An Efficient Association Rule Mining Algorithm Based on Animal Migration Optimization Processing of Unknown Incidents in Crime Analysis Branch

Jagadeesh K R\(^1\) and Kumanan T\(^2\)

\(^1\) Department of Computer Science, Meenakshi Academy of Higher Education and Research, Chennai, India.
\(^2\) Faculty of Engineering and Technology, Meenakshi Academy of Higher Education and Research, Chennai, India

\(^1\)jaganbeema@gmail.com, \(^2\)kumananvetri@gmail.com

Abstract. A new protection scheme was proposed to avoid this problem. It is intended that each node uses a salted and a not salted HLL. If their estimates differ considerably, an attacker attempts to manipulate the estimates of HLL. In addition to avoiding manipulation, the proposed salted and unsalted (SNS) regime can also detect attempts at manipulation. A practical configuration showing how manipulation attempts can be detected in a low false positive probability has been shown to be applicable to this SNS scheme. Therefore, if merge ability is to be preserved it can be an interesting approach to protect HLLs from avoidance. In this paper the proposal for a new mining algorithm based on Animal Migration Optimization is made to decrease the number of Association Rules called ARM-AMO. The idea is to remove from the data rules which are not highly supportive and unnecessary. First of all, common item sets and association rules are generated with an Aprior algorithm. AMO also reduces the number of association rules incorporated in a new fitness function. In here, we provide a well-organized mechanism for incident derivation under the unwanted incident. This mechanism very useful for measure the heavy load of an incoming incident and exact calculation of the probability. In additional method is a Select-ability mechanism, which performs an important responsibility in incident derivation under the unwanted incident in both the settled and the unknown incident. A model for signifying derivative incident introduced jointly with an Advanced Sampling Technique that come close to the derived incident probabilities. This augmentation executed the prioritization techniques. In this prioritization techniques, recognize such cases in which the order of incident finding is strong-minded and mechanism for the definition of a settled detection execution.

Keywords: Unknown incident prediction, crime analysis, prioritization, computation overhead, rule reduction.

1. Introduction

Although crimes are a thought with different definitions, the definition of crime, generally accepted, consists of actions beyond the moral values of society which are fated by society as well as which require lawful penalty from administration [1]. In studying various scientific fields such as sociology,
psychology and criminology, the concept of crime is considered [2]. In order to address crime and detect offenders, furthermore, the advantages of technology as well as the analysis of assistance of informatics is important. Due to its increasing complicated nature, new approaches to case analysis and the detailed identification of similarities among crime records are a major problem for criminology [3] as criminal records have become more complicated than before. Big data stacks which include before verified crimes can be investigated using big data approach and data mining techniques to disclose the linkages and the relationships between the characteristics. This offer significant advantages for offenders, resolving criminal cases as well as taking appropriate defences before illegal cases are taken.

Although an illegal case is generally instantaneous as well as unpredictable, a possible prevention exists in certain areas if corruptions of the same type are dedicated by offenders of the similar profile in the same manner [4]. Analysis of crimes may be used to identify a wide range of links and deductions on records of crime, including the characteristics of offenders, victims, characteristics of cases, weapons and type of behavior of crime, geographical distribution and causal relations Various studies on the analysis of crime and profiling of offenders [5] have been carried out. Neuronal reproduction network as well as social media analytics are used to analyze crime-based research on data mining techniques [6].

The main contribution of this research work is to detect the unwanted incidents happening on the scenario through association rule mining algorithm. In this research work crime analysis is considered where unwanted incidents are happening through database analysis. This is attained by generating the association rules using apriori algorithm and reducing the number of generated rules using animal migration algorithm. The performance assessment is carried out in the matlab to ensure the performance improvement in the proposed techniques.

The complete organization of the research effort is given as follows: In this section, thorough overview about the crime analysis, unwanted incidents happening in crime department and the need of association rules are given. In section 2, discussion of different related research works is given. In section 3, detailed discussion about the planned technique along with suitable examples as well as diagrams. In section 4, performance assessment of the research work is mentioned. Lastly in section 5, complete conclusion of the research work is given.

2. Related Works
Sudhakar in the process of unifying crime data, Chau et al. [7] have industrialized a new method based on police reports as well as dictionary search, rules, statistics as well as machine learning. Smearing the inputs from the extraction model built on the artificial neural network, the report attempted to recognize the usable objects.

Lin and Brown's research [8] seeks identify criminals by using rules on the crime of theft. This study uses the Apriori algorithm as well as association rules are found to produce fine results on data from crime. Their study shows that the true rate of prediction increases according to the increase in data in the dataset.

