A Survey and Current Research Challenges in Multi-Label Classification Methods

Purvi Prajapati, Amit Thakkar, Amit Ganatra

Abstract*— Classification is used to predict class of unseen instance as accurate as possible. Multi label classification is a variant of single label classification where set of labels associated with single instance. Multi label classification is used by modern applications, such as text classification, functional genomics, image classification, music categorization etc. This paper introduces the task of multi-label classification, methods for multi-label classification and evolution measure for multi-label classification. Also done comparative analysis of multi label classification methods on the basis of theoretical study and than on the basis of simulation done on various data sets.

***Keywords*—Classification, Single label problem, Multi label problem**

I. Introduction

A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one class label and several attributes. The goal of classifier is to produce a model which predicts label of the test data given only the test data attributes.

In classification problems, each instance of a dataset is associated with just one class label that is single label classification. (As shown in fig. I)

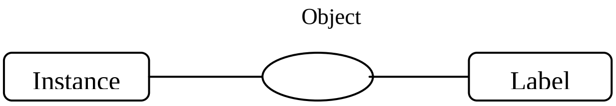

Figure I: Single Label Classification

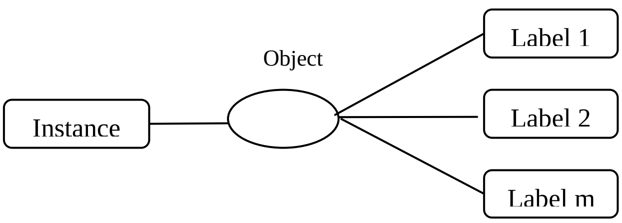

Figure II: Multi Label Classification

Manuscript received February, 2012.

Purvi Prajapati, Department of Information Technology, Charotar University of Science and Technology Changa, Anand, Gujarat, India. (purviprajapati.it@ecchanga.ac.in)

Amit Thakkar, Department of Information Technology, Charotar University of Science and Technology Changa, Anand, Gujarat, India. (amitthakkar.it@ecchanga.ac.in)

Amit Ganatra, U and P U Patel Department of Computer Engineering, , Charotar University of Science and Technology Changa, Anand, Gujarat, India. (amitganatra.ce@ecchanga.ac.in)

However, there are many classification tasks where each instance can be associated with one or more class labels. This group of problems represents an area known as Multi-Label Classification. (As shown in fig. II)

Multi-label classification methods are increasingly required by modern applications, such as text classification, gene functionality, music categorization and semantic scene classification. The number of class labels is predicted for each instance.

This paper is organized as follows. Start discussion of multi label classification in Section II. In Section III and IV explains the working of multi label classification methods with its comparative study. Evaluation measures for multi label classification are discussed in Section V. Section VI presents experimental analysis of multi label classification methods on different dataset. Finally current research challenges and concluded work in Section VII and Section VIII respectively.

II. MULTI-LABEL CLASSIFICATION

In single label problems, set of instances is D, set of labels is L. For each instance $d \in D$, select label set $l \in L$. So the Single label representation is (d, l). [1, 2]

In multi-label problems, set of instances is D, set of labels is L. For each instance $d \in D$, select label sub set $S \subseteq L$. So the Multi label representation: (d, S). [1, 2]

There are mainly two methods for multi-label classification problems: (1) problem transformation method and (2) algorithm adaptation method. Problem transformation method transfers multi-label problems into single label problems. And algorithm adaptation method extends specific learning algorithm to handle multi-label problems.

Below table shows example of multi label problem, with five class labels. L = {rec, sport, swim, auto, run}

Table I: Example of Multi Label Problem

Attributes		**Class Labels**				
A	**B**	**rec**	**sport**	**swim**	**auto**	**run**
A	1	✓	✓	✓		
A	2	✓	✓			✓
A	2	✓	✓	✓		✓
B	1	✓	✓			
B	2	✓			✓	

III. Problem Transformation Method

In this method, the main idea is to transfer multi label problem into a set of single label problems. It is an algorithm independent method so any traditional classification algorithm can be used to deal with multi label problems. There are several problem transformation methods available for transferring multi label problems into single label problems. [1, 3, 4, 5, 7, 8]

For example, L = {l1, l2, l3, l4} where L is number of labels.

Table II: Multi Label Example

Example	**Label set**
1	{l1, l4}
2	{l3, l4}
3	{l1}
4	{l2, l3, l4}

A. BINARY RELEVANCE (BR)

This method is basically binary classification of labels. So it transforms original multi label dataset into |L| single label dataset. It builds binary classifier for each label. For the classification of new instance, BR gives union of the labels that are positively predicted by |L| classifier.

