A COMPARISON OF MULTI-LABEL CLASSIFICATION METHODS USING MEKA ON BENCHMARK DATASETS

Kavitha C.R[1], Mahalekshmi T[2]

[1]*Research Scholar, R&D, Bharathiar University, Coimbatore, Tamil Nadu, India*
[2]*Principal, Sree Narayana Institute of Technology, Kollam, Kerala, India*

Abstract

Recent research interest of many researchers is multi- label classification, where each instance is assigned a set of multiple class labels simultaneously. It is used to solve problems in different application domains such as text categorization, semantic scene classification, music categorization and protein function classification. This paper gives an overview of multi-label classification and its methods. This paper also presents a comparative analysis of multi-label classification methods using MEKA on various data sets such as Genbase and Enron.

Keywords: *Machine learning, Classification, Multi-Label Classification, MEKA, Binary Relevance (BR)*

---***---

1. INTRODUCTION

In traditional classification problems, each example is associated with just one of two or more classes. [1] Classification is a technique which is used to predict class of unseen instance as accurate as possible. [2] In traditional approaches, single-label classification [3] deals with the association of each instance of a dataset with just only one class label. If the instances are classified into one of the two classes is called binary classification [4]. The problem of classifying instances into one of the more than two classes is called multiclass classification [5]. For example, a set of images of fruits can be classified into oranges, apples or pears. Multiclass classification assumes that each instance is associated to one and only one label that is, a fruit can be either an apple or a pear but not both at the same time. In latest approaches, multi-label classification [6] deals with the association of each instance with one or more class labels. For example, a text can be associated with any of religion, politics, finance or education at the same time. The difference between the single-label and multi-label classification is that single-label are mutually exclusive whereas multi-label are interdependent from one another. [7] This paper is organized as follows. An overview of multi-label classification methods is given in Section 2. The MEKA tool is described in Section 3 followed by the description of the data set used in this study in Section 4. Performance Evaluation metrics for multi-label classification are discussed in Section 5. Section 6 presents experimental analysis of multi label classification methods [8] on different dataset which is followed by the discussion. Finally the conclusion is given in section 8 followed by the references.

2. MULTILABEL CLASSIFICATION METHODS – AN OVERVIEW

There are two categories of multi-label classification methods: Problem Transformation (PT) methods [2] and Algorithm Adaptation (AA) methods [2]. In PT methods, the problem is transformed into the single-label classification problem [9] and in AA methods; the existing algorithms are extended to handle the multi-label data directly. Here the PT methods are discussed. Though there are many methods under DT methods, Binary Relevance (BR) [8], Classifier Chains (CC) [10], Classifier Trellis (CT) [11] and Label Combination (LC) [8] which has been used for the study are described here.

2.1 Binary Relevance (BR)

BR is a problem transformation method that learns q binary classifiers; one for each labels in L. It transforms any multi-label problem into L binary problems. [12] Each binary classifier is then responsible for predicting the association of a single label. [13]

2.2 Classifier Chains (CC)

CC maintains the computational efficiency of the BR method and also considers the label dependencies into account for classification.CC involves $|L|$ binary classifiers. These are linked along a chain where each classifier deals with the binary relevance problem associated with label lj $\in$L. The feature space of each link in the chain is extended with the 0/1 label associations of all previous links [10].

2.3 Classifier Trellis (CT)

In CT, a fixed structure called a lattice or trellis is used which avoids the complexity involved in identifying a structure. A structure a-priori is used instead that improves the order of labels within that structure. [11] It uses label-frequency based pair wise mutual information in order to place the labels into the trellis rationally i.e. one that tries to maximize label dependence between parents and children.[11] An efficient hill climbing method is used to insert nodes into this trellis according to marginal dependence information, in a manner similar to the FS method. [11]

2.4 Label Combination (LC)

In Label power set, also known as label power set method the multi-label problem can be transformed into one multi-class single-label learning problem, using as target values for the class attribute all unique existing subsets of multi-labels present in the training instances (the distinct subsets of labels).[8] This method considers label dependency. It considers each unique occurrence of set of labels in multi-label training dataset as one class for newly transformed dataset. [12] For example, if an instance is associated with three labels L1, L2, L4 then the new single-label class will be L1,2,4. So the new transformed dataset is a single-label classification task and any single-label classifier can be applied to it.[12] For a new instance to classify, LP outputs the most probable class, which is actually a set of labels. Thus it considers label dependency and also no information is lost during classification. If the classifier can produce probability distribution over all classes, then LP can give rank among all labels using the approach of [14].

