Multilabel Classification Exploiting Coupled Label Similarity with Feature Selection

Prajakta Chaudhari
Department of Computer Engineering
KKWIEER
Nashik, India

S. S. Sane, PhD
HOD, Department of Computer Engineering
KKWIEER
Nashik, India

ABSTRACT
In multilabel classification each example is represented with features and associated with multiple labels. Multilabel classification aims to predict set of labels for unseen instances. Researchers have developed multilabel classification using both the problem transformation approach and algorithm adaptation approach. An algorithm called ML-kNN that follows algorithm adaptation approach has been developed and being used to perform multilabel classification. However it does not considers label correlation and thus results in lesser prediction accuracy. A new approach called CML-kNN reported in the literature exploits label correlation using both Intra-Coupling and Inter-Coupling label similarities between the labels to provide better accuracy than that of ML-kNN, but curse of dimensionality is the great challenges in multilabel data. So to address this problem a new approach called CML-kNN with feature selection is presented in this work.The basic idea of this work is to investigate the performance of CML-kNN with and without feature selection. The experiments indicate that proposed CML-kNN with feature selection method achieves superior performance than existing CML-kNN method.

General Terms
Data mining

Keywords
Algorithm adaption, k nearest neighbor, label correlation, ML-kNN, Multilabel classification

1. INTRODUCTION
Traditionally each real world entity is represented by single instance and single label is associated with this instance. Since only one label from a set of disjoint labels is assigned to the instance, this classification is called as single label classification. The basic assumption adopted by traditional supervised learning is that each object belongs to only one semantic concept. However, there are several situations where real world object may be complicated and have multiple semantic meanings where the instance may belong to multiple classes. Above problem can be solved by using multilabel classification. Consider example in which an image of Taj Mahal may be tagged with ancient architecture, cultural heritage, and India.The multilabel classification aims to find the set of labels for unseen instances. In multilabel classification each instance is associated with multiple labels.

Due to the large amount of possible sets of labels, the process of learning from multilabel instances is quite difficult. So, success of multilabel classification depends on how effectively we exploit the label correlations. To improve existing multilabel classification it is important to consider label correlation.

Multilabel classification can be used in various applications such as gene function prediction, text categorization, bioinformatics, image annotation, direct marketing, Medical diagnosis, Tag recommendation and Query categorization [1] [2].

Rest of the paper is described as follows: Section I provides introduction while literature survey has been discussed in section II. Section III deals with implementation details while proposed experimental setup has been discussed in section IV. Finally section V provides conclusion.

2. EXISTING METHODS
Several multilabel classification algorithms are reported in literature [15]. Multilabel classification approaches are divided into two categories namely problem transformation approach and algorithm adaption approach [4].

The problem transformation approach, which is independent on algorithm, converts the multilabel problem into a set of single label problem. Some problem transformation methods reported in literature such as Binary relevance [3], Label powerset [4], Classifier chain (CC) [5].

An algorithm adaptation approach updates the traditional machine learning algorithms to make them suitable for handling multilabel data. The various algorithm adaptation methods reported in literature such as multilabel k nearest neighbor [6], binary relevance k nearest neighbor (BR-kNN) [7], IBLR [8].

Zhang and Zhou [6] introduced multilabel k nearest neighbor (ML-kNN) algorithm. ML-kNN extends existing k nearest neighbor (kNN) algorithm so as to handle multilabel data. First it identifies k nearest neighbors for every instance in training data and also for unseen instance. Then prior probabilities, frequency arrays, statistical information for labels are calculated. After that set of labels for an unseen instances are predicted with the help of MAP (Maximum a posterior) rule which is based on Bayes theorem. Advantage of ML-kNN is class imbalance issue can be reduced due to consideration of prior probabilities. ML-kNN does not consider label correlations.

G. Tsoumakas , E. Spyromitros, I. Vlahavas, [7] proposed binary relevance k nearest neighbor (BR-kNN) algorithm which enhances k nearest neighbor (kNN) machine learning algorithm in combination with binary relevance problem transformation method.

Two extension of BR-kNN algorithms such as BR-kNN-a and BR-kNN-b are depicted in [7] are depend on value of confidence score for each class label which are obtained from BR-kNN. For the Datasets with low cardinality BR-kNN predicts the empty set. So this problem can be solved with the help of BR-kNN-a. BR-kNN-b provides improvement for datasets which have larger cardinality. BR-kNN and extensions of BR-kNN does not consider label correlations.

BSVM [3] follow binary decomposition method to deal with multilabel classification problems. Initially the multilabel dataset is transformed into single label data using one vs. all binary decomposition method. Binary classifier SVM is used as base classifier. Final prediction is done by combining predictions of all SVMs. BSVM does not consider label correlations.

