Educational Administration: Theory and Practice
2024,30(5), 1395-1401
ISSN:2148-2403
https://kuey.net/

Research Article

A Survey On Analyzing Crime Patterns Using Data Mining Techniques

S. Imthiyas Ahamed[1*], Dr. J. Chockalingam[2], Dr. A. Shaik Abdul Khadir[3]

[1*]Assistant Professor of Computer Science, Khadir Mohideen College, Adirampattinam, (Affiliated to Bharathidasan University, Trichy-24), Thanjavur, Tamilnadu, India.
[2]Research Supervisor, Associate Professor of Computer Science (Retd), Khadir Mohideen College, Adirampattinam, (Affiliated to Bharathidasan University, Trichy-24), Thanjavur, Tamilnadu, India
[3]Research Co-Supervisor, Head & Associate Professor of Computer Science, Khadir Mohideen College, Adirampattinam, (Affiliated to Bharathidasan University, Trichy-24), Thanjavur, Tamilnadu, India

Citation: S. Imthiyas Ahamed et.al (2024), A Survey On Analyzing Crime Patterns Using Data Mining Techniques, *Educational Administration: Theory And Practice*, *30*(5), 1395-1401 Doi: 10.53555/kuey.v30i5.3097

ARTICLE INFO

ABSTRACT

The data mining is information looking at systems that used to research wrongdoing information as of late taken care of from various sources to track down models and examples in violations. In extra, it will in general be applied to construct efficiency in settling the violations faster and moreover can be applied to illuminate the wrongdoings normally. Anyway, there are various information mining procedures. To grow efficiency of wrongdoing area, it is critical to pick the information mining techniques properly. This paper reviews the compositions on various information mining applications, especially applications that applied to handle the wrongdoings. Review moreover enlightens research openings and troubles of wrongdoing information mining. In extra to that, this paper gives understanding with respect to the information digging for tracking down the models and examples in wrongdoing to be used appropriately and to be a help for novices in the assessment of wrongdoing data mining.

I. INTRODUCTION

Crime prevention and identification become a significant pattern in wrongdoing and an exceptionally difficult to tackle violations. A few investigations have found different procedures to tackle the wrongdoings that used to numerous applications. Such examinations can assist with accelerating the most common way of settling wrongdoing and assist the mechanized frameworks with recognizing the lawbreakers consequently. Also, the quickly propelling advancements can assist with resolving such issues. In any case, the wrongdoing designs are continuously changing and developing [1]. The wrongdoing information recently put away from different sources tends to consistently increment. As a result, the administration and investigation with colossal information are truly challenging and complex. To take care of the issues recently referenced, information mining procedure utilize many gaining calculations to remove concealed information from tremendous volume of information. Information mining is information breaking down procedures to track down examples and patterns in wrongdoings. It can assist with tackling the wrongdoings all the more expediently and furthermore can assist with alarming the criminal identification naturally.

This paper gives the short surveys of explores on different execution of information mining and the rules to settle the wrongdoings by utilizing information mining procedures. It likewise talks about research holes and difficulties in the space of wrongdoing information mining. In the following segment, the foundation and the issues of information mining are examined. Segment III extravagantly examines about the purposes of information mining methods to address the violations. The examination issues and difficulties are displayed in Segment IV. At last, the review is finished up in Segment V.

II. DATAMININGFUNDAMENTALS

Data mining is the analysis process used to analyze the historical data to find trends, patterns and knowledges. To extract the hidden knowledges, there are the initial important factors for analysis as follows: 1) the data used for analysis require the accuracy and sufficiency. 2) Knowledges and experiences of specialists.Fig.1 the knowledge results obtained from data mining processes are used to assist in decision making and to solve the problems. In the data mining, the analyzing techniques are explained in the following subsections

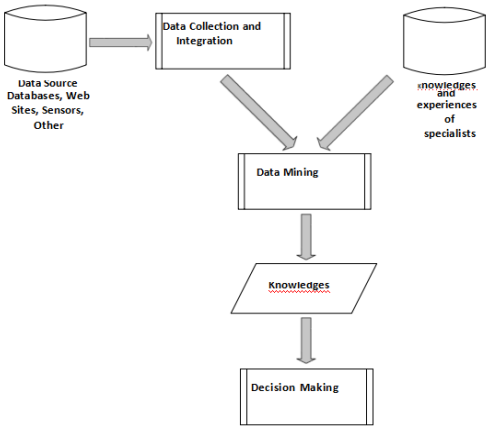

Fig.1. the background of data mining.

