

Implementation of Association Rule with Modified Genetic Algorithm for Loan Approval in Banking

Sekhar Kuchi¹ Mrs. I. Madhavi Latha²

¹Student ²Assistant Professor

^{1,2}Department of Computer Applications

^{1,2}KMM Institute of PG Studies, Tirupati, India

Abstract— In the banking sector one of the crucial task is approval of loan. Depending on the customer’s credit card score and report, employment status, EMI income ration, payment history and lot of activities are involved in the approval process of the loan. For detecting who is genuine or who is not correct person to get the loan is not a simple tasks for bank peoples. In this process it consumes lot of time and money to verify for the genuinity of the loan seekers. In this paper it provides the solution for the problems in loan approval processing. By applying association rule along with genetic algorithm, it gives high accurate data about the loan seeker. It reduce the consuming of time and in the short time bank peoples can approve the loan. By implementing genetic algorithm in association rule, it gives high accuracy of data about the loan seeker. So that it reduce effort of the bank peoples as a result loan can approved with in the stipulated time.

Key words: Association Rule, Modified Genetic Algorithm, Frequent Patterns

I. INTRODUCTION

A. Data Mining (DM):

In the concept of Big Data (BD), Data Mining (DM) is a technology by which knowledge Discover (KD) is processed by DM tasks such as Summarization, classification, association and statistical techniques so on. Data Mining is a intersection of machine learning, Artificial Intelligence (AI), database systems.

B. Association Rules:

Association Rules was introduced in the year 1993. It is one of the task in Data Mining Process. In the Association rule it follows the Apriori algorithm. On the item sets apriori algorithm is used to find out frequent items (Association rule). Association rule is a expression of $A \Rightarrow C$, where ‘A’ and ‘C’ are the item sets and antecedent, consequent respectively. The meaning of the expression is, in the Dataset (D) it contains N fields (assume) and for instance T be the transaction in the data set then C (item) consists high probability when the item A is purchased. This probability is also considered as “confidence”. This confidence is represents in the transaction (T). In the mathematical expression it can be write as $p \{(C \text{ belongs to } T) / (A \text{ belongs to } T)\}$. One of the best real time examples is present market scenario that is butter have the highest probability to purchase when the combination of product of (milk and bread) purchased.

C. Genetic Algorithm (GA):

Genetic Algorithm (GA) was proposed by “John Holland” in the year 1960 by implementing the concept of “Drawin’s theory of Evolution. After that, his student Goldberg

enhanced genetic algorithm in the year 1989. Genetic Algorithm it is a tool to solve various optimization problems like integer nonlinear problems. It is frequently utilized for finding better optimal solution for all combination of problems. The functionality of the GA depends on three operators.

By converting each and every chromosomes in the form of binary format (0 and 1), next generation will made. It is a activity of random selection of chromosomes.

- 1) Selection
- 2) Crossover
- 3) Mutation

1) Selection:

Selection is a process of finding next generation (population) through the “ Fitness-function”. By the fitness function it gives more fitted generation (population) among all the population. The main purpose of the fitness function is to find quality chromosomes (population) so that it contributes to reproduce new chromosomes . Fitness function rides by the Cross over and mutation operators.

2) Cross Over:

In the cross over, reproduction of chromosomes will be made on Cross Over that is by changing random bit (s) as it produces new population.

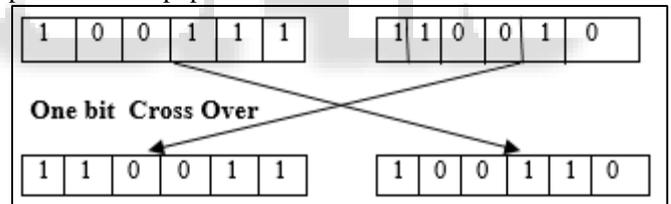


Fig. 1: cross over operator

3) Mutation:

In the mutation process randomly bit (s) will select and swapped that selected bit into new population. Mutation process may be one or two any size of bit. Mutation is applied on the crossed over population. Example for mutation as follows.

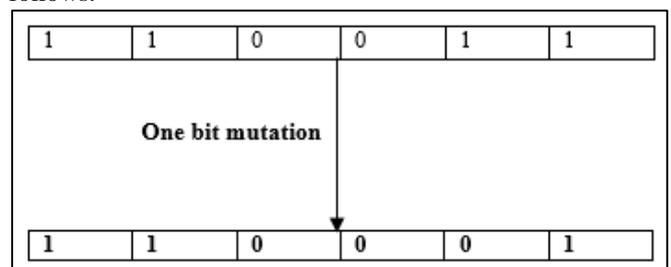


Fig. 2: Mutation operator

D. Association Rule with Genetic Algorithm:

The main idea behind association rule with genetic algorithm is to fine best predictive frequent patterns in the Data Mining by using apriori approach. To reduce the complexity in

finding of association rule, genetic algorithm is used. Genetic Algorithm replaces the methods of finding support count and confidence with the 'confusion matrix'.

