IJCSN - International Journal of Computer Science and Network, Volume 7, Issue 6, December 2018
ISSN (Online): 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Patterns Quality Improvement Using Logical Analysis of Data and Mixed Integer-Linear Programs

[1]Abdulkareem Owehan Alresheedi; [2]Mohammed Abdullah Al-Hagery

[1]Master Degree Student, Computer Science Department, College of Computer, Qassim University
KSA

[2]Computer Science Department, College of Computer, Qassim University
KSA

Abstract - Logical analysis of data (LAD) is an important subfield of supervised machine learning and data mining. It is a methodology for data analysis, which uses concepts of optimization, combinatorics and Boolean functions. LAD is a binary classification that used for Boolean data with high explanatory power. Because patterns are the most important building blocks in LAD, they must be selected carefully. One of the main drawbacks in LAD, which needs to be addressed, is the quality of the generated patterns and extraction of positive and negative patterns. By these quality patterns, we can classify new observations with high accuracy. The proposed methodology developed to address this issue. It studied the LAD method, its refinements, and define quality measures for pattern generation. Then, contribute to improving the pattern selection procedures using an optimization technique called Mixed Integer-Linear Programs (MILP) and the General Algebraic Modelling System (GAMS) tools using MIP solver. Using this technique for generating an optimized set of patterns aims at selecting the most important patterns to improve pattern quality, and get very strong results with a high accuracy. Experiments carried out on the SPECT dataset, it shows the efficiency of the proposed method in regards to minimize the number of generated patterns and increase the accuracy of the classification model.

Keywords - *Logical Data Analysis, Optimization Techniques, Patterns Reduction, Machine Learning, Classification Accuracy, Set Covering Problem.*

1. Introduction

Artificial Intelligence (AI) is a promising field, which has a lot of hot topics and an open area for the researchers. Machine Learning (ML) is known as one of the branches of Artificial intelligence. In scientific and industrial fields, it has raised considerable attention to people. Rapidly growing data, extraction meaningful insights from datasets and learning predictive models, all are key reasons triggering ML in numerous areas. ML is typically divided into two common fields, supervised ML as classification and unsupervised ML as clustering. Both fields of ML are concerned with the analysis of datasets. The most important application of ML is in Data Mining (DM). Humans are generally likely to make mistakes during analysis or, possibly, when working to create relationships between multiple attributes. This makes it hard for them to find solutions to certain problems. The process of extraction of knowledge from datasets, in order to generate patterns for data interpretation in a clear way called knowledge discovery.

Relatively, a new supervised ML methodology is LAD, it depends upon Boolean Logic, combinatorics and optimization. LAD performs analysis on data. The goal behind is to discover logical hidden patterns that distinguish one class from other based on labelled training data. Patterns are primary building units in LAD and the total number of positive and negative patterns called Pandect. From a primary group of pandect, which contain all negative and positive patterns, some patterns with high quality are selected to create a classification model (decision rules) in order to classify new observations correctly is one of the main goals of LAD method. On the selected dataset, experiments results show how successfully the goal met using the MILP optimization technique. The importance of this contribution is to improve the LAD outputs in regards to the patterns reduction and accuracy improvement.

The rest of this paper organized as follows: Section 2 presents the LAD method for data classification, Section 3 explains a background of previews contributions, whereas, section 4 describes the research methodology, then section 5 focuses on the application of LAD and MILP, while section 6 describes the experiments and results. Section 7 concentrates on results discussion then section 8 presents the conclusion. Finally, section 9 highlights the future work.

2. LAD Method

Numerous digital data were generating each day using the Internet and other smart sensor devices. Data analysis finds hidden structural information in datasets and thus assist decision making through ML algorithms and DM concepts. Training set comprises input observations; the observation is one row of data and called an example, an instance, records, and points. The observation is a special set of vectors built using attributes and columns value of features for some instances. LAD methodology comprises four main steps data binarization, support set, pattern generation, and classification model (theory formulation) [1], [2].

2.1 Patterns Generation and Optimization in LAD

Human experts understand patterns that are central to LAD. A positive pattern can be interpreted as a collection of values of a small number of attributes. They appeared in a large number of positive observations and seen in a few numbers of negative ones [3]. Hence, new observation under positive pattern can be an indication of its positive character. Likewise, duality and indications of negative patterns give negative patterns. Patterns are reliable. One can rely upon them to build theory, as much as he can due to the dependencies it provides inherent to the observations. Pattern generation for positive and negative patterns are symmetric procedures. The generation of both patterns is almost similar hence, only three approaches to generate positive patterns are discussed: Top-down, Bottom-up and hybrid [4]. The Hybrid approach consists of the above two approaches (Top-down and Bottom-up). Firstly, the bottom-up technique is used to generate patterns and then the top-down technique is used to observe positive patterns that were left behind. Simplicity and comprehensiveness are used as prime logical objectives in a hybrid approach. Both focus on short patterns and covering all positive and negative observations.

2.2 Patterns Optimization Criteria

There are many optimization criteria for pattern generation. The pattern optimization criteria used in this paper are Degree, Homogeneity and number of covered observations [5]. Where, a good LAD model has a small number of features, a small number of patterns, and high-quality patterns, which have (a small Degree and High Homogeneity).

The top of this paragraph illustrates a sub-subheading.

3. Background

Binary classification problems are solved using various algorithms For examples, Decision Tree Classifiers (DTCs) [6], Support Vector Machines (SVMs) [7], and Neural Networks (NNs) [8] as well as logical analysis of data (LAD).
Peter Hammer was the first person discussed the idea of LAD in Rutgers Center for Operations Research at Rutgers University, USA[2]. Although the idea of LAD was in 1988, applications of LAD started in 2002-2003, since some medical researchers publish their studies about risk assessment among cardiac patients [9], [10]. There are many practical usages of LAD classifier include as discussed by [11], [12], [13].
LAD is an inductive method that lies under the ML field. It identifies the useful part of information and aids in making decisions or building models. Similar to an inductive process, it groups the observations into classes. Class of unknown examples will be predicted after formulating the rules [14], [15]. After discriminating observations from one class to opposite classes, LAD identifies positive (negative) patterns as conjunctions of the attribute that cover a large number of positive (negative) observations and a few numbers from negative (positive) observations[16], [17].
Since the official publication by [18], there has been a rise in the progress of LAD. New theoretical developments and applications are found too. Further, it is extended via experiences with new extensions and implementations especially with the combination of other techniques. Moreover, the contribution of [19] deals with the issue of binary classification of data records to solve limited problems by LAD. Probability considerations are also, represented with improvements in terms of standard LAD procedures.

