

A Multi-Label Classification Approach Based on Correlations Among Labels

Raed Alazaidah

Computer Science Department
Philadelphia University
Amman- Jordan

Fadi Thabtah

E-Business Department
Canadian University of Dubai
Dubai-UAE

Qasem Al-Radaideh

CIS Department
Yarmouk University
Irbid- Jordan

Abstract—Multi label classification is concerned with learning from a set of instances that are associated with a set of labels, that is, an instance could be associated with multiple labels at the same time. This task occurs frequently in application areas like text categorization, multimedia classification, bioinformatics, protein function classification and semantic scene classification. Current multi-label classification methods could be divided into two categories. The first is called problem transformation methods, which transform multi-label classification problem into single label classification problem, and then apply any single label classifier to solve the problem. The second category is called algorithm adaptation methods, which adapt an existing single label classification algorithm to handle multi-label data. In this paper, we propose a multi-label classification approach based on correlations among labels that use both problem transformation methods and algorithm adaptation methods. The approach begins with transforming multi-label dataset into a single label dataset using least frequent label criteria, and then applies the PART algorithm on the transformed dataset. The output of the approach is multi-labels rules. The approach also tries to get benefit from positive correlations among labels using predictive Apriori algorithm. The proposed approach has been evaluated using two multi-label datasets named (Emotions and Yeast) and three evaluation measures (Accuracy, Hamming Loss, and Harmonic Mean). The experiments showed that the proposed approach has a fair accuracy in comparison to other related methods.

Keywords—Classification; Data mining; Multi-label Classification

I. INTRODUCTION

Data classification is a form of data analysis that can be used to extract models describing important data classes. The classification task concentrates on predicting the value of the decision class for an object among a predefined set of classes given the values of some given attributes for the object. In general, data classification is a two-step process. In the first step (learning), a model that describes a predetermined set of classes or concepts is built by analyzing a set of training database objects. Each object is assumed to belong to a predefined class. In the second step, the model is tested using a different data set.

Classification problems can be divided into three main categories: Binary classification, Multi-Class classification and Multi-Label classification. In binary classification, there are only two possible values for the class label (X, Y). However, most real world application domains contain several

classes and therefore several multi-class approaches have been proposed.

Formally, the traditional classification problem can be defined as follows: "let D denotes the domain of possible training instances, and Y be a list of class labels, let $H: D \rightarrow Y$ denotes the set of classifiers. Each instance $d \in D$ is assigned a single class label y that belongs to Y . The goal is to find a classifier $h \in H$ that maximize the probability that $h(d) = y$, for each test case (d, y) . In multi-label problem, however, each instance $d \in D$ can be assigned multiple labels y_1, y_2, \dots, y_k for $y_i \subseteq Y$, and is represented as a pair $(d, (y_1, y_2, \dots, y_k))$ where (y_1, y_2, \dots, y_k) is a list of ranked class labels from Y associated with the instance d in the training data [1].

Multi label classification is concerned with learning from set of instances that are associated with a set of labels, that is, an instance could be associated with multiple labels at the same time. This task occurs frequently in application areas like text categorization, multimedia classification, bioinformatics, protein function classification and semantic scene classification. An Example of a multi label dataset is presented in Table1. In practice, most of the current classification approaches do not consider the generation of rules with multiple labels from multi-class or multiple label data [2].

TABLE I. MULTI-LABEL DATA

A1	A2	A3	A4	Class
5	A	2	R	X, Y
3	B	0	A	X, W, Z
3	B	2	A	Z
3	B	6	T	Y, Z

This paper proposes a guided multi-label classification approach based on correlations among labels in class label attribute and then applying a classical classification algorithm to learn rules from the training dataset. Most of multi-label classifications methods, both problem transformation methods and algorithm adaptation methods depend, for its classification task, on a function that maps between the attributes and the labels in the training data. The proposed approach introduces a new approach to solve the problem of multi-label classification. This approach is based on correlations among labels learned by predictive classification.

II. RELATED WORK: MULTI-LABEL CLASSIFICATION METHODS

Existing methods for handling multi-label classification can be grouped into two main groups. The first group, which is an algorithm independent, is called problem transformation methods, while the second group is an algorithm dependent, and is called algorithm adaptation methods. The first group transforms multi-label classification problem into one or more single classification problem, while the second group extends a specific learning algorithm, in order to handle multi-label data directly [3].

