International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 12 | Dec-2017 www.irjet.net p-ISSN: 2395-0072

Multi-label Classification Methods: A Comparative Study

Jadon Mayurisingh Nareshpalsingh[1], Prof. Hiteshri N. Modi [2]

[1]*M.Tech Student, Dept. of Computer Engineering, U.V Patel College of Engineering, Gujarat, India*
[2]*Assistant Professor, Dept. of Information Technology, U.V Patel College of Engineering, Gujarat, India*

---***---

Abstract - *Classification is a vital task in data mining. There are two extensive categories of classification problem: Single-label classification and Multi-label classification. Traditional binary and multi-class classifications are two sub-categories of single-label classification. Multi-label classification methods are falling under two extensive categories of Problem Transformation Methods, Algorithm Adaptation Methods. Nowadays, Multi-label classification methods are getting more prominent now a days because of their expanding demand in various application domains. This paper introduces various methods of multi-label classification in depth.*

Key Words: **Problem Transformation Methods, Algorithm Adaptation Methods, Ensemble Pruned Methods, Binary Relevance, Label Power-Set, Pruned Set, Adaboost, Decision Tree, Random k-label sets, Multi-label classification.**

1. INTRODUCTION

Classification is a data mining function that assigns items in a collection to target categories or classes. Classification is used to predict categorical class labels*.* The goal of classification is to accurately predict the target class for each case in the data. There are two broader categories of classification methods: Single-label classification methods and Multi-label classification methods. Traditional binary and multi-class classifications are two sub-categories of single-label classification. Single label classification is one in which training examples are associated with only one label from a set of disjoint labels. But applications such as text categorization, semantic scene classification, protein function classification, music categorization may belong to more than one class label [2]. These applications require multi-label classification. In multi-label classification training examples are associated with more than one label from a set of disjoint labels. For example, in medical diagnosis a patient may suffer from diabetes and cancer both at the same time. Two major methods of multi-label classification exists: The first one is Problem Transformation(PT) Methods, in which multi-label classification problem is transformed into single-label classification problem and then classification is performed in the same way as in single-label classification. The second one is Algorithm Adaptation Methods in which, single-label existing algorithm is modified then applied directly on Multi-label data. Various problem transformation methods such as simple problem transformation methods (copy, copy-weight select-max, select-min, select-random, ignore), Binary methods Binary Relevance is belonging to this category. Label-combination methods are Label Power-Set method (LP), Pruned Problem Transformation Methods (also known as Pruned Set), Classifier Chains (CC). Pair wise Methods are Ranking via Pair wise Comparison (RPC), Calibrated Label Ranking (CLR), Ensemble Methods are Ensemble of Classifier Chains (ECC), Random k-label sets (RAkEL), Ensemble of Pruned Sets (EPS). Various Algorithm Adaptation Methods are C4.5 (ML-DT), Multi-Label k-Nearest Neighbors (MLKNN), AdaBoost.MH, AdaBoost.MR, Back-Propagation Multi-label Learning (BPMLL), Support Vector Machine with Heterogeneous Feature Kernel (SVM-HF), Ranking Support Vector Machine (Rank-SVM), and Multi-label Naive Bayesian (ML-NB).

Classification is an important theme in data mining. It is a process to assign a class to previously unseen data as accurately as possible. The unseen data are those records whose class value is not present or not predicted and using classification. In order to predict the class value, training set is used. Training set consists of records and each record contains a set of attributes, where one of the attribute is the class. From training set a classifier is created. Then that classifier's accuracy is determined using test set. If accuracy is acceptable then and only then classifier is used to predict class value of unseen data. Classification can be divided in two types: Single-label classification and Multi-label classification.

Single-label classification is to learn from a set of instances, each instance associated with a unique class label from a set of disjoint class labels B. Multi-label classification is to learn from a set of instances where each instance belongs to one or more classes in B. For example, a text document that talks about scientific contributions in medical science can belong to both science and health category, genes may have multiple functionalities (e.g. diseases) causing them to be associated with multiple classes, an image that captures a Sunset, tree, Wood, Elephant, Lion, Forest etc. a movie can simultaneously belong to action, crime, comedy, and drama, thriller categories, an email message can be tagged as both work, Study and research project; such examples are Plentiful. In text or music categorization, documents may belong to multiple types, such as government and health, or rock and Hip hop, Jazz [2].

