

Review on BAT Algorithm in IDS for Optimization

Aliya Ahmad*, Bhanu Pratap Singh Senger

Department of Computer Science & Engineering

All Saint's College of Technology, Bhopal

*Email: aliyasrkhan@gmail.com

Abstract – The data mining is an approach to extract the important or meaning full information from the large data set as per the need of end users. The IDS is another approach to identified the attacker from the system. The both concept works together for the best results. The IDS have to self sufficient in order to identify the attacker as soon as possible to that some time optimization approach gives the batter results. When there is a optimized data set so that execution time may be reduced. This paper is a brief summary of the all the above approaches by which a efficient IDS can be designed.

Keywords – IDS, Data mining, Optimization, Bat Algorithm.

I. INTRODUCTION

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Data mining is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics. And, it enables them to determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transactional data. With data mining, a retailer could use point-of-sale records of customer purchases to send targeted promotions based on an individual's purchase history. By mining demographic data from comment or warranty cards, the retailer could develop products and promotions to appeal to specific customer segments.

II. ISSUES PRESENTED

One of the key issues raised by data mining technology is not a business or technological one, but a social one. It is the issue of individual privacy. Data mining makes it possible to analyze routine business transactions and glean a significant amount of information about individuals buying habits and preferences.

Another issue is that of data integrity. Clearly, data analysis can only be as good as the data that is being analyzed. A key implementation challenge is integrating conflicting or redundant data from different sources. For example, a bank may maintain credit cards accounts on several different databases. The addresses (or even the names) of a single cardholder may be different in each. Software must translate data from one system to another and select the address most recently entered.

A hotly debated technical issue is whether it is better to set up a relational database structure or a multidimensional one. In a relational structure, data is stored in tables, permitting ad hoc queries. In a multidimensional structure, on the other hand, sets of cubes are arranged in arrays, with subsets created according to category. While multidimensional structures facilitate multidimensional data mining, relational structures thus far have performed better in client/server environments. And, with the explosion of the Internet, the world is becoming one big client/server environment.

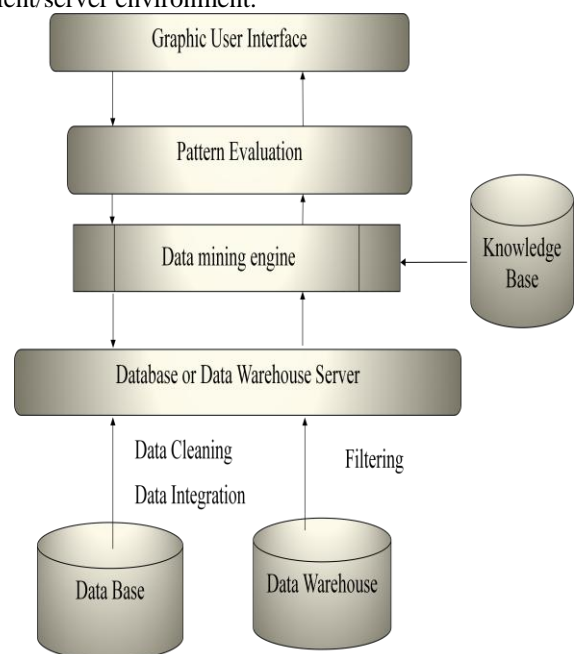


Fig.1. Data mining approach

III. INTRUSION DETECTION SYSTEM

Computer security is important for all users of IT systems issue. The rapid growth of the Internet, cyber attacks are increasing and can easily cause millions of dollars in damage to an organization. The detection of these attacks is a major topic of IT security. The intrusion detection systems (IDS) technology is effective in treating problems of intrusion detection system security. The main objective of the network approach is to detect unauthorized use and abuse of computer systems both in systems and external intruders. There are several methods used to implement intrusion detection systems expertise in statistical analysis and state transition approaches, etc., and these approaches are based on the immune system have been proposed over the years [2].

Intrusion prevention requires a well-chosen combination of "bait and trap" for the front two surveys. Divert attention from the intruder protected resource is another task of IDS. The data generated by intrusion detection systems are carefully examined for possible attacks. [2]

An IDS is an element of security policy. Among the various tasks IDS identifying the intruder it is more fundamental. It may be useful in forensic incident investigation and install the appropriate patches to allow detection of future attempted attacks directed at specific people or resources. Intrusion detection can sometimes produce false alarms, for example, as a result of malfunctioning network interface or send the description of the attack or email signatures [2]. The general operations performed by the intrusion detection system generally shown in Figure 2.

