Eurasia Journal of Science and Technology
Print ISSN: 2663-1024
Online ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3098

FROM CONCEPT TO MECHANISM: THE MODERN TRANSFORMATION PATH OF BIG DATA CRIME INVESTIGATION MODEL

Yao Zhang
Science and Technology Department, Yunnan Police College, Kunming 650223, Yunnan, China.
Corresponding Email: 46741256@qq.com

Abstract: The rapid development of big data technology has profoundly changed the landscape of criminal investigation. As a result, the big data criminal investigation model has emerged and become a key force in promoting the modernization of criminal justice. This paper conducts an in-depth analysis of the big data crime investigation model, elaborates on the characteristics of diversity, predictability, and correlational thinking it presents in the evolution of data thinking, and introduces key technologies such as crime data retrieval, collision, extraction, profiling, and relationship analysis that underpin this model. It also proposes establishing and improving compliance mechanisms (e.g., data collection, retrieval, review, supervision, and privacy protection) to ensure effective operation. The big data crime investigation model drives innovation in investigative concepts and methods, providing strong data support for investigative activities. Through optimizing concepts, technologies, and mechanisms, it will advance investigative work toward modernization, legalization, and intelligence.
Keywords: Big data criminal investigation; Data thinking; Key technologies; Compliance mechanism

1 INTRODUCTION

"Big data does not simply refer to large-scale data volumes, but rather to significant changes in epistemology and methodology[1]." Thanks to the rapid evolution of computer network technology and the exponential growth of network data, the "big data +" model is becoming increasingly mature, exerting a profound influence on various social fields. Notably, "big data + criminal justice" and "big data + criminal investigation" have become mainstream trends in the contemporary judicial field[2]. The benefits of applying big data technology in criminal investigations are increasingly prominent[3], and it has become a key driver of China's current criminal judicial reform. Thus, optimizing the big data crime investigation model is imperative in the new era.

2 THE EVOLUTION OF DATA-DRIVEN THINKING

Chinese police departments have widely adopted big data technology in criminal investigations, achieving significant results. The traditional information-driven investigation model is gradually transitioning to a big data-driven one, characterized by shifts from passive to proactive investigation, from individual to collaborative efforts, and from retrospective to predictive practices. This transformation has fundamentally changed investigative philosophy: traditional concepts such as "from case to person" and "from confession to evidence" are fading, while the new "from data to case" philosophy, reliant on big data platforms, is emerging as a trend. The big data-driven model demonstrates advantages in early warning, precise targeting, and comprehensive control. Thus, investigative departments and personnel should recognize the historical inevitability of this shift, break free from traditional constraints, and actively apply big data thinking to guide practice.

2.1 Diverse Thinking in Big Data Crime Investigation

In the era of big data, diverse thinking in crime investigation is reflected in the following three aspects: First, diversity in data acquisition. In traditional crime investigation models, investigators rely on the collection of physical information from crime scenes, such as "bodies, fingerprints, footprints, bloodstains, and DNA"; In contrast, big data-driven criminal investigations place greater emphasis on extracting crime-related information, with data acquisition shifting from physical crime scenes to virtual ones. Investigators can now collect additional crime-related information through sources such as "online shopping data, video footage, audio recordings, photographs, and social media platform information." Second, the diversity of data types. Traditional criminal investigations focused on the accuracy of crime-related information; in the big data-driven model, investigators emphasize the complexity and diversity of data types. Third, the diversity of data relationships. In traditional criminal investigations, causal relationships are a key logical basis for proving criminal facts; however, in the big data era, criminal investigations rarely consider causal relationships, instead emphasizing data associations. This approach mitigates the negative pursuit of causal relationships in traditional criminal investigations and reduces the adverse effects of unclear causal relationships on investigative activities.