1.1 Single-label Classification

Single-label Classification is categorized mainly into two parts: 1. Traditional Binary Classification, 2. Multi-class Classification. As you can see in figure 1, horse, plant, Person, at a same time each picture belongs to only one class-label.

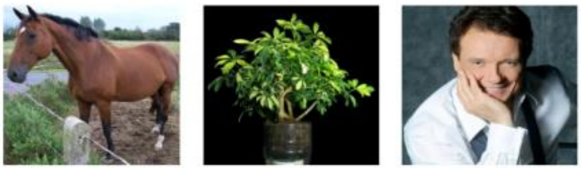

Fig -1: Example of Single label Classification

1.1.1 Traditional Binary Classification

In Traditional Binary Classification we can consider e.g. person may be smoker or non-smoker, e.g. Student may be present or absent, means two values are there 0 or 1.

1.1.2 Multi-class Classification

In Multi-class Classification we can consider e.g. Image of banana, orange, apple, each fruit image is belong to only one class label of fruit at same time.

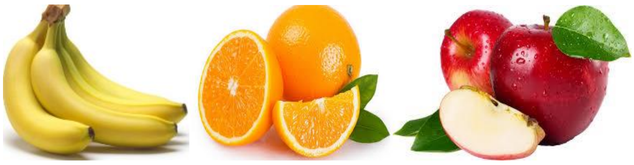

Fig -2: Example of Multi-class Classification

1.2 Multi-label Classification

In multi-label classification each image belongs to multiple labels at same time as shown in figure 3, e.g. Beach image belongs to trees, sky, water, land, chairs etc, e.g. Natural scene image have water, mountains, flowers, sky, greenery.

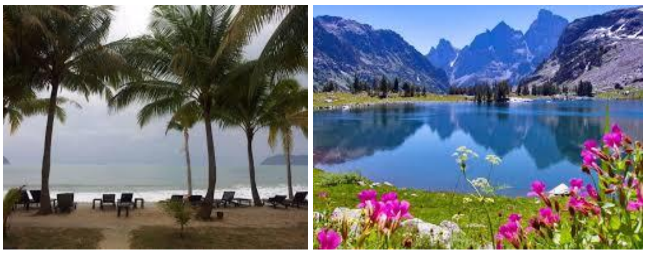

Fig -3: Example of Multi-label Classification

2. Methods for Multi-label Classification

Multi-label classification methods can be grouped in two categories:

(1) Problem Transformation Methods.
(2) Algorithm Adaptation Methods.

2.1 Problem Transformation methods

2.1.1 Simple Problem Transformation methods

Instance	Attribute	Label sets
1	A1	{B1,B2}
2	A2	{B1,B2,B3}
3	A3	{B4}
4	A4	{B1,B2,B5}
5	A5	{B2,B4}

Table- 1: Example of multi-label dataset

There exist many simple problem transformation methods that transform multi-label Dataset into single-label dataset so that existing single-label classifier can be applied to Multi-label dataset [2]. The Copy Transformation method replaces each multi-label instance with a single class label for each class-label occurring in that instance. A variation of this method, dubbed Copy-weight, associates a weight to each produced instances. These methods increase the Instances, but no information loss is there. The Select Transformation method replaces the Label-Set (B) of instance with one of its member. Depending on which one member is selected from B, there are several versions existing, such as, select-min (select least frequent label), Select-max (select most frequent label), select-random (randomly select any label). These methods are very simple but it loses some information. The Ignore Transformation method simply ignores all the instances which has multiple labels and takes only single-label instances in training. There is major information loss in this method.