Kaza et al. [9] developed the predicting illegal relationship algorithms that used to predict automatically the buses that are a co-offender to prevent the upcoming attacks. They used the lively social network analysis (SNA) methods as well as multivariate survival analysis.

Sharma [10] proposed an enhanced ID3 algorithm, an improved feature selection technique as well as an attribute-importance feature to classify e-mails as either maybe-suspicious or non-suspicious e-mails. Also they used a tool that is named as zero crime to help the system identify e-mails in relation to illegal actions.
Li et al. [11] has presented comparing the performance of the event ontology method as the priori information and the method based on Support Vector Machine to analyze the attributes as well as relations in web pages. Also these methods are used to reconstruct the scenario for crime mining.

Fogel et al. [12] proposed the use of changed neural networks and evolved rule-based classifiers. Both methods are useful to distinguish between toxic via narcotic and reactive mechanisms of action (MOAs) of small molecules.

Sathyadevan [13] anticipated the use of naive bayes algorithm through the concept of named entity recognition (NER), also recognized as entity or element removal, to classify the news trainings into the crime type and to create a crime model. In addition to that, Apriori algorithm is used to find and create frequent patterns in crime by training crime data from the different web sites. For prediction in crime, they used the decision tree concept.

Thongsatapornwatana and Chuenmanus [14] proposed the suspect vehicle recognition system using association rule to analyze the vehicles with forged license plate crossing the barrier that potentially involved the illegal activity.

3. **Unknown crime incident prediction using optimization procedure**

This work proposes to decrease the number of association instructions by proposing a new mining algorithm founded on the Animal Migration Optix (AM O) known as ARM-AMO. The idea is that rules that are not highly supportive as well as needless are removed from the data. The first thing to be done is to produce common item sets and association rules using Apriori algorithm.

3.1. **Apriori algorithm to generate frequent item sets plus association rules**

An algorithm to learn relational databases as well as rule. In 1994, Agrawal as well as Srikant planned the Apriori algorithm. Apriori is intended for use in transaction databases. Other procedures are planned to discovery association rules for data that do not have transactions or time marks (DNA sequencing). (Winepi and Minepi). A set of items (a set of items) are perceived each transaction. The Apriori algorithm uses a threshold C to identify the articles that constitute a subset of C transactions at least.

The applicant length k of the item k-1 sets is produced. The candidates with a pattern that is unfamiliar are then prunes. According to the lower end lemma, the candidates set comprise all the frequent k-length sets. The database is then scanned for common item sets between the applicants.

For an operation database T, in addition to for a provision threshold of μ, the pseudo code for the procedure is assumed below. The common set notation is used, but note that T is a multi-set. Ck is the level k applicant. The procedure is supposed to produce the applicant sets from the huge item sets of the previous level, which are responsible for the lower closure of the lemma. Count [c] accesses the field of the candidate's data structure c assumed to be zero in the beginning. Many details are omitted below, the data structure used for storing candidate sets and for counting their frequencies is typically the main part of implementation.

Apriori (T, $\mu$)

\[ L_1 \leftarrow \{\text{large 1 - itemsets}\} \]
\[ K \leftarrow 2 \]

While \( L_{k-1} \neq \emptyset \)

\[ C_k \leftarrow \{c = 1 \cup \{b\} | a \in L_{k-1} \land b \notin a, \{s \subseteq c|s| = k - 1 \subseteq L_{k-1}\}\} \]

For transaction \( t \in T \)

\[ D_t \leftarrow \{c \in C_k | c \subseteq t\} \]

For candidates \( c \in D_t \)

\[ \text{Count } [c] \leftarrow \text{count } [c] + 1 \]
This procedure follows three steps:
1. For l after 1 to I do
2. For each set JI such that for both h ∈ JI happens in at least k bags do
3. Inspect the data to determine whether the set JI occurs in at least k bags

Greatest of its period was spent in accessing the database for this algorithm pending it results in a common association match. On this basis, two quantities of the probability model were calculated:
1. Success rate: the possibility that the set is a success
2. Failure rate: the possibility that the set is a failure

With this possibility model, it brings out the main performance features of the procedure.

