Big Data For Criminal Investigation

Chitra Lakhani, Vikas Janu

JK Lakshmipat University Jaipur 302026 India

ARTICLE INFO

Keywords:
Big data
Criminal Investigation
Legal
Data Mining

ABSTRACT

In this paper, we will be giving a brief idea about the work that is done related to big data in the field of legal and law enforcement. Starting with a brief introduction to big data and big data in legal and law enforcement. After that, background information about big data and survey on big data in the field of legal and law enforcement. We discuss some of the progress that has happened in the field until now, as per the data mining techniques and also what are the problems faced regarding the same.

1. Introduction

On a daily basis, people from all over the world post a huge amount of information on various social media platforms. Police have learnt to utilize social media to obtain information on suspects, witnesses, or events in hopes of resolving crimes and stopping planned criminal behavior. A certain amount of this intelligence is of value to authorities in law enforcement. Although police can use search warrants to gain access to non-public social media data, they frequently find the material they need by searching public forums, where access is not restricted by the need for a search warrant.

The nature and scope of police usage of social media are evolving as we go into the big data era. Police departments have decided to experiment with analytics on social media for the sake of surveillance, profiling, and predictive analytics, helped by start-ups providing data analytics targeted to law enforcement demands. These actions bring up fresh concerns about social fairness and privacy. They also bring up crucial questions of accountability and transparency. By doing thus, they draw attention to the expanding connections among private and other legal "disciplines." The reports in the fall of 2016 that exposed the widespread utilization of social media analytics by American police forces serve as the backdrop for this paper's examination of these concerns. It identifies the social justice and privacy concerns created by these actions and looks at normative and transparent frameworks for using such technologies.

2. Background

2.1. Big Data

Big data refers to extremely large sets of data that can be analyzed for patterns, trends, and associations using advanced computing technologies. The term "big data" has emerged as a result of the massive growth in the volume, velocity, and variety of data generated by various sources such as social media, mobile devices, sensors, and other digital technologies.

Big data analytics is the process of examining large and complex data sets to uncover hidden patterns, unknown correlations, and other useful information that can be used to make informed decisions. It involves the use of advanced tools and techniques such as data mining, machine learning, and predictive analytics to extract insights from big data.

One of the main challenges of big data is its management, processing, and analysis. Traditional data processing tools and techniques are often inadequate for handling big data, which requires specialized software and hardware infrastructure, as well as skilled data scientists and analysts.

Another challenge of big data is ensuring its quality and accuracy, as well as protecting its privacy and security. As big data becomes increasingly important in various industries, there is a growing need for standards, regulations, and best practices to ensure the ethical and responsible use of data.

Despite the challenges, big data has the potential to revolutionize various fields such as healthcare, finance, marketing, and government. It can help organizations to make better decisions, improve their operations and services, and create new opportunities for innovation and growth.

2.2. Data Mining

Data mining is a rapidly expanding field that comes at the intersection of many disciplines like statistics, database, machine learning, etc. It combines statistical modelling, data storing and approaches in AI. The main goal is to learn or predict the future, from the current structured or unstructured data, analysing the trends in the present data, and learning from it. There are models introduced by different statisticians that can be used to examine these trends or relationships between the variables in a database in order to create prediction models and other useful models.

AI being one of the most important of those sub-fields to set the base for data mining, has many algorithms developed to automate learning from data. These algorithms can lay a solid foundation for the development of various prediction models such as criminal activity detection, criminal behavior detection, etc. in the field of criminal activity. There are indeed steps being taken for the same to try and stop criminal activities from occurring.

2.3. Legal and Law Enforcement

The application of big data in law enforcement has seen a significant increase in recent years as the vast amount of digital data available continues to grow exponentially.

ORCID(s):

Law enforcement agencies can access various types of data, including surveillance footage and social media activity, to aid in their decision-making processes. This development has led to exciting possibilities such as the use of predictive analytics to identify potential criminal activity and advanced algorithms for facial recognition and biometric identification. However, the use of big data in law enforcement is not free from challenges and controversies. Utilizing advanced data analytics techniques can enhance the effectiveness and efficiency of investigations, allowing law enforcement agencies to quickly process large amounts of information to detect patterns and anomalies that may have gone unnoticed through traditional investigative methods. Yet, one of the main concerns with using big data is the potential for biased algorithms that may perpetuate systemic injustices if improperly designed or trained on biased data. Moreover, the collection and use of personal data raise concerns about privacy and civil liberties, as well as the possibility of data misuse by law enforcement agencies. Therefore, it is crucial that the ethical and legal implications of the use of big data in law enforcement are examined. Policymakers must weigh the potential risks and benefits of using big data in law enforcement and establish safeguards to ensure responsible and ethical usage. This paper will discuss the main ethical and legal considerations linked to the use of big data in law enforcement and propose recommendations to responsibly and effectively use big data in the criminal justice system.

