FUZZY RULE CLASSIFIER FOR GENERALIZED k-LABELSET ENSEMBLE

V.B.BANSODE[1], Prof. Dr. S.S.SANE[2]

[1] PG Student, Department of Computer Engineering, KKWIEER, Nashik, Maharashtra, India
[2] HOD of Computer Dep., Department of Computer Engineering, KKWIEER, Nashik, Maharashtra, India

ABSTRACT

In multi-label classification, set of labels are associated with each example. An algorithm called Random k-labelsets (RAkEL) is an algorithm for multi-label classification that follows problem transformation approach. RAkEL algorithm uses Label powerset (LP) classifier and it assumes equal weightage for each label set. To overcome this drawback, a new approach is reported in the literature that is GLE. GLE performs the basis expansion method to train LP classifier on random k labelsets. To decrease the global error between the estimated and ground truth, the expansion coefficients are learned. GLE uses SVM classifier which uses crisp vales as the base classifier. Fuzzy rule classifier (FURIA) as reported in literature gives the better results compared with other rule based classifiers, for problem transformation methods such as Binary Relevance, Classifier Chain, and LP. It would be interesting to observe the performance of GLE with FURIA. This work aims at implementation of GLE with FURIA algorithm and compares its performance using SVM as a base classifier. Experimental results shows that GLE using fuzzy rule classifier FURIA provides better performance in terms of hamming loss, ranking loss, subset 0/1 loss, one error, average precision.

Keyword:- *multi label classification, LP, RAkEL, GLE, fuzzy rule classifier*

---

1. INTRODUCTION

Classification is the process where a problem is related to single label or group of labels of each example. A single label is associated with an instance in case of single label classification whereas multiple labels are associated with an instance simultaneously in case of multi label classification.

Label powerset (LP) is a method of problem transformation approach [1]. LP considers a new different class for every different set of labels available in the training dataset. LP method has limitation that the number of classes increases as the number of labels in the labelset increases, where each class may be associated with very less training data.

To overcome this limitation, a new method Random k-Labelsets (RAkEL) [2] is proposed, where k specifies the size of the each labelsets. RAkEL method divides the original labelset randomly in different subsets of size k and after that LP method is applied to train every subset. For the final prediction of RAkEL method voting of the LP classifiers is performed in ensemble. This method decreases number of classes; also, each class can have more training instances. RAkEL have limitation that, it equally assigns importance to the every base classifier in the ensemble. This is problematic because each LP classifier is trained on different randomly selected k-labelsets where some classifiers may give worst performance compared to others or it can even be redundant.

In the case of multi-label classification (MLC), there could be seen some degree of uncertainty among the labels boundaries, which could not be properly captured by the non-fuzzy i.e. crisp classifiers [6]. With the help of fuzzy rule classifier (FURIA) the performance of GLE method could be improved.

In the rule base classification, fuzzy rules are used instead of conventional crisp rules. Compared to conventional crisp rules, fuzzy rules have many advantages also are more general. The boundaries of fuzzy rules are soft and are potentially more flexible than conventional rules i.e. non-fuzzy [6]. Conventional rules produces models having sharp decision boundaries which results in sudden or abrupt changes in different classes, this characteristic is questionable. That is why it could be expected that rather than representing boundaries in abrupt way, it could be represented in gradual way. The models developed with conventional rules with sharp decision boundaries results in abrupt changes in different classes, which is questionable. So, it could be expected that the boundaries should be represented in gradual way rather than abrupt way. Using the fuzzy rules could be a way to achieve this.

2. RELATED WORK

In supervised learning Multi label classification (MLC) and label ranking (LR) are two important tasks for multi-label data. MLC model output is a bipartition into irrelevant and relevant of the set of labels with respect to given input query instance. LR output is a learned model where class labels are ordered according to their relevance to a given query instance.

The LR and MLR methods [1] are categorized as: i) problem transformation approach, that transforms the MLC into single label classification; and ii) algorithm adaptation approach, that adapts or extends the existing algorithms.

