ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 6, Issue 6, June 2017

FORECASTING CRIMINAL BEHAVIOR USING GA AND BCO TECHNIQUE IN BIG DATA

Mohini Arya, email id : mohinarya10@gmail.com , Amit Asthana ,SUBHARTI UNIVERSITY(U.P) ,INDIA

ABSTRACT

Through data processing of massive population crime, data mining algorithm is applied to the public security system. When dealing with large amount of data, Genetic and Bee colony algorithm optimization results are greatly influenced by the choice of initial grouping center. If the selection of initial data center is improper, the optimization results may be trapped in local optimal solution, and get better mining effect. So, adaptive genetic algorithm and BCO is used to optimize criminal behavior system. The experiment results show that the proposed scheme can get the valuable information and ideal mining effect.

Keywords-criminal behavior, BCO, genetic algorithm

I.INTRODUCTION

Nowadays, a lot of crimes in all countries of the world are expanding. The cost of living is rising, while the main reason unemployment is growing. In our daily life, we face Crime, an essential part of the hazards. Crime is going with harm & violence in public outlook; harm to individual, demolition of property and the rejection of reverence to people. Crime is not widening equally across maps [1]. It is of significance, that the incorporation of dissimilar statistics, such as Pearson's sample coefficient correlation, a coefficient of multiple correlations, and an incomplete correlation coefficient, improves viable benefits in the study of obscured situations and conclusion support [2]. From the early 20th century, Forecasting has been a significant part of criminal justice practice. Our contemplation is focused on the intelligence based on conceptual knowledge and reasonable logics of dissimilar kinds of statistical models associated with the assessment of a vast quantity of data that imitate criminal event and their correlations [3]. A large amount of the first work was carried out for parole forecast, which soon after developed into reoffending further commonly. A number of key methodological concerns were lifted in a way that is still insightful early in the post-war era, [4]. Police did not usually followed Crime forecasting. As there are many econometric studies of crime or integrating crime in the literature, one is hard-pressed to locate police departments or other police associations making standard use of forecasting econometric or predictive for consumption of restricted resources [5]. One of the causes that crime forecasting may have been judged to be not possible in the past is that the preferred scale for examination is too little for dependable model estimation [6]. Police must identify crime in regions as small as feasible for strategic purposes, at the patrol district level or smaller [7].

II. RELATED WORK

In [6] Malathi. A. et. al., they discuss that, a major challenge facing all law-enforcement and intelligence gathering organizations is accurately and efficiently analyzing the growing volumes of crime data. There has been an enormous increase in the crime in the recent past. they look at MV algorithm, DB Scan and PAM outlier detection algorithm with some enhancements to aid in the process of filling the missing value and identification of crime patterns. They applied these techniques to real crime data and use semi-supervised learning technique here for knowledge discovery from the crime records and to help increase the predictive accuracy.

In [7] Mande U. et. al., introduce binary clustering and classification techniques have been used to analyze the criminal data. The crime data considered in this paper is from Andhra Pradesh police department this paper aims to potentially identify a criminal based on the witness/clue at the crime spot an auto correlation model is further used to ratify the criminal.

In [8] Malathi. A et. al., they use a clustering/classify based model to anticipate crime trends. The data mining techniques are used to analyze the city crime data from Police Department. The results of this data mining could potentially be used to lessen and even prevent crime for the forth coming years.

In [9] Sathyaraj S. R. et. al., they studies to integrate a large volume of data sets into useful information by adopting a various information techniques in the hottest technology world. The adopted approaches of Single variate Association Rule for Crime to Crime based on the knowledge discovery techniques such as, clustering and association-rule mining. It reveals with an inherent of patterns of information into a fruitful exploratory tool for the discovery of spatio-temporal patterns. This tool is an autonomous pattern detector to reveal plausible cause-effect associations between layers of point and area data. Also present REPA, VBEA and TEA algorithms with an exploratory analysis for the effectively explore geo-referenced data. The present study of this paper was focuses through the real crime dataset by using various algorithms.