### 3.2. Animal migration optimization to reduce number of association rules

The above section will be resultant with the set of association rules for the crime incident patterns. These generated rules will be huge in numbers which will increase the computation overhead of crime pattern detection. It is resolved in this research work by reducing the number of rules generated considerably. This is achieved through the introduction of the method of optimizing animal migration. Relocation is a common wonder in ecology of animal behavior, which has been intensely studied in the animal kingdom. The migration is persistent and streamlined and influenced by the locomotive efforts of the animal to new habitations. It be contingent on an impermanent inhibition of the response of the station, but encourages its eventual disinhibition and recurrence. The relative prolonged movement of individuals is animal migration, usually on a cyclical basis. This is an omnipresent wonder found in all main groups of animals, such as birds, fish, mammals, reptiles, insects, amphibians plus crustaceans. Local weather, local supply of food as well as the season of the year can be the trigger for the relocation. The typical picture of relocation is that of northern landslides, like swallows plus birds of prey, which take long tropical journeys.

There are two processes for optimizing animal migration: migration and updating the population. At the beginning, the procedure simulates how the animal collections move. The person is referred to as a true coded vector of the D dimensions. Started that uses the vector of the NP D dimension parameter to the minimum as well as maximum limits prescribed, where NP indicates the population size:

\[ \bar{X}_{\text{min}} = \{x_{1,\text{min}}, x_{2,\text{min}}, ..., x_{D,\text{min}}\} \]

\[ \bar{X}_{\text{max}} = \{x_{1,\text{max}}, x_{2,\text{max}}, ..., x_{D,\text{max}}\} \]

Therefore, we can initialize the jth component of the ith vector as

\[ x_{i,j,0} = x_{j,\text{min}} + \text{rand}_{i,j}[0,1].(x_{j,\text{max}} - x_{j,\text{min}}) \]

Where, rand_{i,j}[0, 1] is a uniformly distribution random number between 0 as well as 1, i = 1, ..., NP and j = 1, ..., D.

The animal should follow three rules during the migration phase: (1) to escape crashes with its neighbours; (2) to move in the similar direction as its neighboring animals and (3) to stay near to the neighbours. For the main rule, we need to change each person's position in the group. We require the latter two rules that the person be placed in new positions according to his neighbours' current positions. Particular animals leave the collection as well as then new animals joint the new population during the population update process. We assume that the number of animals available is determined and that a new person with probability Pa will replace the animals. Depending on the fitness quality the probability is used. The Pa probability is 1. For best fitness. The possibility is 1/NP for the nastiest fitness.

**Algorithm**: Animal migration optimization algorithm
1. Begin  
2. Set the generation counter \( G=1 \); and randomly initialize a population of \( NP \) animal \( X_i \)  
3. Evaluate the fitness for each individual in \( P \)  
4. While stopping criteria is not satisfied do  
   5. For \( i=1 \) to \( NP \) do  
      6. For \( j=1 \) to \( D \) do  
         7. \( X_{i,G+1} = X_{i,G} + \delta \cdot (X_{\text{neighbourhood},G} - X_{i,G}) \)  
      8. End for  
   9. End for  
10. for \( i=1 \) to \( NP \) do  
11. Evaluate the offspring \( X_{i,G+1} \)  
12. If \( X_{i,G+1} \) is better than \( X_i \) then  
13. \( X_i = X_{i,G+1} \)  
14. End if  
15. End for  
16. for \( i = 1 \) to \( NP \)  
17. For \( j=1 \) to \( D \)  
18. Select randomly \( r1 \neq r2 \neq i \)  
19. If rand > \( P_a \) then  
20. \( X_{i,G+1} = X_{r1,G} + \text{rand} \cdot (X_{\text{best},G} - X_{i,G}) + \text{rand} \cdot (X_{r2,G} - X_{i,G}) \)  
21. End if  
22. End for  
23. End for  
24. For \( i=1 \) to \( NP \) do  
25. Evaluate the offspring \( X_{i,G+1} \)  
26. If \( X_{i,G+1} \) is better than \( X_i \) then  
27. \( X_i = X_{i,G+1} \)  
28. End if  
29. End for  
30. Memorize the best solution achieved so far  
31. End while  
32. End

By using the above algorithm number of rules generated will be reduced, thus the unwanted incidents can be predicted accurately with lesser computation overhead.

4. Results and discussion

In this segment, the proposed model of associative classifier is evaluated in MATLAB. The performance metrics that are considered in this research method for the estimation of the planned and existing research method are “Accuracy, Precision, Recall, F-measure, Number of rules and Error rate”.

**Accuracy:** Precisions are the most intuitive measurements of performance and are simply a ratio of accuracy to the total observations.

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}
\]

The graphical comparison representation for the above simulation values are illustrated in the following figure 2.
In figure 1, graphical comparison illustrated for the performance metrics accuracy between the proposed ARM-AMO method and the existing ARM method. From this comparison it can be proved that the planned method ARM-AMO is increased in its accuracy value which is 14.42% better than the existing ARM method.