Development has been observed in the theoretical basis of binarization [20], revealing a collection of optimization problems associated with reducing the number of binary variables. Pair of sets of positive and negative examples for modelling may comprise of missing components. This essential issue is seen in many areas like theory and pattern recognition, LAD and example-based knowledge, data mining and knowledge discovery [21]. This modelling reveals how the missing bits are dealt. Three extensions are formulated for different Boolean function classes naming consistent, robust and most robust. These Boolean functions include general, positive, horn, threshold, decomposable and k-DNF, which is a polynomial time algorithm to generate spanned patterns of the dataset. The proposed algorithm is tested in terms of its accuracy using datasets available on a public platform. Results of the experimental series reveal a higher degree of robustness in a spanned pattern [22]. A type of genetic algorithms was developed by [23] giving a solution to the minimal covering of Boolean matrix that tends to use a non-standard crossing, operators, solution admissibility recovery and mutation. About 65 large sized complex problems were used to evaluate the accuracy of algorithms. It was observed that 61 problems give optimal solutions while the length of coverings in four problems are different from one by unity.

Some standard techniques for the logical analysis of numerical data are explored in [24] with an ambition to quantify the accuracy of prediction of methods in the probabilistic model of ML. We have particularly obtained results involving robustness of cut observations, selected for data binarization. The bounds keenly suggest that reducing the combination of cut observations in number and their robustness is beneficial. This also suggests changes in greedy cut observation's selection for robustness.

New approach lies its foundation upon integer programming and network flows. It determines an important pattern in order to generate the LAD model. User-specified essential requirements are also enabled under this approach like hamming distance, coverage, statistical significance, the prevalence of patterns, and Homogeneity. Selection-based network flows and utility of integer programming are experimental results to datasets [25].

An optimum replacement model [26] developed to minimize the long-term average cost of equipment maintenance using LAD. LAD’s pattern generation procedure employed using the generated patterns. The equipment’s survival probability estimated. The probabilities are then used in a dynamic programming model to make an optimum decision to replace a piece of equipment under condition monitoring. Analysis of performance when done, introduced some methods giving a reliable result. Results are advantageous to maintenance practitioner.

A Mixed 0 -1 Integer and Linear Programming (MILP) represented to determine LAD patterns, which are optimal in terms of pattern selection preferences and previously studied patterns too. Performing extensive experiments on large datasets, the proposed approach gives efficiency in generating LAD patterns [5].

MILP models as developed in [27] is for many different optimal and Pareto-optimal LAD patterns. Hence, it was expected that the model would generate useful LAD patterns in an efficient way. After experiments performed on ML datasets, the identified efficiency given by new MILP models followed by their usage in revealing various utilities of strong prime and spanned patterns, improving the entire accuracy of classification of LAD decision theory.

A proposed algorithm created based on MILP to expand LAD in order to solve multi-class classification problems, where One-vs-All learning models are built in an efficacious manner for the classification of observations in predefined classes. Various experiments performed using benchmark datasets suggests the utility of the approach [15].

Multilinear Programming (MP) model, is the triumphant outcome of development done by [28]. MP is for Pareto optimal patterns for LAD. It is being revealed that the existing and old pattern model generation can be attained even naturally using MP through linearization for 0-1 multi-linear functions. 0 -1 MP enables the understanding of how independently developed models are associated with one another. 0 -1 MP gives a unifying theory for the generation of LAD patterns. Results highlight the efficiency of MILP/MP pattern generation. In addition to this, Hammer and Alexe were able to introduce a totally polynomial algorithm to enumerate patterns of a certain degree and also to efficient variants for enumeration of patterns having certain coverage and sign requirements, demonstrated in public datasets [29].

Alexe et al., presented a number of algorithms to generate strong spanned and strong prime patterns. These patterns run linearly for possible conjunctions. Their research reveals LAD strong prime patterns as a classifier with higher accuracy compared to strong spanned patterns. However, the difference level may be small enough that it becomes insignificant statistically [30].

Examples of optimal patterns being identified in MILP and are studied include; strong patterns, strong prime patterns, strong spanned patterns. Other types of patterns are maximum prime patterns and maximum spanned patterns. Art formulation, a MILP method, can also be made to determine optimal patterns, stratifying prevalence, complexity and Homogeneity as per requirements specified clearly by the user [5]. In this research, the MILP is employed to improve the quality of LAD outputs.

The application of MILP is to improve the quality of the generated patterns. It focuses on the most important patterns and ignores the low-quality patterns. The optimized patterns achieve an equal accuracy compared to the accuracy of using all patterns, application of this method during the learning and testing process of the LAD classification model is a perfect strategy in regards cost-effective of the system.

IJCSN - International Journal of Computer Science and Network, Volume 7, Issue 6, December 2018
ISSN (Online): 2277-5420
www.IJCSN.org
Impact Factor: 1.5

4. Methodology

The methodology provides a solution for patterns reduction. It focuses on how to select a subset of patterns from their total number to establish an accurate classification model with a minimum number of patterns and high classification accuracy. The steps are shown in Figure (1 and 2) and include the following steps:

1. Data selection, data pre-processing, binarization for non-binary data
2. Generation of all patterns for a specific Degree, using LAD depends on the bottom-up algorithm [4], then divided the generated patterns into positive patterns and negative patterns.
3. For each generated pattern, get the total number of positive, negative covered observations, Degree, and Homogeneity percentage and then get the accuracy for all initial patterns.
4. Using MILP (GAMS) to reduce the number of patterns and get the optimal patterns that cover all observations in the dataset.
5. Getting the accuracy for optimized patterns by MILP (GAMS), then compare the results of steps 4 and 5 with that of step 3.
6. Developing RStudio scripts (functions code) for all steps.