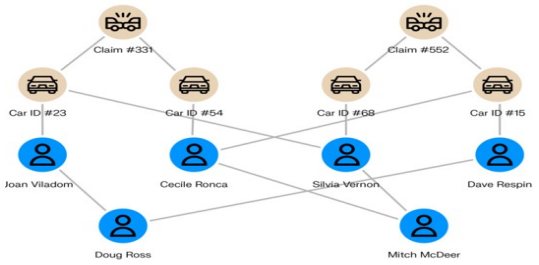
3. Data Mining Technologies

As the amount of crime occurring nowadays increases, there is a requirement for better technology and innovative ideas to fight off these crimes. This section explains some of the technologies in data mining that are or can be used in the field of crime.

3.1. Link Analysis

Link analysis is a data mining technique that shows the structure of the data in the form of a collection of connected or linked objects. Linking the data in an interconnected form is only the beginning. These links between the data can be used to identify relationships between objects that would otherwise look unlinked from a simple point of view. This technique can be used to find links in relation to terrorism, fraud, drug dealings and other matters. This is the first method in the list of techniques in data mining that can be used in exploring, assembling, detecting and analysing huge networks of pupil data through their jobs, vehicles, bank accounts, addresses, connections and other features. Linked data is visualised in the form of a graph with several nodes being people of interest and the links between them defining their relationship with each other.

Over the past few years, various research has used link analysis to examine huge social networks. The volume of data does not matter when using link analysis to find relations between all the suspects. What matters is their relation to each other, how they are connected, what they are doing to each other.



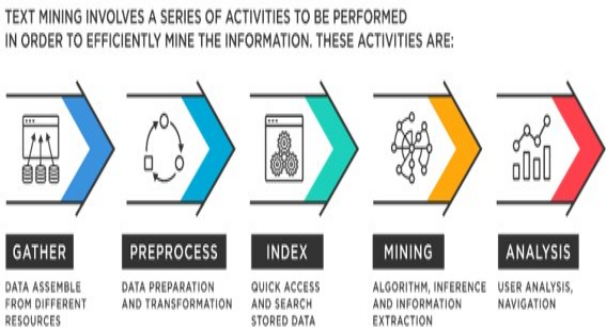
A prototype system was introduced by Jennifer Schroeder and her group called CrimeLink Explorer for enabling automatic crime link analysis. The approaches suggested in their system by them are co-occurrence analysis, a heuristic technique for discovering links between crime items, and shortest path algorithm for reaching the goal in minimum time. These approaches aim to integrate the investigative domain knowledge in the system for automatically determining the strength of association. For this, the system will use event similarity, time-based relation and connection, and other variables to identify relations between terrorist attacks by extracting a criminal's behavior and any other pattern.

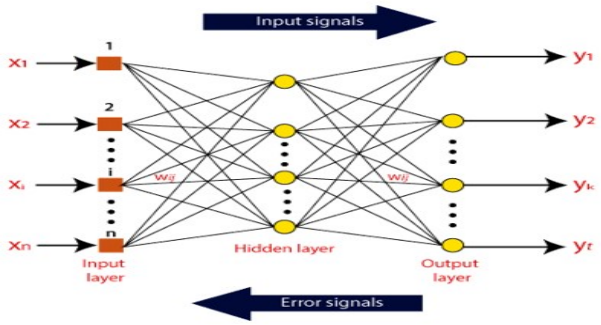
Another main element that helps criminals to easily widen their horizon of impact is through social media. Regarding this, another tool is introduced, social network analysis (SNA). This is also a tool with a link analysis approach implemented in it. It uses the level with which the suspects on the social network are related to each other in proximity, thinking, etc.

On the basis of current research and practice, it is found that we can use link analysis to explore associations among a large number of nodes and with a large number of features to link them. Link analysis is usually useful for the analysis of structured data but can also be used to examine unstructured data by integrating some text mining methods.