A. ASSOCIATIONRULE MINING

This procedure is unaided learning strategy that used to track down the secret proficiencies in unlabeled information. It is utilized to settle the issues assuming the students get the unlabeled model information. In extra, affiliation rule can find the fascinating co-events of items with regards to huge informational collections. In the fundamental of affiliation rule, the standard comprises of two sections. 1) The forerunner, which is on the left side or called the left hand side (LHS). 2) The resulting, which is on the right side or called the right hand side (RHS). A type of general affiliation rule is LHS → RHS, where LHS and RHS are disjoint thing sets. On the off chance that the LHS thing set happens, the RHS thing set will probably happen. For the effective revelation of affiliation leads, the significant factual estimations, the help and certainty measures, ought to be utilized together. A worth of such measures is in the scope of 0-1. In the event that an affiliation rule has exceptionally low help, this standard is probably going to be dull. As a result, the help measure is utilized to arrange the dull affiliation rules. The certainty measure is utilized to check the dependability of affiliation rules. For a given rule An and B in an exchange set T with higher the certainty, B is probably going to be available in T that contain A. To find co-events between two informational indexes, backing and certainty results ought to be more prominent than client indicated limits [3].

B. CLUSTERING

Clustering is a data analyzing technique in unsupervised type. This technique is used to divide the same data into the same group and the different data into the other group. The clustering techniques have a variety of concepts. The use of clustering techniques depends on applied fields [9].For the simple and effective clustering techniques, there are several algorithms such as K-means, Hierarchical Clustering and Expectation-Minimization that are discussed below.

C. CLASSIFICATION

This procedure is regulated learning strategy that used to appoint objects to one of numerous pre-decided classifications. The calculations of arrangement have been generally applied to the few issues that incorporate numerous different applications. For instance, it is utilized to address the distinguishing of the suspect vehicles and interlopers, the expectation of coronary illness, the sorting the record, and so on. The fundamental idea of grouping is portrayed as the accompanying: A gather information, otherwise called an info information, is utilized to deal with in a characterization task. Each record comprises of the property set and a class name. The class mark is pre-resolved class. A gather information is partitioned into two sets. 1) Train set is parceled haphazardly that is utilized to make a grouping model, otherwise called a classifier, to foresee the class of the new obscure record. 2) Test set is an excess set that is utilized to assess the exhibition of the order model. For building the grouping models, there are numerous deliberate methodologies, for example, choice tree, closest neighbor, Bayes' Hypothesis and brain organization, and so on.

III CLOUD STORAGE

Cloud storage is the on-request conveyance of process power, data set capacity, applications, and other IT assets through a cloud administrations stage by means of the web with pay-more only as costs arise pricing[1].Cloud stockpiling offers different types of assistance in which information capacity is the fundamental cloud administration. Distributed storage works behind the scene in our everyday exercises, for example, to watch films, mess around, sending sends and pay attention to music and so on, With Distributed storage, we can store, recuperate and reinforcement information, make new applications, convey programming on request, have sites, etc. Whenever there is an interest, client can get to the administrations of cloud progressively through internet[2].

IV. THEDATAMININGTECHNIQUESFORANALYZINGCRIMEPATTERNS

Nowadays, the various data mining techniques are used for different objectives such as: criminality, science, finance and banking, email filtering, healthcare and other industries. However, this survey focuses on the following crime types[14].

Traffic Violation and Border Control

Police Eyes [15] is the real-time traffic surveillance system that is developed to enhance the automatic detection capability of traffic violations. To extract the foreground from the back-ground in the scene obtained from IP cameras, they used the Gaussian mixture model. Then the foreground extracted is used to analyze the traffic violations by using violation conditions. Cheng et al. [16] used the rough set theory and association rules to find closer relationships with the traffic offense and regular traffic violating data of huge hidden data.