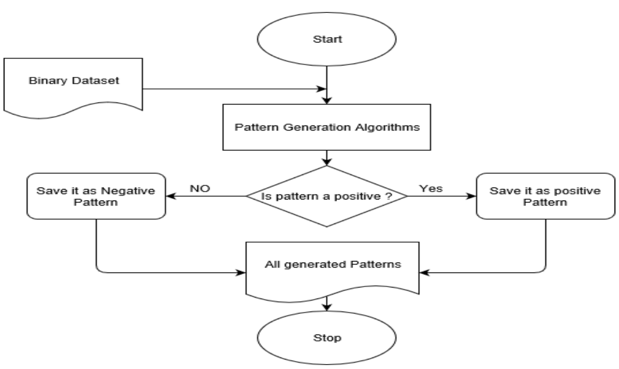


Fig. 1 Pattern generation

4.1 Pattern Generation for a Specific Degree

The essential part of LAD is concentrating on pattern generation. The best algorithm is which combines two main objectives; simplicity and comprehensiveness [6]. In simplicity principle, low Degree patterns are preferred. In another principle, each observation should cover by at least one pattern. Pattern generation step for a specific Degree illustrated in Figure (1), which describes the chart of that algorithm for generating all patterns using the bottom-up algorithm. The main idea to generate all patterns (negative and positive patterns) in a specific Degree is to achieve flexibility in pattern generation.

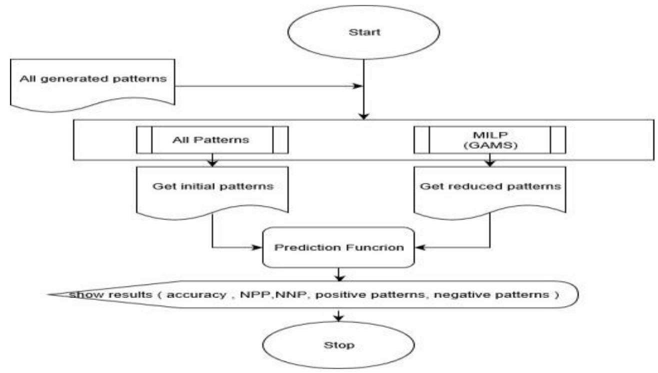


Fig.2 Main methodology

In practice, it is sometimes difficult to define prime patterns, which cover only positive or negative observations. In this case, we look for positive (negative) patterns that cover a larger number of positive (negative) observations and very few negative (positive) observations. In this type of pattern, a characteristic called Homogeneity, which is equal to the proportion of the positive (negative) observations covered compared to all observations covered by this pattern. Clearly, a pure pattern has a Homogeneity equal to 100%.

4.2 Pattern Evaluation

At each research stage, a pattern can be in one of three situations:

a. Pattern with a Degree is smaller than D, it covers some positive observations and its Homogeneity is greater than H.

b. It is a candidate for the next stage, where, its Degree is smaller than D and it covers some positive observations, but its Homogeneity is smaller than H.

c. It is not a candidate. It does not cover any positive comments.

d. To evaluate each pattern, we developed a number of functions code, for instance, Evaluate Function and EvaluateN Functions, which developed by the R language, included in (Appendix A and B) that will return the observations (Positive/Negative) covered by the pattern, then we will use this information to calculate Homogeneity and prevalence.

4.3 Choose the Degree and Homogeneity

To choose the maximum Homogeneity and the minimum Degree for used patterns, the start Function was developed as the main function to achieve this task and the function

IJCSN - International Journal of Computer Science and Network, Volume 7, Issue 6, December 2018
ISSN (Online): 2277-5420
www.IJCSN.org
Impact Factor: 1.5

code is presented in Appendix C. This function has three arguments; the first argument is that a Boolean value indicates if we want to generate positive patterns (True) or negative patterns (False). The second argument for Degree reduces our generation of patterns that have at most this value as Degree and the third argument is for Homogeneity that we accept for patterns.

5. Application of MILP (GAMS)

MILP is a subfield of linear programs in which some variables are required to take integer values. We have generated linear programs for minimizing the number of used patterns. For that, an optimizer function has developed to generate the linear programming model. Because of the pattern generation step, their number is usually too large. It needs to choose a subset that is of a reasonable size and at the same time will be sufficient to separate positive and negative observations.
The method used frequently to minimize the number of patterns to solve the set covering problem presented in this section. Therefore, it is useful to use this method in our case. Patterns that cover a large number of observations have likely to be chosen by MILP and give us guaranty all observations will be covered by this method [5]. We proposed a global model for the set of positive and negative generated patterns from S+ and S-, subsets of positive and negative observations. For that, Boolean Matrix is created that has a column of pattern names and row for the observations. For each pattern and each row, the value is set to 1 if the patterns cover the observation and otherwise set to 0. The Appendix, part D, shows the GAMS code to solve the set covering problem using MIP solver, as presented in appendix part D, E, and F.

6. Patterns Generation

Based on the proposed methodology, a large number of experiments carried out and the results generated from the SPECT dataset. If we want to generate patterns in an exhaustive way, we need to generate all patterns with 1 Degree, make all combination with these patterns one by one and go through this algorithm until we make a combination with all patterns. In SPECT dataset, the last pattern will have a Degree equal to 22. Because SPECT dataset contains 22 attributes and 80 observations.

By calling the function Start(), shown in the appendix, part C, to give three arguments the first argument will True or False. It depends on what we desire to storage (if the value is TRUE then the patterns stored a positive pattern, else negative patterns). The second argument defines the maximum Degree for these patterns, and then the last argument is the minimal threshold for the Homogeneity. In the first running of the script, we have chosen 4 as the maximum Degree and 50% for Homogeneity (50% means that we will generate all patterns with at most 4 Degree)
The SPECT dataset generates 25,962 positive patterns and 10,658 negative patterns. In order to identify the smallest subset of patterns that cover all observations. The LAD uses a linear model that minimize the number of used patterns to cover those observations. However, the objective is not minimizing the number of used patterns to cover all observations, but the search is to find the optimal sets of patterns that will help to predict unknown observations. In order to do that, we will compare the result of the accuracy of using all patterns (LAD method) and MILP model results in regards the reduced patterns and accuracy.
To understand how we get the predicted value, we just calculate the number of positive and negative patterns got and calculate the percentage of existing for each pattern then choose the highest value as our prediction. For example, if we have a new input that is covered by 20% of positive patterns and 30% of negative patterns we conclude that the input is a negative observation. A concrete example (one observation) for the results from MILP method.

