IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 79-82

Anomalous Detection Using Association Rule Mining

S. SenthilKumar[1], Dr. S. Mythili[2]
[1]*Ph.D. Research Scholar, Department of Computer Science*
[2]*Associate Professor & Head, Department of Information Technology,*
[1&2]*Kongunadu Arts and Science College,(Autonomous), Coimbatore, Tamil Nadu, India*

Abstract: *Association rule mining is Given a number of itemsets, find frequent subsets which are common to at least a minimum number s of the itemsets. Association rule mining are used to find rare events that are suspected to represent anomalies. In this research paper, association rule mining to find and summarize anomalous flows. Experimental results on the UCI repository datasets show that the proposed provide higher performance than the existing models. To extract anomalous flows, one could build a model describing normal flow characteristics and use the model to identify deviating flows.*
Keywords: *Anomaly detection, Association rule mining*

I. INTRODUCTION

An anomaly detection system may provide meta-data relevant to an alarm that help to narrow down the set of candidate anomalous flows [1]. For example, anomaly detection systems analyzing histograms may indicate the histogram bins that an anomaly affected [2], e.g., a range of IP addresses or port numbers. Such meta-data can be used to restrict the candidate anomalous flows to these that have IP addresses or port numbers within the affected range [3]. To extract anomalous flows, one could build a model describing normal flow characteristics and use the model to identify deviating flows [4]. However, building such a microscopic model is very challenging due to the wide variability of flow characteristics [5].

Similarly, one could compare flows during an interval with flows from normal or past intervals and search for changes, like new flows that were not previously observed or flows with significant increase/decrease in their volume [6]. Such approaches essentially perform anomaly detection at the level of individual flows and could be used to identify anomalous flows. Anomaly detection techniques are the last line of defense when other approaches fail to detect security threats or other problems [7]. They have been extensively studied since they pose a number of interesting research problems, involving statistics, modeling, and efficient data structures. Nevertheless, they have not yet gained widespread adaptation, as a number of challenges, like reducing the number of false positives or simplifying training and calibration, remain to be solved [8].

Identifying network anomalies is critical for the timely mitigation of events, like attacks or failures that can affect the security and performance of network [9]. Traditional approaches to anomaly detection use attack signatures built in an Intrusion Detection System (IDS) that can identify attacks with known patterns [10]. Significant research efforts have focused on building IDS's and, therefore, related production systems are presently employed in many networks [11]. Although signature-based detection finds most known attacks, it fails to identify new attacks and other problems that have not appeared before and do not have known signatures [12].

II. LITERATURE SURVEY

Thabtah et al [13] investigated the problem of producing rules with multiple labels and proposed a new associative classification approach called multi-class, multi-label associative classification (MMAC). The proposed model has many distinguishing features over traditional and associative classification methods in that it (1) produces classifiers that contain rules with multiple labels, (2) presents three evaluation measures for evaluating accuracy rate, (3) employs a new method of discovering the rules that require only one scan over the training data, (4) introduces a ranking technique which prunes redundant rules, and ensures only high effective ones are used for classification, and (5) integrates frequent items set discovery and rules generation in one phase to conserve less storage and runtime.

Zhang & Jiao [14] proposed an associative classification-based recommendation system for personalization in B2C e-commerce applications. Knowledge discovery techniques are applied to support personalization according to an inner established model that anticipates customer heterogeneous requirements. The framework and methodology of the associative classification-based recommendation system have been addressed.

Veloso et al [15] assessed the performance of lazy associative classification. Greedy (local) search may discard important rules while the global search, however, may generate a large number of rules. Further, many

of these rules may be useless during classification, and worst, important rules may never be mined. Lazy (non-eager) associative classification overcomes this problem by focusing on the features of the given test instance, increasing the chance of generating more rules that are useful for classifying the test instance.

In [16], Veloso et al proposed an approach which deals with small disjoints while exploring dependencies among labels. To address the problem with small disjoints, adopted a lazy associative classification approach. Instead of building a single set of class association rules (CARs) that is good on average for all predictions, the proposed lazy approach delays the inductive process until a test instance is given for classification, therefore taking advantage of better qualitative evidence coming from the test instance, and generating CARs on a demand-driven basis. A novel heuristic called progressive label focusing is employed, which makes feasible the exploration of associations among labels.

