

# Application of Big Data Technology in Criminal Investigation

Ma Ning\*

Criminal Investigation Police University of China, Shenyang, Liaoning, China

\*Corresponding Author.

## **Abstract:**

In order to improve the efficiency of criminal investigation and ensure the objectivity of investigation behavior, big data is widely used in the field of investigation. Through the search, query, comparison and analysis of various big data in the public security organ's own data system and social data system, criminal clues can be found. This will be able to detect criminal evidence, capture suspect and detect crime. However, the application of big data in criminal investigation also has some problems, such as "data island", backward big data technology, lack of big data experts, violation of citizens' privacy and threat to national security. Therefore, we should break the data barrier and break the dilemma of "data island". Therefore, the judiciary should develop big data technology and train big data talents in order to strengthen big data information legislation and protect citizens' privacy and national security.

**Keywords:** *Data analysis, Crime Investigation, Big Data, Data Island..*

---

## I. INTRODUCTION

In the era of big data, the number of crimes with major social dangers such as computer network crime, terrorism crime and drug crime has increased sharply. In the 21st century, the criminal activities as the object of investigation have undergone fundamental changes [1-2]. The number, type, subject and form of crimes are very different from traditional crimes; The diversity of criminal means, the complexity of crime time and space and the social harmfulness of crime are different from traditional crimes. Due to its inherent defects, China's traditional investigation model is inadequate in the face of crimes in the era of big data. Therefore, it is particularly necessary to realize the modernization and upgrading of the traditional criminal investigation mode [3]. In the era of big data, the combination of big data and investigation behavior has become an irresistible trend. Using big data to improve crime investigation ability and crime control ability is the future development direction. Criminal investigation in the era of big data should comply with the development trend of the times, explore a reasonable path of the combination of big data and investigation behavior, and realize the modernization and upgrading of investigation mode.

## II. CONSTRUCTION SCHEME OF DATA-DRIVEN INVESTIGATION MODE

### 2.1 Build investigation data platform

Building a unified and efficient investigation data platform is the basic construction of building a data-driven investigation model [4-7]. In the era of big data, data fills our world. Only when the data are collected and managed purposefully, can we use the data correctly and reasonably. Building an investigation data platform can effectively integrate investigation data, expand investigation vision and provide investigation direction for investigators. Building an investigation data platform not only needs to improve the basic construction of the data application platform, but also needs to achieve the interoperability and sharing of data resources.

The data application platform of investigation organs is not only a huge database, but also an operable platform with the functions of data search, data collision and data mining. The construction and improvement of data application platform provides hardware support for the development of data-driven investigation mode [8]. The improvement of data application platform needs to integrate all-round information data and enrich the amount of data, including the integration of public security information data and social information data. It is certain that the basic establishment and improvement of the data application platform will greatly improve the efficiency of the investigation organ in solving cases. The construction of data application platform is inseparable from the effective integration of public security information data. Social information has many characteristics, such as wide source, wide radiation range and large amount of information resources, which makes social information contain great utility value. Social information includes all kinds of traces produced by people's production and life. The effective use of social information can greatly improve the investigation efficiency of investigators to a certain extent [9]. Social information ranges from Internet data information, video communication data information and GPS positioning information (Figure 1) generated in people's daily life to personal social security information, online registration information, hotel room opening information, etc. Personal information stored on microblog websites, QQ websites, e-mail and other personal information during shopping [10]. Video communication data information is more extensive. Video data is generated all the time in streets, units and even citizens' homes. The popularity of mobile phones and wechat has led to the explosive growth of communication information. These huge amounts of social information can be used to produce great value.

### 2.2 Construction and operation mechanism

The criminal behavior of criminal offense is recorded by a large amount of data. The investigation organ can rely on the data application platform, search, analyze and judge the criminal data by using big data technology, and quickly lock the criminal suspect, and improve the efficiency of case detection. Crime monitoring mechanism, crime detection mechanism and crime prediction mechanism complement each other and rely on each other to form the operation mechanism of data-driven investigation mode.