Target	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18	a19	a20	a21	a22
1	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	1

Positive Patterns	Negative Patterns
a5-a7-a8-a11-a20-	|a8|-|a12|-

Fig.3 Positive and Negative patterns from MILP

As shown in Figure 3, this observation is covered by a8 and by |a12| as shown in Table (4.3). As a result, 1⁄5 of positive patterns covered this point and 1⁄2 of negative patterns covered this point. As 0.5 is greater than 0.2 this point will be considered as a negative observation.

7. Results Discussion and Evaluation

The results are discussed and evaluated based on different values of Degree (D) and Homogeneities (H). Every result must be under inspection, to get an idea of how the criteria work. For that, we compare the result by changing only one constant variable. In this case, we change the Degree and let the Homogeneity, 60%, as a constant. The objective of this processes accomplished for continuous improvement of results as shown in Table 1: rows 1 and 2.
The results in Table 1: rows 1 and 2 show that when the Degree of patterns changed from 3 to 2, the results of MILP still fixed (62.5%), without changes. The results of the MILP method for Degree 2 and Homogeneity 60, still a

good choice. Because the accuracy result is not changed in some experiments but in others, it changed positively, see Figure 4. Although, the results of MILP in some cases are equal to the results of LAD, with MILP the accuracy generated based on a small set of patterns in contrast with LAD which give equivalent accuracy results but with all patterns. So the accuracy of MILP is the best.

Table 1: Results of Different Degrees and Different Homogeneities

Experiemnt. No	Method Type	Accuracy	Number.PP	Number.NP	PP	NG	H	D
1	LAD method	56.25	3576	47	-	-	60	3
	MILP	62.5	5	1	a5-a7-a8-a11-a20-	|a13|-	60	3
2	LAD method	56.25	398	10	-	-	60	2
	MILP	62.5	5	1	a5-a7-a8-a11-a20-	|a13|-	60	2
3	LAD method	81.25	67	27	-	-	85	3
	MILP	81.25	3	2	a16-a17-a18-	|a8|-|a12|-	85	3
4	LAD method	68.75	433	28	-	-	80	3
	MILP	75	4	2	a7-a14-a16-a19-	|a8|-|a12|-	80	3
5	LAD method	68.75	1238	47	-	-	75	3
	MILP	75	6	2	a7-a11-a13-a14-a19-a20-	|a8|-|a12|-	75	3
6	LAD method	62.5	2867	176	-	-	65	3
	MILP	68.75	5	2	a5-a7-a8-a11-a20-	|a8|-|a12|-	65	3
7	LAD method	56.25	3576	47	-	-	60	3
	MILP	62.5	5	1	a5-a7-a8-a11-a20-	|a13|-	60	3
8	LAD method	68.75	3783	912	-	-	55	3
	MILP	68.75	5	2	a5-a7-a8-a11-a20-	|a13|-|a16|-	55	3
9	LAD method	68.75	3864	1740	-	-	50	3
	MILP	62.5	5	1	a5-a7-a8-a11-a20-	|a17|-	50	3
10	LAD method	81.25	140	48	-	-	85	5
	MILP	81.25	3	2	a16-a17-a18-	|a8|-|a12|-	85	5
11	LAD method	81.25	81	48	-	-	90	5
	MILP	81.25	3	2	a16-a17-a18-	|a8|-|a12|-	90	5
12	LAD method	81.25	55	27	-	-	90	3
	MILP	81.25	3	2	a16-a17-a18-	|a8|-|a12|-	90	3
13	LAD method	81.25	77	35	-	-	90	4
	MILP	81.25	3	2	a16-a17-a18-	|a8|-|a12|-	90	4
14	LAD method	81.25	104	35	-	-	85	4
	MILP	81.25	3	2	a16-a17-a18-	|a1|-|a13|-	85	4
15	LAD method	68.75	1590	39	-	-	80	4
	MILP	75	4	2	a7-a14-a16-a19-	|a1|-|a13|-	80	4
16	LAD method	75	4989	135	-	-	75	4

	MILP	75	6	2	a7-a11-a13-a14-a19-a20-	|a1|-|a13|-	75	4
17	LAD method	62.5	17490	569	-	-	65	4
	MILP	75	5	2	a11-a13-a14-a19-a7 |a15|-	|a8|-|a13|-	65	4

On the other hand, when the Degree is 3, the MILP model reduces the number of patterns from 3576 to 5 for positive patterns and from 47 to 1 for negative patterns. This means that we have decreased the number of patterns by ((3576+47)-(5+1))/(5+1)= 602.83 times. In addition, when the Degree is 2, the MILP model reduces the number of patterns from 398 to 5 for positive patterns and from 10 to 1 for negative patterns. This means that we have decreased the number of patterns by ((398+10)-(5+1))/(5+1)= 67 times. Also, we increase the accuracy by (62.5-56.25)=6.25%. In these two cases, the accuracy kept stable and better than the accuracy of using all initial patterns in LAD method, as shown in Table 1: rows 1 and 2, as in Figure 5.

The generated patterns for Homogeneity 85 and Degree 3 are shown in Table 1: rows 3, these patterns include 67 positive patterns and 27 negative patterns. Because the Homogeneity is high, 85%, accuracy for MILP decreased 3 positive patterns and 2 negative patterns, respectively. Whereas, the accuracy still without changes.

The MILP satisfied same accuracy and best reduction of patterns because it has 5 patterns (3 positives and 2 negatives).

When the Homogeneity is 80 and Degree = 3, the number of generated patterns was 433 positives and 28 negatives produced by LAD method. The accuracy for MILP is increased by 6.25% (from 68.75% to 75%, using LAD method), as shown in Table 1: row 4. On the other hand, the MILP method in this experiment achieved a high degree of accuracy and a high percentage of patterns reduction because it uses just 6 patterns (4 positives and 2 negatives) and the accuracy increased to 75% where the number of used patterns decreased by ((461-6)/6) = 75.83%. According to these results, a large number of used patterns effect on the classification model. The accuracy of 461 patterns is 68.75. Whereas, the accuracy of 6 patterns is 75. Besides that, we found the numbers of generated patterns for Homogeneity 75 and Degree 3, are 1238 positive patterns and 47 negative patterns by the LAD, while, the MILP method decreased the number of patterns from 1285 (1238 positives and 47negatives) to 8 patterns (6 positives and 2 negatives). The reduction rate is ((1285-8)/8) = 159.6%). In addition, the accuracy of MILP increased by 6.25% (from 68.75% using LAD method to 75%), as illustrated in Table 1: row 5.

