlation in our approach, but we have also done some dimensionality reduction in the feature space of our multi-label data, thereby reducing the search space as well as the feature space for prediction.

3 Multiple Regression and Traditional kNN

Multiple regression is an extension of simple linear regression for incorporating the dependencies of more than one variable. In our approach we considered the dependence property of labels in the label space.

The equation for multiple regression is

$$Y_{\text{pred}} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_n X_n \tag{1}$$

where 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.

If we want to predict the class of a newly arrived data, traditional kNN will perform a similarity computation in the entire input space. The most similar ***k*** neighbors from the training data will be used for testing. Among various measures, we used cosine similarity for similarity computation. Considering two feature vectors x and y, cosine similarity is computed using [3];

$$\cos(\Theta) = \text{similarity}(x, y) = \frac{x.y}{\|x\| * \|y\|}. \tag{2}$$

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 less similar. Cosine similarity will generate a metric that says how two vectors are related by looking at the angle instead of the magnitude.

4 System Architecture

Our objective is to classify a data into more than one class instead of a single class. We combined hybrid kNN and multiple regression for prediction label set of a data. When a test data is given hybrid kNN it will reduce the feature space as well as the search space.

Multiple Regression and kNN is the main algorithm of our proposed multi-label classifier. The input to our algorithm is the training data with its associated label set and the test data.

Algorithm 1 - Multiple Regression and Hybrid kNN

Input: { Train data with Label set, Test data }
Output: { Label set of test data }

1: Generate models using multiple regression on the label set of training data.
2: Compute Principle components of train data.
3: Select some eigen vectors with largest eigen value.
4: Project entire training data and the test data along each selected PC.
5: Perform binary search over each projected space to find ***L*** nearest neighbors .
6: Select the most similar *k* neighbors.
7: *for i :1 to k*
Each Label value prediction using generated model
8: *end*
9: Average predicted value along each label
10: *for i :1 to C*
if (average < threshold) then
Test data has no label named *i*
else
Test data has label named *i*
11: *end*

12:Predict the Label set

Multiple regression is used to obtain the information from the dependent labels, i.e., from the label space. We assume that for the occurrence of one label, all other labels will contribute. A weight is associated with the labels that give the information about its dependency. A linear model is generated for each label by using the labels from the training set.

We used hybrid kNN for finding the most similar k neighbors from our training set. We limit the kNN search space by finding the L nearest neighbors along each principal component in the dimensionally reduced feature space. The data that are clustered in the original space will also be closer in the projected space and along each of the principal components. A threshold value is used for predicting the label.

The architecture of the training phase of multi-label classifier is shown in Fig. 1. It is composed of two phases, model generation and reduced search space generation phase.

The architecture of the testing phase of multi-label classifier is shown in the Fig. 2. When a new data arrives, we project the newly arrived data along the principal components. In order to reduce the classification time and increase the efficiency of kNN, project the vector along each component and select the ***L*** nearest neighbors. These neighbors are the input dataset for kNN. Prediction is based on the most similar ***k*** neighbors from the nearest neighbors.