2.1 Problem transformation approach

In this approach, the MLC problem is transformed into one or more single-label classification problems. Problem transformation methods are independent of learning algorithms [2].
1. Binary Relevance (BR) [1] method is a simple and also popular method. Separate classifier one for each label is trained in the labelspace. To classify new instance all classifiers are executed parallely, and then positively predicted labels are assigned to the new instance.
2. Classifier Chains (CC [3] is the linked chain of the classifiers. Every classifier in this chain uses Binary Relevance problem method for each label.
3. Label Powerset (LP) [4] method considers a new possible class for different combinations of labels in training data. For example, if an instance is associated with the labels l1 and l2, then a new class named l12 is created, representing the class of this instance. The classifier predicts the new labels that is set of labels, for the given instance.
4. Pruned problem transformation (PPT) [1] The LP method is extended by PPT. The labelsets which occurs less number of times than the threshold are pruned and also replaces their data by adding disjoint sub sets of these labelsets which exists more times than the given threshold. Threshold value is user defined.
5. RAkEL [4] method is an ensemble of LP classifiers. It considers the label correlation, and overcomes the LP problem. Overcomes drawback of LP that number of classes increases in proportion to the number of labels. RAkEL restricts the size of each labelset to k. It randomly selects a number of labelsets from the original set of labels. Then LP learning is applied to the training set. Labelset for an unseen instance is computed from voting of the LP classifiers in the ensemble.

2.2 Generalized k-labelset ensemble

Hung-Yi Lo, Shou-De Lin, Hsin-Min Wang proposed a new Generalized k-labelset ensemble [6] method which learns and makes use of model of expansion for MLC. The expansion coefficients are learned which reduces the global error between the prediction and ground truth. The experiments performed by author shows that the performance of LP-based ensemble method is improved by assigning different weights to the classifiers in the ensemble. Limitation of RAkEL is overcomed by GLE method by assigning different weight to each classifier.

2.3 Fuzzy classifier

In [7], the performance of MLC using fuzzy rule classifier is evaluated. For the problem transformation methods FURIA (Fuzzy Unordered Rule Induction Algorithm) is used as a base classifier. Experiment was conducted on four problem transformation methods, using eight different base classifiers and six datasets. Performance is evaluated using multilabel classification performance metrics. Experimental results shows that the FURIA outperforms compared to other eight different base classifiers.

3. SYSTEM ARCHITECTURE

The architecture of the proposed system is shown in fig.1.

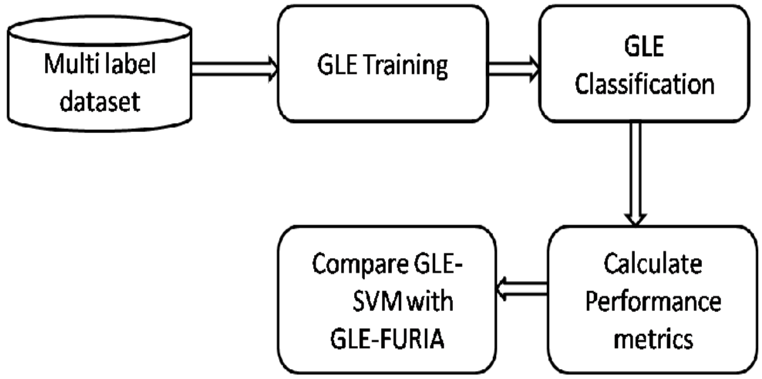

Fig -1: System Architecture

The blocks that are depicted in the system architecture of figure 1 can be elaborated using the following steps.

1. **Read input multi label dataset**
   System reads the input dataset which labels, attributes and its values stored. The dataset file is stored in ARFF (Attribute Relation File Format) file format. The first line shows the name of the dataset relation. The attributes are specified using attribute keyword. The data is started using data keyword. The coefficients are learned to decrease the global error between the actual truth and prediction of labels.

2. **GLE Training**
   Training process is done in GLE Training block. User enters the value of number of models M and size of subset k also gives dataset training file as input for training process. After training process classifier file is created which is used for the classification process.