Likewise, the results of Homogeneity 65 and Degree 3 produce 2867 positive patterns and 176 negative patterns and the MILP has just 7 patterns (5 positives and 2 negatives), the number of used patterns decreased by ((3043-7)/(7)= 433 times and satisfy an accuracy equal to 68.75 %, by increase equal to 6.25%, as displayed in Table 1: row 6.

As well, the result of generated patterns for Homogeneity 60 and Degree 3 were 3576 positive patterns, 47 negative patterns by LAD. The MILP has just 7 patterns (5 positives and 2 negatives), the number of patterns decreased by ((3623-7)/(7) = 516.57 times, as well, the MILP increased the accuracy by 6.25%, as in Table 1: row 7

In addition, when the Homogeneity is 55 and Degree is 3, LAD method produced 3782 positive patterns and about 912 negative patterns. It found that the accuracy kept as it is 68.75%. The MILP use just 7 patterns (5 positives and 2 negatives), thus the number of used patterns by MILP decrease by ((4694-7)/(7)= 669.6 times.

Likewise, if the Homogeneity is 50 and Degree 3, LAD generates 3864 positive patterns and 1740 negative patterns. In contrast, we find that the numbers of generated patterns by LMIP are 5 positives and 2 negatives. Besides, the MILP methods decreased the accuracy with 6.25%, as shown in Table 1: rows 8 and 9.

The generated patterns for when Homogeneity = 85 and Degree = 5 include 140 positive and 48 negative patterns. MILP produces 5 patterns; 3 positives and 2 negatives. It produces a better reduction of patterns. Because the Homogeneity is high, 85%, the accuracy of MILP stayed at the same level 81.25 without changes.

It noticed that in Table 1: rows 10, 11, 12, the results are similar in regards to accuracy, whereas the differences only in the number of generated patterns by these methods. The MILP is the best in reduction rate, although the Homogeneity value is different in rows 10 and 11. So, when Homogeneity= 90, we got the best quality patterns that have a small Degree and high Homogeneity [5]. Furthermore, using a Homogeneity 90 and Degree 4 give the same accuracy when compared with Degree 3 and 5. According to the results in Table 1: row 13, it found that a small Degree of patterns has efficiently and utility to select high-quality patterns as well as the power of explanations and justifications. Also, it found that when the Homogeneity 85 and Degree 4, the patterns generated are given in Table 1: row 14, which produce 104 positive patterns and 36 negative patterns. The accuracy for MILP is still same results. The MILP uses only 5 patterns (3

IJCSN - International Journal of Computer Science and Network, Volume 7, Issue 6, December 2018
ISSN (Online): 2277-5420
www.IJCSN.org
Impact Factor: 1.5

positives and 2 negatives), so the number of reduction equal to ((139-5)/5)=26.8 times.

Moreover, the generated patterns for Homogeneity 80 and Degree 4, produced about 1590 positive patterns and 39 negative patterns, as presented in Table 1: row 15. The accuracy increased by 6.25% (from 68.75% to 75%, using LAD method). The patterns reduced from 1590 to 4 for the positive patterns and from 39 to 2 for the negative patterns, as well, the reduction percentage of used patterns = ((1629-6)/6) = 270.5 times.

By looking at Table 1: row 16, we find that the generated patterns when the Homogeneity = 75 and Degree = 4, the number of patterns bigger than the number of patterns when Homogeneity is 80. The number of positive patterns was 4989 positive and 135 negative. This number reduced to 6 positives and 2 negatives by MILP. The accuracy for the proposed method is still the same. The number of reductions of used patterns = ((5124-8)/8) =639.5 times.

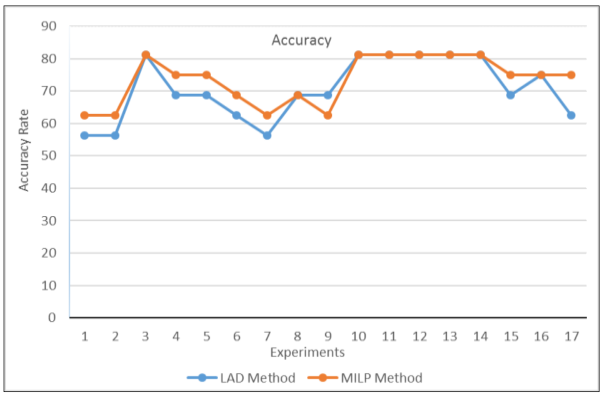


Fig.4 Comparison of Accuracy Reduction

In case the Homogeneity = 65 and Degree = 4, the generated patterns are bigger than the number of patterns when Homogeneity = 75. Here the number of original patterns are 17490 positive patterns and 569 negatives, these patterns reduce to 5 positives and 2 negatives respectively, as shown in Table 1: row 17. The accuracy increased by at least 12.5% (from 62.5% using LAD method to 75% using the MILP, so the number of used patterns decreased by ((18059-7)/7) 2578.9 times.

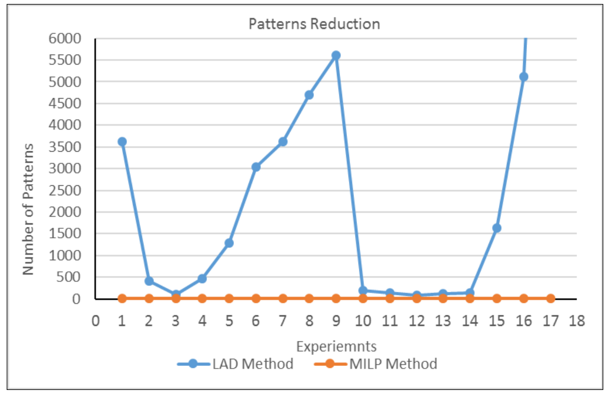


Fig.5 Comparison of Patterns Reduction

8. Conclusion

This paper introduced voluble contribution regarding the optimization and selection of high-quality patterns. The results show a high efficiency to improve the quality patterns selection procedures and helps researchers in future for the continued development of optimization techniques related to classification problems.

