Fuzzy Rule Classifier for Generalized k-labelset Ensemble

Vaishali Balasaheb Bansode
PG student, Dept. of
Computer engg.,
K.K.W.I.E.E.R.,
Savitribai Phule Pune University,
Nashik(MH), India

S. S. Sane, PhD
H.O.D of Computer Dept.,
K.K.W.I.E.E.R.,
Savitribai Phule Pune University,
Nashik(MH), India

ABSTRACT
An algorithm called Random k-labelsets (RAkEL) follows problem transformation approach of multi-label classification and uses Label powerset (LP) classifier. RAkEL assumes equal weightage for each labelset. This drawback is overcome by Generalized k-labelset ensemble (GLE) method that advocates the basis expansion model to train LP classifier on random k labelset. To reduce global error between ground truth and estimate, expansion coefficients are learned by GLE. GLE is further extended to solve multi label misclassification problem. As reported in literature, using Fuzzy rule classifier (FURIA) as a base classifier for problem transformation methods gives the competitive results compared with other rule based classifiers. The base classifier used in GLE is LIBSVM, it uses crisp values. This work aims at implementation of GLE and tests its performance using crisp classifier and fuzzy rule classifier as a base classifier. It is expected that using fuzzy rule classifier performance of GLE would be improved.

General Terms
Data mining, Multi-label classification

Keywords
multi-label classification, RAkEL, GLE, fuzzy rule classifier.

1. INTRODUCTION
Classification is a method in which a problem associates a single label or set of multiple labels to each example or instance. There are many classification tasks where each instance could be associated with one or more labels. When a single label is assigned to an instance then this type of classification is single label classification, whereas when multiple labels are simultaneously assigned to an instance then this type of classification is multi label classification. LP [1], the problem transformation approach method, considers every set of different labels in training dataset as a new unique class. The limitation of LP method is that the number of labels in the labelset increases as the number of classes increases, where every class may be represented with very few numbers of training instances.

This limitation of LP method is overcome by the Random k-Labelsets (RAkEL) [2] method, where k depicts the size of the labelsets. RAkEL method randomly breaks the original labelset into different k-sized subsets and then applies LP method to train each subset. For final prediction of RAkEL method voting of the label powerset classifier is done in ensemble. RAkEL reduces the number of classes, as well as, allows every class to have more training instances.

The limitation of RAkEL method however is that, it gives equal importance to the each base classifier in the ensemble, which is problematic as each LP classifiers is trained using randomly selected k-labelsets where some of them may provide worst performance than others or could be even redundant.

In multi label classification, often there is some degree of uncertainty among the boundaries of labels, which cannot be captured properly by the crisp i.e. non fuzzy classifiers [5]. Using fuzzy rule classifier for GLE method may help to improve performance by minimizing one error, hamming loss, and ranking loss whereas improving average precision, subset accuracy.

In rule base classification, instead of using conventional crisp rules, fuzzy rules are used. When compared with conventional rules, fuzzy rules have many advantages and are more general. Fuzzy rule boundaries are soft which are potentially more flexible than non fuzzy or conventional rules [5]. The models produced by conventional rules with sharp decision boundaries results in abrupt changes in different classes, which is questionable. So, it could be expected that the boundaries may be represented in gradual. This could be achieved using the fuzzy rules.

This paper is organized as follows: Section 1 provides introduction while literature survey has been discussed in section 2. Section 3 deals with proposed system and implementation details. Experimental setup of proposed system is presented in section 4. Finally section 5 provides conclusions.

2. LITERATURE SURVEY
The methods of multi label classification are categorized in two different approaches, first called problem transformation approach, and second called algorithm adaptation approach.

2.1 Problem Transformation approach
In this category, the multi-label classification problem is transformed into one or more single-label classification problems. These methods are not dependent of learning algorithms.

Binary Relevance (BR) method is a simple and popular which trains the separate classifier for one for each label. New instance is classified by executing all classifiers in parallel, and assigning positively predicted labels to the new instance.

Classifier Chains (CC) is the linked chain of the classifiers. In this chain, every classifier is the binary relevance problem which is associated with each label.

Label Powerset (LP) method treats different combinations of labels in training data as a new possible class. For example, if an instance have the labels l1 and l2, it creates a new class named l1;2, which represents the class of this instance. The classifier predicts the new labels that is set of labels, for the given instance.

Pruned problem transformation (PPT) extends the LP method. This method prunes the labelsets which occurs less number of times than the threshold which is user defined and also replaces their data by adding disjoint sub sets of these labelsets which exists more times than the given threshold.

RAkEL method is an ensemble of LP classifiers. It takes into account the label correlation, and avoids the LP problem.