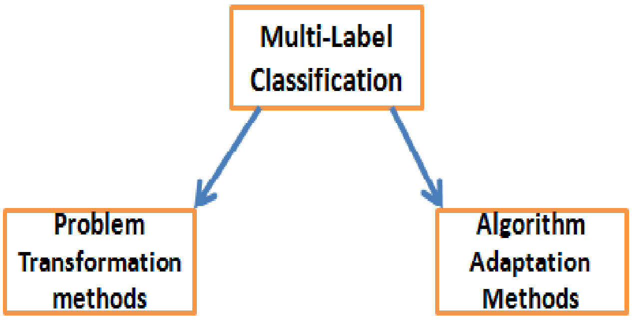

Fig-1. Categories of multi-label classification methods

3. MEKA

In this experiment, results were found out using MEKA (Multi-label Extension to WEKA [15]) [16], based upon the WEKA framework [15]. It provides the support for development, running and evaluation of multi-label and multi-target classifiers. MEKA is released under the GNU GPL v3 license. It can be downloaded from: http://sourceforge.net/projects/ meka/files/. MEKA uses Weka's ARFF_File format. It uses JAVA to provide multi-label classification methods.

4. DATASETS

In this research, two different benchmark datasets such as 'Genbase.arff' and 'Enron.arff' were used. The characteristic of the data sets are summarized in the Table. 1. Enron is a data set that includes the e-mails from 150 senior Enron officials grouped into many categories. The labels can be further grouped into four categories: coarse genre, included/forwarded information, primary topics and messages with emotional tone. [17] Genbase is a biological data set that has been downloaded from Mulan: A java library for multi label learning. [18]

Table -1: Benchmark datasets

Dataset	No of Instances	No of Attributes	No of Labels
Genbase	662	1186	27
Enron	1702	1001	53

5. PERFORMANCE EVALUATION METRICS

Hamming Loss (HL) is an example-based evaluation measure, which is defined as the measure of accuracy in a multi-label classification problem. It finds out how many times on an average, the relevance of an example to a class label are incorrectly predicted. [9]Hamming Loss is defined as follow in

$$\text{Hamming-Loss} = \frac{1}{m}\sum_{i=1}^{m}\frac{|Y_i \triangle Z_i|}{M}$$

where Δ stands for the symmetric difference of two sets, which is the set-theoretic equivalent of the exclusive disjunction (XOR operation) in Boolean logic. [9]

Accuracy (AC) is an example –based evaluation measure, which is defined as the proportion of the predicted correct labels to the total number (predicted and actual) of labels for that instance. [9]

$$\text{Accuracy A} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{|Yi \cap Zi|}{|Yi \cup Zi|}\right]$$

Exact Match (EM) is defined as the percentage of examples that have all their labels classified correctly. [9]

F1 Macro (FM) is a label based measure, which is the harmonic mean between precision and recall, where the average is calculated per label and then averaged across all labels. [19] [20] If p_j and r_j are the precision and recall for all λ_j ε h(xi) from λ_j ε Y_i, the macro-F1 is

$$macro\_F_1 = \frac{1}{Q}\sum_{j=1}^{Q}\frac{2 \times p_j \times r_j}{p_j + r_j}$$

One error (OE) is a ranking measure that evaluates how many times the top-ranked label is not in the set of relevant labels of the example. The metric one_error (f) takes values between 0 and 1. The smaller the value of one_error(f), the better the performance. [20] This evaluation metric is defined as

$$one\_error(f) = \frac{1}{N}\sum_{i=1}^{N}\left[\!\left[\ \left[\arg\max_{\lambda \in \mathcal{Y}} f(\mathbf{x}_i, \lambda)\right] \notin \mathcal{Y}_i\ \right]\!\right]$$

Where $\lambda \varepsilon L = \{\lambda 1, \lambda 2, \ldots, \lambda Q\}$ and $[\Pi]$ equals 1 if Π holds and 0 otherwise for any predicate Π.

