Extraction of Association Rules Using Chemical Reaction Optimization

Sudhir Kumar Mohapatra, Bimal Prasad Kar, Befkadu Belete, Tarini Prasad Panigrahy

Abstract: This paper explores the applicability of chemical reaction optimization in association rule mining. We apply CRO on transactional database. Our algorithm generates N number of rules from the given database. The proposed algorithm is tested on real-life data from friendship mall, Addis Ababa, Ethiopia. From the results, we find it to be the best alternative to the existing popular algorithm like apriori algorithm and the FP-growth algorithm.

Keywords: CRO, Association rule mining, Apriori, FP-growth, Chemical Reaction optimizations.

I. INTRODUCTION

One of the Market basket analysis is association rule mining. In association rule min-ing, we analyze the purchasing habit of the customer. This analysis is processed by discovering association among the different items placed in the shopping basket of the customer. The analysis is about to find frequently purchase items together. The process of chemical reaction is used in Chemical Reaction Optimization (CRO) [7]. This recent metaheuristic is applied to solve many combinatorial problems successfully. In many problems, CRO outperforms many existing metaheuristics in most of the test cases. We propose an algorithm using CRO to solve association rule mining. The propose CRO approach is tested on books dataset and different algorithms, namely the Apriori algorithm and FP-growth algorithm.

II. RELATED WORK

To infer knowledge from the database or big data, we require data mining [1, 2, 3]. Using the customer’s market basket business analysis is possible. This market basket of the customer helps in future business predictions. One of the uses of data mining on these type of data is association rule discovering. Association rule mining is one of the methods in an unsupervised system known as pattern recognition. This is a method of searching undiscovered rules in the database. Discovery of huge rules may sometimes give negative potential because of lack of proper analysis.

Use of an evolutionary algorithm for association rule mining is first done by Sag-gar et al.[4]. He uses a genetic algorithm(GA) to optimize the rule generated by the Apriori algorithm. Waiswa and Baryamureeba[5], mine the association rules based on GA method. For this, they used the Pareto-based multi-objective evolutionary algorithm. The concept of a fuzzy association rule is used by Kay and Alhajj [2]. To extract semi optimal rules they proposed multi-objective. The use of GA for extrac-tion of the rule having a negative attribute is proposed by Anandhavalli et al. [8]. Cluster-based multi-objective GA is present by Hadian et al.[9]. Some of the researchers also tried Particle swarm optimization(PSO) for extracting association rules[6]. Kuo et al.[10] is among the first researcher to use PSO for association rules mining. The concept of weighted particle swarm optimization is used by Gupta[11]. Asadi et al.[12] use PSO to derive the threshold value for a priory algo-rithm. Nandhini et al.[13] use PSO and domain ontology for extraction of the association rule.

III. ASSOCIATION RULE MINING USING CRO

The algorithm starts with the initialization of CRO parameters PopSize, KElossRate, MoleColl, buffer, InitialKE, α and β. Set of solutions(Association Rules) are also generated at the beginning which is present as a set of molecule represented as M. In the next different elementary reaction of CRO is applied. We will discuss them one by one. Finally, our algorithm gives the best rules. We have to execute the algorithm for N number of times to get N dif-ferent rules. Now we will discuss the algorithm, initial population, and operators of CRO.

3.1. CRO Algorithm

M←set of molecules
AR← association rules
Begin
Initialize PopSize, KElossRate, MoleColl, buffer, InitialKE, α and β in the initial stage.
M←GenMole(PopSize)
for M≠ɸ do
calculate PE for each molecule and set InitialKE for each molecule
Initialize value for KE, NumHit, MinPE, MinHit and Minstruct
end for
AR←assign random molecule from M
while(PE of molecule remain constant for successive iteration) generate a random number b<0,1
if b>MoleColl
m1←select a molecule randomly from M
if (number of hits - minimum hit number) > α
Apply decomposition(m1)
Else
Apply On-wall-Ineffective-Collision(m1)
end if
end while
End Apply On-wall-Ineffective-Collision(m1)
else
endif

Revised Manuscript Received on January 25, 2019.