A. Problem Transformation Methods

Several problem transformation methods exist in the literature that is used to convert multi-label classification problem into one or more single label classification problem. To exemplify these methods, we will use the dataset of Table2 which consists of four examples that belong to the following class set {Reading, Swimming, Painting, TV Watching}

TABLE II. MULTI-LABEL DATA SET

Inst #	Reading	Swimming	Painting	TV Watching
1		X	X	
2	X		X	
3		X		X
4		X		

The first problem transformation method discards every multi-label instance from the data set. Therefore, in the previous example, instances 1, 2, 3 will be discarded. Another problem transformation method selects one of the multiple-labels of each multi-label instance either randomly or subjectively. The transformed version of the previous example instances is presented in Table3.

TABLE III. MULTI-LABEL DATA SET

Inst #	Reading	Swimming	Painting	TV Watching
1		X		
2			X	
3				X
4	X			

The copy transformation method transforms every multi-label instance to a single label instance by replacing the multi-label instance (x_i, y_i) with $|y_i|$ instances. Several transformation methods could be then chosen such as: (1) copy-weight which associates a weight of $(1/|y_i|)$ to each of the transformed examples, (2) select-max (most frequent), (3) select-min (least frequent), (4) select-random, and (5) the ignore transformation option.

One of the most popular transformation methods, that learn single binary classifier for every label in the label set, is called Binary Relevance (BR) [3]. This method transforms the original data set into $|L|$ data sets, which contain all the instances from the original data set. It then gives a positive sign for a label, if it exists in the data set and a negative sign otherwise. To classify new instance, the BR method returns the union of all labels that are predicted by the $|L|$ classifiers.

Although Binary Relevance is a simple transformation method, it is based on implicit assumption of labels independence which might be completely incorrect in the data.

Another method called the Label Power Set (LP) is a straight forward method that works as follows: it considers each unique set of labels that exists in the data set as a new single label in single – label classification task as shown in Table4.

TABLE IV. MULTI-LABEL DATA SET

Inst #	Label
1	{Swimming, Painting}
2	{Reading, Painting}
3	{Swimming, TV Watching}
4	{Swimming}

To predict the class label of a new instance, the LP method returns the most probable class which actually could be a set of labels in the original data set [4]. The Computational complexity of LP is upper-bounded by $(\min(|L|, 2^k))$ where k : is the total number of classes in the data set before transmission, and usually it is much less than 2^k . LP has an advantage of taking labels correlations into account, on the contrary of BR, but it has a disadvantage when a large number of classes in the original data set associated with small number of instances, which may cause an imbalance problem for learning.

The previous mentioned problem of LP was addressed by the pruned problem transformation methods [5] which used a user- defined threshold to prune some label sets that occur less than this threshold. The pruned set could be replaced by disjoint subsets of these labels that are more frequent in the data set.

The RAKEL (Random K label sets) method is an effective transformation method that breaks the initial set of labels into a number of small random subsets called label-sets and then employs the LP method to train a corresponding classifier, where k is a parameter that determines the size of the subsets [4]. RAKEL offers advantages over LP for the two reasons: (a) The resulting single label classification tasks are computationally simpler, and (b) The resulting single label classification tasks are characterized by much more balance distribution of class values. In RAKEL, the parameter K which is used to determine the size of the subsets and specified by the user should be small to avoid the problems associated with the LP method.

The Ranking by Pair wise Comparison (RPC) approach by [6] transforms the multi-label classification problem into a single label classification problem through performing pair wise comparisons of labels. RPC learns $(|L| * (|L| - 1)) / 2$ binary classifiers, one model for each different pair of labels. For predicting new instance, all models are invoked and ranking is obtained through counting the votes received by each label. An extension of RPC called Calibrated Label Ranking (CLR) [7] which introduces a virtual label (often called calibration label, L_0) that aims to separate relevant labels from irrelevant ones.

Another problem transformation method called the Classifier Chains (CC) method tries to enhance the BR method through taking label correlations into account [8]. CC builds $|L|$ binary classifier for each label as in BR. Then Classifiers are linked along a chain where each classifier deals with the binary relevance problem associated with label $l_j \in L$. The feature space of each line in the chain is extended with 0/1 label association of all previous links. The CC method counteracts the disadvantages of the binary method while maintaining acceptable computational complexity.

The Ensemble of Classifier Chains (ECC) method is an enhancement version of CC which in turn is an enhancement of BR. ECC trains m Classifier Chains C_1, C_2, \dots, C_m , Where each C_k is trained with a random chain ordering of L and a random subset of D . Each C_k model is likely to be unique and able to give different multi label predictions. These predictions are then summed by label so that each label receives a number of votes. A threshold is used to select the most popular labels which form the final prediction of multi label set [8].