6. COMPARISON STUDY OF DIFFERENT CLASSIFICATION METHOD USING MEKA

In this study, four multi-label classification methods such as Binary Relevance (BR), Classifier Chains (CC), Classifier Trellis (CT) and Label Combination (LC) are used. And four classifiers such as Naive Bayes (NB), JRip, SMO and J48 were used as the base classifiers. Multi-label classification is performed on two datasets such as Enron and Genbase using MEKA and their performance were evaluated based on accuracy (AC), Exact Match (EM), Hamming Loss (HL), F1 Macro (FM), One error (OE) and total time (TT). Figure 2 shows the evaluation of the classifiers based on accuracy. It is clear from the figure that when CT method was used with NB classifier, it gives less accuracy. When SMO method was used with BR, CC, CT and LC, it gives more accuracy, followed by J48 and JRIP when used with CC, CT and LC classifiers. First the experiment was conducted on Genbase dataset. Figure 3 displays the performance of classifiers based on Exact Match. Exact match measure is higher when J48 was used with LC, CT, &CC, When SMO was used with all classifiers and when JRIP was used with CC, CT & LC. Figure 4 shows the performance of classifiers based on hamming loss. When SMO was used with all the methods shows less hamming loss followed by J48 used with LC,CT and CC. Figure 5 give the performance of Classifiers based on total time. From the figure, it is clear that SMO method used with all the classifiers has taken long time.

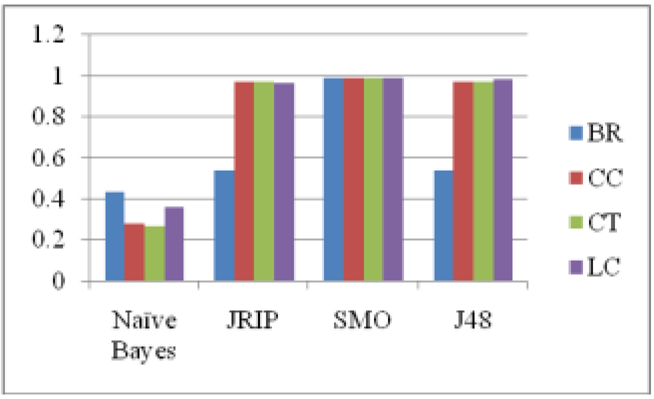

Fig 2 Performance of the classifiers based on accuracy

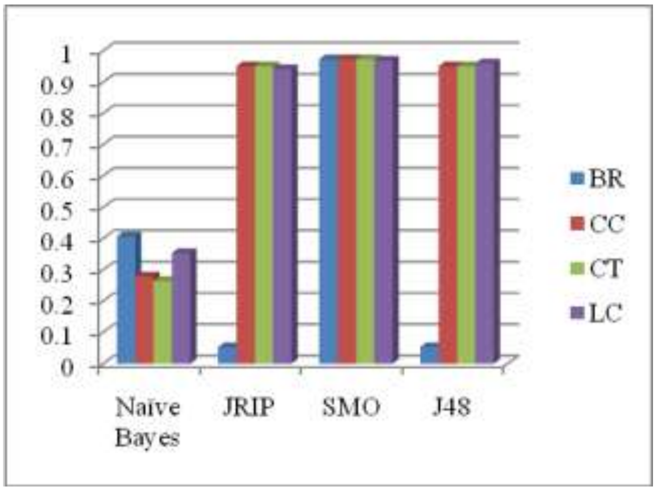

Fig - 3 Performance of classifiers based on Exact Match

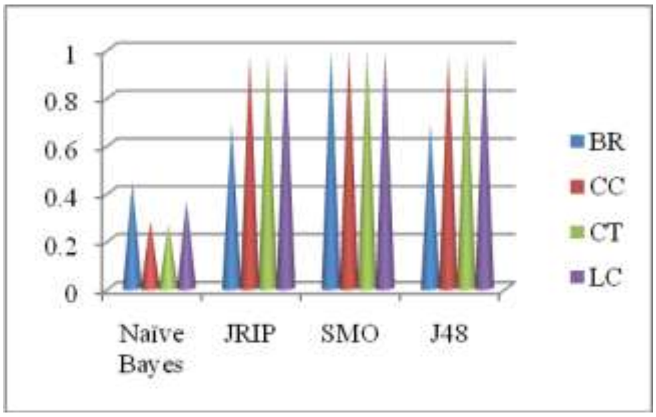

Fig - 4 Performance of classifiers based on hamming loss.

