Volume 8, No. 5, May-June 2017
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
REVIEW ARTICLE
Available Online at www.ijarcs.info

Systematic Review of Crime Data Mining

Sarpreet Kaur
Student, Department of Computer Engineering,
Punjabi University, Patiala, India

Dr. Williamjeet Singh
Assistant Professor, Department of Computer Engineering,
Punjabi University, Patiala, India

Abstract: Crimes are a social nuisance and cost the society extremely in many ways. Any research that can help in analyzing and solving crime faster pays for itself. Crime data mining has the ability of extracting useful information and hidden patterns from the large crime data sets. The crime data mining challenges are becoming stimulating opportunities for the coming years. Since the literature of crime data mining has increased briskly in recent years, it becomes mandatory to develop an overview of the state-of-the-art. This systematic review focuses on crime data mining techniques and technologies used in previous studies. The existing work is classified into different categories and is presented using visualizations. This paper also indicates some challenges related to crime data mining research.

Keywords: Crime data mining; crime data analysis; systematic review; systematic study

I. INTRODUCTION

Crime is one of the prevailing and concerning facet in any society. The increase in crime rates is one of the cause for the alarm. Law enforcement agencies, intelligence agencies and police maintain crime databases[1,2]. The crime data be analyzed to gain insights and to extract knowledge from it. Several studies have discovered number of techniques to analyze the crime data [3]. The crime data analysis can provide the crime statistics of a region, country or world [4]. The law enforcement agencies can take better decisions for prosperity of the citizens by understanding the various parameters that influence the crimes [5].

A. Crime data mining information

Crime data mining information can be of different types as shown in the Figure 1.

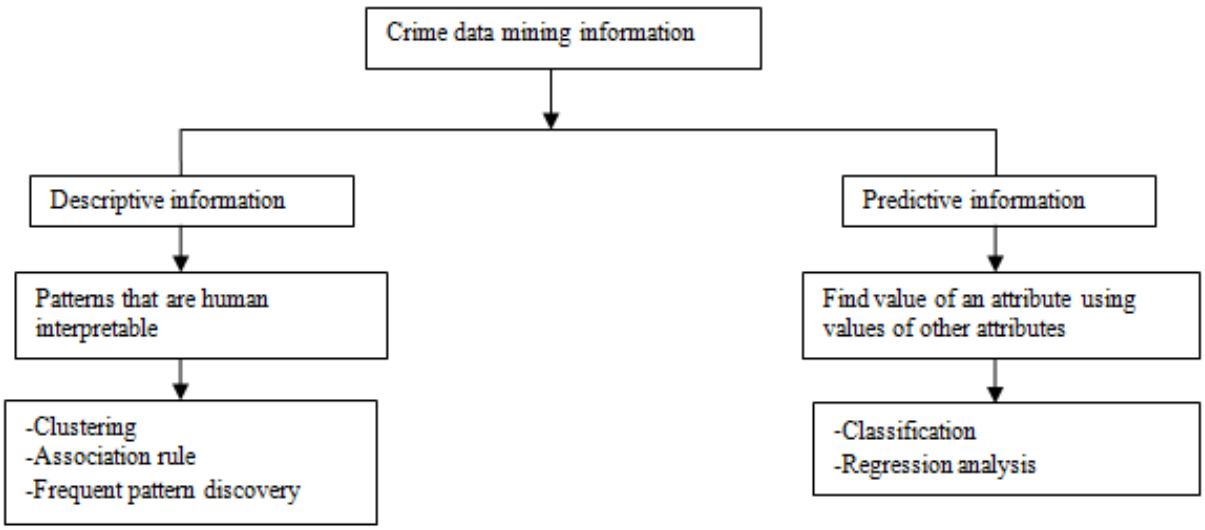

Figure 1. Crime data mining information types

B. Motivations and Objectives

This study is carried out in order to explore the crime data mining techniques, challenges and to know the technologies which are relevant. This analysis is compulsory to make it possible to know which categories of crime data mining technologies, techniques and challenges have been covered in past research and helps to identify gaps.

This study aims at systematically reviewing the crime data mining techniques, issues and challenges and technologies used in existing studies. The results may help the researcher to get an overview of the status of crime data mining and highlight the research gaps.

This paper is structured as follows. Section II describes the research methodology used in this study. Section III gives the classification of crime data mining and analysis papers considering the following criteria: (1) techniques used, (2) technologies used and (3) challenges addressed. Section IV presents the sources of crime data and type of crime. Section V presents the summary of researches in crime data mining. Section VI discusses the paper. Section VII presents the conclusion and future directions.