In.	Label
1a	B1
1b	B2
2a	B1
2b	B2
2c	B3
3a	B4
4a	B1
4b	B2
4c	B5
5a	B2
5b	B4

(a)

In.	Label	Weight
1a	B1	0.33
1b	B2	0.25
2a	B1	0.33
2b	B2	0.25
2c	B3	1.00
3a	B4	0.50
4a	B1	0.33
4b	B2	0.25
4c	B5	1.00
5a	B2	0.25
5b	B4	0.50

(b)

In.	label
1	B2
2	B2
3	B4
4	B2
5	B2

(c)

In.	label
1	B1
2	B3
3	B4
4	B5
5	B4

(d)

In.	label
1	B1
2	B2
3	B4
4	B2
5	B4

(e)

In.	label
1	B4

(f)

(a) Copy (b) Copy-weight (c) Select-max (d) Select-min (e) Select-random (f) Ignore

Table -2: Transformation using Simple Problem Transformation Methods

2.1.2 Binary Methods

2.1.2.1 Binary Relevance:

A Binary Relevance is one of the most popular transformation methods which learns q binary classifiers (q=|B|, total number of classes (B) in a dataset), one for each label. BR transforms the original dataset into q datasets, where each dataset contains all the instances of original dataset and trains a classifier on each of these datasets. If particular instance contains label Bj (1≤ j ≤ q), then it is labeled positively otherwise labeled negatively. The instances are labeled positively if they have the existing label, otherwise they are labeled negatively. Table- 3 shows dataset that are constructed using BR from dataset of Table-1. From these datasets, it is easy to train a binary classifier for each dataset. For a new instance to classify, BR outputs the union of the labels that are predicted positively by the q classifiers. BR is used in many practical applications, but it can be used only in applications which do not hold label dependency in the data. This is the major limitation of BR. It is an algorithm independent method.

In.	Label	In.	Label	In.	Label	In.	Label	In.	Label
1	B1	1	B2	1	-B3	1	-B4	1	-B5
2	B1	2	B2	2	B3	2	-B4	2	-B5
3	-B1	3	-B2	3	-B3	3	B4	3	-B5
4	B1	4	B2	4	-B3	4	-B4	4	B5
5	-B1	5	B2	5	-B3	5	B4	5	-B5

Table -3: Transformation using Binary Relevance

2.1.3 Label-combination Methods

2.1.3.1 Label Power-set (LP):

These methods remove the limitation of the previous methods by taking into account the correlation and dependencies among labels. It is a simple and less-common problem transformation method. Label Power-set considers each unique occurrence of set of labels in multi-label training dataset as one class for newly transformed dataset. For example, if an instance is associated with three labels B1, B2, B4 then the new single-label class will be B1,2,4. So the new transformed dataset is a single-label classification task and any single-label classifier can be applied to it (Table- 4). 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 [2]. Given a new instance x with unknown dataset, (Table- 4) shows an example of probability distribution by LP. For label ranking, for each label calculates the sum of probability of classes that contain it. So, LP can do multi-label classification and also do the ranking among labels, which together called MLR (Multi-label Ranking). As discussed earlier, LP considers label dependencies during classification. But, it's computational complexity depends on the number of distinct label-sets that exists in the training set. This complexity is upper bounded by min (m, 2q). The number of distinct label is typically much smaller, but it is still larger than q and poses important complexity problem, especially for large values of m and q. When large number of label-set is there and from which many are associated with very few examples, makes the learning process difficult and provide class imbalance problem. Another limitation is that LP cannot predict unseen label-sets. LP can also be termed as Label-Combination Method (LC) [2].

Instance	Attribute	Label sets
1	A1	{B1,B2}
2	A2	{B1,B2,B3}
3	A3	{B4}
4	A4	{B1,B2,B5}
5	A5	{B2,B4}

c	P(c/x)	B1	B2	B3	B4	B5
B1,2	0.4	1	1	0	0	0
B1,2,3	0.2	1	1	1	0	0
B4	0.1	0	0	0	1	0
B1,2,5	0.0	1	1	0	0	1
B3,4	0.3	0	1	0	1	0
	∑P(c/x)Bj	0.6	0.9	0.2	0.4	0.0

c	P(c/x)	λ1	λ2	λ3	λ4
1001	0.1	1	0	0	1
0011	0.3	0	0	1	1
1000	0.6	1	0	0	0
0111	0.2	0	1	1	1
	∑P(c/x)λj	0.7	0.2	0.5	0.6

Table- 4: Transformation using Label Power-Set and Example of obtaining Ranking from LP

2.1.3.2 Pruned Set (PS):

It is also called as Pruned Problem Transformation (PPT). This method follows the same paradigm of LP. The pruned problem transformation method extends LP to remove its limitations by pruning away the label-sets that are occurring less time than a small user-defined threshold [1].