Weiwei Cheng [8] stated Instance based Logistic Regression (IBLR) which is combination of instance based learning and logistic regression. IBLR overcomes the drawback of existing instance based multilabel classification method such as ML-kNN. IBLR consider interdependencies between labels. But ML-kNN is better in terms of Hamming loss.

Boriah [9] stated a various similarity measure such as overlap similarity, frequency based cosine similarity, different Goodall's measure for categorical data but it does not consider co-occurrence information. The drawback of these similarity measures does not consider inter relationship between labels. Coupled behavior between two entities stated in [10], [11], [12].

Longbing Cao and Chunming Liu [13] presented a new coupled k nearest neighbor algorithm for multilabel classification (CML-kNN) which is based on lazy learning and considers correlation between labels. CML-kNN extends ML-kNN algorithm so as to handle label correlation. First coupled label similarity is calculated with the help of inter and intra coupling similarity. Then it estimates k nearest neighbor for every instance in training data as well as for unseen instance. Prior probabilities, posterior probabilities, frequency arrays and likelihood for labels are calculated with the help of k nearest neighbors information and coupled label similarity. Finally labels for unseen instances are predicted via MAP (Maximum a posterior) rule. CML-kNN algorithm handles correlation between labels but it is more complex than ML-kNN. To improve the existing CML-kNN algorithm feature selection technique is incorporated in this algorithm which can improve the prediction performance.

3. IMPLEMENTATION DETAILS

The proposed system is a systematic approach for multilabel classification. Due to large number of features the complexity increases, this problem can be solved by incorporating feature selection technique in the CML-kNN algorithm. Due to feature selection redundant and irrelevant features will be eliminated that will provide better prediction accuracy.

The proposed multilabel classification system is shown in the Fig.1. The modules of proposed multilabel classification system are as follows:

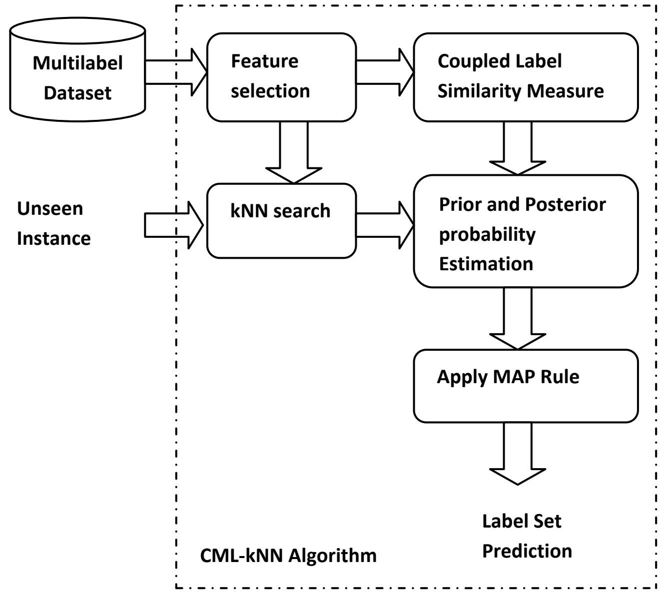

Fig. 1. Proposed System Block Diagram

3.1 Coupled Label Similarity

Coupled label similarities (CLS) is obtained by combining intra and inter coupling similarity between labels. A coupled label similarity is calculated as follows:

$$CLS\left(w_i^x, w_j^y\right) = \delta^{ia}\left(w_i^x, w_j^y\right) . \sum_{k=1}^{n} \delta^{ier}\left(w_i^x, w_j^y | w_k\right) \qquad (1)$$

Where n is number of feature attributes.

3.1.1.1 *Intra Coupling label similarity:*

Intra coupling similarity between two different labels captures interaction between labels with the help of occurrence frequency. It is calculated as:

$$\delta^{ia}\left(w_i^x, w_j^y\right) = \frac{F(w_i^x).F(w_j^y)}{F(w_i^x)+F(w_j^y)+F(w_i^x).F(w_j^y)} \qquad (2)$$

Where $F(w_i^x)$ and $F(w_j^y)$ represents the occurrence frequency of label.