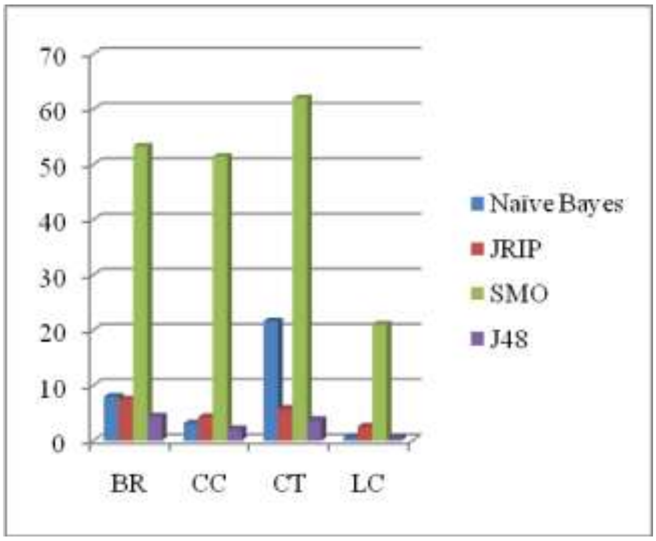

Fig - 5 Performance of Classifiers based on total time

From the above figure, it is clear that SMO method used with all the classifiers has taken long time.

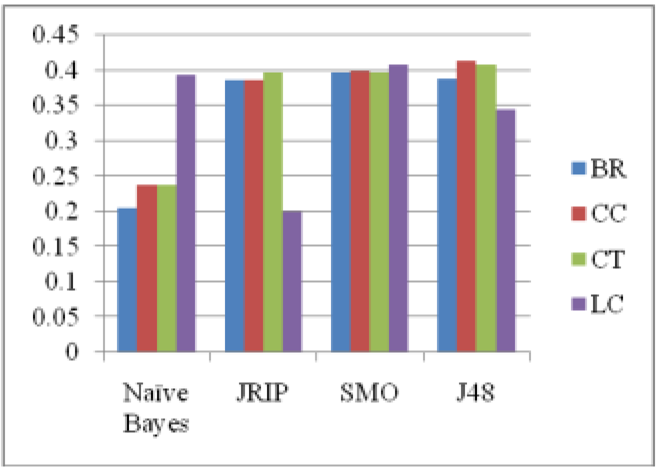

Fig - 6 Performance of the classifiers based on accuracy

When Naïve Bayes method used with BR classifier, it gives the less accuracy followed by JRIP method used with LC also gives less accuracy. From the above figure 6, it was found that J48 with CC classifier shows the highest accuracy followed by J48 with CC, SMO with LC, CT, CT and BR, JRIP with CT and NB with LC.

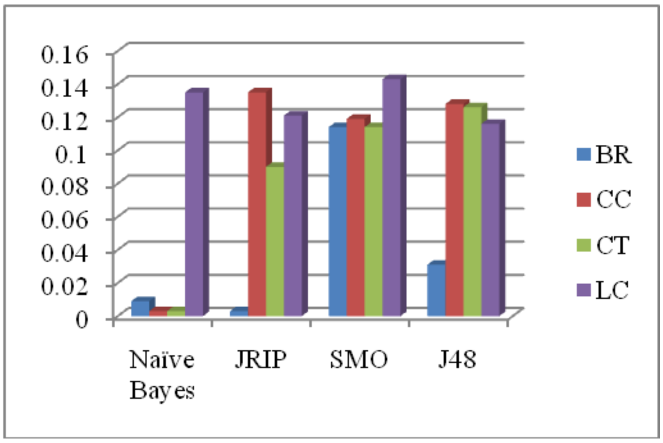

Fig - 7. Performance of classifiers based on Exact Match

From figure 7, it is clear that when SMO is used with LC classifier, it gives the highest exact match.