Another problem transformation method called Pruned Sets (PS) is an enhancement of Label Power-set (LP) which treats every unique subset of labels as a single label, and suffers from label imbalance specially, when number of training examples is small and number of labels is too large [5]. PS try to solve this problem by focusing only on the most important correlations, which reduce complexity and improve accuracy [8].

B. Algorithm Adaptation methods

Algorithm Adaptation methods extend a specific single label learning algorithm in order to handle multi-label data directly. In this section, we introduce a brief plethora of algorithm adaptation methods grouped by the learning concept that they extend.

Reference [9] developed a re-sampling technique and modified the C4.5 algorithm to deal with a gene hierarchy multi-label classification problem.

Reference [1] proposed a Multi-class, Multi-label Associative Classification algorithm (MMAC) which is an associative rule learning based covering algorithm that recursively learns a new rule and each time removes the examples associated with that rule. Labels for the test instances are ranked according to confidence, support, and rule's cardinality (number of conditions in the left hand side (LHS) of the rule).

Reference [4] proposed the AdaBoost.MH and AdaBoost.MR as two extensions of AdaBoost for multi-label data, where AdaBoost.MH aims to reduce Hamming loss and AdaBoost.MR aims to increase accuracy.

Reference [10] proposed a K - nearest Neighbors (KNN) lazy learning based method for multi label data. In general, the KNN based methods share the same first step with KNN (retrieving the K nearest example) and differ from each others on the aggregation of the label sets of these examples.

III. THE PROPOSED APPROACH FOR MULTI LABEL CLASSIFICATION

The general structure of the proposed approach consists of three phases: (a) Transforming multi-label dataset into single label dataset and discovering correlations among labels. (b) Applying a rule-based classification algorithm on the transformed dataset. (c) Generating the multi-label rules based on the output of the rule-based classifier and the correlations among labels. Fig.1 shows the general structure of the proposed approach and the main steps of the approach are described in Fig. 2.

As shown in Fig. 1, the input of the algorithm is a multi-label dataset, and then two operations are performed on the multi-label dataset: the first operation is transforming multi-label dataset into a single label dataset; in this step there are several methods to choose from such as: selecting the most frequent label, selecting the least frequent label or select any label randomly.

For the proposed approach we choose to select the least frequent label as transformation criteria. The second operation is to find all positive association among labels using the predictive Apriori method [11]. This operation tries to associate each label with labels from the label set; if that is possible. The output after performing these two operations will be:

1) A single label dataset which has been extracted or transformed from multi-label dataset using the least frequent label criteria.

2) Rules between labels with different rule's cardinality, starting from cardinality 1 up to rule's cardinality which is equal to the dataset cardinality -1, (i.e, Association rule's cardinality = Label Cardinality - 1).

In the next step, a single rule-based classifier is applied on the transformed dataset. Several rule-based classifiers could be used in this stage such as RIPPER, IREP, PART or Prism. The output of any single rule based classifier will be set of "IF-THEN" rules with one consequent on the right-hand-side of the rule like the following rule:

IF (con₁ and con₂ and ... con_n) THEN Label.

Using both, the output of the single rule based classifier and the rules based on the correlations among labels previously discovered, we will be able to build a multi-label rules classifier in the form:

*IF (con₁ and con₂ and ... con_n) THEN
{Label₁, Label₂, ..., Label_n}.*

The Learning Phase

The learning Phase in the proposed approach consists of two different tasks. The first task is an unsupervised learning task, which aims to discover the correlations among labels using Predictive Apriori. While the second task is a supervised learning task that aims to predict the class label of unseen instance as accurate as possible using a rule based classifier.