In.	Labels
1	{B1,B2}
2	{B3,B2,B4}
3	{B1}
4	{B3,B2}
5	{B2}
6	{B1}
7	{B3,B2}
8	{B4}
9	{B4}
10	{B2}

In.	Labels
1	{B1,B2}
2	{B3,B2,B4}
3	{B1}
4	{B3,B2}
5	{B2}
6	{B1}
7	{B3,B2}
8	{B4}
9	{B4}
10	{B2}

In.	Labels
3	{B1}
4	{B3,B2}
5	{B2}
6	{B1}
7	{B3,B2}
8	{B4}
9	{B4}
10	{B2}

In.	Labels
1	{B1,B2}
2	{B3,B2,B4}

It removes the infrequent label-sets. Then, it replaces these label-sets by the existing disjoint label-sets that are occurring more times than the threshold [2].

This method is done in two steps as stated earlier,

1. Pruning
2. Sub Sampling.

In.	Labels
1	{B1}
2	{B2}
3	{B3,B2}
4	{B4}

In.	Labels
1	{B1}
2	{B2}
3	{B3,B2}
4	{B4}
5	{B1}
6	{B3,B2}
7	{B2}
8	{B1}
9	{B3,B2}
10	{B4}
11	{B4}
12	{B2}

Table- 5: Transformation using Pruned Problem Transformation (PPT)

2.1.3.3 Classifier Chains (CC):

It involves Q-binary classifiers as in a BR method. It resolves the BR limitations, by taking into account the label correlation task. The classifiers are linked along a chain where each classifier deals with the BR problem associated with the label. Each link in the chain is expressed with the 0/1 label associations of all previous links.

2.1.4 Pair-wise methods

2.1.4.1 Ranking via Pair-wise Comparison (RPC):

Ranking via pair-wise comparison transforms the multi-label datasets into q(q-1)/2 binary label datasets, one for each pair of labels (Bi, Bj), 1≤i<j<q. Each dataset contains those instances of original dataset that are annotated by at least one of the corresponding labels, but not by both (Table-6).

In.	Label
5	B-1,2

In.	Label
1	B1,-3
4	B1,-3

In.	Label
1	B1,-4
2	B1,-4
3	B-1,4
4	B1,-4
5	B-1,4

In.	Label
1	B1,-5
2	B1,-5

In.	Label
1	B2,-3
4	B2,-3
5	B2,-3

In.	Label
1	B2,-4
2	B2,-4
3	B-2,4
4	B2,-4

In.	Label
1	B2,-5
2	B2,-5
5	B2,-5

In.	Label
2	B3,-4
3	B-3,4
5	B-3,4

In.	Labels
2	B3,-5
4	B-3,5

In.	Labels
2	B3,-5
4	B-3,5

In.	Labels
3	B4,-5
4	B-4,5
5	B4,-5

new instance x'

Label	Votes
λ1	2
λ2	1
λ3	3
λ4	0

Ranking: λ3 λ1 λ2 λ4

Table- 6: Transformation using RPC, New instance is created then; Ranking is obtained by counting votes

A binary classifier is then trained on each dataset. For a new instance, all binary classifiers are invoked and then ranking is obtained by counting votes received by each label.

2.1.4.2 Calibrated Label Ranking (CLR):

CLR is the extension of Ranking by Pair-wise Comparison (RPC). This method introduces an additional virtual label (calibrated label), which acts as a split point between relevant and irrelevant labels. Thus CLR solves complete MLR (Multi-label Ranking) task. Each example that belongs to particular label is considered positive for that example does not belong to particular label is considered negative for that particular label and positive for virtual label. Thus CLR corresponds to the model of Binary Relevance. When CLR applied to the dataset of Table-1, it constructs both datasets of Table-6, 6.1 and Table-3.

2.1.5 Ensemble Methods:

The ensemble methods are developed on top of the common Problem transformation and Algorithm adaptation methods, they construct a set of classifiers and then classify new data points by taking a weighted vote of their predictions. They are used for further augment predictive performance and high accuracy results. They aim to aggregate the predictions of several base estimators built with a given learning algorithm [10].