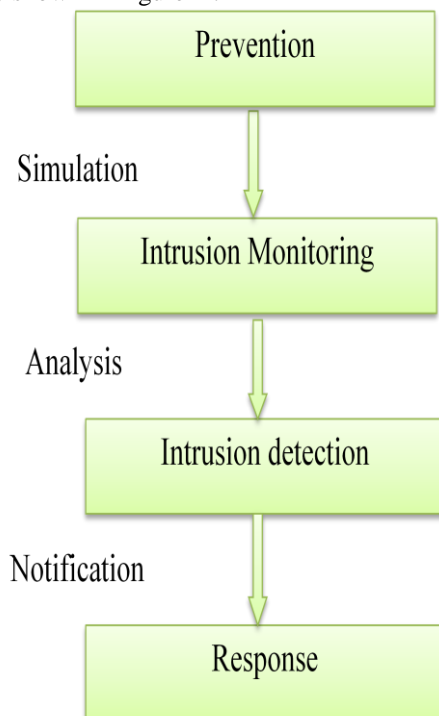


Fig.2. Flow chart for activities performed by IDS

Table 1 explains the summary of differences between strengths and weaknesses of the two approaches.

Many traditional IDS techniques are able to detect and prevent intrusion known, and are not able to react to new network intrusion recently using biological templates such as artificial immune system (AIS), networks and genetic algorithms, neural helps solve this problem.

Several types of techniques used for the design of an intrusion detection system. These anomalies include statistical techniques, techniques of fuzzy logic, rule-based technical faults; identification of penetration based on standards, technical state transition based on the network neurons, the techniques of data mining, etc. William Stallings [6] classified IDS various parameters based on the detection, rule-based and statistical anomaly detection.

Table 1: Summary

Misuse Based Detection	Anomaly Based Detection
Require Continuous updates	No updates required
No initial training	Training is required
Needs tuning as per Environment	Tuning is a part of training Itself
Cannot detect Novel attacks	Detect any novel attacks
Accurate alerts	Vague alerts
Very few false positives	Huge numbers of false Positives
Easy to design	difficult to design

IV. BAT ALGORITHM

The Bat algorithm is a metheuristic algorithm for global optimization. It was inspired by the echolocation behavior of microbats, with varying pulse rates of emission and loudness. The standard bat algorithm was based on the echolocation or bio-sonar characteristics of microbats

The standard bat algorithm has many advantages, and one of the key advantages is that it can provide very quick convergence at a very initial stage by switching from exploration to exploitation. This makes it an efficient algorithm for applications such as classifications and others when a quick solution is needed. However, if we allow the algorithm to switch to exploitation stage too quickly by varying A and r too quickly, it may lead to stagnation after some initial stage. In order to improve the performance, many methods and strategies have been attempted to increase the diversity of the solution and thus to enhance the performance, which produced a few good variants of bat algorithm.

V. LITERATURE REVIEW

The bat algorithm (BA) [6] is a nature-inspired algorithm, which has recently been applied in many applications. BA can deal with both continuous optimization and discrete optimization problems. The literature has expanded significantly in the past few years; this paper provides a timely review of the latest developments. We also highlight some topics for further research.

Optimization problem [7] relates to finding the best solution from all feasible solutions. Over the last 30 years, many meta-heuristic algorithms have been developed in the literature including that of Simulated Annealing (SA), Genetic Algorithm (GA), Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Harmony Search Algorithm (HS) to name a few. In order to help engineers make a sound decision on the selection amongst the best meta-heuristic algorithms for the problem at hand, there is a need to assess the performance of each algorithm against common case studies. Owing to the fact that they are new and much of their relative performance are still unknown (as compared to other established meta-heuristic algorithms), Bacterial Foraging Optimization Algorithm (BFO) and Bat Algorithm (BA) have been adopted for comparison using the 12 selected benchmark functions. In order to ensure fair comparison, both BFO and BA are implemented using the same data structure and the same language and running in the same platform (i.e. Microsoft Visual C# with .Net Framework 4.5). We found that BFO gives more accurate solution as compared to BA (with the same number of iterations). However, BA exhibits faster convergence rate.

Feature selection [8] is selecting a subset of optimal features. Feature selection is being used in high dimensional data reduction and it is being used in several applications like medical, image processing, text mining, etc. Several methods were introduced for unsupervised feature selection. Among those methods some are based on filter approach and some are based on wrapper approach. In the existing work, unsupervised feature selection methods using Genetic Algorithm, Particle Swarm Optimization with Relative Reduct, Quick Reduct and Ant Colony Optimization have been introduced. These methods yield better performance for unsupervised feature selection. In this paper we proposed a novel method to select subset of features from unlabeled data using binary bat algorithm with sum of squared error as the fitness function. The proposed method is then tested with various classification algorithms like decision tree, multilayer perception, support vector machine and clustering quality measures like sum of squared error. The results show that our proposed method gives more accuracy when compared with other optimization algorithm.