3.1.1.2 *Inter Coupling Label Similarity:*

Inter Coupling Label Similarity is used to capture the interaction of two different label values according to the co-occurrence value of labels and features. Inter coupling similarity between label values w_i^x *and* w_j^y according to feature value w_p^z of feature a_z is calculated as:

$$\delta^{ier}\left(w_i^x, w_j^y | w_p^z\right) = \frac{min(CF(w_p^{zx}), CF(w_p^{zy}))}{max(F(w_i^x), F(w_j^y))} \qquad (3)$$

Where $CF(w_p^{zx})$ and $CF(w_p^{zy})$ are the co-occurrence frequencies of the label according to feature value.

3.2 k nearest neighbor estimation

For each training instances k nearest neighbors are calculated. Also for unseen instance k nearest neighors are calculated.

3.3 Prior and Posterior probability estimation

Prior probability P(Hj) and P(-Hj) can be calculated as:

$$P(Hj)= \frac{s+\sum_{i=1}^{n} \delta_{L_i^*|j}}{s\times 2+n\times N} \qquad (4)$$

$P(-Hj) = 1-P(Hj)$ (5)

where $\delta_{L_i^*|j}$ is addition of the Coupled Label Similarity values of i[th] neighbors label set to j[th] label lj.

s is smoothing parameter (here s=1).

n represents number of records in training set.

N represents total number of labels.

CML-kNN maintains frequency arrays $\alpha_j[r]$ and $\beta_j[r]$ calculated as:

$$\alpha_j[r] = \sum_{i=1}^{n} \delta_{L_i^*|j} \mid C_j(x_i) = r \quad (6)$$

$$\beta_j[r] = \sum_{i=1}^{n} (N - \delta_{L_i^*|j}) \mid C_j(x_i) = r \quad (7)$$

Instance with label j have $\delta_{L_i^*|j} \geq 0.5$ and instance with $\delta_{L_i^*|j} < 0.5$ does not have label j. $\alpha_j[r]$ counts the addition of CLS values to label j of training example which have label lj and have exactly r number of neighbors. Where ($0 \leq r \leq k \leq N$). k represents number of neighbors and n is number of labels in training set. Where $\beta_j[r]$ counts the addition of Coupled Label Similarity values to label j of training example which do not have label lj and have exactly r number of neighbors. N is total number of labels in training set.

$$C_j = Round\left(\sum_{i=1}^{k} \delta_{L_i^*|j}\right) \quad (8)$$

Where $\delta_{L_i^*|j}$ is the addition of the Coupled Label Similarity values of the i[th] neighbor's label set to the j[th] label lj. L_i is neighbor set of ith neighbor and $L_i \in T(x)$, T(x) is set of k nearest neighbor of unseen instance x. Round() represents rounding function. Using these frequency arrays $\alpha_j[r]$ and $\beta_j[r]$ likelihood P(Cj|Hj) and P(Cj|-Hj) can be calculated as:

$$P(Cj|Hj) = \frac{s + \alpha_j[Cj]}{s \times (k \times N + 1) + \sum_{r=0}^{k \times N} \alpha_j[r]} \quad (9)$$

$$P(Cj|-Hj) = \frac{s + \beta_j[Cj]}{s \times (k \times N + 1) + \sum_{r=0}^{k \times N} \beta_j[r]} \quad (10)$$

Posterior probability P(Hj|Cj) and P(-Hj |Cj) calculated as follows:

$$\frac{P(Hj|Cj)}{P(-Hj|Cj)} = \frac{P(Hj).\ P(Cj|Hj)}{P(-Hj).\ P(Cj|-Hj)} \quad (11)$$

3.4 MAP rule

According to maximum a posteriori (MAP) rule, the predicted label set (U) for unseen instance is determined by deciding P(Hj|Cj) is greater than P(-Hj|Cj) or not.

$$U = \{lj | \frac{P(Hj|Cj)}{P(-Hj|Cj)} > 1,\ 1 \leq j \leq N\} \quad (12)$$

Various modules described above are as per the implementation reported in [13]. We propose to use feature selection method depicted in [14].

3.5 Feature selection

Feature selection technique [14] eliminates the redundant and irrelevant features. The steps for feature selection are as follows:

Step I. For each subset calculate information gain between features and labels. The information gain between label L and feature fi is given as:

$$IGS(fi, L) = \frac{2 * IG(fi, L)}{H(fi) + H(L)} \quad (13)$$

Information Gain(IG) can be calculated as:

$$IG(fi, L) = H(L) + H(fi) + H(fi\ .\ L) \quad (14)$$

H(.) represents information entropy.

Step II. According to threshold (T) value relevant features are obtained.

$$T = \frac{1}{m} \sum_{i=1}^{m} IGS_i \quad (15)$$

where n is total number of features.

4. RESULTS AND DISCUSSION

The hamming loss and one error will be computed to evaluate the performance of multilabel classification system.