3. **GLE Classification**
   Classification process is done in GLE Classification block. For the classification, user gives testing file as input.

4. **Calculate Performance Metrics**
   Performance metrics of multi label classification hamming loss, ranking loss, subset 0/1 loss, one error, average precision are calculated for each dataset and stored in a text file taken from user.

5. **Compare Performance metrics**
   Performance metrics of GLE-SVM are compared with GLE-Training to find which method gives the improved results.

4. RESULTS AND DISCUSSION

4.1 Experimental Setup

In experiments, Generalized k-labelset ensemble algorithm reported in literature (called here as GLE-SVM) is compared with our implementation of Generalized k-labelset ensemble using FURIA (called here as GLE-FURIA). Experiments were performed on i3 processor with 3GB RAM, 500GB HDD and on Windows 7 operating system. Implementation of GLE-SVM and GLE-FURIA was done in java 8 and development tool is Netbeans 8.0.2. The parameters k, M used for experiments are listed in Table 1, as was used in [11]. The experiments were performed and then average was calculated. Three-fold cross validation was performed on above dataset, as was used in [11].

Table -1: Selected parameters k and M

Dataset	K	M
Scene	4	15
Enron	16	250
Cal500	10	250
Medical	14	250
Bibtex	24	250

4.2 Performance Evaluation

Five performance metrics for multi-label classification are considered for the evaluation of performance of the system. These performance metrics are as follows:
1) Hamming loss: It calculates the total percentage of labels which are predicted incorrectly.
2) Ranking loss: It calculates the average proportion of pairs which are ordered incorrectly.
3) Subset 0/1 loss: It calculates the percentage of predicted label subset which does not match with actual label subsets.
4) One error: It calculates the number of times best ranked label is not in the set of correct labels
5) Average precision: It calculates the average proportion of labels which are ranked above a particular desired label.

4.3 Results

Experimental results are shown in Table 2 to Table 6.
For each evaluation metric "↓" indicates "smaller value has better results" and "↑" " indicates "bigger value has better results".

Table 2 shows comparison of GLE-SVM with GLE-FURIA with respect to hamming loss. The GLE-FURIA provides slightly better hamming loss when compared with GLE-SVM.

Table -2: Comparison of GLE-SVM and GLE-FURIA in terms of Hamming Loss

Hamming Loss "↓"		
Dataset	**GLE-SVM**	**GLE-FURIA**
Scene	0.0993	0.0893
Enron	0.0595	0.0495
Cal500	0.1704	0.1501
Medical	0.0119	0.0109
Bibtex	0.0135	0.0114
Average	**0.0709**	**0.0622**

Table 3 shows comparison of GLE-SVM with GLE-FURIA with respect to ranking loss. The GLE-FURIA provides slightly better ranking loss when compared with GLE-SVM.

Table -3: Comparison of GLE-SVM and GLE-FURIA in terms of Ranking loss

Ranking Loss "↓"		
Dataset	**GLE-SVM**	**GLE-FURIA**
Scene	0.1251	0.1001
Enron	0.0991	0.0813
Cal500	0.1571	0.1451
Medical	0.0495	0.0421
Bibtex	0.1925	0.1905
Average	**0.1246**	**0.1118**

Table 4 shows comparison of GLE-SVM with GLE-FURIA with respect to ranking loss. The GLE-FURIA provides slightly better ranking loss when compared with GLE-SVM.

Table -4: Comparison of GLE-SVM and GLE-FURIA in terms of Subset 0/1 loss

Subset 0/1 Loss "↓"		
Dataset	**GLE-SVM**	**GLE-FURIA**
Scene	0.2733	0.2281
Enron	0.7911	0.7565
Cal500	0.9572	0.9571
Medical	0.1560	0.1420
Bibtex	0.7501	0.7301
Average	0.5855	0.5613

Table 5 shows comparison of GLE-SVM with GLE-FURIA with respect to one error. The GLE-FURIA provides slightly better one error when compared with GLE-SVM. For the dataset cal500, GLE-FURIA gives slight worst performance compared to GLE-SVM. However GLE-FURIA provides slightly better average one error.