In the field of data mining [9] association rule is a very popular and efficient technique while it has different technique such as classification, clustering, sequential pattern etc to extract or optimize the large database. The

main aim of data mining is to find an effective or an optimized set of data from the big database. An association rules mining technique is used in various applications such as in banking, department stores etc. The Genetic Algorithm (GA) system can expect the rules which include negative attributes in the created rules mutually more than one attribute in resulting part. In this paper, we explore a review to optimize association rules using K- map and genetic algorithm (GA).

The potential to extract actionable [10] insights from Big Data has gained increased attention of researchers in academia as well as several industrial sectors. The field has become interesting and problems look even more exciting to solve ever since organizations have been trying to tame large volumes of complex and fast arriving Big Data streams through newer computing paradigms. However, extracting meaningful and actionable information from Big Data is a challenging and daunting task. The ability to generate value from large volumes of data is an art which combined with analytical skills needs to be mastered in order to gain competitive advantage in business. The ability of organizations to leverage the emerging technologies and integrate Big Data into their enterprise architectures effectively depends on the maturity level of the technology and business teams, capabilities they develop as well as the strategies they adopt. In this paper, through selected use cases, we demonstrate how statistical analyses, machine learning algorithms, optimization and text mining algorithms can be applied to extract meaningful insights from the data available through social media, online commerce, telecommunication industry, smart utility meters and used for variety of business benefits, including improving security. The nature of applied analytical techniques largely depends on the underlying nature of the problem so a one-size-fits-all solution hardly exists. Deriving information from Big Data is also subject to challenges associated with data security and privacy. These and other challenges are discussed in context of the selected problems to illustrate the potential of Big Data analytics.

Data mining [11] has been widely studied and applied into many fields such as Internet of Things (IoT) and business development. However, data mining techniques also occur serious challenges due to increased sensitive information disclosure and privacy violation. Privacy-Preserving Data Mining (PPDM), as an important branch of data mining and an interesting topic in privacy preservation, has gained special attention in recent years. In addition to extracting useful information and revealing patterns from large amounts of data, PPDM also protects private and sensitive data from disclosure without the permission of data owners or providers. This paper reviews main PPDM techniques based on a PPDM framework. We compare the advantages and disadvantages of different PPDM techniques and discuss open issues and future research trends in PPDM.

REFERENCES

- [1] Kamber "Data Mining Concepts and Techniques"
- [2] H. Afrabandpey, M. Ghaffari, A. Mirzaei, and M. Safayani. A novel bat algorithm based on chaos for optimization tasks. In Intelligent Systems (ICIS), 2014 Iranian Conference on, pages 1-6. IEEE, 2014.
- [3] G. Wang and L. Guo. A novel hybrid bat algorithm with harmony search for global numerical optimization. Journal of Applied Mathematics, 2013, 2013.
- [4] I. Fister Jr, X.-S. Yang, I. Fister, J. Brest, and D. Fister. A brief review of nature-inspired algorithms for optimization. Elektrotehniski vestnik, 80(3):116-122, 2013.
- [5] L. Dali et al., "A survey of intrusion detection system," Web Applications and Networking (WSWAN), 2015 2nd World Symposium on, Sousse, 2015, pp. 1-6.
- [6] I. Fister, I. Fister, X. S. Yang, S. Fong and Y. Zhuang, "Bat algorithm: Recent advances," Computational Intelligence and Informatics (CINTI), 2014 IEEE 15th International Symposium on, Budapest, 2014, pp. 163-167.
- [7] Y. A. Alsariera, H. S. Alamri, A. M. Nasser, M. A. Majid and K. Z. Zamli, "Comparative performance analysis of bat algorithm and bacterial foraging optimization algorithm using standard benchmark functions," Software Engineering Conference (MySEC), 2014 8th Malaysian, Langkawi, 2014, pp. 295-300.
- [8] A. S. S. Rani and R. R. Rajalaxmi, "Unsupervised feature selection using binary bat algorithm," Electronics and Communication Systems (ICECS), 2015 2nd International Conference on, Coimbatore, 2015, pp. 451-456.
- [9] G. Dhanore and S. K. Chaturvedi, "An optimization of association rule mining for large database using K- map and Genetic Algorithm: A review," Electronics and Communication Systems (ICECS), 2014 International Conference on, Coimbatore, 2014, pp. 1-5.
- [10] D. Arora and P. Malik, "Analytics: Key to Go from Generating Big Data to Deriving Business Value," Big Data Computing Service and Applications (BigDataService), 2015 IEEE First International Conference on, Redwood City, CA, 2015, pp. 446-452.
- [11] X. Li, Z. Yan and P. Zhang, "A Review on Privacy-Preserving Data Mining," Computer and Information Technology (CIT), 2014 IEEE International Conference on, Xi'an, 2014, pp. 769-774.