3.2. Text Mining

Also known as text data mining, is the process of extracting important and hidden knowledge from text documents. Text mining focuses on the concept or theme of the data given instead of specific words, classifying them accordingly, and using these classifications to help analysts and investigators in finding useful data or information within a large amount of documents. Text mining tools and procedures provide appropriate links among a variety of crucial textual information for learning new information and choosing the best course of action. Authors in [1] suggested using text mining to uncover criminal networks from the gathered textual materials. The suggested method primarily finds relevant information for criminal research before visualising the retrieved criminal network for analysis. Similar research is shown in [2] where information retrieval is done from news reports regarding crimes. Another work in [3], to police investigations, a comparison of the usability of emergent self-organizing maps (ESOM) and multidimensional scaling (MDS) as text exploration tools was done for finding hidden information from unstructured data of police





records. Likewise, another technique was presented in [4], primarily using NLP technique to find entities like occupation, people, addresses etc, by matching words or sentences from the already available text data. Then, those words or sentences are compared within the criminal network to find any associations or trends. It is feasible to recognise dishonest people using text mining techniques based on linguistic characteristics. For instance, text mining and data mining techniques were used in [5] to find bogus items in a textual content. A rule-based approach was created by Yin to capture sentiment features, where contextual elements were found by comparing textual postings to a window of nearby textual posts in [6]. Additionally, SVM was used in the text mining technique to categorise abusive posts on online social media. Furthermore, perilog is a tool introduced by NASA that can extract relevant data by context from any text document or phrases in [7]. It was originally created as part of FAA's Aviation Safety Reporting System, whose implementation was done to look for the primary reason for the crash of an airline in [8].

3.3. Artificial Neural Network

An artificial neural network imitates the working of the human brain's neurons, as a model of several hundreds and thousands of computational units. Some of these inbuilt, through software, abilities of ANN are memory, learning and decision making. Forecasting possible criminal patterns on the basis of observations on ongoing criminal activity using big data analytics, using the current sample to get trends and predict new trends. A NN can be used to build a tool which can help in behavioural and psychological examination of a suspect. An example is the project in [9] which is a psychological profiling system built on ANN. ANN models are more capable than standard database-oriented approaches in link analysis. Investigators can use it to retrieve important information from the police narrative report textual data. The Coplink project in [10] is the extraction of name entities based on ANN. Another recent study in [11] based on ANN is the retrieval of associations based shortest path algorithms.

The shortest path technique was used to find the strength of an association in the criminal network using ANN. Despite the limitations imposed by the complex properties of criminal data on conventional security and criminal investigation techniques, ANNs can significantly advance criminal network analytics because they can effectively integrate the elusive qualities of human reasoning with the perfect memory of computers, compulsive thoroughness, precise logic and commutation.

Using ANN, a model was proposed in [12], where the previous knowledge of the behavior of a criminal can be used to predict the future crime location. The authors used criminal classification to retrieve the most accurate pattern. A number of ANN tools have been developed over the years, having user friendly interfaces and having the ability to automatically change, according to their environment, their internal structure.

Technique	Identity	Opportunity	Challenges	Potential
Link analysis	using connections to retrieve knowledge	Group and subgroup as per characteristics	Intelligence capability	Fraud and money laundering, feature extraction, searching, browsing, visualization
Text mining	Working on unstructured and textual data	Retrieve information from texts	Veracity and velocity of big data	Clustering, classification and categorization of manner, data extraction and prediction.
Neural network	Human brain neurological functions	Statistical information processed automatically	Privacy, confirmation crisis	Bio-terrorism, forensic investigation

4. Criminal Investigation Implementation

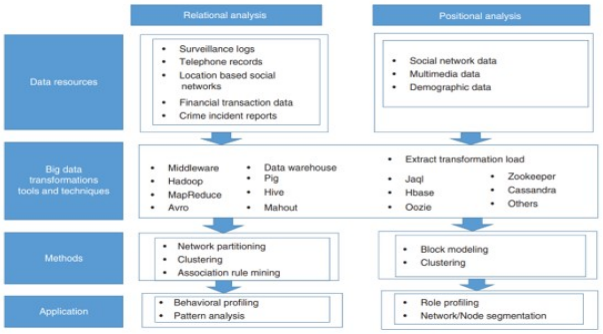
4.1. Introduction

In recent years, the importance of big data in the field of law enforcement has grown significantly. Big data refers to vast amounts of information generated from various sources, including social media, surveillance cameras, and other technologies. Law enforcement agencies are utilizing big data to enhance their ability to prevent and investigate crimes. For example, they can utilize data analytics tools to identify patterns in crime data, such as the time, location, and nature of crimes, enabling them to allocate resources more efficiently. Furthermore, they can monitor social media and other online sources for potential threats or criminal activity. Nevertheless, the application of big data in law enforcement raises concerns regarding privacy and civil liberties. Critics argue that collecting and analyzing enormous amounts of data can lead to profiling and discrimination, and that using automated tools to analyze data may lead to false positives and other errors. To address these concerns, law enforcement