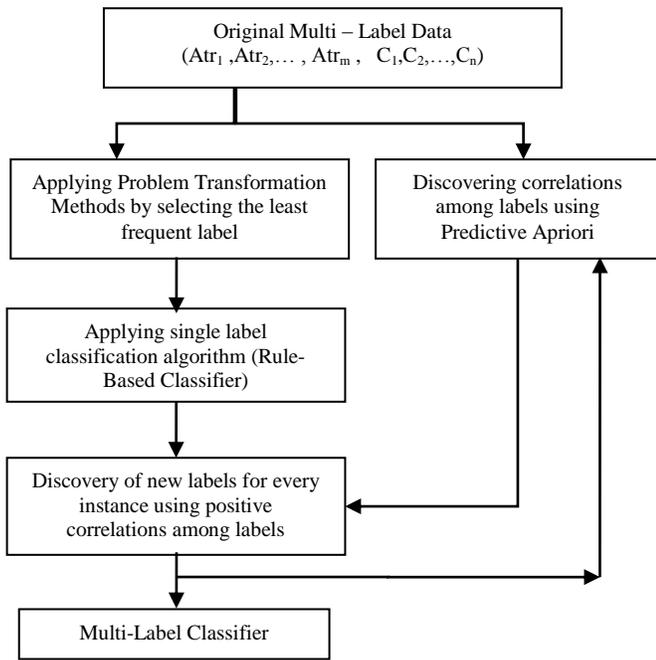


Fig. 1. The General structure of the proposed approach

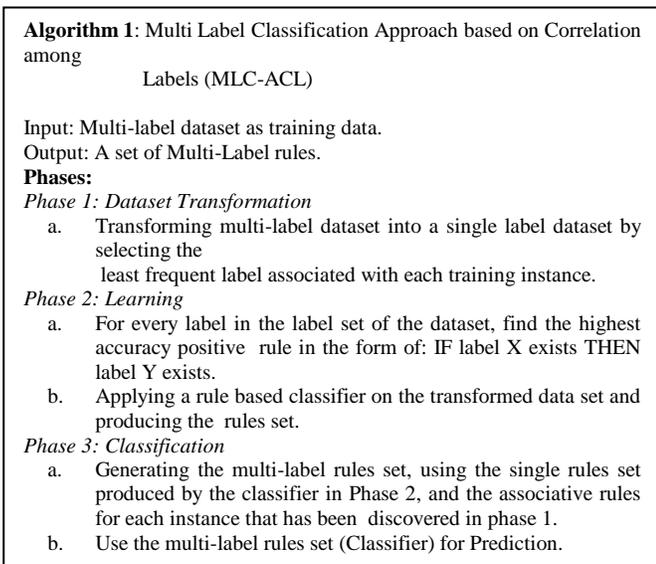


Fig. 2. The main phases of the Proposed Algorithm

A. Discovering of Positive Correlations among Labels.

Suppose we have the itemsets (Labels) C1, C2, and C3. We are interested in having association rules with good confidence between every possible Pair-wise of the three previous labels. For the first two labels C1, C2 we may have the following rules for example:

- 1- If C2=1 Then C1=0
- 2- If C1=1 Then C2=1

In the proposed approach, we are interested in rules like the second rule, we are looking for a rule in a form of (IF Label x exists THEN label y exists). For each label (x) in the dataset we want to find another label (y) that has a positive

correlation with it, i.e. label (x). In case we have more than one label positively associated with the label in the antecedent, we select the rule with the highest confidence or accuracy. For example suppose that we have the following association between C1, C2 and C3:

- 1- If C1 =1 Then C2=1 (Accuracy = 0.80)
- 2- If C1=1 Then C3=1 (Accuracy = 0. 71)

In the previous case, we choose the rule with the highest accuracy, so rule one will be selected, and rule two is ignored. In fact ignoring such a rule with a meaningful confidence such 0.71 may cause too much information loss but let us stuck on the choice of selecting the best rule, and leave ignoring other rules with meaningful confidence to be discussed later in the future work section.

After having all positive associations of length "1" between labels in the dataset , we move forward to find all positive associations of cardinality "2" as the following rule (If C1=1 and C2=1 Then C3=1) and so forth.

For the proposed approach, we will choose the rule with the highest accuracy without any pre specified condition about the value of accuracy, such as the accuracy should be greater than or equals to a predefined user threshold. For example, suppose we have the following rules:

- 1- If C1=1 Then C2=1 (Accuracy = 0.27)
- 2- If C1=1 Then C3=1 (Accuracy = 0.19)

B. Applying Rule-Based Classifier

After having the transformed data set, and finding the highest positive association rules among labels, we are ready to apply any single rule-based classification algorithm to the transformed data, and we choose PART classifier.

PART is a rule-based classification algorithm that combines between two approaches. The first one is creating rules using decision tree, and the second one is separate and conquer learning method [12]. The algorithm produces accurate rules in the same size as those generated by decision tree C4.5 algorithm. PART algorithm has been chosen for being accurate, efficient and fast.

The Prediction Phase

Finally, the classifier consists of a Set of multi-label rules that have been learned from both correlations among labels and rule-based classifier. This classifier will be used in the prediction step to predict the class label / labels of a new instant.