2.2 Generalized k-labelset ensemble (GLE)

H.Y. Lo, H.M. Wang, S.D Lin, proposed a new Generalized k-labelset ensemble [4] method. This method is based on the concept of LP method. GLE learns and makes use of expansion model for the multi-label classification. GLE method learns the expansion coefficients and helps to reduce the global error between the ground truth and prediction. The output of the experiment conducted by author shows that by assigning different weights to the classifiers in the ensemble, the performance of LPbased ensemble methods is improved.

In cost-sensitive classification, for each instance a misclassification cost is coupled with each label associated with the label of that instance. The main aim of the cost-sensitive multi label classification is to train the classifier which may help to minimize the misclassification cost of a new instance. GLE method is also extended for cost-sensitive multi-label classification [7] and uses social tagging by considering tag counts information as the misclassification costs. Social tags are the text labels which may be assigned by the domain experts or the users. These social tags may consist of errors or noise. Tag count is the number of users, who have assigned tag to the given resource. This tag count shows the confidence degree associated with the tag. The researchers have observed that, if the same tag is annotated by many users to the visually same images, then these tags may represent the semantic concept of the image. The information of tag count is used for training the cost sensitive classifier which may help to minimize the training error related with tag counts

2.3 FURIA (Fuzzy Unordered Rule Induction Algorithm)

In [5], performance of multi label classification using fuzzy rule classifier is evaluated. FURIA is used as a base classifier multi label problem transformation approach. For experiment, six datasets, eight different base classifiers, four problem transformation methods are considered, and five different performance metrics are measured to evaluate the performance of fuzzy rule classifier (FURIA). Experiments shows that the FURIA outperforms when compared with other eight different base classifiers.

3. IMPLEMENTATION DETAILS

3.1 Problem Definition

To design, develop and implement Generalized k-labelset ensemble method (GLE) for the multi-label classification and cost-sensitive classification using fuzzy rule classifier.

3.2 Mathematical Model

The system takes multi label training dataset (D), number of models (M), set of labels (L)and size of labelset (k) as the input. GLE training process trains the LP classifiers using randomly selected k labelsets from the given original set of labels. For each base classifier weight coefficients are learned. After learning coefficients, GLE classification process is performed. This GLE method will predict the labels for new instance by minimizing the global error between ground truth and prediction. Using fuzzy rule classifier as the base classifier will improve the performance of the GLE system. Proposed system S is defined as follows,

S = { D, I, TP, Q, CP, FZ,O }

D is the multi label dataset.

I is the new data instance.

TP is the training process of GLE.

Q is the weight coefficient of the base classifier.

CP is is the classification process of GLE.

FZ is fuzzy rule classifier (FURIA) used as base classifier for GLE in training process.

O is the set of predicted labels for new data instance.

3.3 Proposed System

Figure 1 shows the block diagram of the proposed system. As shown in the figure 1, GLE makes use of fuzzy classifier FURIA.

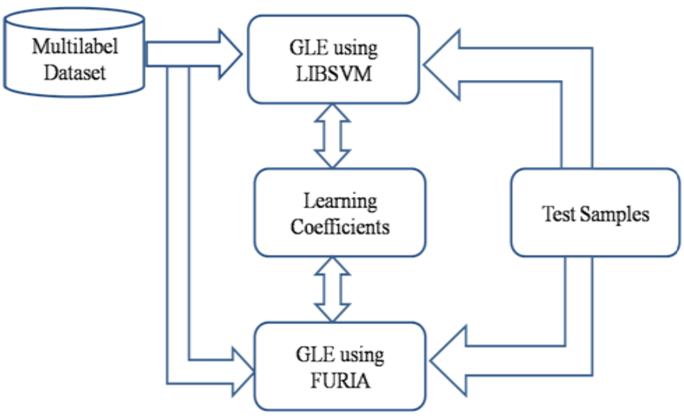

Figure 1.Proposed system block diagram

1. GLE Training Process

For training process the input is , training set D, the number of models M, set of labels L, and size of labelset k.

Initially the L^k subsets is assigned to the S. Then for each model m, the k-labelset selected randomly from S and assigned to the Rm. The LP classifier gm is trained based on training set D and Rm. Transformed prediction of LP classifier gm is calculated using following formula, where g_m is the LP classifier, x_i is the instance, and j is label,

$$q(g_m(x_i), j) = \begin{cases} 1, if\ j\ \in\ R_m, and\ j\ is\ positive\ in\ g_m(x_i), \\ -1, if\ j\ \in R_m, and\ j\ is\ negative\ in\ g_m(x_i), \\ 0, if\ j\ does\ not\ belongs\ to\ R_m \end{cases}$$

After this the Rm is removed from S. Finally the coefficients are learned.

Output of training process is ensemble of LP classifier gm, corresponding k-labelsets Rm, coefficients Bm.