In the field of border control and security, Thongsataporn-watana and Chuenmanus [17] proposed the suspect vehicle detection system using association rule to analyze the vehi-cles with for gedlicensep late crossing the check point that potentially involved the criminal activity. Reference [6] has applied association analysis by using mutual information (MI)and modified the MI formulation with the time heuristic [3] to identify the potential criminal/suspect vehicles at the border. One of the important tools for collecting data is the sensors. The data obtained from the various sensors are analyzed to detect the criminal at the borders. In addition, Geographic Information Systems (GIS) is used to help generate geographic data from the sensors. However, if the system use only the GIS techniques, the geographic data will cannot be extract useful hidden knowledge. Hence, Kondaveeti et al. [18] used the GIS techniques with data mining techniques (spatial and association data mining techniques) to model the crime patterns and trends.

Violent Crime

Reference [1] proposed the use of naïve bayes algorithm with the concept of named entity recognition (NER),also known as entity or element extraction, to classify the news articles 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 websites. For prediction in crime, they used the decision tree concept. As a tested results, their system can classify and predict the crimes more than 90% accuracy. For a crime predicting model implemented in collaboration with the police department of a United States city in the Northeast crime, the hotspots are the best method for crime for ecasting proposed by [19].To improve the accuracy of clustering technique, the segmented multiple metric similarity measure(SMMSM) is proposed by [20] that used to find the crime suspects.

The Narcotics

In the opiates organizations, the fundamental part comprises of hubs or entertainers and associations or connections among them. The opiates network is portrayed which changes after some time that may be from the evacuation and addition of the hubs and connections. As an outcome, Kaza et al. [21] fostered the foreseeing criminal relationship calculations that used to foresee naturally the vehicles that are a co-guilty party to forestall what's in store assaults. They utilized the unique interpersonal organization examination (SNA) techniques and multivariate endurance investigation by utilizing the risk proportions of Cox relapse investigation. Reference [22] proposed the utilization of advanced brain organizations and developed rule-based classifiers. The two strategies are helpful to recognize harmful by means of opiate and responsive systems of activity (MOAs) of little atoms. The Fresh TDMn approach with help for worldly information mining, proposed by [23], is utilized to recognize relating the pulse fluctuation (HRV) with the respiratory rate inconstancy (RRV) to distinguish the patients getting opiates or different medications and the patients with unavoidable sepsis. They utilized making passing deliberations of hourly briefs to dissect connections among HRV and RRV. Chau et al. [24] has zeroed in on information assortment and text extraction which these information handling is a significant test. Subsequently, they proposed a brain network-based substance extractor by utilizing named-element extraction strategies, for example, lexical query, AI, and insignificant hand-created rules.

Cyber Crime

For the detection and prevention on cyber crime for Chinese web pages, Reference [25] has presented comparing the performance of the event ontology method as the prior knowledge and the method based on Support Vector Machine (SVM) to analyze the attributes and relations in web pages. Also these methods are used to reconstruct the scenario for crime mining. A web based crime analysis system is proposed by [26]. This system can extract the news article entities from news website, blog, etc. Then the newspaper articles entities are classified as crime and non-crime articles. It has a duplicate detector used to identify exact or near duplications of news paper articles and remove them from the database. For the crime analysis processes, the system used hot spot detection to identify the crimes and the crime frequencies. Sharma [27] proposed animprovedID3algorithm, an enhanced features election method and an attribute-importance factor to classify e-mails as either maybe-suspicious or non suspicious emails. Also they used a tool that is named as zero crime to help the system detect emails in relation to criminal activities. Frame work of Marketing or Newsletter Sender Reputation System (FMNSRS) [28]is developed from applying classification method called as sender reputation algorithm with the centralized user feedback database.