Fig. 1. GPS positioning information system

The original intention of monitoring is to pay close attention to the objects to be monitored, including people, things, places, etc. In traditional criminal investigation, crime monitoring mainly depends on manpower, which is time-consuming, laborious and inefficient. In the era of big data, we can rely on various monitoring equipment for data monitoring. This "panoramic open monitoring" covers both space and time. Data monitoring is an important means of crime monitoring. The premise of data monitoring is that the data application platform stores a sufficient amount of crime related data. In the data-driven investigation mode, investigators rely on the data application platform and use big data technology to analyze and judge the crime related information, so as to reduce the dependence on causality. Turn attention to the correlation analysis of crime data, and the processing results of data application platform analysis provide direction guidance for investigators to carry out investigation activities. Using big data technology to analyze data information can get rid of the subjective defects of crime analysis in traditional crime investigation, highlight the objectivity and stability of big data analysis results, and improve the accuracy of investigation decision-making.

The crime detection mechanism under the data-driven investigation mode refers to the process in which the investigation organ adheres to the concept of data-driven investigation, collects crime related data, makes full use of data analysis methods, excavates the potential value of data, defines the investigation ideas and finally detects cases. Data preparation (Figure 2) mainly includes data collection, data cleaning, data conversion and data integration. This process needs to be completed on the data application platform. The most important thing of data collection is to select the data related to criminal behavior. Relevant crime data should be collected not only in the crime scene, but also in the crime scene. Criminological theories such as daily activity theory, rational choice theory and crime mode theory believe that the perpetrator will consider the factors of space and time when choosing to commit a crime, and analyze the appropriateness of the target and the risk of being caught. Therefore, the behavior of criminals is not completely random, but follows the usual life pattern and temporal and spatial characteristics, which is predictable to a certain extent.

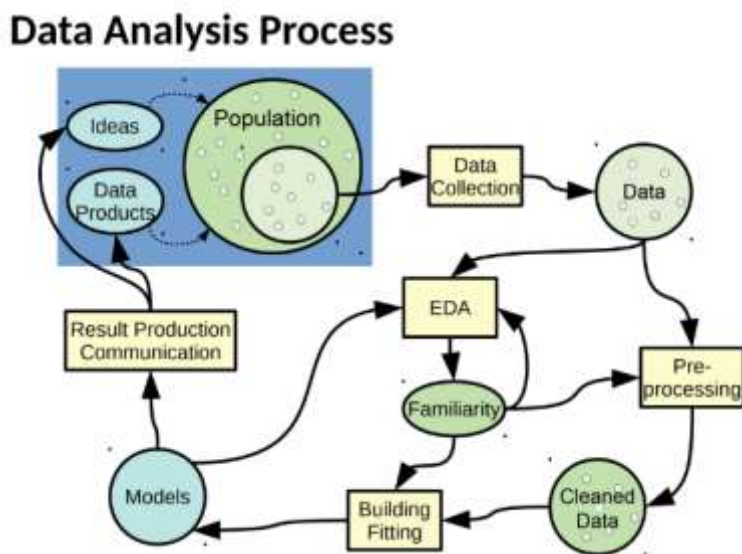


Fig. 2. Data preparation process.

### III. THE APPLICABLE PATH OF DATA-DRIVEN INVESTIGATION MODE

#### 3.1 Determine the investigation ideas

In the traditional criminal investigation, investigators regard finding the cause of the criminal case as the key to the case investigation. The reverse thinking of tracing the cause of the crime from the criminal result occupies an important position in the traditional investigation. Investigators always focus on the crime scene, gradually deduce the reasons for the occurrence of criminal cases, and then have an overall grasp of the constituent elements of the case such as the perpetrator's criminal motivation, criminal mode and criminal tools.