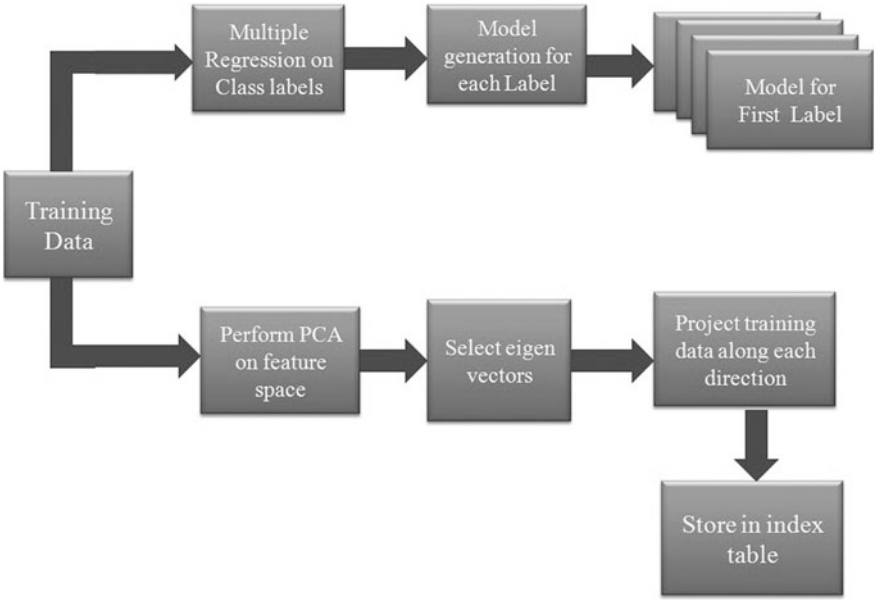

Fig. 1 Training phase of multi-label classifier

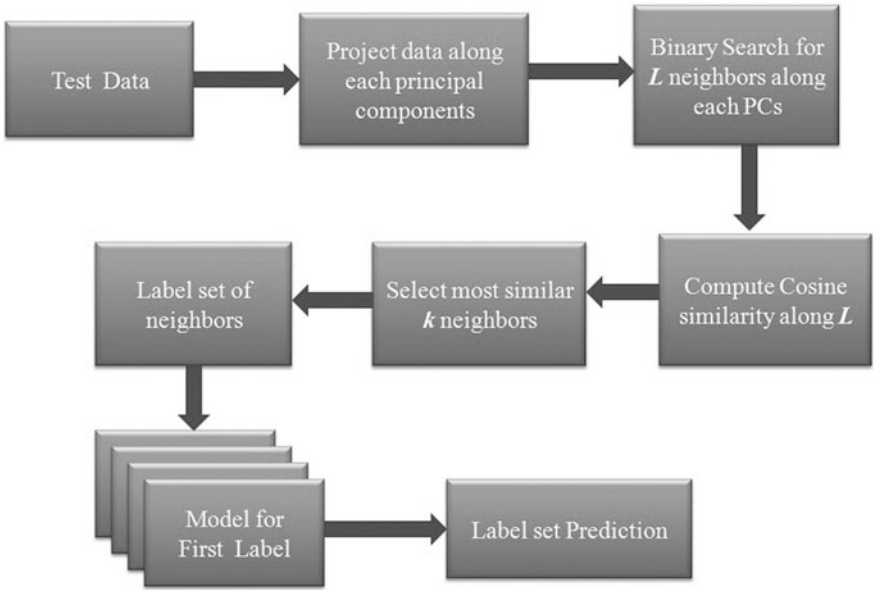

Fig. 2 Testing phase of multi-label classifier

5 Experimental Evaluation

We cannot evaluate a multi-label learning system with the common measures such as precision, recall, etc., since they are used for evaluating the performance of single-label learning system.

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 smaller the value of hamming loss, the better the performance and ranges between 0 and 1 and Δ is the symmetric difference between two sets.

$$\text{hloss}_s(h) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{Q}|h(x_i)\Delta Y_i|. \tag{3}$$

We have used a multi-label dataset from Knowledge Extraction based on the Evolutionary Learning (KEEL Dataset) data repository. The ***L*** value chosen is 80 for both the datasets.

The hamming loss computed for our proposed method is one regression-based method and ML-kNN on the Yeast dataset with 103 attributes and 14 labels that are shown in Table 1.

Table 1 shows our approach outperforms Regression and ML-kNN at ***k = 15 and k = 45***. At ***k = 6*** hamming loss of our approach is equal to ML-kNN and outperform Regression. Table 2 shows the hamming loss computed for the Emotions dataset that are 72 attributes and 6 labels. At ***k = 5 to k = 7*** and for ***k = 10*** our approach

Table 1 Hamming loss on yeast data set compared with ML-kNN, regression, hybrid multi-label

Yeast	Hamming loss					
	$k = 5$	**$k = 6$**	$k = 14$	**$k = 15$**	$k = 40$	**$k = 45$**
ML-kNN	0.2042	**0.2126**	0.2074	**0.2063**	0.2144	**0.2123**
Regression	0.2154	**0.214**	0.2108	**0.2077**	0.2097	**0.2134**
Combined	0.2169	**0.2126**	0.2088	**0.2057**	0.2103	**0.2111**

Table 2 Hamming loss on emotions data set compared with ML-kNN, regression, hybrid multi-label

Emotions	Hamming loss					
	$k = 5$	**$k = 6$**	**$k = 7$**	$k = 8$	$k = 9$	**$k = 10$**
ML-kNN	**0.294**	**0.3081**	**0.2959**	0.2978	0.2828	**0.2846**
Regression	**0.2912**	**0.2846**	**0.2781**	0.279	0.279	**0.2856**
Combined	**0.2753**	**0.2781**	**0.2753**	0.2809	0.2884	**0.2846**

(Combined) outperforms ML-kNN and Regression. Since both data sets have high-dimensional feature space, we can guarantee that our method will give high accuracy in the high-dimensional feature space.