Performed Researches	Techniques	Tasks	Research Gaps	Research Challenges
[1]	Apriori algorithm	Extracting elements from data sources and analyzing the crime patterns	False detection	Improve precise detection
[6],[17]	Association analysis concept	Analyzing the crime patterns	No concerning with solving processing time and visualization	Improve performance of processing time and visualization
[18]	Spatial and association rule techniques	Analyzing the crime patterns and trends in geo graphic data	No concerning with performance	Improve performance
[20]	Segmented Multiple Metric Similarity Measure(SMMSM)	Classifying attributes into similar and related groups for detecting the crime suspects	No data collection and visualization	Collect data and improve visualization
[21]	Cox regression	Analyzing the co-offending relationships using dynamic social network analysis (SNA) method and multivariate survival analysis	No data collection and visualization	Collect data and improve visualization
[24]	Neural network-based entity extractor	Collecting and extracting the data obtained from police reports	No crime model and visualization	Model the crime future attacks and improve visualization
[25]	Event ontology and SVM algorithm	Analyzing the crime patterns	No flexibility of crime model and visualization	Model the crime future attacks and improve visualization
[26]	Crawler, document classifier, entity extractor, duplicate detector, data base handler, analyzer and graphical user interface	Extracting element data and analyzing the crime and frequency	No crime prediction and crime model creation	Improve performance and model the crime future attacks
[27]	An enhanced Decision Tree Algorithm and an improved ID3 Algorithm with enhanced feature selection method and attribute importance factor	Classifying emails in relation to crime activities	No data collection, crime prediction and crime model creation	Collect data and model the crime future attacks
[28]	Sender reputation algorithm based on the centralized user feedback database	Classifying unwanted emails sent from attackers or spammers	No crime model creation	Model the crime future attacks

Table 1.Summary of Researches In Crime

V. ISSUESANDCHALLENGESONCRIME

The summaries of research gaps and challenges in crime are shown in Table I.

Data Collection and Integration

In the crime analysis processes, input data is very important to use in training process and testing process. The training process is used to conduct the crime model and the testing process is used to validate the algorithm. Input data can be obtained from various sources such as news, social Medias, different sensors, criminal records obtained from the government agencies, etc. As a consequence, the collected data is large volumes of data. In additional, these data are in many formats that may be unstructured data. The collected data is stored into different databases. The issues of data collection lead to the challenge of preparation, transformation and integration of data. The many researches are concerning with solving these issues.

However, one challenge is the difficulty and complication in analyzing and extracting hidden knowledge from large volumes of data. The methods may be useful to collect and integrate data such as entity extraction [1] or grouping and filtering method [29].

Crime Pattern

The issues of crime pattern are concerning with finding and predicting the hidden crime. Nowadays, the crime rate is increase continuously and the crime patterns are always changing. As a consequence, the behavior's in crime are difficult to be explained and predicted. The research interests on crime prevention and detection are concerning with finding and conducting the crime model to detect crimes. The challenge is modeling the crime attack behavior's that support crime detection although the crime patterns are changing. The predictive and statistical methods may be useful to find and conduct the crime model.

Performance

The issues on performance are concerning with precision, reliability and processing time. The uncertainty in crime patterns effects the precision of crime detection. Besides that, the algorithms used properly and the transformed data also effects the processing time. Many researches attempt to develop algorithms to detect crimes efficiently. Most of them used a combination approach. However, the challenge on performance is developing the detecting algorithms to increase the crime detection accuracy although crime patterns are always changing or the crime data increases continuously.

Visualization

The main responsibility of the data visualization is to create images, diagrams, or animations to provide data summarization. It can help the text data and mining results provide more interesting and more easily understood. The current issue is that the amount of data is growing rapidly, which leads to the difficulty and complication to display the hidden knowledges. One of the greatest challenges is finding out how to display the data summaries of important crime patterns and trends from huge data. To visual the low-dimensional data, there are many visualization methods used for visualization such as chart, maps, scatter diagram, coxcomb plot, etc. Additionally, the visualization for multidimensional data needs to use the visualization methods such as geometric projection, image-based visualization technology, pixel-oriented visualization methods, distortion techniques, etc.[30].

V.CONCLUSION

Crime are characterized which change over time and increase continuously. The changing and increasing of crime lead to the issues of understanding the crime behaviour, crime predicting, precise detection, and managing large volumes of data obtained from various sources. Research interests have tried to solve these issues. However, these researches are still gaps in the crime detection accuracy. This leads to the challenges in the field of crime detection. The challenges include modeling of crimes for finding suitable algorithms to detect the crime, precise detection, data preparation and transformation, and processing time.