2.2 Predictive Thinking in Big Data Crime Investigation

As a product of the big data era, the big data crime investigation model has driven criminal investigation activities to transition from labor-intensive to technology-intensive and data-intensive approaches. Compared to traditional crime investigation models, predictive crime investigation thinking is one of the core concepts of the big data crime investigation model. The greatest value of big data lies not in retrospectively analyzing the past or explaining the present, but in predicting the future[4]. As such, big data-driven criminal investigations have transcended traditional passive investigative models to achieve a functional transformation toward proactive criminal investigations. By applying machine learning to data regarding the "modus operandi, characteristics, and tools" of criminal activities or suspects within a specific spatio-temporal context, these investigations can predict and proactively identify potential criminal cases or activities, thereby realizing the predictive warning capabilities inherent in the big data-driven criminal investigation model.

2.3 Associative Thinking in Big Data Crime Investigation

In the era of small data, causal logic was the core thinking logic of traditional crime investigation models; in the era of big data, associative thinking is the core logic of big data crime investigation models. The associative thinking approach in the big data crime investigation model requires investigators to collect all data related to criminal offenses and analyze criminal cases using a systematic and holistic approach, breaking away from the traditional sampling data statistical methods in criminology, thereby enhancing the efficiency of criminal case resolution. Therefore, in the process of investigating criminal cases, the big data crime investigation model not only emphasizes causal logic but also places greater emphasis on the application of associative thinking. Through associative thinking, it further expands the scope of criminal investigation evidence collection, obtains additional criminal leads, and integrates criminal investigation resources, thereby promoting the healthy development of criminal investigation activities.

3 KEY TECHNOLOGIES

The big data crime investigation model differs from traditional crime investigation models in that it has a distinct "technical" characteristic, namely, it treats technology as the primary element throughout the entire crime investigation process. By searching, collating, and analyzing data, it further clarifies the facts of the crime and provides leads for crime investigation. In summary, the big data crime investigation model can rely on the following technologies or methods to carry out crime investigation activities.

3.1 Crime Data Retrieval Technology in Big Data Crime Investigation

Crime data retrieval technology is one of the most commonly used methods in the big data crime investigation model. Under this model, based on the type of data carrier, big data crime investigation can be divided into three categories: "database retrieval, internet retrieval, and electronic data retrieval." First, database retrieval primarily involves linking and retrieving data from various internal police department databases to identify criminal information. Additionally, other databases from the broader societal context can be appropriately utilized to seek out criminal information. This approach emphasizes data retrieval and criminal information mining within closed environments; Second, internet data retrieval. As is well known, the internet stores vast amounts of information available for investigators to extract[5]. Compared to database retrieval, internet data is more extensive, comprehensive, and open-source, with no restrictions on access. Investigators can utilize, extract, and save data at any time; Finally, electronic data retrieval. During the investigation and evidence collection process, when faced with a vast amount of electronic data, investigators can also utilize data retrieval to uncover new criminal information.

3.2 Crime Data Collision Technology in Big Data Crime Investigation

Crime data collision technology refers to investigators using computer algorithms and models to compare two or more databases or data sets closely related to criminal behavior and criminal information. The cross-referenced data and overlapping data generated by the collision are largely information related to criminal activity. Typically, crime data collision can be divided into the following stages: First, identifying the data objects and setting data tags. It is necessary to clearly define the data tags for obtaining the criminal information, such as "the identity characteristics, activity trajectories, and transportation information of the criminal suspect"; Second, screen the databases. By setting data tags, themes, and conditions, investigators further screen out data or information suspected to be related to criminal activities; Third, perform collision comparisons. Investigators match overlapping and cross-referenced data related to criminal activities based on the preliminary screening, reflecting the specific circumstances of the criminal suspect's criminal activities; Fourth, conduct in-depth analysis and judgment. Investigators Manually analyze and judge the aforementioned data to enhance data accuracy, thereby obtaining criminal leads and guiding the direction for subsequent criminal investigations[6].

3.3 Crime Data Extraction Technology in Big Data Crime Investigation

Data extraction and mining are core technologies in the current big data crime investigation model. Generally speaking, data mining technology is a key technology for automatically discovering criminal information. Investigators use

various big data technologies such as machine learning and distributed computing to automatically identify and extract information related to crimes from massive amounts of data. There are various analysis methods and technologies, including “anomaly analysis, correlation analysis, spatio-temporal analysis, and clustering analysis.”