© 2015-19, IJARCS All Rights Reserved

II. RESEARCH METHODOLOGY

The research methodology is composed of three stages. The first stage involves the research of works related to crime data mining and analysis. The second stage is concerned with establishing a classification scheme described in Section III. The third stage involves the presentation of summary of researches in crime data mining and analysis and the report of detailed literature review.

(1)RQ.1: What are the different techniques for the crime data mining and analysis?

To answer this question bar chart in Section III describes the frequency of publications corresponding to the various techniques in this area.

(2)RQ.2: What are the technologies used in crime data mining and analysis?

To order to be able to answer this question, Table IV in Section III describes the technologies used in crime data mining and analysis.

(3)RQ.3: What are the various challenges involved in crime data mining and analysis?

To answer this question pie chart in Section III describes the various challenges and issues involved. The research is initialized with these queries and then follows the steps described.

A. Search Strategy and Screening

1. Sources of Information

To increase the probability of relevant articles, a set of appropriate databases must be chosen. For this review, the major databases of electronic journals are searched. The digital libraries considered are:

- IEEE Xplore (http://ieeexplore.ieee.org)
- ACM Digital Library (www.acm.org/dl)
- Science Direct (www.sciencedirect.com)
- Springer (www.springerlink.com)
- Wiley Interscience (www.interscience.wiley.org)

2. Study Selection

Research papers published by journals, conference proceedings and workshops are thought to be worthy and reliable. Keyword based search is employed to select the most relevant works. The keywords used are "crime data mining", "crime data analysis". The works not directly related to crime data mining and analysis are discarded. The criteria used for omission of a research paper includes:

- Unpublished papers, non-english papers, text-books, Master and Doctoral dissertations, non-peer-reviewed papers.
- Research of big data mining is relatively current, recent works published between 2012 to 2016 are searched.

The Table I shows the defined search strategy and number of results obtained. From the returned studies, firstly irrelevant studies are excluded on the basis of title. Certain studies could not be estimated from the title, then their abstract is considered. If even abstract is not evident then after reading the full text of papers, irrelevant studies are excluded.

Table I. Search selection criteria

S. no.	E-resource	Studies returned	Excluded			keywords used	Studies repeated
			based on title	based on abstract	based on full text		
1.	ieeexplore.ieee.org	10	1	3	----	crime data mining	7
		12	2	1	----	crime data analysis	
2.	www.acm.org	5	3	1	----	crime data mining	----
		1	----	----	----	crime data analysis	
3.	www.sciencedirect.com	13	4	2	3	crime data mining	3
		29	12	6	4	crime data analysis	
4.	www.springerlink.com	73	38	14	13	crime data mining	8
		27	10	7	4	crime data analysis	
5.	www.interscience.wiley.com	9	2	1	2	crime data mining	6
		18	10	3	3	crime data analysis	

B. Establishing a classification scheme

The selection process resulted in 183 papers (discarding repetition of papers) from five different digital libraries as on 31st March 2017, 10:30 AM (IST). Each paper is carefully assessed and classified. The selected research papers are classified according to the criteria established in Section III of this article.

C. Distribution of papers and summary of researches in crime data mining

The results of the classification offer important guidelines for future research on crime data mining and analysis. Literature related to crime data mining and analysis has increased enormously. Papers from 2004 to 2016 are reviewed. The distribution of reviewed papers over the years is depicted in Table II.

© 2015-19, IJARCS All Rights Reserved

Table II. Distribution of papers over the years

Year / E-resource	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
IEEE	1		1		1		1	1	1		2	1	1
ACM							1	1					
Science direct				1		1		1	2		1	1	1
Springer			1	2			1	1			1	2	
Wiley				1				3			1		1

The Table VI in Section V presents the summary of researches in crime data mining and analysis according to the contribution of previous studies.

III. CLASSIFICATION METHOD

The research papers are classified by considering the following criteria: (1) techniques used, (2) technologies used and (3) challenges and issues addressed.

A. Classification based on techniques used

On the basis of papers analyzed, the crime data mining techniques are illustrated in the Table III. The frequency of publications corresponding to the various techniques used in crime data mining and analysis are shown using a bar chart in Figure 2.