agencies are establishing policies and procedures to ensure that the use of big data is transparent, ethical, and accountable. This includes implementing measures to protect the privacy and civil liberties of individuals and guaranteeing the accuracy and reliability of data analytics tools. To sum up, while the use of big data has the potential to improve the efficiency of law enforcement agencies, it should be balanced against concerns about privacy and civil liberties. Law enforcement agencies must consider the ethical and legal implications of utilizing big data and ensure that its usage is responsible and accountable.

4.2. Tools and Resources

Agencies can gather information from many resources for examining criminal networks, which can include service logs, location based networks, call records, financial statements and previous crime reports. Primary means for law enforcement agencies are call records, bank statements and call records. Recently, messages and emails, under text data, have also been used for examining to disclose close relations and links between groups. Several social media applications have also been used for positioning analysis, which has the motive of finding closeness between two participants rather than between their geographical location. Some of the tools used for big data are HDFS, pig, hive, MapReduce, Cassandra, Avro, etc.

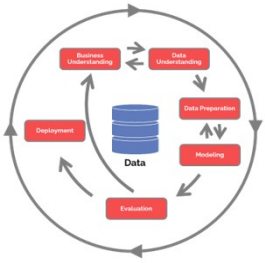


4.3. Method

The above mentioned data mining techniques are few of the ones that can be used to get information about crimes and criminals from the available data. Cluestering, block-mining, association rules are some of the techniques that can be used for this purpose. Some authors have proposed solution that implement most of these techniques and can be practically implemented after overcoming their respective fields' challenges. We can build systems with these technologies, feed data to it, and then the system will train to predict future trends.

Conclusion

Big data applications in law run the risk of being characterized as either inevitable or impractical. Understanding the techniques used, the biases introduced when selecting a method for machine learning, the type of inferences formed,



and the appropriateness of using results in decision-making are some of the issues. Even though all these issues affect all empirical methodologies, big data analytics has the potential to be more effective, opaque, and mythical than traditional approaches. The main objective of this article is to demystify some tactics so that we may understand how and when to use them properly. While we cannot ban people from using statistics, we can actively discuss the kind of interpretations that can and should be made as well as the impact that these inferences should have on various types of decisions.

The methods used by attorneys and law enforcement to draw conclusions and reach judgements are not unbiased, which is also the case with many technologies. It is notable that there has been a shift away from using court decisions as precedents to build doctrinal arguments and towards seeing them as data that can be analyzed using statistical methods. The transition to "predictive policing," in which deployment choices are informed by statistical forecasts, is also acceptable. Some people may find a trend in this direction to be appealing due to the strength of big data analysis and its prospective efficacy in discovering correlations. But, one would need to go through the police and lawyers' historic resistance to changing their ingrained ways of thinking.

4.4. Progress

The integration of big data in law enforcement has been a rapidly advancing field, increasingly utilized by law enforcement agencies globally. The following are some examples of how big data has been implemented in law enforcement, as well as the advancements made:

- Crime Prevention: Law enforcement agencies are leveraging big data to prevent crimes before they happen. They use data analytics to study previous crimes, predict where and when future crimes are likely to occur, and then allocate resources more efficiently to proactively prevent crimes.

- Investigation: Big data is also used for investigations. By analyzing data from social media, surveillance cameras, and other sources, law enforcement agencies can collect evidence and identify suspects more efficiently and accurately.

- Predictive Policing: Predictive policing is a method that uses data analysis to identify areas with a high likelihood of crime. Law enforcement agencies use big data to build predictive models that determine where and when crimes are most likely to occur, enabling them to deploy resources more effectively.

- Sentencing: Big data is also used to inform sentencing decisions. By studying data on past sentencing decisions,

law enforcement agencies can identify disparities and biases and develop more equitable sentencing guidelines.

- Real-Time Analytics: Real-time analytics involves analyzing data as it is produced to provide law enforcement agencies with current information on possible threats and criminal activity. This allows them to respond faster and more effectively to emerging situations.

In conclusion, the application of big data in law enforcement has made significant strides, and its usefulness will only continue to grow with advancements in technology and data analytics. However, it is critical that ethical and legal considerations are made to ensure responsible and accountable use of big data in law enforcement.