3.4 Criminal Data Profiling Technology in Big Data Crime Investigation

By employing big data analysis methods, this technology identifies the identity characteristics, activity patterns, and interpersonal relationships of criminal suspects, and describes them in the form of data. This provides criminal intelligence and leads to assist in solving cases, which is the essence of criminal data profiling technology in big data crime investigation[7]. Criminal data profiling is a step-by-step analytical process. On one hand, investigators must extract foundational information and data related to criminal activities from massive datasets. On the other hand, investigators must apply the aforementioned analytical techniques to summarize the foundational data, transforming it into more valuable insights. This enables the big data-driven criminal investigation model to characterize the basic circumstances of criminal suspects, thereby driving substantial progress in big data-driven criminal investigations.

3.5 Criminal data Relationship Analysis Technology in Big Data Crime Investigation

The big data crime investigation model emphasizes data interconnectivity as a foundation, using data analysis to reveal criminal relationship networks. The world today is an era of rapid information technology development, with internet technology reaching new heights. Social networks, applications, and other technologies have deeply integrated online activities into personal life. Criminal suspects' online activities leave traces, such as “interaction information with others, shared friends, and communication frequency” as data indicators. Through the analysis of these data technologies, the relationships among members involved in criminal activities can be further revealed. This analysis technology is more widely applied to organized criminal activities such as “terrorism, organized crime, and cross-border smuggling of firearms and ammunition.” Taking cross-border cybercrime as an example, an increasing number of suspects are now using the internet to communicate with one another, which provides a solid data foundation for the operation of criminal investigation models in the big data era.

4 ENSURING COMPLIANCE MECHANISMS

4.1 Establishing and Improving Data Collection Mechanisms for Big Data Crime Investigation

First, from the perspective of data collection sources in big data crime investigation, currently, criminal investigation departments primarily rely on commercial data models and platforms, leading to certain discrepancies in terms of “data collection standards, data updates, and data quality.” In line with the specific needs of criminal investigation departments: First, it is essential to establish and improve proprietary professional databases, standardize data standards, and ensure data quality to ensure that data effectively supports criminal investigation activities; Second, criminal investigation activities should make greater use of other official data to provide authentic data support for criminal investigations, avoiding the waste of investigative resources due to data verification; finally, data collection must balance comprehensiveness and authenticity. Big data crime investigations require data support; the more comprehensive the data, the more accurate the intelligence analysis, and the better the investigative outcomes.

Second, from the perspective of data collection standards in big data-driven criminal investigations, data collection standards vary across different regions, levels, police units, and departments, and data barriers inevitably exist[8]. Given the characteristics of different police units and departments, it is not feasible to enforce uniform standards and regulations for data sharing platforms, as this is a purely rational goal with an enormous engineering scope that is typically impossible to achieve. However, unified standards can be established based on common issues across police units to facilitate data sharing. Uniform data formats can also be established to convert all data into a standardized format, enabling data sharing and exchange between different police units. Additionally, common data security regulations can be established to conduct security reviews on shared data to prevent unauthorized use. Furthermore, common data update rules can be established to ensure timely data updates, maintaining data reliability and accuracy.

Third, from the perspective of data collection algorithms and models in big data crime investigation, the design of big data crime investigation data collection models is based on the summary of practical experience in crime investigation activities. These models are only deployed and applied in actual crime investigation and law enforcement activities after being validated through a large number of crime investigation cases and data, thereby possessing a certain degree of reliability. At the same time, once big data crime investigation algorithms and models are designed, they are not set in stone but require continuous improvement, supplementation, and revision through practical crime investigation activities. Additionally, to enhance crime investigation efficiency, it is necessary to continuously optimize big data crime investigation algorithms, integrate them with practical crime investigation activities, and further refine data algorithm models to better apply them to practical crime investigation activities.