In [10] Mande U. et. al., they aims towards the construction of new methodologies based on Data mining concepts and serves as a decision support system. Given a set of available clues, from the forensic labs and the clues collected at the crime spot, a methodology is presented to map the evidence and identify a criminal.

In [11] Chen N. et. al., their model, firstly, we predict the residence of the offender based on the locations of the last crime scenes with three methods (distance analysis: the location that has the shortest distance to each crime site, circle fitting, probability theory); secondly, they predict the time of the next crime based on previous data with the method of fitting a straight line; next, predict the location of the next crime based on the locations of the last crime scenes and the time predicted in the second step with the method of weighted average; finally, generate a predicted location based on the three predicted locations with the method of weighted average.

In [12] Yu C. H. et. al., they discuss the preliminary results of a crime forecasting model developed in collaboration with the police department of a United States city in the Northeast. They first discuss approach to architecting datasets from original crime records. The datasets contain aggregated counts of crime and crime-related events categorized by the police department. The location and time of these events is embedded in the data. Additional spatial and temporal features are harvested from the raw data set. Second, an ensemble of data mining classification techniques is employed to perform the crime forecasting.

III. OVERVIEW OF CRIME

Crime occurs in a variety of forms which police informally categorizes as being either major or volume. Major crime consists of the high profile crimes such as murder, armed robbery. These crimes can either be one-offs or serial. Serial crimes are relatively easy to link crimes together due to clear similarities in terms of modus operandi or descriptions of offenders. This linking is possible due to the comparatively low volume of such crimes. Major crimes usually have a team of detectives allocated to conduct the investigation. In contrast volume crimes such as burglary and shoplifting are far more prevalent. They are usually serial in nature as offenders go on to commit many such crimes. Property crimes, such as domestic burglary offences, committed by different individuals are highly similar and it is rare to have a description of the offenders. The police officers or detectives use free text to record most of their observations that cannot be included in checkbox kind of pre-determined questions. While the first two categories of information are usually stored in the computer databases as numeric, character or date fields of table, the last one is often stored as free text. The challenge in data mining crime data often comes from the free text field. While free text fields can give the newspaper columnist, a great story line, converting them into data mining attributes is not always an easy job.

II.PROBLEM STATEMENT

Crime and violence are social perpetual problem and their impacts and consequences are devastating. Crime and violence are extremely detrimental to the moral order and relationships within society. Crime rates, either violent or property crime; are often used as a barometer in reflecting the safety level of a nation. The fluctuating stream of crime rate worldwide seen as public perplexing problem as it fosters) public fear, distrust, anger, and perceptual errors, and b) causes grief among family members and friends of the crime victim. Across the world, the horrific nature of crime has prompted in depth studies concerning the causes and factors that underlie criminal behavior. For these problems we proposed a new hybrid algorithm GABCO to find best or predicting criminal behavior.

IV. PROPOSED METHOD

A. Apriori algorithm

Apriori algorithm is the most classical and important algorithm for mining frequent itemsets This algorithm relies on generate and test approach and an important property: theApriori property. This

property is also known as anti-monotone property, and it is a basic pillar of the Apriori algorithm. It states that all non-empty subsets of a frequent itemset must be frequent. For example, it itemset 1,2,3 is a frequent itemset, then all of its subsets 1,2,3,1-2,2-3 and 1-3 must be frequent. In the other view, if an itemset is not frequent, then none of its supersets can be frequent. As a result, the list of potential frequent itemsets eventually gets smaller as mining progresses.