1) For each pattern, calculate and extract the main criteria that help in pattern selection procedures. For that, R codes are developed to extract the hidden information (patterns) from binary datasets and its characteristics such as Degree, Homogeneity, and a number of positive (negative) covered observations and show all index of all covered observations.

2) Adapt MILP method in LAD methodology to improve patterns quality. This adaption is a new contribution regarding to the optimization of patterns.

3) The proposed method is flexible to show various results depend on parameters sets such as Degree and Homogeneity. This is very important for decision makers.

4)The MILP method shows a higher efficiency in decreases the patterns and an equal percentage of accuracy, but using a small number of patterns and this counted as a stunning achievement, this is based on the analysis of the data and the two given criteria to find out the percentage of decrease in patterns or percentage of increase in accuracy. Pattern selection procedures used the optimization technique (MILP), that illustrated great results with a high percentage of decrease in patterns and get equal accuracy of the classification model using a very small number of patterns. These results based on different experiments (17 x 2) with changing both Degrees and Homogeneity, as shown in Figure 4 and 5 as well in Table 1 above.

9. Future work

From a special perspective, some ideas are suggested as an extension of this work for further research. It will be

IJCSN - International Journal of Computer Science and Network, Volume 7, Issue 6, December 2018
ISSN (Online): 2277-5420
www.IJCSN.org
Impact Factor: 1.5

good topics in future for those who have an interest in this field, as follows:

1- All results can be used separately as functions to apply several machine learning algorithms and data mining techniques to improve the current accuracy results.

2- Application of other techniques such as Multiobjective Evolutionary Algorithms (MOEA), Pareto on the same data for the same purpose (increase the accuracy and finding the minimum number of patterns.

3- Adapt data mining and machine learning techniques, e.g. genetic algorithms, to get the best selection of quality patterns. In additions, improves the main steps of LAD.

Appendix

The appendix includes samples of the code for three types of functions developed; start function evaluate and evaluateN functions.

A-Evaluate Function

```
evaluate<-function(table){
  s<-table
  table<-data.table(table)
  if(ncol(table)>1){
    s<-apply(table,1,FUN=prod)
  }
  invisible(list(value=length(s[which(s==1)]),points=paste0(which(s==1),sep="-"
,collapse="")))
}
```

B- EvaluateN Function

```
evaluateN<-function(table){
  s<-table
  table<-data.table(table)
  if(ncol(table)>1){
    s<-apply(table,1,FUN=prod)
  }
  if(length(which(s==1))>0){
    x<-which(s==1)+length(Pos_pnt[,1])
  }else{
    x<-which(s==1)
  }
  invisible(list(value=length(s[which(s==1)]),points=paste0(x,sep="-",collapse=
"")))
}
```

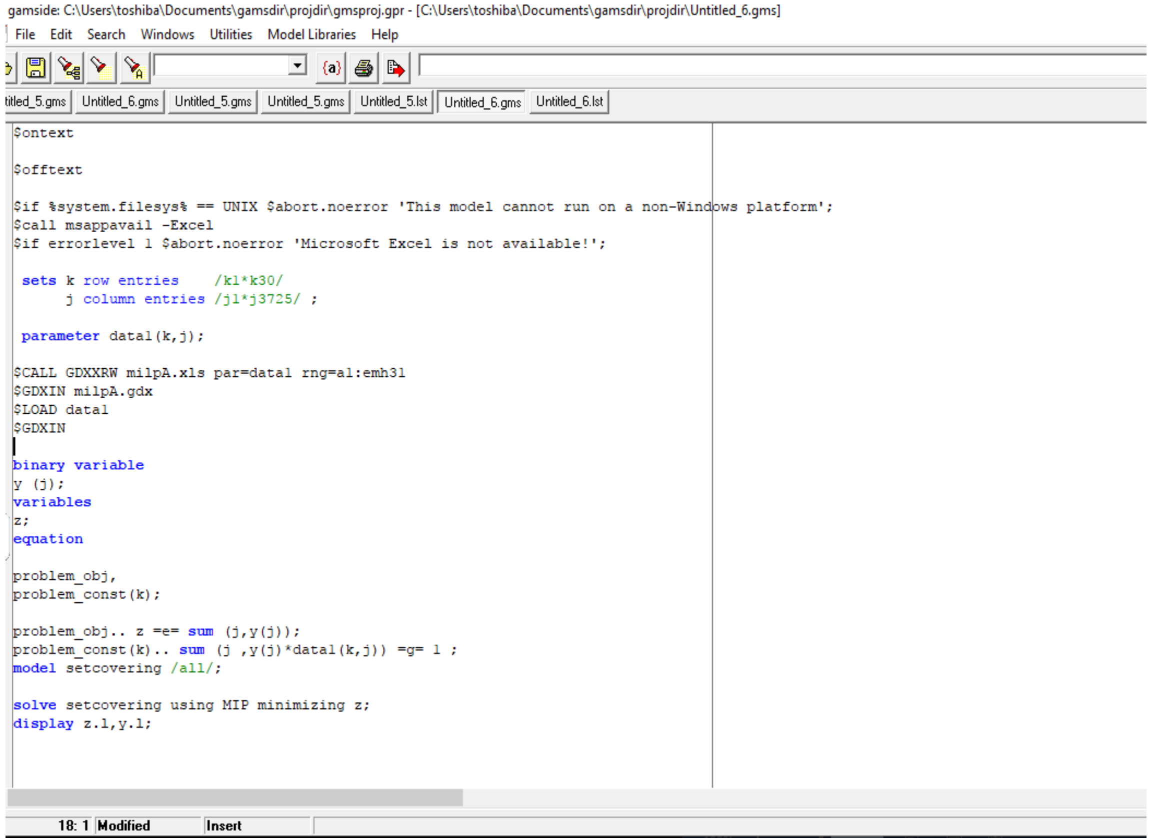
C- Start Function for Degree and Homogeneity

```
Start<-function(p=TRUE,degree,Homoginity){
  outputs1<-Init(p)
  outputs1 <-unlist(outputs1 )
  outputs1 <-outputs1 [which(outputs1 !="")]
  outputs1 <-getTable(outputs1 ,p)
  e<-outputs1
  if(max(as.numeric(unlist(e[,5])))>=as.numeric(Homoginity)){
    f<-which(as.numeric(unlist(e[,5]))>=as.numeric(Homoginity))
    e<-e[f,]
  }else{
    f<-which(as.numeric(unlist(e[,5]))>=max(as.numeric(unlist(e[,5])))*95/100)
    e<-e[f,]
  }
  outputs1<-e
  for(i in 1:(degree-1)){
    e<-Next_Degree(e,p)
    if(max(as.numeric(unlist(e[,5])))>=as.numeric(Homoginity)){
      f<-which(as.numeric(unlist(e[,5]))>=as.numeric(Homoginity))
      e<-e[f,]
    }else{
      f<-which(as.numeric(unlist(e[,5]))>=max(as.numeric(unlist(e[,5])))*95/100
)
      e<-e[f,]
    }
    outputs1<-rbind(outputs1,e)
  }
  invisible(outputs1)
}
```