Table III. Various crime data mining techniques

Technique	Description
Clustering	-It is the process of making a group of objects into classes of similar objects. -A cluster of data objects can be treated as one group.
Classification	- Classification is used to predict group membership for data instances. -It is the derivation of model which determines the class of an object based on its attributes.
Association rule mining	-It looks for relationship between variables or objects. -It is a popular and well researched method for discovering interesting relationships between variables in large databases.
Frequent pattern mining	- It is used to find patterns that occur frequently in a data-set. -It forms the foundation for finding inherent regularities (associations) in data.
Outlier analysis	-It is the identification of events, items or observations that do not conform to an expected pattern. -It is also known as anomaly detection.
Regression	- It is used to predict a range of numeric values (continuous values), given a particular dataset.

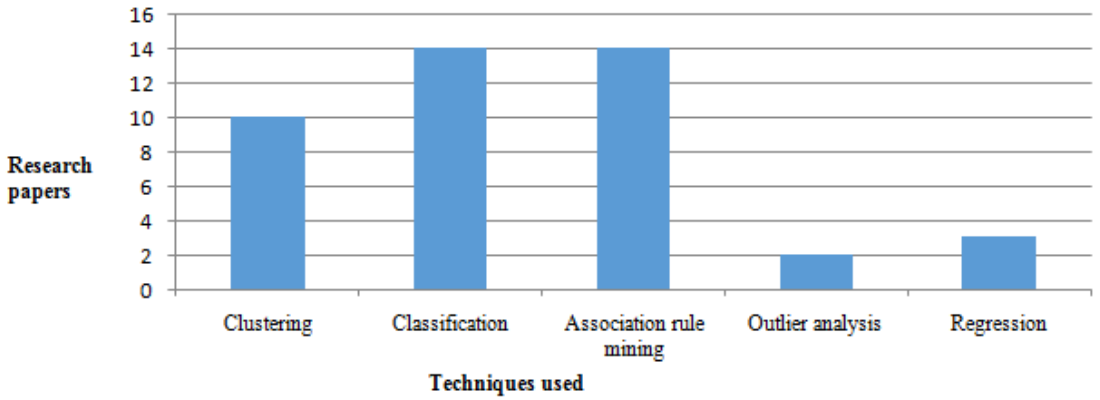

Figure 1. No. of research papers according to the techniques used

© 2015-19, IJARCS All Rights Reserved

B. Classification based on technologies used

The major objectives for crime data mining and analysis technology implementation are to minimize the hardware costs and reduce processing costs [6]. On the basis of papers analyzed various important crime data mining and analysis technologies are presented in the Table IV.

Table IV. Various technologies used in crime data mining

Technologies	Description	Features
SPSS	-acronym of Statistical Package for the Social Science -can perform highly complex data manipulation and analysis	-detailed labeling of variables and data values -flexible definition of missing data codes -reading raw data files in wide variety of formats(numeric, alphanumeric, binary, data and time formats) -reading hierarchical and non rectangular raw data files -ability to read and write to compressed files
MongoDB	-NoSQL database, shuns the relational database's table-based structure to adapt JSON-like documents that have dynamic schemas	-ad hoc queries -indexing -replication -aggregation -load balancing -file storage
Neo4j	-ACID-compliant transactional database with native graph storage and processing	-easy query language -support indexing -support unique constraints -represents semi-structured data easily
ESOM tool	-perform data mining tasks with Emergent Self-Organizing Maps (ESOM)	-animated visualizations -creation of the ESOM classifier and automated application to new data
Weka	-provides machine learning algorithms for data mining tasks	-data preprocessing -classification -clustering -regression -association rules -visualizations
Pig	-high-level mechanism for the parallel programming of MapReduce jobs to be executed on Hadoop clusters	-no complex codes -multi-query approach -PigLatin is SQL-like language
Hadoop	-core component of any modern architecture -allow organization to collect, store, analyze and manipulate large amount of data	-ability to handle variety of data(structured, semi-structured and unstructured) -visualizations

C. Classification based on challenges addressed

Challenges addressed are used as the third criteria for classification of papers. Difficulties lie in data collection and integration, data preprocessing, data storage, performance, visualization. [7,8]. Other challenges involved dealing with data that have no specified structure, selection of appropriate techniques, selection of appropriate technologies and collecting data from reliable sources [9,10]. The challenges addressed by various research papers are presented using a pie-chart in Figure 3.

1. Data collection and integration

In the crime data mining and analysis process, input data is important [11,12].Crime data is collected from various heterogeneous and autonomous sources such as news, social medias, criminal records, police reports (FIRs) etc. The collection of such a large volume of data is a challenging task[13].