• Apriori property [14]: ―All nonempty subsets of a frequent itemset must also be frequent‖.
FREQUENT ITEMSETS WITH APRIORI

1. Apriori algorithm is –
-A basic algorithm for finding frequent itemsets for Boolean association rules
-Based on levelwise search iteratively find frequent itemsets with size from 1 to k (k-itemset)
2. Basic idea is to reduce the search space by using the Apriori principle:
- any subset of a frequent itemset must be frequent
- that is, if {AB} is a frequent itemset, both {A} and {B} should be frequent itemsets

B. Genetic algorithm

A genetic algorithm (or GA) is a search technique used in computing to find true or approximate solutions to optimization and search problems. Genetic algorithms are categorized as global search heuristics. Genetic algorithms are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover (also called recombination).
Algorithm works like this:
1) [Start] a population is created with a group of individuals created randomly. The individuals in the population are then evaluated.
2) [Caused Factors] The evaluation function gives the individuals a score based on how well they perform at the given task.
3) [New Population] Create a new population by repeating following steps until the new population is complete
4) [Selection] Two individuals are then selected based on their fitness, the higher the fitness, the higher and the chance of being selected.
5) [Crossover and Mutation] These individuals then "reproduce" to create one or more offspring, after which the offspring are mutated randomly.
6) [Accepting] Place new offspring in a new population
7) [Replace] Use new generated population for a further run of algorithm
8) [Test] If the end condition is satisfied, stop, and return the best solution in current population
9) [Loop] Go to step 2

B. Bee colony optimization

The Bee Colony Optimization algorithm, one of the Swarm Intelligence techniques, is a meta–heuristic method inspired by the foraging behavior of honeybees. It represents a general algorithmic framework applicable to various optimization problems in management, engineering, control, etc., and should always be tailored for a specific problem. The BCO method is based on the concept of cooperation, which increases the efficiency of artificial bees. BCO has the capability to intensify search in the promising regions of the solution space through information exchange and recruiting process. The diversification process is realized by restricting the search within different iterations.

C. Hybrid BCO

After the completion of a forward pass, each bee decides whether to stay loyal to the previously discovered solution or not. This decision depends on the quality of its own solution related to all other existing solutions. The probability that b-th bee (at the beginning of the new forward pass) is loyal to its previously generated partial/complete solution is expressed as follows:

$$p_b^{u+1} = e^{-\frac{O_{max}-O_b}{u}}, \quad b = 1, 2, \ldots, B \qquad (1)$$

Where:
Ob - denotes the normalized value for the objective function of partial/complete solution created by the b-th bee;

Omax - represents maximum over all normalized values of partial/complete solutions to be compared;

u - counts the forward passes (taking values 1, 2, . . . , NC).

The normalization is performed in two ways, depending on whether a minimization or maximization of the objective function is required. If Cb (b = 1, 2, . . . , B) denotes the objective function value of b-th bee partial/complete solution, normalized value of the Cb in the case of minimization is calculated as follows:

$$O_b = \frac{C_{max} - C_b}{C_{max} - C_{min}}, \quad b = 1, 2, \ldots, B \tag{2}$$

Where C_{min} and C_{max} are the values of partial/complete solutions related to minimal and maximal objective function value, respectively, obtained by all engaged bees. From equation (2), it could be seen that if b-th bee partial/complete solution is closer to maximal value of all obtained solutions, C_{max}, than its normalized value, Ob, is smaller and vice versa.

Initialization: Read problem data, parameter values (B and NC),
and stopping criterion.
Do
(1) Assign a(n) (empty) solution to each bee.
(2) For (i = 0; i< NC; i + +)
//forward pass
(a) For (b = 0; b < B; b + +)
For (s = 0;s< f(NC);s + +)//count moves
(i) Evaluate possible moves;
(ii) Choose one move using the roulette wheel;
//backward pass
(b) For (b = 0; b < B; b + +)
Evaluate the (partial/complete) solution of bee b;
(c) For (b = 0; b < B; b + +)
Loyalty decision for bee b;
(d) For (b = 0; b < B; b + +)
If (b is uncommitted), choose a recruiter by the roulette wheel.
(3) Evaluate all solutions and find the best one. Update xbest and f(xbest)
while stopping criterion is not satisfied.
return (xbest, f(xbest))