Key aspect in the genetic algorithm is to represent the rules. Mainly two approaches are there in GA that is Michigan and Pittsburgh. How the rules are encoded in the population (chromosomes) is depends on the two approaches (Michigan and Pittsburgh). When compare Michigan, Pittsburgh approach have more complexity to find fitness computation. If want apply Pittsburgh approach on the individual (chromosomes) then it need some requirements to modify genetic operators. So that in this paper we apply Michigan approach, in this, individuals are simpler and syntactically shorter. This tends to reduce the time taken to compute the fittest chromosome by roulette wheel method. And it is simple to design the genetic operators. In this approach fitness function evaluate the quality of each individual separately. Encoding can be done in mainly two ways i.e, binary encoding or expression encoding for example let consider a rule

"If milk and bread then butter"

When applying Michigan approach, binary encoding to represent the rule as follows as 00 01 01 01 11 01 ,where bold numbers indicates id of the product in binary form like 00 for milk, 01 for bread, 11 for butter and bits (01 or 00) for appearance or absents of the product. Now based on the encoded bits further computation follows that is selection, cross over, mutation.

II. LITERATURE SURVEY

In this paper, the concept of combining association rule with genetic algorithm has inherited from the research paper "Performance based Association rule-mining technique using genetic algorithm", proposed by Frarah hanna, , Al-Zawaidhan[1] in that research paper not considered negative occurrence of attribute (item) in the transaction but by applying genetic algorithm it gives good results even present of negative occurrence.

In this paper tried to apply genetic algorithm on the association rule for optimizing the frequent item sets it also consider negative occurrence of the item [2]

IS Dehuri, A K jagadev, A Ghosh and R. Mall[3] also proposed genetic algorithm with association rule with minor changed in algorithm, that concept also imported for research present paper . In the paper' optimization of association rule its simplifies the calculation of finding fitness of the function by using confusion matrix.

Multi object evolutionary algorithm along with genetic algorithm proposed by the Ruplali Haldulakuar and prof.jitendra Agrawal [4], in that authors explained how encode and decode the chromosomes at initial population as well as reproduced chromosomes. Author simplifies how chromosomes are reproduce with the parameters such as total positive, false negative and so on. Author proposed association rule mining approach that can efficiently discovered particularly in the association rule mining in large data sets.

Surjay kumar sudhanshu and ayush kumar and ghose m.k [5]., in their thesis explained detail on how to implement genetic algorithm in Apriori algorithm. The

proposed various simple steps for implement genetic algorithm

III. METHODOLOGY

In the proposed methodology we apply genetic algorithm to find frequent patterns that are generated by the apriori algorithm. By considering the complexity in genetic algorithm, little bit of modification are imposed in the genetic algorithm (modified genetic algorithm) i.e., modify the formula to find the fitness function for optimize fit chromosomes. And there is restriction in the selection of consequent (that should be 1). The above two major modification are applied on the frequent item sets to generate the association rule.

A. Association Rule mining (Apriori)

From set of transaction from the database all the frequent items are generated by the Apriori algorithm. For instance, the following transactions that are frequently appeared in the transaction as follows

```
IF Mathematics AND Stats THEN computers.  
IF Mathematics AND physics THEN chemistry.  
IF Zoology AND physics THEN chemistry.
```

The above patterns explain, computer subject has the highest Probability when a student selects the combination of the subject mathematics and stats. Similarly chemistry subject has the highest probability when student select combination subjects (mathematics and physics) or (zoology and physics).By applying the modified genetic algorithm we can find most frequent association rule.

B. Genetic Algorithm with Modification

Each individual of pattern is represented in the binary form by applying Michigan's approach, i.e., each individual encodes single rule.

1) Representing the Rule for Antecedent

If there are 'n' products then we allocate (log n/log 2) size of bits for every pattern. One extra bit is assign for every subject so that it indicates the absence or present of the subject in the transaction. By applying this binary representation on the example that as follows.

00 1 10 1 10 1 11 0

Here two bit representation indicates the code of the subject and single bit (0 and 1) indicates the appearance and presence of the subject in transaction i.e., 00 for mathematics, 10 for physics, 10 for computers, 11 for chemistry and so on.

2) Genetic operators

For selecting fittest chromosome we apply roulette wheel sampling method in this phase crossover and mutation operators are selected based on their fitness (quality). If a chromosome (candidate) has more fitness then it will be more chance to be selected. The implementation of roulette wheel sampling is done by normalizing the candidate values so that probability lies between (0 and 1).