IV. AN ILLUSTRATIVE EXAMPLE FOR THE PROPOSED APPROACH

For more clarification, this section presents a complete step by step example for the proposed approach using the "Emotions" dataset which has been downloaded from the following address (<http://mulan.sourceforge.net/datasets.html>). The characteristics of the dataset are presented in Table5. Table6 shows the frequency of the six labels in the "emotions" dataset. It is clear that the Most Frequent Label (MFL) is "Relaxing" and the Least Frequent Label (LFL) is "Quite-still".

TABLE V. MULTI-LABEL DATASET INFORMATION

Dataset name	Domain	# of Instances	# of Numeric Attributes
Emotions	Music	593	72

TABLE VI. "EMOTIONS" DATASET LABELS STATISTICS

Label	Amazed	Happy	Relaxing	Quite-still	Sad	Angry
Frequency	173	166	264	148	168	189

A. Approach Phases

Here we describe the main phases of the approach as the following:

Phase1 (a): Transform the dataset ("Emotions") into a single label dataset using least frequent label. Sample of the transformed dataset is presented in Table7. As we can see in Table7, the first example is associated with three labels at the same time (*Relaxing, Quite-Still, Sad*), and since "*Quite-Still*" has frequent 148 which is less than the frequent of "*Relaxing*" (264) and "*Sad*" (168), it will be transformed to the single label "*Quite-Still*". The second example is associated with two labels: "*Amazed*" with frequent equals to 173 and "*Angry*" with frequent 189, so it was transformed to the least frequent label which is "*Amazed*", and so on for the rest of examples.

TABLE VII. TRANSFORMING "EMOTIONS" DATASET INTO SINGLE LABEL DATASET

In s #	Amaze d	Happ y	Relaxin g	Quite -still	Sa d	Angr y	Class
1	0	0	1	1	1	0	Quite-still
2	1	0	0	0	0	1	Amazed
3	0	0	0	0	1	0	Sad
4	0	1	1	0	0	0	Happy
5	0	0	0	0	1	0	Sad
6	0	0	1	0	1	0	Sad

Phase1(b): The second step is to find positive correlations among labels using predictive Apriori. Best correlations are chosen without determining any threshold value in this stage, and since "Emotions" dataset is of cardinality "2"; association rules will be with "1" condition only in the antecedent. Table11 shows the complete positive correlations among labels in "Emotions" dataset.

As notices in Table8, Rule #5 has the lowest accuracy, in this case we will stuck in the choice of having the highest positive association among labels, and since no other rule could be found to be associated with the label "angry", and has accuracy greater than this rule, this rule is chosen.

TABLE VIII. POSITIVE CORRELATIONS AMONG LABELS IN "EMOTIONS" DATASET

Rule #	Rule	Accuracy
1	IF amazed THEN angry	0.53
2	IF happy THEN relaxing	0.44
3	IF Quite-still THEN sad	0.71
4	IF Sad THEN Relaxing	0.57
5	IF angry THEN Relaxing	0.03
6	IF Relaxing THEN Relaxing	1.00

Phase (2): The third step in the proposed approach is to apply a rule based classification algorithm on the transformed dataset. Table9 shows some of the learning rules discovered after applying the PART classifier.

TABLE IX. LEARNING RULES DISCOVERED AFTER APPLYING THE PART CLASSIFIER

Rule #	Rule Condition	Consequence
1	IF AQ > 0.217678 AND B <= 0.090652 AND V > 0.580398 AND AZ > 3.787686 AND AX > 0.060033 AND BD <= 0.173826	Sad
2	IF AQ <= 0.215792 AND BJ <= 0.108461 AND J <= 1.021892 AND BO <= 0.066288	Angry
3	IF AS > 0.206592 AND AI > 0.010202 AND D > -76.700621	Amazed
4	IF AS > 0.206592 AND AI > 0.010202 AND B <= 0.191563	Quit-Still
5	IF AS > 0.208738 AND B <= 0.119991 AND AP > 0.213677 AND BN <= 102 AND D > -75.367339	Relaxing
6	IF G > 2.024609 AND E > 3.112653	Happy

Phase (3): The last step is to build multi-label classifier based on correlations among labels and rules discovered from applying a rule based algorithm on the transformed dataset. Table10 summarizes some of the multi-label rules discovered from "Emotions" dataset.