Hamming loss: It calculates number of times the <instance, label> pair is misclassified. Lesser value of hamming loss represents better performance.

One error: It calculates number of times the top ranked class labels predicted by classification system are not in the set of ground truth label set of instances. Smaller value of this metric represents better performance.

4.1 Experimental Setup

In experimentation Enron and medical datasets are considered that have been used by earlier researchers for multilabel classification. Enron dataset contain 1702 number of instances, 1001 number of features and 53 numbers of labels. Medical dataset contain 978 number of instances, 1449 number of features and 27 numbers of labels.

4.2 Results

Table 1. Results of CML-kNN Algorithm

Dataset	Hamming loss	One error
Enron	0.087	0.30
Medical	0.013	0.15

Table 2. Results of CML-kNN Algorithm with feature selection

Dataset	Hamming loss	One error
Enron	0.054	0.22
Medical	0.012	0.14

Table 1 describes results of CML-kNN algorithm. Table 2 describes results of CML-kNN algorithm with feature selection. Lesser value for hamming loss and one error represents better performance. The experiments indicate that proposed CML-kNN with feature selection method achieves superior performance than existing CML-kNN method.

5. CONCLUSION

Researchers have developed multilabel classification using both the problem transformation approach as well as algorithm adaptation approach. A novel lazy learning approach CML-kNN has been reported in the literature for multilabel classification using inter and intra coupling similarity within labels which handles label correlation, but curse of dimensionality is the great challenge in multilabel data. So to address this problem a new approach called CML-kNN with feature selection is presented in this work. The experiments indicate that proposed CML-kNN with feature selection method achieves superior performance than existing CML-kNN method.

6. REFERENCES

[1] Gjorgji Madjarov ,DragiKocev , Dejan Gjorgjevikj , Saso Dzeroski, "An extensive experimental comparison of methods for multi-label learning", Pattern Recognition 45 (2012) 3084-3104

[2] Tsoumakas, G., Katakis, I."Multi-label classification: An overview" International Journal of Data Warehousing and Mining (IJDWM) 3(3), 1-13 (2007)

[3] Boutell, M.R., Luo, J., Shen, X., Brown, C.M. "Learning multi-label scene classification", Pattern recognition 37(9), 1757-1771 (2004)

[4] Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas, "Mining Multi-label Data", Data Mining and Knowledge Discovery Handbook, Springer, Berlin Heidelberg, 2010, pp. 667-685.

[5] Read, J., Pfahringer, B., Holmes, G., Frank, E.: Classifier chains for multi-labelclassification. Machine Learning 85(3), 333–359 (2011).

[6] Zhang, M.L., Zhou, Z.H., "ML-kNN: A lazy learning approach to multi-label learning", Pattern recognition 40(7), 2038-2048 (2007) pp. 401-406. Springer (2008)

[7] Spyromitros, E., Tsoumakas, G., Vlahavas, I., "An empirical study of lazy multilabel classification algorithms", Artificial Intelligence: Theories, Models and Applications, pp. 401-406. Springer (2008)

[8] Cheng, W., HÂlullermeier, E., "Combining instance-based learning and logistic regression for multi-label classification",Machine Learning 76(2-3), 211-225 (2009)

[9] Boriah, S., Chandola, V., Kumar, V., "Similarity measures for categorical data: A comparative evaluation", red 30(2), 3 (2008)

[10] Cao, L., "Coupling learning of complex interactions",Information Processing and Management (2014)

[11] Cao, L., Ou, Y., Yu, P.S., "Coupled behavior analysis with applications", Knowledge and Data Engineering, IEEE Transactions on 24(8), 1378-1392 (2012)

[12] Liu, C., Cao, L., Yu, P.S., "Coupled fuzzy k-nearest neighbors classification of imbalanced non-iid categorical data", Neural Networks (IJCNN), 2014 International Joint Conference on. pp. 1122-1129. IEEE (2014)

[13] Chunming Liu and Longbing Cao, "A Coupled k-Nearest Neighbor Algorithm for Multi-label Classification", Springer PAKDD 2015, Part I, LNAI 9077, pp. 176- 187, 2015

[14] Shining Li, Zhenhai Zhang, Jiaqi Duan "An Ensemble Multi-Label Feature Selection Algorithm Based on Information Entropy" The International Arab Journal of Information Technology, Vol. 11, No. 4, July 2014.

[15] Prajakta Chaudhari , Prof. Dr. S. S. Sane, " Review on Multilabel Classification Algorithms ",IJSRD – International Journal for Scientific Research and Development Vol. 3, Issue 11, 2016 ISSN (Online) 2321-0613