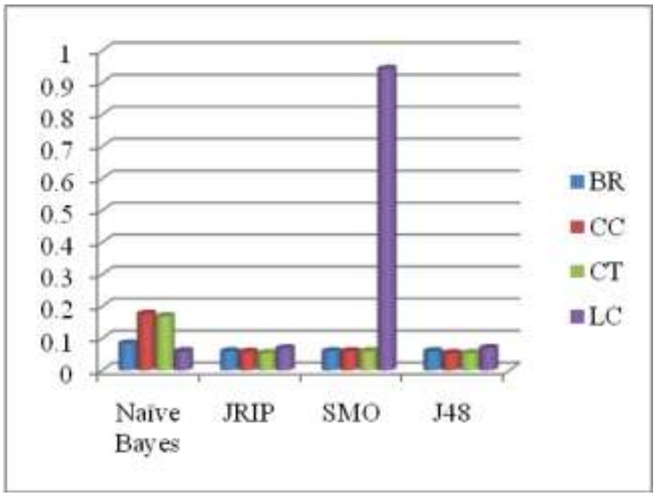

Fig - 8. Performance of classifiers based on hamming loss.

From the above figure 8, it was found that JRIP when used with all classifiers shows less hamming loss where as when SMO method used with LC classifiers shows the highest hamming loss.

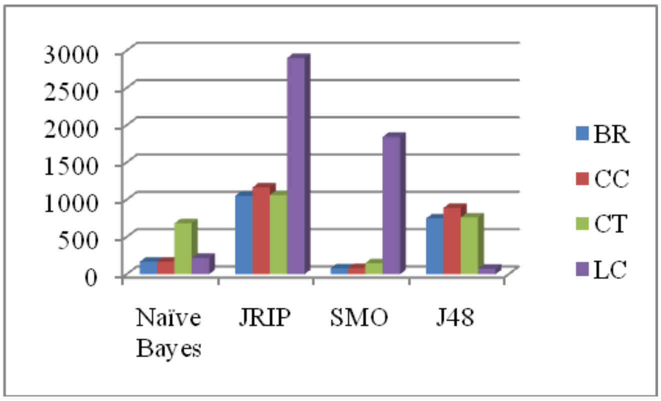

Fig - 9. Performance classifiers based on total time

It is clear from figure 9 that when JRIP clasifier is used with LC method, it has taken highest total time and when SMO is used with BR&CC; it has taken lowest total time followed by J48 used with LC method.

Table -2: Results from Genbase

Genbase/BR	AC	EM	HL	FM	OE	Time
Naïve Bayes	0.433	0.407	0.033	0.444	0.407	7.989
JRIP	0.536	0.053	0.036	0.691	0	7.504
SMO	0.989	0.973	0.001	0.993	0	53.297
J48	0.536	0.053	0.036	0.691	0	4.402
Genbase/CC	**AC**	**EM**	**HL**	**FM**	**OE**	**Time**
Naïve Bayes	0.28	0.279	0.034	0.281	0.593	3.16
JRIP	0.969	0.951	0.002	0.973	0.018	4.316
SMO	0.989	0.973	0.001	0.993	0	51.459
J48	0.969	0.951	0.002	0.973	0.018	2.149
Genbase/CT	**AC**	**EM**	**HL**	**FM**	**OE**	**Time**
Naïve Bayes	0.267	0.265	0.035	0.268	0.606	21.645
JRIP	0.969	0.951	0.002	0.973	0.018	5.787
SMO	0.989	0.973	0.001	0.993	0	62.027
J48	0.969	0.951	0.002	0.973	0.018	3.772
Genbase/LC	**AC**	**EM**	**HL**	**FM**	**OE**	**Time**
Naïve Bayes	0.361	0.354	0.054	0.364	0.658	0.661
JRIP	0.963	0.942	0.003	0.968	0.022	2.578
SMO	0.987	0.969	0.001	0.991	0	21.081
J48	0.981	0.96	0.002	0.986	0.004	0.543