TABLE X. MULTI-LABEL RULES DISCOVERED FROM "EMOTIONS" DATASET

Rule #	Multi-Label Rules	Consequence
1	IF AQ > 0.217678 AND B <= 0.090652 AND V > 0.580398 AND AZ > 3.787686 AND AX > 0.060033 AND BD <= 0.173826	{Sad, Relaxing}
2	IF AQ <= 0.215792 AND BJ <= 0.108461 AND J <= 1.021892 AND BO <= 0.066288	{Angry, Relaxing}
3	IF AS > 0.206592 AND AI > 0.010202 AND D > -76.700621	{Amazed, Angry}
4	IF AS > 0.206592 AND AI > 0.010202 AND B <= 0.191563	{Quite-Still, Sad}
5	IF AS > 0.208738 AND B <= 0.119991 AND AP > 0.213677 AND BN <= 102 AND D > -75.367339	{Relaxing}
6	IF G > 2.024609 AND E > 3.112653	{Happy, Relaxing}

To illustrate how this step is performed, let us give a sample rule from the rules set that are obtained after applying PART algorithm on the transformed dataset. The sample rule is:

IF AQ > 0.217678 AND B <= 0.090652 AND V > 0.580398 AND AZ > 3.787686 AND AX > 0.060033 AND BD <= 0.173826 THEN Sad.

Using Association rules among labels that have been discovered earlier, and since there is a rule indicates that (*IF Sad THEN Relaxing*), the rule is rebuilt from the rule based classifier as following:

IF AQ > 0.217678 AND B <= 0.090652 AND V > 0.580398 AND AZ > 3.787686 AND

$AX > 0.060033$ AND $BD \leq 0.173826$ THEN {Sad, Relaxing}

We repeat the previous process for all rules extracted from the rule based classifier and using the association rules discovered in the first step. The outcome will be the complete set of multi-label rules, which will be used to classify the test instances.

V. EXPERIMENTS AND RESULTS

In this paper, we used two different application domains data sets which they are: Biological, and Musical. For each application domain, one multi-label dataset has been used, as shown in Table11. The datasets are available at (<http://mulan.sourceforge.net/datasets.html>). The first dataset is called "Emotions" and it is concerned about songs according to the emotions they evoke. This data set contains six labels, with label cardinality (LC) equal to 1.869and label density (LD) equal to 0.311. There are 27 distinct label-sets (DLS) in a total number of 593 examples in this dataset. As mentioned earlier, label cardinality (LC) is the average number of labels per example; while label density is the same number (LC) divided by number of labels in the dataset (6 in the emotion dataset as an example).

The second dataset is called "Yeast" which is concerned about protein function classification. This dataset contains 2417 examples with 198 distinct label-sets. The Yeast dataset has 14 different labels with cardinality equals to 4.327 and density equals to 0.303.

TABLE XI. MULTI-LABEL DATASETS STATISTICS

Dataset	Domain	# of Instances	# Attributes	# of Labels	DLS	LC	LD
Emotions	Music	593	72	6	27	1.869	0.311
Yeast	Biological	2417	103	14	198	4.327	0.303

Based on the statistics presented in Table14, we are more interested in LC to determine the association's cardinality which is equal to Label Cardinality – 1. Table6 and Table12 summarize the labels that could be found in the datasets which will be used in the evaluation process and the frequency of each label.

TABLE XII. FREQUENCY OF "YEAST" DATASET LABELS (C1 – C14)

Label	C1	C2	C3	C4	C5	C6	C7
Frequency	762	1038	983	862	722	597	428
Label	C8	C9	C10	C11	C12	C13	C14
Frequency	480	178	253	289	289	1799	34

An extensive evaluation process has been made using three evaluation measures, five problem transformation methods, and two algorithm adaptation methods. All multi-label classification methods and all supervised learning algorithms

which are used in this paper are implemented using Mulan tool [13] [14] which is a WEKA-based Java package for multi-label classification. All experiments were conducted using the 10-fold cross validation method. The proposed approach is evaluated using different evaluation measures which are: Accuracy, Hamming Loss, and Harmonic Mean (F1 Measure).

A. Experiments on "Emotions" Dataset

- **Accuracy:** In term of accuracy and as noticed from Fig.3, the proposed approach has the highest accuracy (0.767) among all the multi-label classification methods. The second best accuracy is 0.592 achieved by RAKEL. This indicates that using correlations among labels increase accuracy in a great way.
- **Hamming Loss:** As notices from Fig.4, the proposed approach has the lowest Hamming Loss (0.155) among all the multi-label classification methods. The second best hamming lost is achieved by RAKEL method (0.186), which indicates that the proposed approach decreases both incorrect labels classification and missing labels classification in a good way.
- **The Harmonic Mean (F1 Measure):** As noticed from Fig.5, the proposed approach has the highest Harmonic Mean (0.837) among all multi-label classification methods.