Sudhir Kumar Mohapatra, Dept of CSE,IT & SE, AASTU, Addis Ababa, Ethiopia
Bimal Prasad Kar, Dept of CSE,GITA, Bhubaneswar, Odisha, India
Befkadu Belete, Dept of CSE,IT & SE, AASTU, Addis Ababa, Ethiopia
Tarini Prasad Panigrahy, Dept of CSE,GITA, Bhubaneswar, Odisha, India

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication
m1, m2 ← select two molecules randomly from M
if KE ≤ β
    Synthesis(m1, m2)
else
    Inter-Molecular Ineffective Collision(m1, m2)
endif
update AR
end
end

3.2. Molecule
The function GenMole(PopSize) in the algorithm is responsible for initial population or molecules. These molecules represent association rules. Two approaches are available to represent association rules. The first approach represents a set of rules in a single molecule called as Pittsburgh approach, and another approach is the Michigan approach, where a single molecule represents a separate rule. Here we followed the Michigan approach. We represent each molecule by a vector. The size of the vector is equal to l where l equals to number of the item. The entries in the vector will be anyone from {0, 1, 2}. The ith entry in the molecule w will follow the following rules.

For example let, us we have 5 item set as {I1, I3, I3, I4, I5} and one rule like I1I4 → I2I5, then the molecule representing this rule is represented in figure 1.

Figure 1: Molecule representing a single rule

3.3. Potential Energy
The potential energy will decide the potential of each molecule in CRO. It is like fitness function in GA. We use the fitness function proposed by Sarath et al. [6]. They justify the use of the following fitness function

\[ \text{Fitness} = \text{support}(A \rightarrow B) \times \text{confidence}(A \rightarrow B) \] (1)

where A → B are association rule.

The author uses this because he wants to maximize. In CRO implementation as it became minimization problem, so we have to represent PE as

\[ \text{PE} = 1 - \text{support}(A \rightarrow B) \times \text{confidence}(A \rightarrow B) \] (2)

where A → B are association rule.

3.4. On-Wall Ineffective Collision
This operator of CRO is like the mutation operator of the genetic algorithm. In the randomly selected molecule, the value of 1 or 2 is replaced with 0. The reason behind the construction of this operator is inspired by a property of support i.e. \( \forall A, B \subseteq I \text{Sup}(A) \geq \text{Sup}(A, B) \).

3.5. Decomposition
This process generates two new molecules from a single molecule. The reason behind it is to increase the search space. The decomposition operator for the association rules mining is designed as fol-low. The source molecule is divided into two from a ran-domly selected point and the vacant space in the newly produce molecule is filled up with any of the value from \{0, 1, 2\}.

3.6. Synthesis Operator
The operator combines two molecules into one molecule with a solution. It decides a point, from the first molecule first portion is copied to the new molecule and rest from the second molecule.

3.7. Inter-Molecular Ineffective Collision Operator
This operator produces new two molecules from the old two molecules. It is something similar to the crossover operator of GA.

IV. EXPERIMENTAL RESULTS
In order to implement the CRO based association algorithm, we use Matlab version R2017a. The experiments are conducted on an HP computer with the windows10 operating system. The architecture of the machine is Intel Core-i7 processor, 2.4 GHz, and 4 GB memory. Different CRO parameter used in our implementation is present in table I.

<table>
<thead>
<tr>
<th>M (Number of Molecule)</th>
<th>KELoss Rate</th>
<th>Molecule Collision buffer</th>
<th>InitialKE</th>
<th>α</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>0.2</td>
<td>0.2</td>
<td>0</td>
<td>1000</td>
<td>500</td>
</tr>
</tbody>
</table>

Table I: CRO Parameter For Association Rule Mining

Apriori algorithm and FP-Growth algorithm are used for result comparisons. The books dataset taken from XLMINER tool (www.solver.com/xlminer-data-mining) containing 11 items and 2000 transactions purchased by customers is used for implementation.

The rule generated by our algorithm is presented in table number II The corresponding potential energy is also represented in the last column.
Table II: Association Rules Generated By CRO Algorithm

Table number III and IV represents the rules generated by Apriori algorithm and FP-Growth algorithm on the same book dataset.

Table III: Association Rules Generated By Apriori Algorithm

Table IV: Association Rules Generated By Fp-Growth Algorithm

By analyzing these tables, it is visible that many rules from Apriori and FP-growth are present in our CRO based algorithm. It indicates that this CRO based association rule mining algorithm can be successfully used as alternatives to this two algorithm.

V. CONCLUSION

In this paper, we propose an algorithm based on CRO to solve association rule mining. The main aim is to extract rules from the dataset. To validate our approach we carried out experiment on book dataset. The results are compared with the Apriori algorithm and the FP-growth algorithm. Our approach gives a better result in terms of efficiency and performance. This algorithm can be used as a better alternative to existing traditional methods.

REFERENCES