As shown in Table III, BR method gives four individual classifier (|L|=4) from Table II.

Table III: Binary Relevance Method

Ex #	l1
1	✓
2	
3	✓
4	

Ex #	l2
1	
2	
3	
4	✓

Ex #	l3
1	
2	✓
3	
4	✓

Ex #	**l4**
1	✓
2	✓
3	
4	✓

B. RANKING VIA SINGLE LABEL

This method transforms the multi label dataset into single label dataset. There are different ways for transformation like ignore multi label instance, find maximum count of labels, find minimum count of labels, random selection of label and assign weight to each labels. A single label classifier outputs a vote (probability) for each class label which produce ranking. (As shown if Table IV)

Table IV: Ranking via Single label

Ex #	Label set
3	{l1}

(a) Ignore

Ex #	Label
1	l4
2	l4
3	l1
4	l4

(b) Maximun

Ex #	Label
1	l4
2	l4
3	l1
4	l4

(c) Minimun

Ex #	Label
1	l4
2	l4
3	l1
4	l4

(d) Random

Ex #	Label	Weight
1	l1	0.50
1	l4	0.50
2	l3	0.50
2	l4	0.50
3	l1	1.00
4	l2	0.33
4	l3	0.33
4	l4	0.33

(e) Copy weight

C. RANKING VIA PAIR-WISE COMPARISON (RPC)

This method performs pair wise comparison of labels. It learns m=k(k-1)/2 binary models, one model for each pair of labels. (Where k is number of labels,k=|L|) Model is trained based on examples that are annotated by at least one of the labels, but not both. So for new instance, all m models are invoked and ranking is obtained by counting the votes received by each label.[9] (see Table V and Table VI)

Table V: one classifier for each pair of labels

Ex #	l1_l2
1	l1
3	l1
4	l2

Ex #	l1_l3
1	l1
2	l3
3	l1
4	l3

Ex #	l1_l4
2	l4
3	l1
4	l4

Ex #	l2_l3
2	l3

Ex #	l2_l4
1	l4
2	l4

Ex #	l3_l4
1	l4

Table VI: Ranking of labels for new instance

New instance x':

l1_l2	l1_l3	l1_l4	l2_l3	l2_l4	l3_l4
l1	l3	l1	l3	l2	l3

Votes for each label:

L1	l2	l3	l4
2	1	3	0

Ranking based on votes: r (l3) > r (l1) > r (l2) > r (l4)

D. CALIBRATED LABEL RANKING (CLR)

This method is extension of RPC method. It introduces one additional virtual label V, with the purpose of separating positive and negative labels. Final ranking is obtained by considering votes of all labels including virtual label V. (As shown in Table VII)

Table VII: Calibrated Ranking of labels

Ex #	l1_V
1	l1
2	V
3	l1
4	V

Ex #	l2_V
1	V
2	V
3	V
4	l2

Ex #	l3_V
1	V
2	l3
3	V
4	l3

Ex #	l4_V
1	l4
2	l4
3	V
4	l4

Table VIII: Ranking of labels for new instance

New instance x':

l1_l2	l1_l3	l1_l4	l2_l3	l2_l4	l3_l4
l1	l1	l1	l2	l2	l4

l1_V	l2_V	l3_V	l4_V
l1	V	V	V

Votes for each label:

l1	l2	l3	l4	V
4	2	0	1	3

Ranking based on votes:

r (l1) > r (lV) > r (l2) > r (l4) > r(l3)

E. LABEL POWER SET (LP)

This method replaces each unique subset (Distinct Label Set) of labels that exists in multi label dataset with single label. So LP introduces new set of class labels. For new instance, base classifier of LP predicts one label which is originally a set of labels in multi label dataset. Below Table IX shows LP method performs on Table II. For first instance label l1, l2 are present and label l3,l4 are absent so LP gives 1001.

Table IX: Label power set

Ex #	Label(l1l2l3l4)
1	1001
2	0011
3	1000
4	0111

F. PRUNED SET (PS)

This method transforms multi label dataset into single label dataset using LP method. Pruning parameter p (user defined threshold) identifies pruned examples in given multi label dataset. Pruned examples are those whose label set occur less time than pruning parameter p. The PS method identifies less important examples from multi label dataset. As shown in below Table X last row is discarded considering pruning parameter 3.