In the case of maximization criterion, normalized value of Cb is calculated as follows:

$$O_b = \frac{C_b - C_{max}}{C_{max} - C_{min}}, \quad b = 1, 2, \ldots, B \tag{3}$$

From equation (3), it is obvious that if the value of the partial/complete solution Cb, is higher, then its normalized value Ob, is larger, and vice versa. Using equation (1) and a random number generator, each artificial bee decides whether to become uncommitted follower or to continue exploring its own solution. If chosen random number is smaller than the calculated probability, then the bee stays loyal to its own solution. Otherwise, if the random number is greater than the probability p u+1 b , the bee becomes uncommitted. Let us discuss equation (1) in some more details. A greater Ob value corresponds to a better generated solution, and a higher probability of the bee staying loyal to the previously discovered solution. The higher index in the forward pass increases the influence of the already discovered solution. This is expressed by the term u in the denominator of the exponent (equation (1)). In other words, at the beginning of the search process bees are "braver" when searching the solution space. The more forward passes are made, the less courage they have: as we approach to the end of the search process, the bees are more focused on the already known solutions. Some other probability functions are examined in [31], indicating that alternative ways to determine loyalty should also is considered since for some combinatorial problems they may result in faster execution of the BCO algorithm.u6u67

V.RESULT

Crime rate is increasing at very fast speed, with that the data also becomes vast. So it is really difficult to handle such a big data manually. Therefore in this paper crime data set has been taken to determine the occurrence of crime. The sample dataset is shown below:

Mining large data set is an important issue to deal with as data is growing as the field grows. Today, crime rate is a menace that each country faces. With the increase in crime rate the data is increasing and it is such a critical field that accuracy is important at the same time. This paper shows the comparison in the results between clustering and the classification.

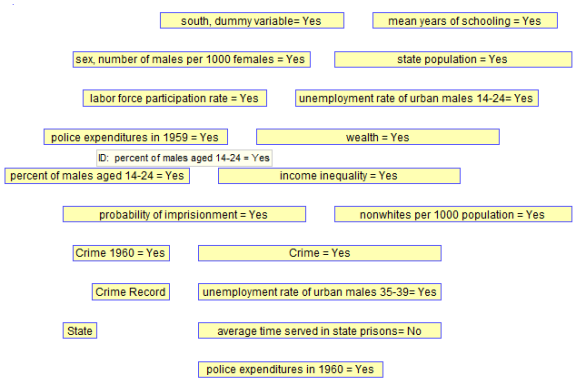

Figure 1: Crime detail with name and state

Itemset: {percent of males aged 14-24, south, income inequality, Ante : {percent of males aged 14-24, Sex}
Ante : {State Population, Wealth}
Conseq : { percent of males aged 14-24}
Conseq : {south, income inequality}

Rule : {percent of males aged 14-24, Sex} => {State Population}
Rule : {Wealth, percent of males aged 14-24} => {south, income inequality}

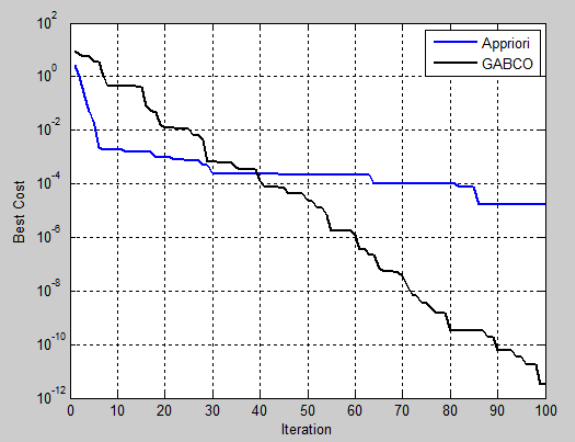

Figure 2: Comparison between Apriori and GABCO to Best cost

Thehybrid GABCO algorithm is more useful and faster in discovering association rules in a transactional database and it also reach the answer in less repetition. It obtains better answers in compare with other algorithms like Apriori. We have compared our model with Aprioriwith hybrid GABCO algorithm. As we can see in the figure 2, GABCO created better rules than Aprioriand Time consumption for creating model using Hybrid algorithms is very low although both Apriori and GABCO algorithms are successful in better situations, it also get less memory. According to this experiment, our method has a good performance in compare with existing algorithm.