2. Data preprocessing

The collected crime data are in many formats such as text, graph, images, relational data, semi-structured data and unstructured data [14]. Transforming different types of data to the understandable and desired format is a major challenge in mining of crime data [15].

3. Data storage

Crime data is characterized by large volume of data. Crime data mining and analysis requires a huge amount of storage as well as it demands new data management techniques on distributed systems [16]. NoSQL databases, such as Apache Hadoop are becoming the new core technologies for crime data mining [17].

4. Performance issues

Performance issues are concerned with processing time, precision and reliability [18,19,20]. Techniques and technologies used effect the processing time. The challenge is to develop algorithms and technologies to increase the accuracy and precision.

5. Visualization

The main challenge of crime data mining and analysis is visualizing the results. Visualization refers to representing the knowledge using graphs, bar-charts, pie-charts, line graphs[21]. It is difficult to find user-friendly visualizations for the massive data [22].

© 2015-19, IJARCS All Rights Reserved

IV. SOURCES OF CRIME DATA AND CRIME TYPES

In order to discover the right insights and successful investigation, it is necessary to recognize available data sources of crime and the various types of crime[23].

A. Data sources for crime data mining and analysis

1. Police reports

Police reports such as FIRs (First Information Report) contain information about the crime, complainant and suspect.

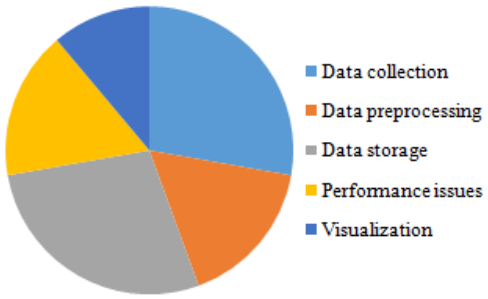

Figure 2. Challenges addressed by research papers

FIRs are written by police staff on paper and have unstructured data format. These are one of the reliable sources for collecting crime data[24].

2. Previous investigation files

After identifying a previously convicted suspect, the police ask for previous investigation files of the intended suspect. These files come in text, photo, video, CCTV video files, bank account, credit card statements, phone call, e-mail send-receive records, forensic reports, witness and victim statements and lawyer statements [25,26].

3. Intelligence reports

Intelligence agencies maintain information about the criminals. Intelligence agencies of India are Research and Analysis (RAW), Intelligence Bureau (IB), Narcotics Control Bureau (NCB) [27,28].

4. Open source intelligence findings

Open source intelligence findings are extracted from the web, search engines, social networking sites (Facebook, Twitter, LinkedIn). This information is in unstructured data format [29,30].

5. Police arrest records

When police officers arrest suspects then their arrest records are maintained. These records are mainly in relational format or text format [31].

B. Type of crimes

Table V presents different crime categories classified by various law-enforcement agencies. Some type of crime, such as arson and traffic violations concern police at the city, state and national level [32,33]. Other crime types such as terrorism, cybercrimes are investigated by local agencies as well as national and international agencies [34].

V. SUMMARY OF RESEARCHES

The summary of researches in crime data mining and analysis is presented in the Table VI.

Table V. Crime types and law enforcements

Crime type	Local law enforcement	National and International law enforcement
Traffic violations	Speeding, causing damage or injury in a collision during under the influence of drugs or alcohol, hit-and-run, reckless driving	----------
Sex crime	Child molestation, sexual abuse, rape, sexual assault, child pornography, prostitution	Human trafficking, prostitution, pornography
Theft	Robbery, burglary, motor vehicle theft	Fraud, corruption, trafficking in stolen software, music, movies
Arson	Intentionally setting fires to damage property	----------
Gang/drug offenses	Possessing, distributing and selling illegal drugs	Drug trafficking, people smuggling
Violent crime	Murder, hate crime, armed robbery	Terrorism, bombings
Cybercrime	Internet fraud, fraudulent websites, illegal online gambling and trading, cyber piracy, network intrusion and hacking, credit card and advance fee fraud	

I. **Table VI. Summary of researches in crime data mining**

Performed researches	Techniques	Tasks	Research Gaps
[1], [2], [16], [21], [24], [25]	Association rule mining	Analyzing the crime patterns	No concern with solving processing time and visualization
[1], [2], [3], [4], [5], [6], [14], [25], [28]	Clustering	Analyzing the similar crimes	No concern with solving processing time
[1], [2], [4], [5], [6], [8], [14], [19], [21], [25], [28], [29]	Classification	Classifying the crimes	No data collection, no crime model creation

© 2015-19, IJARCS All Rights Reserved

[1], [21]	Regression analysis	Analyzing the co-offending relationships	No data collection and visualization
[27]	Neural network based entity extractor	Extracting the data collected/obtained from police reports	No visualization

VI. DISCUSSION

The systematic review is derived from the 34 publications. The Table VI shows the summary of researches in crime data mining and analysis. Any technique such as classification, clustering, association rule mining, regression analysis can be used for mining and analyzing the crime data. Table VI shows that large number of papers used classification techniques, followed by clustering and association rule mining respectively. A large number of studies confirm that application of crime data mining techniques depend on the situation and objectives.