4.2 Establish and Improve Data Retrieval Mechanisms for Big Data Crime Investigation

At present, big data crime investigation involves a vast amount of data information. To facilitate crime investigation departments and investigators in obtaining accurate and effective data services, it is necessary to establish and improve

data retrieval mechanisms for big data crime investigation. On the one hand, data retrieval functions should be improved. Based on the needs of criminal investigation activities, further optimize the distribution of retrieval functions, remove modules with low usage frequency in actual operations, and add other required functions; on the other hand, it is crucial to accurately distinguish the authenticity of data information. The process and outcomes of big data-driven criminal investigations are directly influenced by the release and storage of data information. Therefore, the screening of genuine and fake data is particularly important. Genuine data can enhance the accuracy of predictive and analytical results in big data-driven criminal investigations, whereas fake data has little analytical value for criminal investigations. Therefore, by improving relevant mechanisms, investigators can obtain accurate and reliable case information from large amounts of data, eliminating the burden caused by fabricated information.

4.3 Establish and Improve the Data Review Mechanism for Big Data Crime Investigations

The data review mechanism in big data crime investigations involves third-party professional technical institutions using data cleaning and data verification methods to conduct legal, security, and technical reviews of the “data collection, data extraction, and data comparison” involved in big data crime investigations, thereby determining and verifying the authenticity and reliability of the data. This data review mechanism mainly includes two types: “case-by-case review” and “routine review.” On one hand, case-specific review in big data crime investigations refers to situations where the parties involved in criminal proceedings proactively request a review for a specific case, or where legal supervision departments conduct a review of a case within their statutory authority. In such cases, third-party institutions review the compliance of big data crime investigation techniques and provide an impartial review conclusion for decision-making and reference by relevant departments. On the other hand, routine review is a mechanism based on the need for daily database maintenance, involving regular data cleansing and security checks. In big data crime investigation activities, establishing a sound data review mechanism can effectively maintain and enhance the accuracy of big data crime investigation data, provide technical support to legal supervision departments, and thereby achieve the technological and informatization development of legal supervision.

4.4 Establish and Improve the Data Supervision Mechanism for Big Data Crime Investigation

In big data crime investigation, the importance of data supervision is self-evident. Once data is lost or mismanaged, it will inevitably lead to irreparable errors in criminal investigations. On one hand, it is essential to prioritize database security, enhance security levels, and strengthen technical oversight to prevent cybercriminals from stealing data. On the other hand, in big data-driven criminal investigations, data collection must not infringe upon citizens' personal privacy rights. Personal information must not be collected arbitrarily, nor should citizens' privacy information be disclosed without proper authorization[9]. Criminal investigation departments must strictly manage, approve, and use data during its extraction and utilization processes. At the same time, it is also necessary to improve relevant remedial mechanisms. If violations or illegal disclosures of citizens' personal privacy occur during criminal investigations, the legal responsibility of the relevant parties should be pursued. Therefore, to effectively, safely, and legally constrain criminal investigation departments or investigators in the collection and use of citizens' information, should establish a national-level big data criminal investigation information control platform and set different permissions for criminal investigation departments at different levels, in different regions, and of different police types. Additionally, when collecting citizens' privacy data, technical oversight must be strengthened over criminal investigation departments or investigators during the process of collecting and using public privacy data. The relevant regulatory authorities must use technical means to clearly track the entire process of data collection, use, and deletion, as well as the associated traces, to hold violators accountable for their actions.

4.5 Establish and Improve Mechanisms to Protect Privacy Rights in the Investigation of Big Data Crimes

First, the status of personal information rights in criminal justice must be clarified. With the development of technology, the security of personal information has become increasingly important. Criminal justice practice requires relevant organizations and individuals to protect citizens' personal information, especially their personal privacy information. Therefore, criminal investigation departments need to further strengthen the concept of protecting citizens' privacy rights, effectively increase the protection of citizens' information, and emphasize its status in criminal justice. In summary, protecting personal information rights as a fundamental right of citizens in the criminal justice system, and regulating the collection and processing of citizens' personal information by criminal investigation departments or investigators as statutory investigative measures, is of critical importance for standardizing the operation of big data crime investigations.