2.1.5.1 Random k-label sets (RAkEL):

It constructs an ensemble of LP classifiers. It breaks the large label-sets into m models or subsets, which are associated with random and small sized k-label-sets here take k=3 (parameter that specifies the size of the label-sets), threshold=0.50. It takes label correlation into account and also avoids LP's problems within the large number of distinct label-sets [1]. Given a new instance, it query models and average their decisions per label. And also uses thresholding to obtain final model. Thus, this method provides more balanced training sets and can also predict unseen label-sets [2].

model	3-labelsets	predictions					
		λ1	λ2	λ3	λ4	λ5	λ6
h1	{λ1, λ2, λ6}	1	0	-	-	-	1
h2	{λ2, λ3, λ4}	-	1	1	0	-	-
h3	{λ3, λ5, λ6}	-	-	0	-	0	1
h4	{λ2, λ4, λ5}	-	0	-	0	0	-
h5	{λ1, λ4, λ5}	1	-	-	0	1	-
h6	{λ1, λ2, λ6}	1	0	1	-	-	-
h7	{λ1, λ2, λ6}	0	-	-	1	-	0
average votes		3/4	1/4	2/3	1/4	1/3	2/3
final prediction		1	0	1	0	0	1

Table-7: Transformation using RAkEL

2.1.5.2 Ensemble Pruned Set (EPS):

It will enhance the prediction of Existing PS method which was discussed earlier. This method uses a Pruned Sets method in an ensemble framework, and uses a voting scheme to produce the prediction confidences. It provides a powerful and general framework. EPS's training algorithm can be used with any multi-label-capable classifier [3]. It combines the PS method in an ensemble scheme. PS is specifically suited to an ensemble due to its fast build times. Also, it counters any over fitting effects of the pruning process and allows the creation of new label sets at classification time. Applying the ensembles on PS method increases the predictive performance of the algorithm [1, 10].

2.1.5.3 Ensembles of Classifier Chains (ECC):

It uses the CC method as a base classifier. It trains m models of CC classifiers C1, C2,…,Cm. Each Ck model is trained with a random chain ordering of labels B and a random subset of the datasets shown in (table- 1). Each model is likely to be unique and able to predict different label-sets. After that, these predictions are summed by label so that each label receives a number of votes. A threshold value is applied to select the most relevant labels, which form the final predicted multi-label set.

2.2 Algorithm Adaptation Methods (AAM):

It extends and adapts the existing specific learning algorithm to directly handle the multi-label problem. It is an algorithm dependent method. Many methods belong to this category.

2.2.1 Multi-Label Decision-Tree (ML-DT):

This method is belonging from Decision Tree based method category. It is an adaptation of the well-known C4.5 Algorithm to handle multi-label data. The process is accomplished by allowing multiple labels in the leaves of the tree; the output of C4.5 is a decision tree which is constructed from top-down manner. The formula for calculating the entropy is modified for solving multi-label problems. The modified entropy sums all the entropies for each individual label.

The Key property of ML-DT is its computational efficiency:

$$\text{Entropy (D)} = \sum_{i=1}^{q} -p_j \, log_2 \, p_j - (1 - p_j) \, log_2 \, (1 - p_j)$$

Where D is the set of instances in the dataset and p_j is the fraction of instances in D that belongs to the label j.

2.2.2 Multi-Label k Nearest Neighbors(ML-KNN):

ML-KNN is a lazy learning approach and it is one of algorithm adaptation based method. It extends the traditional K Nearest Neighbor (KNN) method for multi-label classification. In addition to KNN, ML-KNN uses Bayesian reasoning approach. Firstly, k- nearest neighbors of a given test instance will be selected from the training set, then each label occurrence in the neighbor training set will be counted. Finally statistical analysis mechanism called Maximum Posteriori Principle (MAP) will be applied. Prior and posterior probability of a label will be estimated from the training set. This probability estimation will be used for the final prediction of the label set of a test instance [4]. In this Euclidean metric is used to measure distances between instances [3].