In [17], Veloso et al proposed calibration mechanisms based on learning a mapping from original estimates to calibrated estimates. The lazy associative classifiers (LAC) are well calibrated using an MDL-based entropy minimization method. These mechanisms discretize the probability space into a set of bins, and for each bin it is associated a calibrated probability. One of the proposed mechanisms greatly differs from other existing approaches, because instead of using bins with pre-specified boundaries, it automatically finds the boundaries that minimize the entropy in each bin.

Jiang et al [18] proposed a novel associative classification model, which first mines multi-class classification information from need-rating data, then constructs a rating classifier, and finally predicts customer ratings. This research proposes a rating classification model to estimate a potential customer's satisfaction level. It builds a rating classifier for a product by discovering rules from the need-rating database collected for the product. The rules imply the co-relationship between customers' needs, preferences, demographic profile, and their ratings for the products. The proposed is an improved decision tree (DT) algorithm for associative classification to find correlative relationship between power low state and the stability margin of system when specified faults occur and to get some comprehensible rules. Some unrelated features are eliminated through the operation experience, the principal component analysis (PCA) is used to reduce the dimension of features to get the key features which can describe the state of system.

Senthil Kumar et al [19] There exists a great variety of tools used for detecting outliers, exceptions or anomalies: expert systems, neural networks, clustering techniques, and association rules are some of them.

III. PROPOSED METHODOLOGY

In the proposed research method, the accuracy of association rule mining is improved by removing flows that could result in false-positive item-sets. Compared to existing studies, association rule mining can be combined with anomaly detection to effectively extract anomalous. The frequent single items are inputs to the process of finding possible frequent pairs of items, the frequent pairs of items are input to discover frequent triples of items, and so on

.

IV. PERFORMANCE EVALUATION

In this section, the proposed model of associative classifier is evaluated in MATLAB. The testing is carried out on datasets from UCI repository database. The comparisons are made in terms of classification accuracy and number of rules generated. Table 1 shows the accuracy comparison while table 2 shows the number of generated rules.

Table.1. Accuracy

Dataset	Accuracy (%)	
	FP Growth algorithm	Association Rule Mining
Iris	91.87	94.32
Diabetes	78.98	84.5
Glass	76.56	81.21
Pima	71.87	78.92

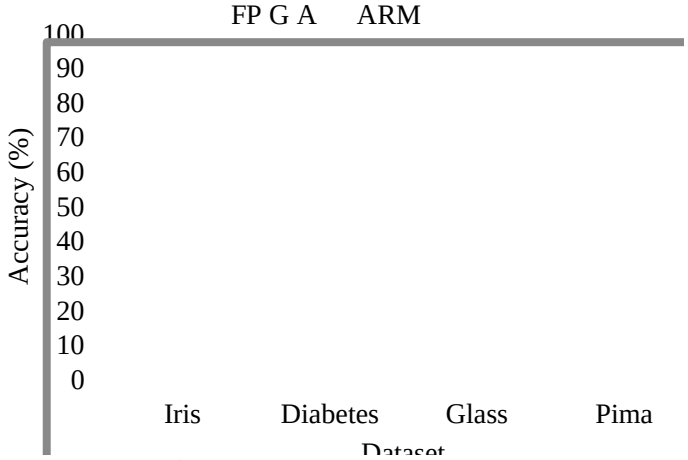

Figure.1. Accuracy comparison

Figure 2 shows the comparison of FP Growth algorithm and Association Rule Mining in terms of accuracy. It can be seen that for all the four datasets considered, the Association Rule Mining based model provides higher values of accuracy than the FP Growth algorithm. This is due to the efficient selection of attributes and improved pruning process.

Table.2. Number of rules generated

Dataset	Number of rules	
	FP Growth algorithm	Association Rule Mining
Iris	90	71
Diabetes	267	212
Glass	860	654
Pima	285	233

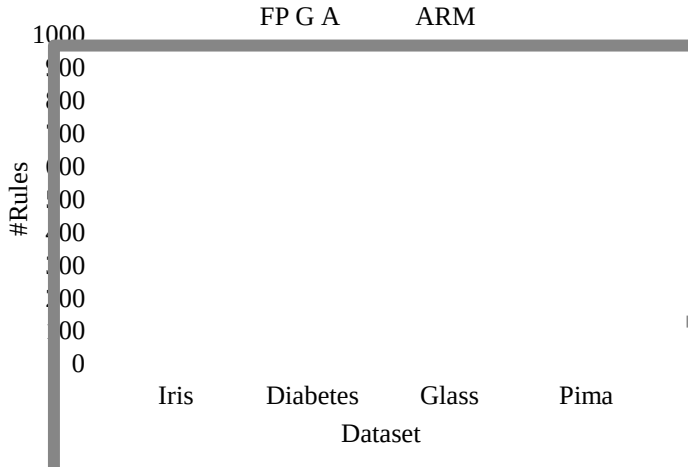

Figure.2. Number of rules

Figure 2 shows the comparison in terms of number of rules generated by as well as Association Rule Mining. Based on the analysis, it is proved that the proposed Association Rule Mining reduces the overall rules significantly thus minimizes the redundancy while improving the accuracy.