Table X: Pruned set method for p=3

Label-set	Count
l1	16
l2	14
l2, l3	12
l1,l4	8
l3,l4	7
l1,l2,l3	2

G. RANDOM K-LABEL SET (RAKEL)

This method randomly breaks a large set of labels into a number n of subsets of small size k, called k-label sets. For training of multi label classifier LP method is used, an average decision is calculated for each label in L. And final decision is positive for a given label if the average decision is larger than threshold t. It considers label correlation ship and avoids LP problems.

Table XI: Comparative Study of Problem Transformation Methods

Method	**Merits**	**Demerits**
BR	Simple binary classification and relatively fast.	Does not consider label correlation ship .
Ranking via single label	Conceptually Simple	Not dealing well with overlapping of labels.
RPC	Flexible method	Consume more prediction time and more memory space.
CLR	It deals with pair wise comparison of each label with virtual label and it also provide ranking.	It is conceptually expensive method. Unlabeled data is not considered during classification.
LP	It considers label correlation ship.	Conceptually complex method and leads to over fitting of training data.
PS	Run faster and considers label correlation ship.	Dependence on predictions of base classifier.
RAkEL	Simpler, considers label correlation ship and more predictive capability.	Consumes more time and Unlabeled data is not considered during classification.

IV. ALGORITHM ADAPTATION METHOD

In this method, single label classifier is extended to develop multi label classifier to handle multi label problems. So this method is an algorithm dependent method. Various algorithm adaptation methods are developed based on different algorithms. [1, 3, 4, 5, 7, 8]

A. MULTI-LABEL DECISION TREE (C4.5)

This algorithm is the extension of basic decision tree algorithm for handling multi label data. In basic decision tree algorithm, entropy formula is modified to handle multiple labels.

Entropy (D) =

$$-\sum_{j=1}^{q} p(\lambda j)\log p(\lambda j)+q(\lambda j)\log q(\lambda j) \qquad (1)$$

Where,

$p(\lambda j) =$ Relative frequency of class λj

$q(\lambda j) = 1 - q(\lambda j)$

B. MULTI-LAYER NEURAL NETWORK (MLNN)

The multi-label neural network uses the multi layer feed forward neural network as its base algorithm. Adapting neural network algorithm to classify multi-label instances requires three key steps: (1) Creating a new error function that captures the characteristics of multi-label learning. (2) Modify the network to minimize this new error function. (3) Using threshold function to determine an output is in the relevant set of labels.

C. BACK PROPAGATION MULTI-LABEL LEARNING (BPMLL)

BPMLL extends basic back-propagation algorithm by introducing a new global error function that captures the characteristics of multi label learning.

D. MULTI-LABEL K NEAREST NEIGHBORS (MLKNN)

This algorithm is the extension of kNN algorithm. It uses the kNN algorithm independently for each label. It finds the k nearest examples to the test instance and considers those that are labeled with positive and negative. (MLkNN has also the capability of producing a ranking of the labels as an output.)

E. MULTI-LABEL BOOSTING (ADABOOST.MH, ADABOOST.MR)

These two algorithms are extensions of basic AdaBoost algorithm for handling multi-label data. Hamming loss is reduced using AdaBoost.MH and accuracy is increased using AdaBoost.MR.

Table XII: Comparative study of Algorithm Adaptation Methods

Method	Merits	Demerits
C4.5	Easy to learn and more informative attributes are used for splitting decision tree.	Does not consider label correlation ship.
BPMLL	Provides better generalization capability to learning system.	Because of neural network complexity becomes high in training phase.
MLkNN	Work well on image and text data. Better performance compared to other algorithms.	Unlabeled data is not considered for classification.
AdaBoost.MH AdaBoost.MR	Improved accuracy and minimized hamming loss.	Unlabeled data is not considered for classification.

Table XIII: Comparative study of MLC methods

Problem Transformation	Algorithm Adaptation
Algorithm Independent	Algorithm Dependent
Multiple model or single model is used	Single model is used
Data Preprocessing is required	Limited preprocessing is required.

V. EVALUATION MEASURE

The evaluation of multi label problems is different than single label problems. Multi label problems are associated with more than one labels therefore classification of an instance may be partially correct or partially incorrect. Mainly there are two types of evaluation measures for multi label classification problems: (1) example based measure and (2) label based measure. [4, 7, 8]

A. EXAMPLE BASED MEASURES

Let (x , Y) be a multi-label example, $Y \subseteq L$. Let h be a multi-label classifier. Let z=h(x) be a set of labels predicted by h for (x , Y).