VI.CONCLUSION

This research explores a new method to understand and analyze behavior of criminals. Several data of criminals was collected from genuine sources; the data is taken from spatial-econometricsfor its effective analysis. Using hybrid GABCOof criminal data set is used to reduce data so as to achieve efficient results; optimization is performed in the behavior analysis. We have achieved greater accuracy using Hybrid algorithm for comparison of Apriori algorithm with the stored data. Thus the objective of the research is achieved and the criminal behavior is analyzed.
In future the analysis can be done adding more parameters like crime patternbehavior.

REFERANCES

[1] Fagbaibi S.O, Ibrahim, Y. L and Longe O.B, "On the Use of Data Mining Techniques for Crime Profiling", Computing Information Systems &Development Informatics Journal, Vol. 3. No. 3, pp. 61-68, 2012

[2] Abraham, T. and de Vel O., "Investigative profiling with computer forensic log data and association rules", In Proceedings of IEEE International Conference on Data Mining, Edinburgh, SA, Australia, pp. 11-18, 2002.

[3] Giles Oatley, Brian Ewart and John Zeleznikow "Decision support systems for police: Lessons from the application of data mining techniques to "soft" forensic evidence", Journal of Artificial Intelligence and Law, Vol. 4, No. 1-2, pp. 35-100, 2006.

[4] Giles C Oatley and Brian W Ewart, "Crimes analysis software: 'pins in maps', clustering and Bayes net prediction", Expert Systems with Applications, Vol. 25, No. 4, pp. 569-588, 2003.

[5] Que ThiNguyet Nguyen, Philip A. Neck and Thanh Hai Nguyen "The Critical Role of Knowledge Management in Achieving and Sustaining Organizational Competitive Advantage", International Business Research, Vol. 2, No. 3, pp. 3-16, 2009.

[6]. Malathi. A and. Baboo S. S., "Enhanced Algorithms to Identify Change in Crime Patterns", International Journal of Combinatorial Optimization Problems and Informatics, Vol. 2, No.3, Sep-Dec, 2011, pp. 32-38, ISSN: 2007-1558.

[7]. Mande U., Srinivas Y. and Murthy J.V.R., "An Intelligent Analysis Of Crime Data Using Data Mining & Auto Correlation Models", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 4, July-August 2012, pp.149-153 149 | P a g e

[8]. Malathi. A ,Baboo S. S. and Anbarasi A., " An intelligent Analysis of a City Crime Data Using Data Mining", 2011 International Conference on Information and Electronics Engineering, IPCSIT vol.6 (2011) © (2011) IACSIT Press, Singapore.

[9]. Sathyaraj S. R., Thangavelu A., Balasubramanian S., Sridhar R., Chandran M. and Prashanthi M. D., "Clustered Spatial Association Rule To Explore Large Volumes Of Georeferenced Crime To Crime

Data", 3rd International Conference On Cartography And Gis 15-20 June, 2010, Nessebar, Bulgaria.

[10]. Mande U., Srinivas Y. and Murthy J.V.R., "Criminal Mapping Based On Forensic Evidences Using Generalized Gaussian Mixture Model", The International Journal of Computer Science &Applications (TIJCSA), Volume 1, No. 4, June 2012 ISSN – 2278-1080, Available Online at http://www.journalofcomputerscience.com/

[11]. Chen N. and Wang Y., " Prediction of Series Criminals: An Approach Based on Modeling", 2010 International Conference on Computational and Information Sciences.

[12]. Yu C.H., Ward M. W., Morabito M., and Ding W., " Crime Forecasting Using Data Mining Techniques", 2011 11th IEEE International Conference on Data Mining Workshops.