VII. REFERENCES

1. S. Sathyadevan, M. Devan, and S. Surya Gangadharan, "Crime analysis and prediction using data mining," in Networks Soft storage(ICNSC), 2019 First International Conference on, Aug 2019, pp. 406–412.
2. T. Pang-Ning, S. Michael, and K. Vipin, Introduction to Data Mining,1sted.Pearson,52005.
3. S. Kaza, Y. Wang, and H. Chen, "Suspect vehicle identification for border safety with modified mutual information," in Proceedings of the 4th IEEE International Conference on Intelligence and Security Informatics, ser. ISI'06. Berlin, Heidelberg: SpringerVerlag, 2018,pp.308–318.
4. V. Vaithiyanathan, K. Rajeswari, R. Phalnikar, and S. Tonge, "Im-proved apriorialgorithm based on selection criterion,"in Computational Intelligence storage Research (ICCIC), 2012 IEEE InternationalConferenceon,Dec2012,pp.1–4.
5. C. Chu-xiang, S. Jian-jing, C. Bing, S. Chang-xing, and W. Yun-cheng,"An improvement apriori arithmetic based on rough set theory," inCircuits, Communications and System (PACCS), 2017 Third Pacific-AsiaConferenceon,July2017,pp.1–3.
6. S. Kaza, T. Wang, H. Gowda, and H. Chen, "Target vehicle identificationfor border safety using mutual information," in Intelligent TransportationSystems,2005.Proceedings.2005IEEE,Sept2005,pp.1141–1146.
7. W.Huang,M.Krneta,L.Lin,andJ.Wu,"Associationbundle-anew pattern for association analysis," in Data Mining Workshops, 2018.ICDM Workshops 2018. Sixth IEEE International Conference on, Dec2018,pp.601–605.
8. N.Sasaki,R.Nishimura,andY.Suzuki,"Audiowatermarkingbasedon association analysis," in Signal Processing, 2018 8th InternationalConferenceon,vol.4,Nov2018.