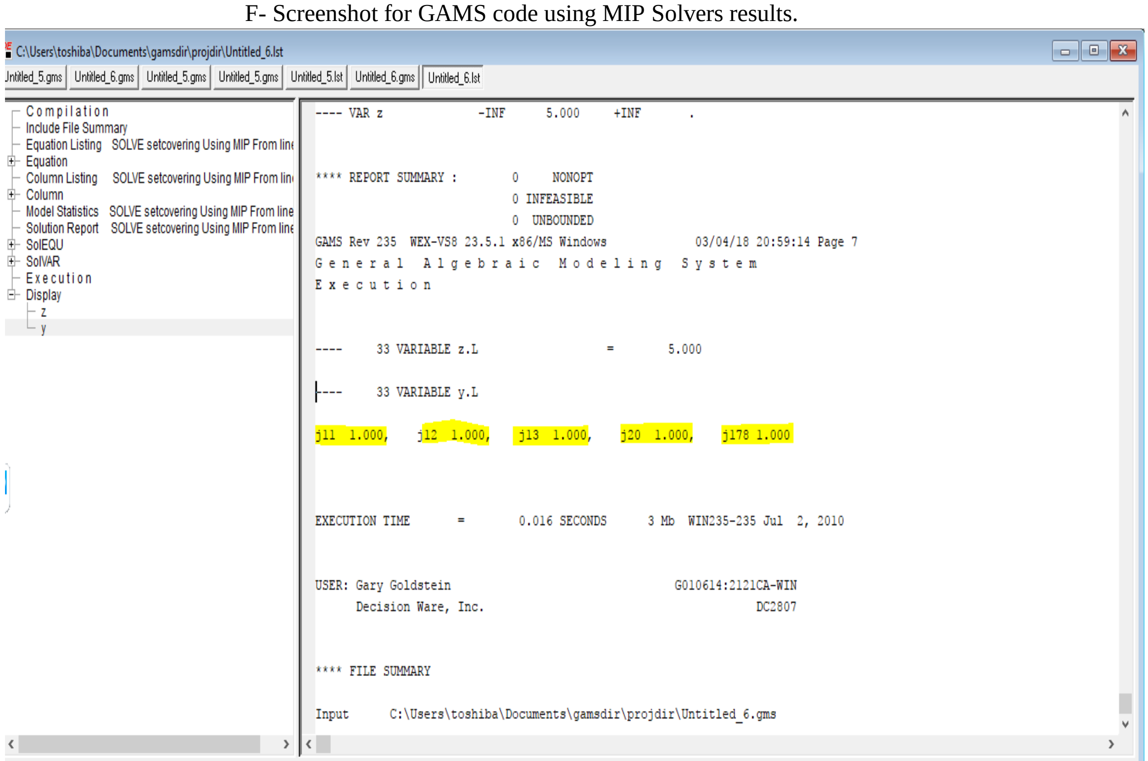
D- GAMS code using MIP solve

```
$ontext
$offtext
$if %system.filesys% == UNIX $abort.noerror 'This model cannot run on a non-Windows
platform';
$call msappavail -Excel
$if errorlevel 1 $abort.noerror 'Microsoft Excel is not available!';
 sets k row entries    /k1*k30/
      j column entries /j1*j3725/ ;
 parameter data1(k,j);
$CALL GDXXRW milpA.xls par=data1 rng=a1:emh31
$GDXIN milpA.gdx
$LOAD data1
$GDXIN
binary variable
y (j);
variables
z;
equation
problem_obj,
problem_const(k);
problem_obj.. z =e= sum (j,y(j));
problem_const(k).. sum (j ,y(j)*data1(k,j)) =g= 1 ;
model setcovering /all/;
solve setcovering using MIP minimizing z;
display z.l,y.l;
```

E- Screenshot for GAMS code using MIP Solvers.



F- Screenshot for GAMS code using MIP Solvers results.



References

[1] H. H. Kim and J. Y. Choi, "Hierarchical multi-class LAD based on OvA-binary tree using genetic algorithm," Expert Syst. Appl., vol. 42, no. 21, pp. 8134–8145, 2015.

[2] Y. Crama, P. L. Hammer, and T. Ibaraki, "Cause-Effect Realtionships and Partially Defined Boolean Functions," Ann. Oper. Res., vol. 16, pp. 299–325, 1988.

[3] A. Ragab, X. de Carné de Carnavalet, S. Yacout, and M.-S. Ouali, "Face recognition using multi-class Logical Analysis of Data," Pattern Recognit. Image Anal., vol. 27, no. 2, pp. 276–288, 2017.

[4] E. Boros, P. Hammer, and T. Ibaraki, "An implementation of logical analysis of data," Knowl. Data …, vol. 12, no. 2, pp. 292–306, 2000.

[5] H. S. Ryoo and I. Y. Jang, "MILP approach to pattern generation in logical analysis of data," Discret. Appl. Math., vol. 157, no. 4, pp. 749–761, 2009.

[6] T. Li, C. Zhang, and M. Ogihara, "A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression," vol. 20, no. 15, pp. 2429–2437, 2004.

[7] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," IEEE Intell. Syst., vol. 13, pp. 18–28, 1998.