Table -5: Comparison of GLE-SVM and GLE-FURIA in terms of One error

One Error "↓"		
Dataset	**GLE-SVM**	**GLE-FURIA**
Scene	0.2911	0.1634
Enron	0.2911	0.2571
Cal500	0.1299	0.1992
Medical	0.2012	0.1502
Bibtex	0.4011	0.3871
Average	0.2628	0.2314

Table 6 shows comparison of GLE-SVM with GLE-FURIA with respect to average precision. The GLE-FURIA provides slightly higher average precision when compared with GLE-SVM.

Table -6: Comparison of GLE-SVM and GLE-FURIA in terms of Average Precision

Average Precision "↑"		
Dataset	**GLE-SVM**	**GLE-FURIA**
Scene	0.8222	0.8752
Enron	0.6512	0.6851
Cal500	0.5992	0.6234
Medical	0.8511	0.8733
Bibtex	0.5014	0.5289
Average	0.6850	0.7171

5. CONCLUSIONS

The concept of multilabel classification has been developed by researchers. Multilabel classification has two approaches as problem transformation and algorithm adaption. In multi-label classification, set of labels are associated with each example. An algorithm called Random k-labelsets (RAkEL) is an algorithm for multi-label classification that follows problem transformation approach. RAkEL algorithm uses Label powerset (LP) classifier and it assumes equal weightage for each label set. To overcome this drawback, a new approach is reported in the literature that is GLE. GLE performs the basis expansion method to train LP classifier on random k labelsets. To decrease the global error between the estimated and ground truth, the expansion coefficients are learned. Fuzzy rule classifier (FURIA) as reported in literature gives the better results compared with other rule based classifiers, for problem transformation methods such as Binary Relevance, Classifier Chain, and LP. GLE uses SVM classifier which uses crisp vales as the base classifier. This paper investigates performance of FURIA classifier as a base classifier for GLE algorithm. Experimental results show that GLE-FURIA provides slightly better performance in terms of hamming loss, ranking loss, subset 0/1 loss, one error and average precision.

6. REFERENCES

[1] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multilabel data", in Data Mining and Knowledge Discovery Handbook, O. Maimon and L. Rokach, Eds. New York, NY, USA: Springer, 2010

[2] Andre C P L F de Carvalho1 and Alex A. Freitas2 "A Tutorial on Multi-Label Classification Techniques", Jul. 2009

[3] Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. 2011. "Classifier chains for multi-label classification" Mach. Learn. 85, 3 (December 2011), 333-359

[4] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Random k-labelsets for multilabel classification," IEEE Trans. Knowl. Data Eng., vol. 23,no. 7, pp. 1079â˘A¸S1089, Jul. 2011

[5] S. Agarwal, K. Branson, and S. Belongie, "Higher order learning with graphs," in Proc. Int. Conf. Mach. Learn., Pittsburgh, PA, USA, 2006.

[6] H.-Y. Lo, S.-D. Lin, and H.-M. Wang, "Generalized k-labelset ensemble for multilabel classification," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2012

[7] Prati, R.C., "Fuzzy rule classifiers for multi-label classification," in Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on , vol., no., pp.1-8, 2-5 Aug. 2015

[8] FURIA ,Machine Learning Algorithms, Mc. Grow Hill

[9] J. C. Huhn and E. Hullermeier, "FURIA: an algorithm for unordered fuzzy rule induction," Data Min. Knowl. Discov., vol. 19, no. 3, pp. 293 - 319, 2009

[10] SVM ,Machine Learning Algorithms, Mc. Grow Hill

[11] H.-Y. Lo, S.-D. Lin, and H.-M.Wang, "Generalized k-labelset ensemble for multilabeland cost-sensitive classification," on Knowledge and data engineering., July 2014

[12] G. Tsoumakas, E. Spyromitros-Xioufis, J. Vilcek, and I. Vlahavas, "Mulan: A Java library for multi-label learning," J. Mach. Learn. Res., vol. 12, pp. 2411 - 2414, Jul. 2011.

[13] D. Ventura, "SVM Example", March 12, 2009