VII. CONCLUSION

Crime data mining and analysis is an active area of research. The results of this study may help new potential users in understanding the range of available crime data mining techniques and technologies. Crime data mining can be used to provide whole crime statistics of a particular region or area that provides benefit to the society by striking the government and law enforcement agencies to understand the various causes that increase the crime rates. The government and law enforcement agencies can take better decisions for better living of the citizens that naturally add up to lot of lives.

REFERENCES

1. Wingyan Chung,JenniferJie Xu,Gang Wang, Yi Qin,Michael Chau Hsinchun Chen, "Crime Data Mining: A General Framework and Some Examples," Computer, IEEE, pp. 50-56, April 2004.
2. S. Srisuk P. Thongtae, "An Analysis of Data Mining Applications in Crime Domain," IEEE 8th International Conference on Computer and Information Technology Workshops, IEEE, pp. 122-126, 2008.
3. Yefang Chen, Hao Huang Lianhang Ma, "AK-Modes: A Weighted Clustering Algorithm for Finding Similar Case Subsets", IEEE 2010, pp. 218-223, 2010.
4. Claus ATZENBECK, Ahmet CELIK, Zeki ERDEM Fatih OZGUL, "Incorporating data sources and methodologies for crime data mining,"IEEE, pp. 176-180, 2011.
5. Murat GOK, Ahmet CELIK, Yakup OZAL Fatih OZGUL, "Mining Hate Crimes to Figure out Reasons Behind," in 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,pp. 887-889, 2012.
6. Ubon Thongsatapornwatana, "A Survey of Data Mining Techniques for Analyzing Crime Patterns," in 2016 Second Asian Conference on Defence Technology (ACDT), IEEE, 2016.
7. Michael W. Packer, Jr., Melissa B. Thomason, John C. Wesley, Patrick J. Hansen,James H. Conklin,Donald E. Brown Andrew M. Lyons, "Uniform Crime Report "SuperClean" Data Cleaning Tool," in Proceedings of the 2006 Systems and Information Engineering Design Symposium,IEEE, pp. 14-18, 2006.
8. Devan M.S,Surya Gangadharan. S Shiju Sathyadevan, "Crime Analysis and Prediction Using Data Mining," in 2014 First International Conference on Networks & Soft Computing, IEEE, pp. 406-412, 2014.
9. Martin Ester, Uwe Glasser,Patricia L. Brantingham Mohammad A. Tayebi, "CRIMETRACER: Activity Space Based Crime Location Prediction," in 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Beijing,China, IEEE, pp. 472-480, 2014.
10. Cheng-Te Li,Shyh-Kang Jeng Yu-Yueh Huang, "Mining Location-based Social Networks forCriminal Activity Prediction," in 2015 24th Wireless and Optical Communication Conference (WOCC),IEEE, pp. 185-189,2015.
11. Christopher M. Gifford Anna L. Buczak, "Fuzzy Association Rule Mining for Community Crime Pattern Discovery," in ISI-KDD 2010, Washington D.C,U.S.A, ACM, 2010.
12. Fenye Bao, Chang-Tien Lu, Ing-Ray Chen Sumit Shah, "CROWDSAFE: Crowd Sourcing of Crime Incidents and Safe," in ACM SIGSPATIAL GIS'11, Chicago, IL, USA, ACM, pp. 521-524,2011.
13. Paul Elzinga, Stijn Viaene, Marc M. Van Hulle, Guido Dedene Jonas Poelmans, "Gaining insight in domestic violence with Emergent Self Organizing Maps," Expert Systems with Applications, vol. 36,Elsevier, pp. 11864–11874, 2009.
14. Ickjai Lee Peter Phillips, "Mining co-distribution patterns for large crime datasets," Expert Systems with Applications, vol. 39, Elsevier, pp. 11556-11563, 2012.
15. Antoine Bagula, Sonia Berman Omowunmi Isafiadea, "A Revised Frequent Pattern Model for Crime Situation Recognition Based on Floor-Ceil Quartile Function," Procedia Computer Science, vol. 55, Elsevier, pp. 251-260, 2015.
16. Jonathan Corcoran, Gary Higgs Chris Brunsdon, "Visualising space and time in crime patterns:A comparison of methods," Computers, Environment and Urban Systems, vol. 31, Elsevier, pp. 52-75, 2007.
17. Gyeong Seok Oh, Seung Yeop Paek Hyeon Ho Park, "Measuring the crime displacement and diffusion of benefit effects of open-street CCTV in South Korea," International Journal of Law, Crime and Justice, vol. 40, Elsevier ,pp. 179-191, 2012.
18. Mostafa Javideh, Mohammad Reza Ebrahimi MohammadReza Keyvanpour, "Detecting and investigating crime by means of data mining: a general crime matching framework," Procedia Computer Science, vol. 3, Elsevier, pp. 872-880, 2011.
19. Gondy Leroy Chih-Hao Ku, "A decision support system: Automated crime report analysis and classification for e-government," Government Information Quarterly, Elsevier, 2014.
20. Vishal Bhatnagar Arushi Jain, "Crime Data Analysis Using Pig with Hadoop," in International Conference on Information Security & Privacy (ICISP2015), Nagpur,INDIA, Elsevier, pp. 571-578, 2015.
21. K.C. Chang, Tai-Ping Hsing, and Shihchieh Chou Patrick S. Chen, "Mining Criminal Databases to Finding Investigation Clues—By Example of Stolen Automobiles Database,"Springer, pp. 91-10, 2006.
22. Xingan Li and Martti Juhola, " Country crime analysis using the self-organizing map, with special regard to demographic factors", Springer, AI & Soc, 2013.