6 Conclusion

Multi-label data classification in high-dimensional space is a new area to explore. In our paper, we proposed a new method for high-dimensional multi-label prediction. A model generation phase fits well in our method to gather label-dependent information from the label space and hybrid kNN is used to compute neighbors from the projected feature space. A hybrid kNN algorithm based on PCA is proposed, to reduce the dimension and efficiently find neighbors along each principal component so as to restrict the kNN search space. This combination is good enough to capture useful information from the label, as well as from the reduced feature space.

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

References

1. Smith, L.I.: A tutorial on principal components analysis. Cornell Univ. USA **51**, 52 (2002)
2. Hotelling, H. Relations between two sets of variates. Biometrika, 321–377 (1936).
3. De Leeuw, J. (2011). History of nonlinear principal component analysis
4. Sammon, J.W.: A nonlinear mapping for data structure analysis. IEEE Trans. Comput. **18**(5), 401–409 (1969)
5. Nedungadi, P., Harikumar, H., Ramesh, M.: A high performance hybrid algorithm for text classification. In 5th International Conference on the Applications of Digital Information and Web Technologies (ICADIWT), IEEE, pp. 118–123 (2014)
6. Nedungadi, P., Haripriya, H.: Exploiting label dependency and feature similarity for multi-label classification. In: International Conference on Advances in Computing, Communications and Informatics (ICACCI) IEEE, pp. 2196–2200 (2014)
7. Zhang, M.L., Zhou, Z.H.: ML-KNN: A lazy learning approach to multi-label learning. Pattern Recogn. **40**(7), 2038–2048 (2007)
8. Chiang, T.H., Lo, H.Y., Lin, S.D.: A Ranking-based KNN Approach for Multi-Label Classification. In: ACML. pp. 81–96 (2012)
9. Clare, A., King, R. D.: Knowledge discovery in multi-label phenotype data. In: Principles of data mining and knowledge discovery, pp. 42–53, Springer Berlin Heidelberg (2001)
10. Schapire, R.E., Singer, Y.: BoosTexter: a boosting-based system for text categorization. Mach. Learn. **39**(2–3), 135–168 (2000)
11. Charte, F., Rivera, A., del Jesus, M.J., Herrera, F.: Improving multi-label classifiers via label reduction with association rules. In: Hybrid Artificial Intelligent Systems, pp. 188–199, Springer Berlin Heidelberg (2012)
12. Zhang, M. L., Wu, L. (2011). LIFT: Multi-label learning with label-specific features

13. Hang, L.I.: A short introduction to learning to rank. IEICE Trans. Inform. Syst. **94**(10), 1854–1862 (2011)
14. Joachims, T.: Optimizing search engines using clickthrough data. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM. pp. 133–142 (2002)
15. Ji, S., Ye, J.: Linear Dimensionality Reduction for Multi-label Classification. In: IJCAI **9**, pp. 1077–1082 (2009)
16. Qian, B., Davidson, I.: Semi-Supervised Dimension Reduction for Multi-Label Classification. In AAAI. **10**, pp. 569–574 (2010)
17. Zhang, Y., Zhou, Z.H.: Multilabel dimensionality reduction via dependence maximization. ACM Trans. Knowl. Disc. Data (TKDD) **4**(3), 14 (2010)
18. Zhou, T., Tao, D.: Multi-label subspace ensemble. In International Conference on Artificial Intelligence and Statistics. pp. 1444–1452 (2012)
19. Wei, Z., Zhang, H., Zhang, Z., Li, W., Miao, D.: A naive Bayesian multi-label classification algorithm with application to visualize text search results. Int. J. Adv. Intell. **3**(2), 173–188 (2011)
20. Cao, Y., Xu, J., Liu, T.Y., Li, H., Huang, Y., Hon, H.W.: Adapting ranking SVM to document retrieval. In: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, ACM pp. 186–193 (2006)
21. Quinlan, J.R.: Induction of decision trees. Mach. Learn. **1**(1), 81–106 (1986)
22. Prajapati, P., Thakkar, A., Ganatra, A.: A survey and current research challenges in multi-label classification methods. Int. J. Soft Comput. **2** (2012)
23. Cheng, W., Hüllermeier, E.: Combining instance-based learning and logistic regression for multilabel classification. Mach. Learn. **76**(2–3), 211–225 (2009)
24. Tsoumakas, G., Katakis, I., Vlahavas, I.: Random k-labelsets for multilabel classification. IEEE Trans. Knowl. Data Eng. **23**(7), 1079–1089 (2011)
25. Rätsch, G., Onoda, T., Müller, K.R.: Soft margins for AdaBoost. Mach. Learn. **42**(3), 287–320 (2001)
26. Aggarwal, C.C., Zhai, C.: A survey of text classification algorithms. In: Mining text data. pp. 163–222, Springer US