7. DISCUSSION

The four multi-label classifiers BR, CC, CT and LC were applied in MEKA. The four multi-label classification methods were experimented in combination with the classifiers learning algorithms such as Naive Bayes, JRIP, SMO and J48. For performance evaluation, the accuracy (AC), Exact Match (EM), Hamming Loss (HL), F1 Macro (FM), One error (OE) and total time (TT) measures are used. The experiments were conducted on multi-label data sets Genbase and Enron which are in ARFF file format of the WEKA library. Genbase is the biological dataset concerned with protein function classification and gene function classification. Enron data set contains email messages. It is a subset of about 1700 labeled email messages. The analytical results on the two data sets are given in the Table 2 and Table 3. From the analysis of the experimented data and the graphs, it was found that with the Genbase data set, When CT method is used with NB classifier, it gives less accuracy. When SMO method is used with BR, CC, CT and LC, it gives more accuracy, followed by J48 and JRIP when used with CC, CT and LC classifiers. Exact match measure is higher when J48 used with LC, CT, &CC. When SMO used with all classifiers and When JRIP is used with CC, CT & LC. When SMO used with all classifiers shows less hamming loss followed by J48 used with LC, CT and CC. When SMO used with all classifiers shows the best performance since the value of one error is zero followed by J48 with LC, CT& CC and JRIP with CC, CT &LC.

In the case of Enron database, when Naïve Bayes method used with BR classifier, it gives the less accuracy followed by JRIP method used with LC also gives less accuracy. From the above figure, it is found that J48 with CC classifier shows the highest accuracy followed by J48 with CC, SMO with LC, CT, CT and BR, JRIP with CT and NB with LC.

When SMO used with LC classifier gives the highest exact match. From the above figure 8, it was found that JRIP when used with all classifiers shows less hamming loss where as when SMO method used with LC classifiers shows the highest hamming loss. When JRIP method used with LC has taken highest total time and when SMO with BR&CC has taken lowest total time followed by J48 used with LC.

8. CONCLUSION

This paper presented the multi-label classification problem. It gave an overview of some of the multi-label classification methods and also provided a comparative study of the experiment results for the above methods. In this work, four different problem transformation methods of multi-label classification were used such as BR, CC, CT and LC on datasets such as Enron and Genbase. It is quite difficult to assess the performance of multi-label classification methods since many other factors can also influence it. From the experiment conducted it is clear that based on performance accuracy of the classifiers, CC, CT and LC are better multi-label classification methods than BR.

Table -3: Results from Enron

Enron/BR	**AC**	**EM**	**HL**	**FM**	**OE**	**Time**
Naïve Bayes	0.205	0.009	0.084	0.301	0.437	166.65
JRIP	0.387	0.003	0.06	0.526	0.28	1051.53
SMO	0.397	0.114	0.06	0.51	0.401	79.407
J48	0.388	0.031	0.06	0.519	0.387	748.601
Enron/CC	**AC**	**EM**	**HL**	**FM**	**OE**	**Time**
Naïve Bayes	0.238	0.003	0.177	0.353	0.648	169.754
JRIP	0.386	0.135	0.058	0.482	0.347	1164.81
SMO	0.399	0.119	0.06	0.508	0.406	83.429
J48	0.414	0.128	0.054	0.519	0.358	889.228
Enron/CT	**AC**	**EM**	**HL**	**FM**	**OE**	**Time**
Naïve Bayes	0.238	0.003	0.169	0.355	0.665	682.159
JRIP	0.398	0.09	0.054	0.51	0.339	1062.341
SMO	0.398	0.114	0.06	0.51	0.409	148.989
J48	0.409	0.126	0.054	0.516	0.349	762.436
Enron/LC	**AC**	**EM**	**HL**	**FM**	**OE**	**Time**
Naïve Bayes	0.394	0.135	0.06	0.494	0.378	218.452
JRIP	0.199	0.121	0.069	0.234	0.667	2902.027
SMO	0.408	0.143	0.943	0.511	0.34	1842.296
J48	0.344	0.116	0.07	0.436	0.465	73.226

REFERENCES

[1] R. Cerri and A. C. P. L. F. de Carvalho, "New top-down methods using SVMs for Hierarchical Multilabel Classification problems," The 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, 2010, pp. 1-8.