On the selected chromosomes (candidates) cross - over and mutation operators applied for reproducing the next generation chromosomes. For finding the support count and confidence, confusion matrix approach is used instead of candidate table (as in Apriori algorithm).

The main motive for using confusion matrix is, there is a possibility to occur over fitness of the function in a genetic algorithm. So that to control and reduce that kind of complexity, confusion matrix is used. The structure of confusion matrix in 2X2 format and it mainly contains 4 parameters i.e., TP, FP, FN, TN

3) *Confusion Matrix:*

Confusion matrix is matrix (2X2) for finding predicted class (consequent C) on the basis of conjunction of condition (antecedent A). The representation of the confusion matrix as follows

		Active class	
		C	not C
Predicted class	C	TP	FP
	Not C	FN	FP

Fig. 3: confusion matrix

The label in each cell has the following meaning

TP = True Positive = number of example satisfies A and C

FP = False Positive = number of examples satisfies A but not C

FN = False negative = number of examples not satisfies A but C

TN = True Negative = number of examples neither satisfies A nor C

IV. PROPOSED ALGORITHM

Mainly two modifications are implemented in the genetic algorithm as follows,

- 1) Step 1 : Start
- 2) Step 2 : load sample transactions from the database into the Memory
- 3) Step 3 : Apply Apriori Algorithm on the dataset to find Frequent with minimum support. For instance set A Be frequent item set.
- 4) Step 4 : For instance, set B (=0) the associate rule set Generated by genetic algorithm.
- 5) Step 5 : Input the terminate condition of genetic Algorithm
- 6) Step 6: Represent binary code for each item in the set A By Using $\log(n)/\log(2)$ for length of bits
- 7) Step 7 : Select the two item from Set A and apply Roulette Wheel sampling for selection of members
- 8) Step 8 : Apply cross-over and mutation operators on Selected Items (from Step 6).
- 9) Step 9 : Find the fitness function by using confusion matrix
 - 1) Step 9.1: calculate Confidence Factor (CF) $CF = TP / (TP + FP)$
 - 2) Step 9.2: calculate Completeness Measure (CM) $CM = TP / (TP + FN)$
 - 3) Step 9.3: find the fitness function for each rule

$$Fitness = w1 * (CP * CM) + w2 * samp // \text{(modified)}$$

- 4) Step 9.4: check $A \rightarrow C$ condition by step 10
- 10) Step 10: if (fitness > minimum confidence)
 - Set $B = B \cup \{A \rightarrow C\}$
 - 1) Step 10.1: otherwise go to Step3 for frequent items
- 11) Step 11 : Stop

Fig. 1: submitting of Loan seeker details in binary Form Covering each and every transaction in the form of binary format (0 and 1) and enter that data in the application

Fig 2: Applying modified genetic algorithm on data set When click on the generate DB button it generates various possibilities of the patterns in the form of binary numbers (0 and 1)

Fig 3: Applying Apriori algorithm By clicking on the button Apriori it will generates Which transaction have highest possibility for selecting that transaction .In the table minimum support and confidence are given by the bank employee.

V. CONCLUSION & FUTURE SCOPE

In this paper, modified genetic algorithm implemented with association rule techniques for find out frequent patterns by applying on loan seeker details. Since there some limitations are involved in this modified genetic algorithm in term of calculation process and it is also difficult to implement tournament method instead of roulette wheel sampling method.

REFERENCES

[1] Frarah hanna, , Al-Zawaidhan, yousef hasan jabar and Marwan Al-Abed Abu-Zanona, "An n improved algorithm for mining Association rule in large datasets", world of computers science and information technology journal (WCSIT) ISSN: 2221-0741 vol.1,no.7,2011,pp.311-316

- [2] Action of interesting Association rule using genetic Algorithms” Advance in System Modeling and ICT Applications,pp.101-110
- [3] M.Ramesh kumar and Dr.k. lyakutti,”IJCA Special Issues on “Artificial Intelligence techniques-Novel Approaches & practical Applications” AIT,2011,pp.35-38
- [4] IS Dehuri,A K jagadev, A Ghosh and R. Mall, “multi-object genetic Algorithm for Association rule mining using a homogenous Dedicated Cluster of WorkStations America Journal of Applied Science 3 (11),2006,pp.2086-2095
- [5] Ruplali Haldulakuar and prof. jitendra Agrawal, “Optimization of Association rule Minimum through genetic Algorithm”, international journal of computer science and Engineering (IJCSE),vol.3 no. 3 mar 2011,pp.1252-1259
- [6] Ansaf salleb-Aouissi, Cheristel Varain and Cryil Nortet, “QuantMiner: A Genetic Algorithm for mining Quantitative Association Rules”,IJCAI- 2007,PP 1035-1040.