4.5. Problems Faced

Working with big data in law enforcement comes with several challenges that need to be addressed to ensure responsible and effective use of the technology. The following are some of the problems that law enforcement agencies face when working with big data:

- Privacy concerns: The collection and use of large amounts of personal data, such as information from social media and surveillance cameras, raise privacy concerns. Law enforcement agencies must balance the need for public safety with individual privacy rights.

- Accuracy and reliability of data: The accuracy and reliability of big data are crucial to its effectiveness in law enforcement. Errors in data collection, processing, or analysis could lead to false positives, which can have severe consequences.

- Data security: The vast amount of data collected by law enforcement agencies needs to be kept secure to protect against data breaches or unauthorized access.

- Bias and discrimination: There is a risk that big data analysis could lead to profiling or discrimination against certain groups based on factors such as race or gender.

- Technological limitations: Law enforcement agencies may face technological limitations in terms of infrastructure, data processing, and data storage capacity.

- Legal and ethical considerations: The use of big data in law enforcement also raises legal and ethical considerations, such as data privacy, transparency, and accountability. Overall, the use of big data in law enforcement presents significant challenges that must be addressed to ensure that its use is responsible, ethical, and effective. Law enforcement agencies need to implement policies and procedures that address these challenges and ensure that the use of big data is transparent and accountable.

References

[1] R. Al-Zaidy, B. C. Fung, A. M. Youssef, Towards discovering criminal communities from textual data, in: Proceedings of the 2011 ACM Symposium on Applied Computing, 2011, pp. 172–177.

[2] Y.-H. Tseng, Z.-P. Ho, K.-S. Yang, C.-C. Chen, Mining term networks from text collections for crime investigation, Expert Systems with Applications 39 (2012) 10082–10090.

[3] J. Poelmans, M. M. Van Hulle, S. Viaene, P. Elzinga, G. Dedene, Text mining with emergent self organizing maps and multi-dimensional scaling: A comparative study on domestic violence, Applied Soft Computing 11 (2011) 3870–3876.

[4] R. Lee, Automatic information extraction from documents: A tool for intelligence and law enforcement analysts, in: Proceedings of 1998 AAAI Fall Symposium on Artificial Intelligence and Link Analysis, volume 23, AAAI Press Menlo Park, CA, 1998.

[5] C. M. Fuller, D. P. Biros, D. Delen, An investigation of data and text mining methods for real world deception detection, Expert systems with applications 38 (2011) 8392–8398.

[6] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, L. Edwards, Detection of harassment on web 2.0, Proceedings of the Content Analysis in the WEB 2 (2009) 1–7.

[7] M. W. McGreevy, Using Perilog to Explore" Decision Making at NASA", Technical Report, 2005.

[8] C. Billings, J. Lauber, H. Funkhouser, E. Lyman, E. Huff, NASA aviation safety reporting system, Technical Report, 1976.

[9] M. Strano, A neural network applied to criminal psychological profiling: An italian initiative, International Journal of Offender Therapy and Comparative Criminology 48 (2004) 495–503.

[10] T. M. Li, M. Chau, P. W. Wong, P. S. Yip, A hybrid system for online detection of emotional distress, in: Intelligence and Security Informatics: Pacific Asia Workshop, PAISI 2012, Kuala Lumpur, Malaysia, May 29, 2012. Proceedings, Springer, 2012, pp. 73–80.

[11] J. J. Xu, H. Chen, Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks, Decision Support Systems 38 (2004) 473–487.

[12] K. Dahbur, T. Muscarello, Classification system for serial criminal patterns, Artificial Intelligence and Law 11 (2003) 251–269.

A. IPR Certificate

We, [Your Name] and [Your Partner Name], with this certify that the project work submitted by us entitled [Project Title] to our supervisors, Dr. Alok Kumar and Dr. Utsav Upadhyay, in partial fulfillment of the requirements for the course is a bonafide work carried out by us and has not been previously submitted to any other course. We further certify that no part of this work shall be published, reproduced, or distributed in any form without the prior permission of our supervisors. We understand that any such unauthorized use of the project work may be considered a violation of academic ethics and result in severe penalties. We also affirm that the project work has been carried out under the ethical standards and guidelines set forth by our supervisors. We acknowledge that our supervisor has the right to make any modifications or revisions to the project work that may be deemed necessary. We also agree to abide by any additional terms and conditions as stipulated by our supervisor.

Date: 20th May, 2023

Signature of Student:
Chitra Lakhani
2020BTECHCSE024
JK Lakshmipat University, Jaipur

Signature of Student:
Vikas Janu
2020BTECHCSE083
JK Lakshmipat University, Jaipur