[2] Alazaidah, Raed, Fadi Thabtah, and Qasem Al-Radaideh. "A Multi-Label Classification Approach Based on Correlations Among Labels", International Journal of Advanced Computer Science and Applications, 2015.

[3] Michielan L, Terfloth L, Gasteiger J, Moro S, “Comparison of multilabel and single-label classification applied to the prediction of the isoform specificity of cytochrome p450 substrates”, J Chem Inf Model. 2009 Nov;49(11):2588-605. doi: 10.1021/ci900299a.

[4] Samuel Mudd, Briggs' Information-processing Model of the Binary Classification Task, University of Illinois Press,NJ,1983

[5] Brian Kolo, Binary and Multiclass Classification, Weatherford Press, USA, 2010

[6] Grigorios Tsoumakas, Ioannis Katakis,” Multi-Label Classification: An Overview”, International Journal of Data Warehousing and Mining (IJDWM), IGI PUBLISHING, Volume 3, Issue 3, 2007

[7] Hua Wang. "Multi-label Feature Transform for Image Classifications", Lecture Notes in Computer Science, 2010

[8] Everton Alvares Cherman, Maria Carolina Monard, Jean Metz, “Multi-label Problem Transformation Methods: a Case Study”, Clei Electronic Journal, Volume 14, Number 1, Paper 4, April 2011

[9] Hiteshri Modi, Mahesh Panchal, 'Experimental Comparison of Different Problem Transformation Methods for Multi-Label Classification using MEKA', International Journal of Computer Applications (0975 – 8887) Volume 59– No.15, December 2012.

[10] Jesse Read, Bernhard Pfahringer, Geoffrey Holmes, Eibe Frank, “Classifier chains for multi-label classification”, In Proceeding(s) of the ECML/PKDD, volume 5782 of Lecture Notes in Computer Science, pages 254–269. Springer, 2009.

[11] J. Read, L. Martino, P. M. Olmos, David Luengo,” Scalable Multi-Output Label Prediction: From Classifier Chains To Classifier Trellises”, arXiv:1501.04870v1 [stat.ML] 20 Jan 2015

[12] Shweta C. Dharmadhikari Maya Ingle Parag Kulkarni,'A Comparative Analysis of Supervised Multi-label Text Classification Methods', International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622

[13] Huaxiang Zhang, Bin Hu, , Xinying Feng,'A Multi-label Learning Approach Based on Mapping from Instance to Label', Pervasive Computing and the Networked World Volume 8351 of the series Lecture Notes in Computer Science pp 743-752

[14] Brinker, K., F¨urnkranz, J., H¨ullermeier, E.: A unified model for multilabel classification and ranking. In: Proceedings of the 17th European Conference on Artificial Intelligence (ECAI ’06), Riva del Garda, Italy (2006) 489–493

[15] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

[16] Read, Jesse and Reutemann, Peter and Pfahringer, Bernhard and Holmes, Geoff,'MEKA: A Multi-label/Multi-target Extension to Weka',Journal of

Machine Learning Research,2016,17,21,1--5,http://jmlr.org/papers/v17/12-164.html

[17] Dragi Kocev, Ivica Slavkov, and SaˇsoDˇzeroski,'Feature ranking for multi-label classification using predictive clustering trees', International workhop on solving complex machine learning problems with ensemble methods in conjunction with ECML/PKDD, pp. 56-68 (2013)

[18] S. Diplaris, G. Tsoumakas, P. Mitkas and I. Vlahavas. Protein Classification with Multiple Algorithms, Proc. 10th Panhellenic Conference on Informatics (PCI 2005), pp. 448-456, Volos, Greece, November 2005.

[19] Madjarov, Gjorgji, Dejan Gjorgjevikj, Ivica Dimitrovski,and Sašo Džeroski. "The use of data-derived label hierarchies in multi-label classification", Journal of Intelligent Information Systems, 2016.

[20] Gjorgji Madjarov1, Ivica Dimitrovski, Dejan Gjorgjevikj,and Saˇso Dˇzeroski, ' Evaluation of Different Data-Derived Label Hierarchies in Multi-label Classification', New Frontiers in Mining Complex Patterns Volume 8983 of the series Lecture Notes in Computer Science pp 19-37.