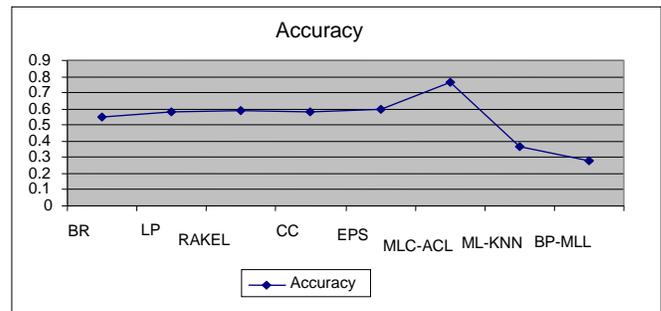


Fig. 3. Difference in accuracy between the proposed approach ((MLC-ACL) and other methods

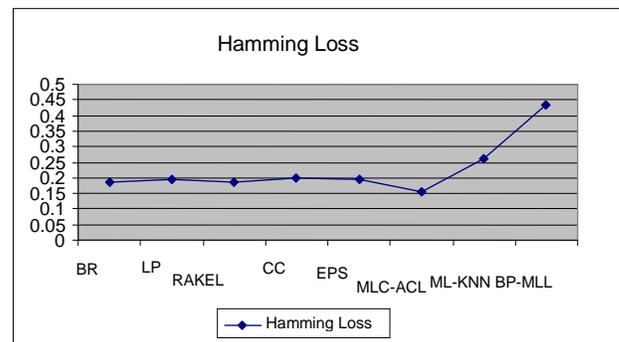


Fig. 4. Difference in Hamming Loss between the proposed approach (MLC-ACL) and other methods

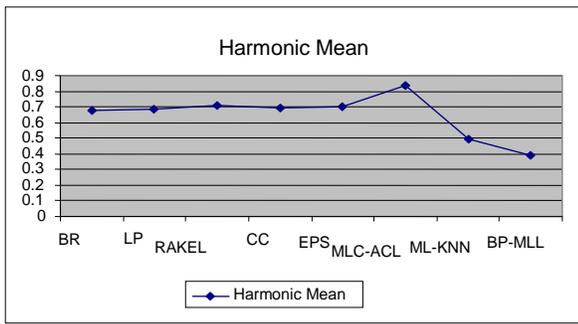


Fig. 5. Difference in Harmonic Mean between the proposed approach (MLC-ACL) and the other methods

B. Experiments on "Yeast" Dataset

Table13 contains the best correlations among labels after applying Predictive Apriori on "Yeast" dataset. Table14 summarizes the results of the evaluation measures on "Yeast" dataset. Table14 shows that the proposed approach has the highest accuracy (0.554), and EPS method has the second highest accuracy (0.537). The proposed approach has the best value for Hamming loss (0.161), while BR and ML-KNN have the second best value (0.193). Finally, the proposed approach has the best value (0.672) of Harmonic mean measure, and ML-KNN has the second best value (0.654) of Harmonic mean.

TABLE XIII. POSITIVE ASSOCIATION RULES USING THE "YEAST" DATASET

Rule #	Rule	Accuracy
1	IF C1 THEN C2	0.49
2	IF C2 THEN C12	0.43
3	IF C3 THEN C12	0.50
4	IF C4 THEN C12	0.51
5	IF C5 THEN C12	0.53
6	IF C6 THEN C12	0.54
7	IF C7 THEN C8	0.63
8	IF C8 THEN C13	0.50
9	IF C9 THEN C8	0.81
10	IF C10 THEN C11	0.82
11	IF C11 THEN C12	0.76
12	IF C12 THEN C12	1.00
13	IF C13 THEN C12	0.80
14	IF C14 THEN C4	0.99

TABLE XIV. EVALUATION RESULTS USING THE "YEAST" DATASET

Method	Accuracy	Hamming Loss	Harmonic Mean
BR	0.522	0.193	0.652
LP	0.530	0.206	0.643
RAKEL	0.493	0.207	0.559
CC	0.521	0.211	0.633
EPS	0.537	0.207	0.654
MLC-ACL	0.554	0.161	0.672
ML-KNN	0.520	0.193	0.654
BP-MLL	0.185	0.322	0.210

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have investigated the problem of multi-label classification, and the benefits from having the correlations among label in building multi-label rules. The outcome of this research is an algorithm for multi-label classification based on correlations among labels. Unlike

previous approaches, this algorithm combines between problem transformation methods with the criteria of selecting least frequent label and unsupervised learning method (Predictive Apriori). The main contributions of this research can be summarized as follows:

- Merging between two different learning tasks, the first task is an unsupervised learning task, which is the task of finding positive association among labels. The second task is a supervised learning task, which is the task of applying any rule-based classifier on the transformed dataset.
- Getting benefits from finding the correlations among labels, in the process of generating multi-label rules. Transforming multi-label dataset into single label dataset causes too loss in information, and by finding correlations among labels, the proposed approach tries to substitute this information loss.
- The proposed approach has much flexibility, since any rule-based classifier could be used in the process of classifying the transformed data set.