Accuracy:

Accuracy for each instance is defined as the proportion of the predicted correct labels to the total number (predicted and actual) of labels for that instance. Overall accuracy is the averages across all instances.

$$\text{Accuracy}(H,D) = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \qquad (2)$$

Precision:

Precision is the percentage of predicted labels that were correct.

$$\text{Precision}(H,D) = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i \cap Z_i|}{|Z_i|} \qquad (3)$$

Recall:

Recall is the percentage of correct labels that were predicted.

$$\text{Recall}(H,D) = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i \cap Z_i|}{|Y_i|} \qquad (4)$$

Hamming loss:

Hamming Loss reports how many times on average, the relevance of an example to a class label is incorrectly predicted. Therefore, hamming loss takes into account the prediction error (an incorrect label is predicted) and the missing error (a relevant label not predicted), normalized over total number of classes and total number of examples.

$$\text{Hamming Loss}(H,D) = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|Y_i \Delta Z_i|}{|L|} \qquad (5)$$

Where, Δ stands for the symmetric difference of two sets.

B. LABEL BASED MEASURES

Calculate a binary evaluation measure separately for each label. Two averaging operations are used across all labels: micro and macro average. Binary evaluation measure is calculated using parameters of confusion matrix. (true positive, true negative, false positive and false negative)

$$M_{macro} = \frac{1}{|L|}\sum_{\lambda=1}^{|L|} M(tp_\lambda, fp_\lambda, tn_\lambda, fn_\lambda) \qquad (6)$$

$$M_{micro} = M\left(\sum_{\lambda=1}^{|L|} tp_\lambda, \sum_{\lambda=1}^{|L|} fp_\lambda, \sum_{\lambda=1}^{|L|} tn_\lambda, \sum_{\lambda=1}^{|L|} fn_\lambda\right) \qquad (7)$$

VI. EXPERIMENT

Four different multi label dataset are used in experiments: Gene base, Yeast, Medical and Scene. The results obtained in below table show that the algorithm adaptation methods has been best option for multi label methods compared to problem transformation methods.

Table XIV: Multi label data set statistics

Dataset	#Instances	#Attributes	#Labels
Genebase	662	1186	27
Yeast	2417	103	14
Medical	978	1449	45
Scene	2407	294	6

Table XV: Experimental results [6]

	Method	Problem Transformation			Algorithm Adaptation	
Dataset	Algorithm	BR	LP	CLR	ML-kNN	J48ML
Genebase	H.Loss (%)	0.1	0.2	0.1	0.5	0.2
	Precision (%)	98.9	98.8	99.0	99.2	98.9
	Recall(%)	98.3	97.2	98.7	90.1	97.7
Yeast	H.Loss(%)	24.5	27.9	22.0	19.4	28.1
	Precision(%)	59.9	54.1	65.2	72.9	53.3
	Recall(%)	57.4	53.7	58.4	57.0	57.2
Medical	H.Loss(%)	1.0	1.3	1.0	1.5	1.3
	Precision(%)	83.4	77.1	83.6	81.2	77.2
	Recall(%)	78.7	74.0	77.7	57.2	74.1
Scene	H.Loss(%)	13.7	14.4	13.8	8.5	14.4
	Precision(%)	61.7	59.8	60.6	82.0	59.7
	Recall(%)	62.2	59.7	65.2	67.2	60.8

VI. RESEARCH CHALLENGES

Following are the research challenges in the field of multi-label classification problem.

- To apply data preprocessing techniques like pruning, feature selection, handle missing value to improve the performance of MLC problem.
- To handle continuous attribute in MLC problem.
- Design a hierarchical structure for multiple labels to manage label correlation ship.
- To extract relevant label set from multiple label set.
- A novel approach is build to use both problem transformation method and algorithm adaptation method for improving performance of multi label classification problem.

VII. CONCLUSION

This paper presented study of different problem transformation methods and algorithm adaptation methods for multi label classification. From comparative study and experimental analysis on four dataset Genebase, Yeast, Medical and Scene concluded that algorithm adaptation method is best option for multi label classification compared to problem transformation method.

REFERENCES

[1] Grigorios Tsoumakas, Ioannis Katakis. Multi-Label Classification: An Overview, International Journal of Data Warehousing and Mining, David Taniar (Ed.), Idea Group Publishing, 3(3), pp. 1-13, 2007.