© 2015-19, IJARCS All Rights Reserved

23. Zeki Erdem, Chris Bowerman, and Julian Bondy Fatih Ozgul, "Combined Detection Model for Criminal Network Detection," in PAISI 2010, Springer, pp. 1-14,2010.
24. Qiben Yan, Lei Zhang, Lizhi Peng,Hongbo Han Zhenxiang Chen, "Using Map-Based Interactive Interface for Understanding and Characterizing Crime Data in Cities," in ICSI-CCI 2015, Springer, pp. 479-490,2015.
25. Manish Gupta, M.P. Gupta B. Chandra, "Adaptive Query Interface for Mining Crime Data," in DNIS 2007, Springer, pp. 285-296,2007.
26. Ickjai Lee Peter Phillips, "Crime analysis through spatial areal aggregated density patterns," Geoinformatica, vol. 15, Springer, pp. 49-74, 2011.
27. Arti Jain,Surbhi Arora, Surbhi Agarwal,Tushar Gupta,Nikhil Tyagi Devendra Kumar Tayal, "Crime detection and criminal identification in India using data mining techniques," Springer, in AI & Soc, 2015.
28. Xu Huang, Emmanuel S. Silva, and Mansi Ghodsi Hossein Hassani, "A Review of Data Mining Applications in Crime," Statistical Analysis and Data Mining: The ASA Data Science Journal, vol. 9, 2016.
29. Brian Ewart Giles Oatley, "Data mining and crime analysis," vol. 1, Wiley Online Library, pp. 147-153, 2011.
30. ondy Leroy Chih-Hao Ku, "A Crime Reports Analysis System to Identify Related Crimes," JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, vol. 62, no. 8, pp. 1533-1547, 2011.
31. Jennifer Xu,Hsinchun Chen, Michael Chau Jennifer Schroeder, "Automated Criminal Link Analysis Based on Domain Knowledge," JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, vol. 58, no. 6, pp. 842-855, 2007.
32. Matthew Quick, Ping W. Chan Jane Law, "Analyzing Hotspots of Crime Using a Bayesian Spatiotemporal Modeling Approach: A Case Study of Violent Crime in the Greater Toronto Area," Geographical Analysis, pp. 1-19, 2014.
33. Michael R. Evans, James M. Kang,Pradeep Mohan Shashi Shekhar, "Identifying patterns in spatial information: a survey of methods," Wiley Online Library, vol. 1, pp. 193-214, 2011.
34. Ickjai Lee Peter Phillips, "Criminal Cross Correlation Mining and Visualization," in PAISI 2009, Springer, pp. 2-13, 2009.

© 2015-19, IJARCS All Rights Reserved