As a future work, we suggest Proposing New Problem Transformation Method based on Accuracy of correlations among labels We may adapt the proposed model as following:

- Step1: Discovery of positive correlations among labels
- Step2: Apply problem transformation method based on correlations among labels and using the highest accuracy criteria, which means to select the label that produces the highest accuracy as being antecedent of the association rule.
- Step3: Applying a rule based classifier on the transformed data set and producing the rules set.
- Step4: Generating the multi-label rules set, using the single rules set produced by the classifier in step 3, and the associative rules for each instance that has been discovered in step 1.

Experiment on "Emotions" dataset shows that the adapted model is promising and need to be studied more. When applying the adapted model in "Emotions" dataset, the accuracy was (0.752) which is really close to the accuracy of the proposed model (0.767).

ACKNOWLEDGMENT

Raed Alazaidah thanks all the professors in faculty of information technology in Philadelphia university, especially prof. Said Ghouli. And deep thanks to my best friend Naela Alsalmán.

REFERENCES

[1] Thabtah, F., Cowling, P. & Peng, Y, " MMAC: A New Multi-Class, Multi-Label Associative Classification Approach". In Proceedings of the 4th IEEE International Conference on Data Mining (ICDM'04), Brighton, UK, 2004, pp. 217–224.

[2] Neda Abdelhamid, Aladdin Ayesh, Fadi Thabtah, Samad Ahmadi, Wael Hadi, " MAC: A multiclass associative classification algorithm", Journal of Information & Knowledge Management , volume 11 , ssue 2, June 2012, pp. 1250011-1 - 1250011-10.

- [3] Boutell, M., Luo, J., Shen, X., Brown, C, "Learning Multi-label Scene Classification". *Pattern Recognition*, 2004, 37:1757–1771.
- [4] Tsoumakas, G., Vlahavas, I, "Random k-labelsets: An ensemble method for multilabel classification". In: *Proceedings of the 18th European Conference on Machine Learning (ECML 2007)*, Warsaw, Poland, 2007, pp. 406–417.
- [5] Read, J., Pfahringer B., Holmes G, "Multi-Label Classification using Ensembles of Pruned Sets", *8th IEEE International Conference on Data Mining*, 2008.
- [6] Furnkranz J. and Hullermeier E, "Pairwise Preference Learning and Ranking", In N. Lavrač, D. Gamberger, H. Blockeel, and L. Todorovski, (eds), *the 14th European Conference on Machine Learning (ECML-03)*, Cavtat, Croatia, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, 2837, 2003, pp. 145–156.
- [7] Jabez Ch, "A Statistical Approach for Associative Classification", *European Journal of Scientific Research*, 58(2), 2011, pp.140-147.
- [8] Read, J., Pfahringer, Bernhard, Holmes, Geo_rey, and Frank, Eibe. "Classifier chains for multi-label Classification". In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 2009, pp 254-269.
- [9] Clare, A., King, R, " Knowledge discovery in multi-label phenotype data". In *Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2001)*, Freiburg, Germany pp. 42–53.
- [10] Zhang, M. L., Zhou, Z. H, "ML-kNN: A Lazy Learning Approach to Multi-Label Learning", *Pattern Recognition*, 2007, 40, pp. 2038–2048.
- [11] Scheffer, T., "Finding Association Rules that Trade Support Optimally Against Confidence", In *Proceedings of the 5th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'01)*, Freiburg, Germany: Springer-Verlag, 2001, pp. 424-435.
- [12] Neda Abdelhamid, Fadi Thabtah, " Associative Classification Approaches: Review and Comparison", *Journal of Information & Knowledge Management*, 2014 .
- [13] Tsoumakas G. Spyromitros-Xioufis E. Vilcek J. Vlahavas I., "MULAN: A Java Library for Multi-Label Learning", *Journal of Machine Learning Research* 12, 2011, pp. 2411-2414.
- [14] Neda Abdelhamid, Aladdin Ayesh, Fadi Thabtah, " Phishing detection based Associative Classification data mining", *Expert Systems with Applications*, Volume 41, Issue 13 , 2014, pp. 5948-5959