[8] G. P. Zhang, "Neural networks for classification: a survey," IEEE Trans. Syst. Man Cybern. Part C (Applications Rev., vol. 30, no. 4, pp. 451–462, 2000.

[9] M. S. Lauer, S. Alexe, C. E. Pothier Snader, E. H. Blackstone, H. Ishwaran, and P. L. Hammer, "Use of the logical analysis of data method for assessing long-term mortality risk after exercise electrocardiography," Circulation, vol. 106, no. 6, pp. 685–690, 2002.

[10] S. Alexe, E. Blackstone, P. L. Hammer, H. Ishwaran, M. S. Lauer, and C. E. Pothier Snader, "Coronary Risk Prediction by Logical Analysis of Data," Ann. Oper. Res., vol. 119, no. 1–4, 2003.

[11] A. Reddy, H. Wang, H. Yu, T. O. Bonates, V. Gulabani, J. Azok, G. Hoehn, P. L. Hammer, A. E. Baird, and K. C. Li, "Logical Analysis of Data (LAD) model for the early diagnosis of acute ischemic stroke," BMC Med. Inform. Decis. Mak., vol. 8, no. 1, p. 30, 2008.

[12] S. A. Brooks, A. R. Brannon, J. S. Parker, J. C. Fisher, O. Sen, M. W. Kattan, A. A. Hakimi, J. J. Hsieh, T. K. Choueiri, P. Tamboli, J. K. Maranchie, P. Hinds, C. R. Miller, M. E. Nielsen, and W. K. Rathmell, "ClearCode34: A prognostic risk predictor for localized clear cell renal cell carcinoma," Eur. Urol., vol. 66, no. 1, pp. 77–84, 2014.

[13] A. Abd-Elhamed, Y. Shaban, and S. N. Mahmoudi, "Predicting Dynamic Response of Structures under Earthquake Loads Using Logical Analysis of Data," Buildings, vol. 8, no. 4, p. 61, 2018.

[14] M.-A. Mortada, T. Carroll III, S. Yacout, and A. Lakis, "Rogue components: Their effect and control using logical analysis of data," J. Intell. Manuf., vol. 23, no. 2, pp. 289–302, 2012.

[15] J. F. Avila-Herrera and M. M. Subasi, "Logical analysis of multi-class data," in Proceedings - 2015 41st Latin American Computing Conference, CLEI 2015, 2015.

[16] P. L. Hammer, "Logical Analysis of Data: From Combinatorial Optimization to Medical Applications," vol. 08854, pp. 1–18.

[17] P. L. Hammer, A. Kogan, and M. A. Lejeune, "Modeling country risk ratings using partial orders," Eur. J. Oper. Res., vol. 175, no. 2, pp. 836–859, 2006.

[18] Y. Crama, P. L. Hammer, and T. Ibaraki, "Cause-effect relationships and partially defined Boolean functions," Ann. Oper. Res., vol. 16, no. 1, pp. 299–325, 1988.

[19] R. Bruni, "Reformulation of the support set selection problem in the logical analysis of data," Ann. Oper. Res., vol. 150, no. 1, pp. 79–92, 2007.

[20] E. Boros, P. L. Hammer, T. Ibaraki, and A. Kogan, "Logical analysis of numerical data," Math. Program., vol. 79, no. 1–3, pp. 163–190, 1997.

[21] E. Boros, T. Ibaraki, and K. Makino, "Logical analysis of binary data with missing bits," Artif. Intell., vol. 107, no. 2, pp. 219–263, 1999.

[22] G. Alexe and P. L. Hammer, "Spanned patterns for the logical analysis of data," vol. 154, pp. 1039–1049, 2006.

[23] R. M. Sotnezov, "Genetic algorithms for problems of logical data analysis in discrete optimization and image recognition," Pattern Recognit. Image Anal., vol. 19, no. 3, pp. 469–477, 2009.

[24] M. Anthony and J. Ratsaby, "Robust cutpoints in the logical analysis of numerical data," Discret. Appl. Math., vol. 160, no. 4–5, pp. 355–364, 2012.

[25] J. Felix and A. Herrera, "R utcor R esearch R eport A New Approach to Select Significant Patterns in Logical Analysis of Data A New Approach to Select Significant Patterns in Logical Analysis of Data," 2012.

[26] A. Ghasemi, "Optimal Replacement using Logical Analysis of Data (LAD) and Dynamic Programming," p. 2015, 2015.

[27] C. Guo and H. S. Ryoo, "Compact MILP models for optimal and Pareto-optimal LAD patterns," Discret. Appl. Math., vol. 160, no. 16–17, pp. 2339–2348, 2012.

[28] K. Yan and H. S. Ryoo, "0-1 multilinear programming as a unifying theory for LAD," Discret. Appl. Math., vol. 218, pp. 21–39, 2017.

[29] S. Alexe and P. L. Hammer, "Accelerated algorithm for pattern detection in logical analysis of data," vol. 154, pp. 1050–1063, 2006.

[30] G. Alexe, S. Alexe, P. L. Hammer, and A. Kogan, "Comprehensive vs. comprehensible classifiers in logical analysis of data," Discret. Appl. Math., vol. 156, no. 6, pp. 870–882, 2008.

Authors -

Abdulkareem Owehan Alresheedi: Currently is a Master student at Qassim University in Computer college, Computer Science department. Alresheedi received his Bachelor degree in computer Science in 2004. From 2004 until now, I joined the public sector as the manager of the Information Technology management. My research interests lie in data mining, machine learning algorithms, advance optimization techniques and big data.

Mohammed Abdullah Al-Hagery: received his B.Sc in Computer Science from the University of Technology in Baghdad Iraq-1994. He got his MSc in Computer Science from the University of Science and Technology Yemen-1998. Al-Hagery finished his PhD in Computer Science and in Information Techonlogy, (Software Engineering) from the Faculty of Computer Science and IT, University of Putra Malaysia (UPM), 2004. He was a head of the Computer Science Department at the college of Science and Engineering, USTY, Sana'a from 2004 to 2007. From 2007 to this date, he is a staff member at the Faculty of Computer, Department of Computer Science, Qassim University in KSA. He published more than 15 papers in international journals. Dr Al-Hagery was appointed a head of the Research Centre at the Computer College, Qassim University, KSA from September 2012 to October 2018.