2.2.3 Support Vector Machine with Heterogeneous Feature Kernel (SVM-HF):

This method exploits relationship among the classes. It enhances the basic purely text based SVM learner by augmenting the feature set with |C| extra features, one for each label in the dataset. The cyclic dependency between features and labels is resolved iteratively. Cosine similarity measure is used to calculate the similarity between two documents [3].

2.2.4 Ranking Support Vector Machine (Rank-SVM):

It is a ranking approach for multi-label learning that is based on SVM. It is used to minimize the Ranking-loss. The main function they use is the average fraction of incorrectly ordered pairs of labels [1].

2.2.5 Multi-label Tree based Boosting methods:

AdaBoost.MH & AdaBoost.MR:

These two methods are based on Tree Based Boosting. AdaBoost.MH and AdaBoost.MR are two extension of AdaBoost for multi-label data. In this purpose of using concept of boosting is to find a highly accurate classification rule by combining many weak or base hypotheses, each of which may be moderately accurate. In this approach, the goal of the learning algorithm is to predict all of the correct labels. Thus, the learned classifier is evaluated in terms of its ability to predict approximation of the set of labels associated with the given document. AdaBoost.MH is extended in to produce better human related classification rule. It is designed to minimize Hamming loss and Adaboost.MR is designed to find hypothesis which places the correct labels at the top of ranking [1].

2.2.6 Neural Network based:

BP-MLL is an extension of the popular back-propagation algorithm for multi-label learning. It is neural network based method. The main modification is the introduction of a new error function that takes multiple labels into account.
Given multi-label training set,

$$S = \{(xi, Yi) \mid 1 \leq i \leq m\}$$

The global training error E on S is defined as:|

$$E = \sum_{i=1}^{m} Ei = \sum_{i=1}^{m} \frac{1}{|Y_i||\overline{Y_i}|} \sum_{(k,l)\epsilon Y_i \times \overline{Y_i}} \exp(-(c_k^i - c_l^i))$$

Where, Ei is the error of the network on (xi, Yi) and cij is the actual network output on xi on the jth label.
The differentiation is the aggregation of the label sets of these examples.

2.2.7 Multi-label Naive Bayesian (ML-NB):

It extends the Naive Bayesian algorithm to adapt it with the multi-label data. It deals with the probabilistic generation among the labels. It uses MAP to specify the more probable labels and assign them to the new given instance.

3. Comparison between various Problem Transformation Methods (PT Methods):

Method	Merits	Demerits
Copy, Copy-weight	Simple, No information loss	Increased instances
Max, Min, Random	Simple, Decreased instances	Information loss
Ignore	Very simple	Major information loss
BR	conceptually simple and relatively fast,	It can also be affected by class-imbalance Label independency
LP	can take into account label correlations	Can also be computationally complex Leads to over fitting of the training data
PS	Run much faster. It can form new label sets at classification time, and in this way handles irregular labeling. Consider label dependency	A disadvantage, however, is the reliance on prediction confidence distributions of the base classifier. High complexity
CC	Provides increased predictive performance It is scalable, can work with any type of base classifier.	Cannot utilize available unlabeled data for classification
RPC	Flexible. It also covers absolute preferences while giving ranking.	Takes more time and storage capacity
CLR	It deals with multi-label classification as well as ranking. It can be generalized	Computationally expensive. Cannot utilize available unlabeled data for classification
RAkEL	Computationally simpler. More predictive capability. Can take into account label correlations	More time complexity. Cannot utilize available unlabeled data for classification.
EPS	Provides increased predictive performance. It allows parallelism. These are scalable	Not able to utilize available unlabeled data for classification
ECC	Able to predict different label-sets. It is scalable	Not utilize available unlabeled data for

3.1 Comparison between various Algorithm Adaptation Methods:

Method	Merits	Demerits
ML-DT	▪ It allows choosing such attributes which splits the data in the most informative way. ▪ It offers easy learn ability	▪ It does not take into account the correlation among the classes. *It cannot able to utilize the unlabeled data for classification
AdaBoost.MH& AdaBoost.MR	▪ Improved accuracy and minimization of Hamming loss error.	▪ Attempts for generalization results into decrease in performance. ▪ Cannot utilize unlabeled data for classification.
ML-kNN	▪ Improved performance as compared to other algorithms in terms of hamming loss, ranking loss and coverage.	▪ Cannot utilize unlabeled data for classification
(BP-MLL)	▪ Outperforms other counterpart in term of ranking loss. ▪ Gives better generalization capability to learning system. ▪ Time cost of making predictions based on the trained model is trivial.	▪ Computational complexity in training phase is high because of use of neural networks. ▪ Cannot able to utilize unlabeled data for classification.
SVM-HF	▪ Take into account correlation among classes. ▪ Significant improvement in accuracy for multi-label data.	▪ Accuracy reduces with consideration of unlabeled data.