9. Ben Ayed, M. Ben Halima, and A. Alimi, "Survey on clustering methods: Towards fuzzy clustering for big data," in Soft storage and Pattern Recognition (SoCPaR), 2019 6th International Conferenceof, Aug 2019, pp. 331–336.
10. Thammano and P. Kesisung, "Enhancing k-means algorithm for solv-ing classification problems," in Mechatronics and Automation (ICMA),2018 IEEE International Conferenceon, Aug2018, pp.1652–1656.
11. Y.ZhaoandG.Karypis,"Evaluationofhierarchicalclusteringalgorithmsfordocumentdatasets,"inProceedings oftheEleventhInternational Conference on Information and Knowledge Management,ser. CIKM '02.New York, NY, USA: ACM, 2016, pp. 515–524.[Online].Available:http://doi.acm.org/10.1145/584792.584877.
12. C.-N. Hsu, H.-S. Huang, and B.-H. Yang, "Global and componentwiseextrapolation for accelerating data mining from large incomplete datasets with the em algorithm," in Data Mining, 2018. ICDM '06. SixthInternationalConferenceon,Dec2018,pp.265–274.
13. X.-M.Zhao,Y.mingCheung,andD.-S.Huang,"Microarraydataanalysis using rival penalized em algorithm in normal mixture models,"inVLSIDesignandVideoTechnology,2005.Proceedingsof2005IEEEInternationalWorkshopon,May2005,pp.129–132.
14. H.Chen,W.Chung,Y.Qin,M.Chau,J.J.Xu,G.Wang,R.Zheng,andH.Atabakhsh,"Crimedatamining:Anoverviewandcasestudies," in Proceedings of the 2016 Annual National Conference onDigitalGovernmentResearch,ser.dg.o'03.DigitalGovernmentSocietyofNorthAmerica,2016,pp.1–5.[Online].Available:http://dl.acm.org/citation.cfm?id=1123196.1123231.
15. R. Marikhu, J. Moonrinta, M. Ekpanyapong, M. Dailey, and S. Sid-dhichai, "Police eyes: Real world automated detection of traffic viola-tions," in Electrical Engineering/Electronics, Computer, Telecommuni-cations and Information Technology (ECTI-CON), 2018 10th Interna-tionalConferenceon,May2018,pp.1–6.
16. W.Cheng,X.Ji,C.Han,andJ.Xi,"Theminingmethodoftheroad traffic illegal data based on rough sets and association rules," inIntelligent Computation Technology and Automation (ICICTA), 2010InternationalConferenceon,vol.3,May2010,pp.856–859.
17. U.Thongsataporn watana and C.Chuenmanus,"Suspect vehicle detection using vehicle reputation withassociationanalysisconcept,"inTourismInformatics,ser.IntelligentSystemsReferenceLibrary,T. Matsuo, K. Hashimoto, and H. Iwamoto, Eds., vol. 90.SpringerBerlinHeidelberg,2019,pp.151–164.
18. Kondaveeti, G. Runger, H. Liu, and J. Rowe, "Extracting geographicknowledgefromsensorinterventiondatausingspatialassociationrules," in Spatial Data Mining and Geographical Knowledge Services(ICSDM), 2017 IEEE International Conference on, June 2017, pp. 127–130.
19. C.-H. Yu, M. Ward, M. Morabito, and W. Ding, "Crime forecastingusing data mining techniques," in Data Mining Workshops (ICDMW),2017IEEE11thInternationalConferenceon,Dec2017,pp.779–786.
20. G.Yu,S.Shao,andB.Luo,"Miningcrimedatabyusingnewsimilaritymeasure," in Genetic and Evolutionary storage, 2008. WGEC '08.SecondInternationalConferenceon,Sept2008,pp.389–392.
21. S.Kaza,D.Hu,H.Atabakhsh,andH.Chen,"Predictingcriminalrelationshipsusingmultivariatesurvivalanalysis,"inProceedingsofthe8thAnnualInternationalConferenceonDigitalGovernmentResearch:BridgingDisciplines&Domains,ser.dg.o'07.DigitalGovernmentSocietyofNorthAmerica,2007,pp.290–291.[Online].Available:http://dl.acm.org/citation.cfm?id=1248460.1248524.
22. G. Fogel and M. Cheung, "Derivation of quantitative structure-toxicityrelationships for ecotoxicological effects of organic chemicals: evolvingneuralnetworksandevolvingrules,"inEvolutionaryComputation,2005. The 2005 IEEE Congress on, vol. 1, Sept 2005, pp. 274–281Vol.1.
23. McGregor, C. Catley, and A. James, "Variability analysis with ana-lyticsappliedtophysiologicaldatastreamsfromtheneonatalintensive care unit," in Computer-Based Medical Systems (CBMS), 2012 25thInternationalSymposiumon,June2012,pp.1–5.
24. M. Chau, J. J. Xu, and H. Chen, "Extracting meaningful entities frompolice narrative reports," in Proceedings of the 2016 Annual NationalConference on Digital Government Research, ser. dg.o '02.DigitalGovernmentSocietyofNorthAmerica,2016,pp.15.[Online].Available:http://dl.acm.org/citation.cfm?id=1123098.1123138
25. L. Cunhua, H. Yun, and Z. Zhaoman, "An event ontology constructionapproachtowebcrimemining,"inFuzzySystemsandKnowledgeDiscovery (FSKD), 2010 Seventh International Conference on, vol. 5,Aug2010,pp.2441–2445.
26. I.Jayaweera,C.Sajeewa,S.Liyanage,T.Wijewardane,I.Perera,and A. Wijayasiri, "Crime analytics: Analysis of crimes through newspaperarticles," in Moratuwa Engineering Research Conference (MERCon),2019,April2019,pp.277–282.
27. M.Sharma,"Z-crime:Adataminingtoolforthedetectionofsuspicious criminal activities based on decision tree," in Data Miningand Intelligent storage (ICDMIC), 2019 International Conferenceon,Sept2019,pp.1–6.
28. A.Kawbunjun,U.Thongsatapornwatana,andW.Lilakiatsakun,"Framework of marketing or newsletter sender reputation system (fmn-srs)," in Advanced Information Networking and Applications

(AINA),2019 IEEE 29th International Conference on, March 2019, pp. 420–427.
29. L. Alfantoukh and A. Durresi, "Techniques for collecting data in socialnetworks," in Network-Based Information Systems (NBiS), 2019 17th International Conferenceon, Sept 2019, pp.336–341.
30. H. Jin and H. Liu, "Research on visualization techniques in data min-ing," in Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conferenceon, Dec 2009, pp.1–3.