2. GLE Classification Process

For classification process the input is the test sample x, the number of models M, an ensemble of LP classifier gm, corresponding k-labelsets Rm, coefficients Bm.

For each and every label in the label space, the value is initialized to 0. Then for each LP classifier, if label belongs to corresponding labelset Rm, then the value for that label is calculated, consider rj is the jth label. Output of this process is multi-label classification vector r = (r1, r2, …, rk).

3. Learning Coefficients

a. GLE for multi-label classification

In the objective function first term's aim is to reduce the global error between the multi-label ground truth Y and the prediction of LP classifier. The second term shows the two-norm regularization term of the coefficients β. The last term represents the hypergraph regularization.

$$\min_\beta \frac{1}{2}\ ||\ Y - \sum_{m=1}^{M} \beta_m Q_m\ \ ||_F^2 + \frac{\gamma}{2}\ \ ||\ \beta\ ||_2^2 + \frac{\gamma}{2}\ trace\Big(\big(\sum_{m=1}^{M} \beta_m Q_m\big)^T\ L\big(\sum_{m=1}^{M} \beta_m Q_m\ \big)\Big)$$

b. GLE for cost sensitive multi-label classification

The objective function is same as GLE for multi label classification, only one difference is there. The first term is modified to a cost-weighted global error by multiplying the global error with the multi-label misclassification cost matrix.

$$\min_\beta \frac{1}{2}\ ||\ C^\circ(Y - \sum_{m=1}^{M} \beta_m Q_m\ )\ ||_F^2\ \frac{\gamma}{2}\ ||\ \beta\ ||_2^2 + \frac{\gamma}{2}\ trace\Big(\big(\sum_{m=1}^{M} \beta_m Q_m\ \big)^T L\big(\sum_{m=1}^{M} \beta_m Q_m\ \big)\Big)$$

4. EXPERIMENTAL SETUP

4.1 Datasets

To conduct the experiments we use ten datasets. The datasets used are scene, bibtex, majorminer, enron, medical, cal500, four versions of delicious (dlc1, dlc2, dlc3, dlc4).

Scenes dataset: includes characteristics about images and their classes.

Enron dataset: contains e-mail messages dataset containing a subset of about 1700 labeled email messages.

Cal500 dataset: consisting of 500 popular Western songs from 500 different artists.

MajorMiner dataset: MajorMiner [3] is a web based game which collects the music tags. MajorMiner game provides a short length music clip to the player and he has to label the music clip with relevant phrases and words.

medical dataset: is used to classify clinical free text using some terms for describe each text.

BibTex dataset: contains metadata for the bibtex items like the title of the paper, the authors, etc.

Delicious datasets: is generated from the famous social bookmarking website del.icio.us. The delicious data is divided into 4 subset versions: from dlc1 to dlc4.

4.2 Performance Metrics

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 Expected Results

It is expected that using fuzzy rule classifier as a base classifier for GLE method performance metrics hamming loss, ranking loss and one error would be minimized whereas average precision, subset accuracy would be improved.

It is expected that the proposed system will improve the performance of the GLE method.

5. CONCLUSIONS

It is observed by researchers that, Generalized k-labelset ensemble implemented with crisp LIBSVM classifier as a base classifier provides improved performance for multi-label classification as compared with both Label powerset and Random k-labelset. It is also known that fuzzy classifier like FURIA outperforms many of the existing methods used for problem transformation approach of multi-label classification. It is therefore interesting to observe performance of Generalized k-labelset ensemble using fuzzy classifier, with that of LIBSVM. It is expected that GLE using fuzzy classifier such as FURIA to outperform GLE using crisp classifier LIBSVM.

6. ACKNOWLEDGMENTS

I would like to express my special thank to all those people who have helped me to complete this work. I am very grateful to my guide, Prof. Dr. Shirish Sane, Head of Department, Computer Engineering, K.K.W.I.E.E.R., Nashik for his guidance, encouragement and the interest shown in this project He has continuously helped and encouraged me in my work. I am also thankful to Prof. Vaishali S. Tidake, for her expert suggestions and scholarly feedback.

7. 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] G. Tsoumakas, I. Katakis, and I. Vlahavas, Random klabelsets for multilabel classification" IEEE Trans. Knowl. Data Eng., vol. 23,no. 7, pp. 1079A S1089, Jul. 2011.

[3] M. I. Mandel and D. P. W. Ellis, A web-based game for collecting music metadata, J. New Music Res., vol. 37, no. 2, pp. 151165, 2008.

[4] 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.

[5] 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.

[6] 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.

[7] Hung-Yi Lo, Shou-De Lin, and Hsin-Min Wang, Generalized k-Labelsets Ensemble for Multi-Label and Cost-Sensitive Classification, IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 7, July2014.