Second, strictly regulate the system for collecting and using personal information. In big data criminal investigations, criminal investigation departments should establish strict systems for the collection and use of personal information, and collect and use citizens' personal information in accordance with the principle of proportionality in investigations, in compliance with laws and regulations. From the perspective of the legislative framework for investigations, on the one hand, relevant laws and regulations such as the Personal Information Protection Act should be improved to further standardize the collection and use of personal information, ensuring strict approval, management, and use. On the other hand, the entities authorized to use personal information should be strictly defined. In big data-driven criminal investigations, only investigative personnel directly involved in handling criminal cases may use citizens' personal

information, but they must first obtain approval through the proper approval procedures. Only after obtaining approval may they utilize big data technology to query, link, and use personal information. Under no circumstances may they arbitrarily query or use citizens' private information[10].

Third, establish a redress system for citizens' privacy rights. To establish a fair and comprehensive system for protecting citizens' privacy rights, big data criminal investigation personnel must resolutely implement relevant legal regulatory measures, strictly conduct big data criminal investigations in accordance with laws, regulations, and rules, and must not infringe upon citizens' privacy rights through any improper means. Big data crime investigators are the litigation subjects responsible for safeguarding citizens' personal privacy rights. On one hand, it is necessary to strengthen the professional competence of big data crime investigators to ensure that citizens' personal privacy rights are effectively protected. On the other hand, any leakage of citizens' personal information resulting from erroneous investigative actions or activities must be strictly held accountable. Additionally, a remedial system must be established to safeguard citizens' privacy rights.

5 CONCLUSION

As discussed above, the big data-driven criminal investigation model represents a major transformation in China's criminal justice system, fundamentally altering investigative concepts, evidentiary frameworks, and operational measures, and providing robust data-driven support for current criminal investigation activities. By objectively, fairly, and rationally understanding the big data-driven criminal investigation model and continuously improving and optimizing it through "conceptual transformation, technological reliance, and institutional safeguards," we aim to drive the ongoing modernization, legalization, and intelligentization of criminal investigation activities in the current stage.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Pei Wei. The Conflict between Personal Information Big Data and Criminal Due Process and Its Mediation. Legal Research, 2018, 40(02): 42-61.
[2] Jiang S. The Use of Big Data in Criminal Justice and its Challenges. Peking University Law Journal, 2023, 11(1): 105-125.
[3] Brayne S. The Criminal Law and Law Enforcement Implications of Big Data. Annual Review of Law and Social Science, 2018, 14(1): 293-308.
[4] Parise. Big data: A revolution that will transform how we live, work, and think, by Viktor Mayer-Schonberger and Kenneth Cukier. Journal of Information Technology Case and Application Research, 2016, 18(3): 186-190.
[5] Mosso F M, Henrique P B S, Castro L B, et al. Quality in governmental data retrieval: a study of public policy data on the internet. PERSPECTIVAS EM CIENCIA DA INFORMACAO, 2020, 25(2): 103-132.
[6] Kadali K D, Mohan J V N R, Naik C M. Uncertain crime data analysis using hybrid approach. Discover Artificial Intelligence, 2025, 5(1): 15-15.
[7] Zhang Leihua, Zhang Huirong. Research on Issues Related to Investigative Profiling Technology Using Integrated Network Data. Journal of the People's Public Security University of China (Natural Science Edition), 2022, 28(04): 81-86.
[8] Anonymous. Recognizing and Overcoming the Data Barrier. Foundry Management & Technology, 2016, 144(5): 22.
[9] Wang A. Data Security And Privacy Protection: A Comprehensive Guide. World Scientific Publishing Company. 2025. DOI: 10.1142/14027.
[10] Marta P D, Marina Z. On Privacy, the Right to Privacy and its Protection in Croatia: The Criminal Offense of Unauthorized Use of Personal Data. Godišnjak Akademije pravnih znanosti Hrvatske, 2023, XIV(1): 57-85.