4. CONCLUSIONS:

Classification task of the data which has one class-label is known as single-label Classification. Whereas the classification task in which the data has two or more class labels are called multi-label classification. In multi-label classification each class-label has only binary values. It is a special case of multi-target classification. In multi-target classification each instance has more than one class-labels and each class-label has multiple values. Multi-label classification has two main types of methods: Problem Transformation Methods, Algorithm Adaptation Method. In this, Deep/detailed study is done on various Multi-label classification Methods used for Text Data & image Data Classification using various tools such as WEKA, MEKA, MULAN. There is also more general-purpose software that handles multi-label data as part of their functionality. LibSVM is a library for support vector machines that can learn from multi-label data using the binary relevance transformation. Clus9 is a predictive clustering system that is based on decision tree learning. Its capabilities include (hierarchical) multi-label classification. The Boos Texter system, implements the boosting-based approaches. There also exist Mat lab implementations for ML*k*NN7 and BPMLL.

REFERENCES

[1] Asma Aldrees and Azeddine Chikh and Jawad Berri, “Comparative Evaluation of Four Multi-labels Classification Algorithms in Classifying Learning Objects”, David C Wyld et al. (Eds): CCSEA, CLOUD, DKMP, SEA, SIPRO – 2016.

[2] Hiteshri Modi ME-CSE Student, Dept. of Computer Engineering, Mahesh Panchal, Head & Associate Prof., Kalol Institute of Technology and Research Center Kalol, Gujarat, India, “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.

[3] 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 Vol. 1, Issue 4, pp.1952-1961.

[4] Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas,” Mining Multi-label Data”, Dept. of Informatics, Aristotle University of Thessaloniki, 54124 Greece, Apr -2010, survey paper.

[5] Mr. Sudhir M. Gorade, Prof. Ankit Deo, Prof. Preetesh Purohit 3, Dept.Of Computer Science and Engineering, SVCE Indore, M.P, India, ”A Study of Some Data Mining Classification Techniques”, International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 04 | Apr -2017.

[6] Grigorios Tsoumakas, Ioannis Katakis, Dept. of Informatics, Aristotle University of Thessaloniki, 54124, Greece,” Multi-Label Classification: An Overview”, survey paper, April 2014.

[7] Jiang Wang, Yi Yang, Junhua Mao, Zhiheng Huang Chang, Huang Wei Xu,” CNN-RNN: A Unified Framework for Multi-label Image Classification”, IJCA journal, May 2016.

[8] X. Xue, W. Zhang, J. Zhang, B. Wu, J. Fan, and Y. Lu. “Correlative multi-label multi-instance image annotation”, in Computer Vision (ICCV), 2011 IEEE International Conference On, pages 651–658, 2011.

[9] Zhaang, M., & Zhou, Z, “Multi-Label Neural Networks with Applications to Functional Genomics and Text Categorization”, IEEE Transactions on Knowledge and Data Engineering, 18(10), 1338-1351, 2006.

[10] Read, J., Pfahringer, B., & Holmes, G. (2008), "Multi-label Classification Using Ensembles of Pruned-Sets" Proc 8th IEEE International Conference on Data Mining, Pisa, Italy, 995-1000.

BIOGRAPHIES

Jadon Mayurisingh N.
M.Tech Student at U.V. Patel College of Engineering
Ganpat University, Kherva, Mehsana, Gujarat(India)

Prof. Hiteshri N. Modi
Assistant Professor,
Dept. of Information Technology,
U.V. Patel College of Engineering
Ganpat University, Kherva,
Mehsana, Gujarat (India)