**Precision:** Precision is the ratio of positive observations correctly anticipated to the total positive observations forecast. The low fake positive rate refers to high precision.

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

The graphical comparison representation for the above simulation values are illustrated in the following figure 2.

In figure 2, graphical comparison illustrated for the performance metrics precision between the proposed ARM-AMO method and the existing ARM method. From this comparison it can be proved that the planned method ARM-AMO is increased in its precision rate which is 80.74% better than the existing ARM method.
Recall: Remember the ratio of positive observations correctly predicted to those in all actual class observations – yes.

Recall = TP / TP + FN

The graphical comparison representation for the above simulation values are illustrated in the following figure 3.

![Figure 3. Recall comparison](image)

In figure 3, graphical comparison illustrated for the performance metrics recall between the proposed ARM-AMO method and the existing ARM method. From this comparison it can be proved that the planned method ARM-AMO is increased in its recall rate which is 40.83% better than the existing ARM method.

**F-measure:** F1 the correct and weighted average reminder value. This ranking thus takes false positives as well as false negatives into account.

\[
F1 \text{ Score} = \frac{2 \times (\text{Recall} \times \text{Precision})}{\text{Recall} + \text{Precision}}
\]

The graphical comparison representation for the above simulation values are illustrated in the following figure 4 and table representation shown in table 1.

| No of transactions | ARM value | ARM-AMO |
|--------------------|-----------|---------|
| 5                  | 55        | 78      |
| 10                 | 53        | 81      |
| 15                 | 52        | 70      |
| 20                 | 51        | 80      |

**Table 1. comparison of ARM and ARM-AMO value representation**
In figure 4, graphical comparison illustrated for the performance metrics F-Measure between the proposed ARM-AMO method and the existing ARM method. From this comparison it can be proved that the planned method ARM-AMO is increased in its F-Measure rate which is 10.81% better than the existing ARM method.

5. Conclusion
This work proposes to decrease the number of suggestion rules by proposing a new mining procedure based on the Animal Migration Optix (AMO) known as ARM-AMO. The idea is that rules that are not highly supportive and unnecessary are removed after the data. The first thing to be done is to produce common itemsets as well as association rules using Apriori algorithm. The performance assessment of this research has demonstrated that the technology proposed tends to show better results with an unwanted accurate incident detection rate than the existing method.

References
[1] Edelstein A 2016 Rethinking conceptual definitions of the criminal career and serial criminality Trauma, Violence, & Abuse 17 62-71.
[2] Feeley M 2017 The concept of laws in social science: A critique and notes on an expanded view In Crime, Law and Society 19-45.
[3] Sevri M, Karacan H and Akcayol MA 2017 Crime analysis based on association rules using apriori algorithm International Journal of Information and Electronics Engineering 7 99-102
[4] Adolphs R, Gläscher J and Tranel D 2018 Searching for the neural causes of criminal behavior Proc. of the National Academy of Sciences 115 451-2.
[5] Canter D and Alison L 2018 Profiling property crimes. In Profiling property crimes 11-40.
[6] Thongsatapornwatana U 2016 A survey of data mining techniques for analyzing crime patterns Asian Conf. on Defence Technology 123-128.
[7] Chau M, Xu JJ and Chen H 2002 Extracting meaningful entities from police narrative reports.
[8] Lin S and Brown DE 2006 An outlier-based data association method for linking criminal incidents Decision Support Systems 41 604-615.
[9] Kaza S, Hu D, Atabakhsh H and Chen H 2007 Predicting criminal relationships use multivariate survival analysis *Int. conf. on Digital government research: bridging disciplines & domains* 290-291.

[10] Sharma M 2014 Z-CRIME: A data mining tool for the detection of suspicious criminal activities based on decision tree *Int. Conf. on Data Mining and Intelligent Computing* 1-6.

[11] Li C, Hu Y and Zhong Z 2010 An event ontology construction approach to web crime mining *In 2010 Seventh Int. Conf. on Fuzzy Systems and Knowledge Discovery* 5 2441-2445.

[12] Fogel GB and Cheung M 2005 Derivation of quantitative structure-toxicity relationships for ecotoxicological effects of organic chemicals: evolving neural networks and evolving rules *IEEE Congress on Evolutionary Computation* 1 274-281.

[13] Sathyadevan S 2014 Crime analysis and prediction using data mining *Int. Conf. on Networks & Soft Computing* 406-412.

[14] Thongsatapornwatana U and Chuenmanus C 2015 Suspect vehicle detection using vehicle reputation with association analysis concept *In Tourism Informatics* 151-164.