[2} Outline:Multi Label Classification http://www.tsc.uc3m.es/~jesse/talks/mend.pdf

[3] Read, J.; Pfahringer, B.; Holmes, G.; Dept. of Comput. Sci., Univ. of Waikato, Hamilton. A Pruned Problem Transformation Method for Multi-Label Classification. Data Mining, 2008. ICDM '08. Eighth IEEE International Conference, pages: 995 – 1000, 15-19 Dec. 2008.

[4] G. Tsoumakas and I. Vlahavas. Random k-labelsets: An ensemble method for multilabel classification. In Proceedings of the 18th European Conference on Machine Learning (ECML 2007), 2007.

[5] Learning from Multi Label Data http://mlkd.csd.auth.gr/multilabel.html.

[6] Eva Gibaja, Manuel Victoriano, Jose Luis Avila-Jimenez, Sebastian Ventura. A TDIDT Technique for Multi-label Classification, IEEE, 2010.

[7] G. Tsoumakas, I.Katakis, and I. Vlahavas. "Mining Multi-label Data", Data Mining and Knowledge Discovery Handbook, O. Maimon, L. Rokach (Ed.) Springer, 2nd edition 2010.

[8] Araken M Santos, Anne M P Canuto and Antonino Feitosa Neto. A Comparative Analysis of Classification Methods to Multi-label Tasks in Different Application Domains. International journal of computer information systems and idustrail management applications. ISSN 2150-7988 volume 3, pp 218-227, 2011.

[9] Klaus Brinker and Johannes Furnkranz and Eyke Hullermeier. A Unified Model for Multilabel Classification and Ranking. Proceedings of the 2006 conference on ECAI.

Purvi Prajapati has received her B.E degree in Information Technology from Sardar Patel University, Vidhyanagar, Gujarat, India in 2004. Since 2006 she has been with faculty of Engineering and Technology, Charotar University of Science and Technology, Changa, Gujarat, Where she is currently working as an Assistant Professor in the Department of Information Technology. She has joined M.Tech at Charotar University of Science and Technology, Changa, Gujarat, India in 2010. Her current research interest includes multi-label classification.

Amit Thakkar has received his B.E degree in Information Technology from Gujarat University, Gujarat, India in 2002 and master Degree from Dharmsinh Desai University, Gujarat, India in 2007. He has joined his Ph.D in the area of Multi relational Classification at Kadi Sarva Vishvidhalaya University, Gandhinagar, India in June 2010. Since 2002 he has been with faculty of Engineering and Technology, Charotar University of Science and Technology, Changa, Gujarat, Where he is currently working as an Associate Professor in the Department of Information Technology. He has published more than 20 research papers in the field of data mining and web technology. His current research interest includes Multi relational Data Mining, Relational Classification and Associate Classification.

Amit P. Ganatra (B.E.-'00-M.E. '04-Ph.D.* '11) has received his B.Tech. and M.Tech. degrees in 2000 and 2004 respectively from Dept. of Computer Engineering, DDIT-Nadiad from Gujarat University and Dharmsinh Desai University, Gujarat and he is pursuing Ph.D. in Information Fusion Techniques in Data Mining from KSV University, Gandhinagar, Gujarat, India and working closely with Dr.Y.P.Kosta (Guide). He is a member of IEEE and CSI. His areas of interest include Database and Data Mining, Artificial Intelligence, System software, soft computing and software engineering. He has 11 years of teaching experience at UG level and concurrently 7 years of teaching and research experience at PG level, having good teaching and research interests. In addition he has been involved in various consultancy projects for various industries.

After spending almost a year in C.U.Shah college of Engineering, Wadhwan, Gujarat, he joined CITC as a faculty member in 2001. His general research includes Data Warehousing, Data Mining and Business Intelligence, Artificial Intelligence and Soft Computing. In these areas, he is having good research record and published and contributed over 70 papers (Author and Co-author) published in referred journals and presented in various international conferences. He has guided more than 90 industry projects at under graduate level and 47 dissertations at Post Graduate level.

He is concurrently holding Associate Professor (Jan 2010 till date), Headship in computer Engineering Department (since 2001 to till date) at CSPIT, CHARUSAT and Deanship in Faculty of Technology-CHARUSAT (since Jan 2011 to till date), Gujarat. He is a member of Board of Studies (BOS), Faculty Board and Academic Council for CHARUSAT and member of BOS for Gujarat Technological University (GTU). He was the founder head of CE and IT departments of CITC (now CSPIT).