V. CONCLUSION

In the proposed research method Association rule mining gives better results. Experimental results on the UCI repository datasets show that the proposed model provide higher performance than the existing models.

REFERENCES

[1]. F. Silveira and C. Diot, "URCA: Pulling out anomalies by their root causes," in Proc. IEEE INFOCOM, Mar. 2010, pp. 1-9.

[2]. S. Ranjan, S. Shah, A. Nucci, M. M. Munafò, R. L. Cruz, and S.M Muthukrishnan, "Dowitcher: Effective worm detection and containment in the Internet core," in Proc. IEEE INFOCOM, 2007, pp.2541-2545.

[3]. G. Dewaele, K. Fukuda, P. Borgnat, P. Abry, and K. Cho, "Extracting hidden anomalies using sketch and non Gaussian multi resolution statistical detection procedures," in Proc. LSAD, 2007, pp. 145-152.

[4]. A. Lakhina, M. Crovella, and C. Diot,"Diagnosing network-wide traffic anomalies," in Proc. ACM SIGCOMM, 2004, pp. 219-230.

[5]. X. Li, F. Bian, M. Crovella, C. Diot, R. Govindan, G. Iannaccone, and A. Lakhina, "Detection and identification of network anomalies using sketch subspaces," in Proc. 6th ACM SIGCOMM IMC, 2006, pp. 147- 152.

[6]. W. Lee and S. J. Stolfo, "Data mining approaches for intrusion detection," in Proc. 7th USENIX Security Symp., 1998, vol. 7, p. 6.

[7]. R. Vaarandi, "Mining event logs with SLCT and LogHound," in Proc.IEEE NOMS, Apr. 2008, pp. 1071-1074.

[8]. K. Yoshida, Y. Shomura, and Y. Watanabe "Visualizing network status," in Proc. Int. Conf. Mach. Learning Cybern., Aug. 2007, vol.4, pp. 2094-2099.

[9]. X. Li and Z.-H. Deng, "Mining frequent patterns from network flows for monitoring network," Expert Syst. Appl. vol. 37, no. 12, pp.8850-8860, 2010.

[10]. V. Chandola and V. Kumar, "Summarization—Compressing data into an informative representation," Knowl. Inf. Syst., vol. 12, pp. 355-378, 2007.

[11]. M. V.Mahoney and P. K. Chan, "Learning rules for anomaly detection of hostile network traffic," in Proc. 3rd IEEE ICDM, 2003, pp.601-604.

[12]. Joshi, M. G. Anomaly Extraction Using Association Rule Mining.

[13]. Thabtah, F. A., Cowling, P., & Peng, Y. (2004). MMAC: A new multi-class, multi-label associative classification approach. In Data Mining, 2004. ICDM'04. Fourth IEEE International Conference on (pp. 217-224). IEEE.

[14]. Zhang, Y., & Jiao, J. R. (2007). An associative classification-based recommendation system for personalization in B2C e-commerce applications. Expert Systems with Applications, 33(2), 357-367.

[15]. Veloso, A., Meira Jr, W., & Zaki, M. J. (2006). Lazy associative classification. In Data Mining, 2006. ICDM'06. Sixth International Conference on (pp. 645-654). IEEE.

[16]. Veloso, A., Meira Jr, W., Gonçalves, M., & Zaki, M. (2007). Multi-label lazy associative classification. In European Conference on Principles of Data Mining and Knowledge Discovery (pp. 605-612). Springer Berlin Heidelberg.

[17]. Veloso, A., Meira, W., Gonçalves, M., Almeida, H. M., & Zaki, M. (2011). Calibrated lazy associative classification. Information Sciences, 181(13), 2656-2670.

[18]. Jiang, Y., Shang, J., & Liu, Y. (2010). Maximizing customer satisfaction through an online recommendation system: A novel associative classification model. Decision Support Systems, 48(3), 470-479.

[19]. SenthilKumar, S, Mythili, S, (2017). Survey on Exception Rules and Anomaly Detection. International Journal of Scientific Research in Computer Science, Engineering and Information Technology 2(6), 521-525.