Exploring the causality of a case is based on the investigators' large amount of investigation experience and good logical reasoning. The determination of causality is restricted by the investigators' subjective factors. The development trend of crime in the new era makes the causality of the case no longer so close, which increases the difficulty of case detection. In the era of big data, in many cases, there is no strict causal relationship between data. It is no longer realistic to rely on the causal relationship to reveal and speculate on the development trend of things. Big data thinking requires us to get rid of the obsession of exploring causality and turn to the search of correlation. Big data is like a treasure. The key to determining the value of the treasure lies in the attitude towards the transformation of cognition and causality from big data to correlation. The data fusion process is shown in Figure 3.

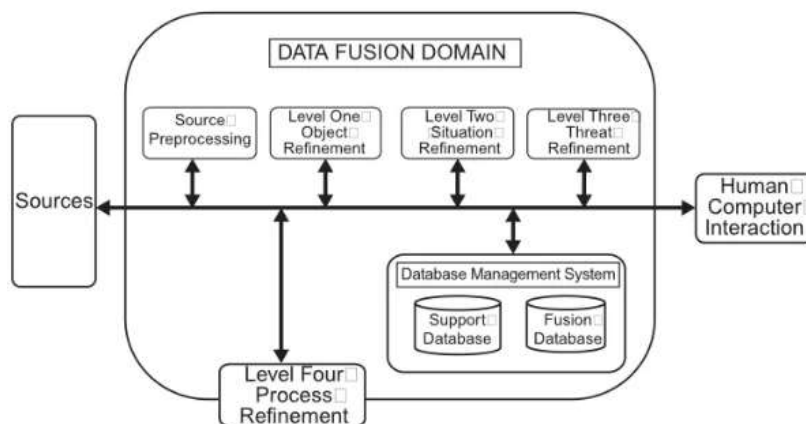


Fig.3. The data fusion process.

In the investigation process of traditional criminal cases, the crime scene is usually the core source of case information and the logical starting point for investigators to understand the case. This is mainly determined by the characteristics of the crime scene: the criminal scene and the information related to the criminal perpetrator are stored in the crime scene. Investigators rely mainly on these residual information to lock the suspect and identify the criminal activities. In the traditional criminal investigation, investigators pay more attention to the investigation of the main scene and ignore the control of the related scene. This is mainly because the crime scene is easy to find and the clues left by the crime scene are more and more direct, while the related scene is usually difficult to find, and the investigation clues are few and not obvious. With the advent of the era of big data, data recording has become the norm. Every information such as individual living habits, behavior hobbies and even physiological data may be recorded by the data and become the object of analysis and judgment of the data application platform, which means that criminal activities in the physical space have a great chance to find a corresponding in the virtual space. Modern video surveillance technology has the opportunity to completely or incompletely record the process of the crime, and the behavior track of the perpetrator before or after the crime may also be captured by video surveillance. If the primary problem that traditional investigation needs to solve is who has committed what kind of criminal behavior, the first problem that data-driven investigation in the era of big data needs to face is not only who has committed what kind of criminal behavior, but also what kind of non criminal behavior that person has also committed.

### 3.2 Pay attention to electronic forensics

In the era of big data, the perpetrator will retain the criminal activities in the form of electronic data in the virtual space, whether in the real space or in the network virtual space. In the investigation stage, these electronic data will become "silent witnesses", waiting for the discovery and extraction of investigators. The emergence of electronic data is the application of the development of modern science and technology in litigation evidence. As a kind of legal evidence, electronic data not only has the basic attributes of

general types of evidence, but also has the characteristics of invisibility, easy collection, easy preservation and repeated reproducibility.

In recent years, it has become a trend for investigative organs to implement "remote forensics" through network video and other information technologies. Remote forensics measures such as remote interrogation, remote inquiry, remote inspection, remote identification, remote search, remote technical investigation and database query have gradually stepped on the historical stage, and forensics measures are increasingly showing the trend of electronization. Database query is an important form of electronic evidence collection measures. Its emergence can not only be used as a supplement to other evidence collection measures, but also an independent investigation and evidence collection measure. China's investigation organs have benefited a lot by establishing databases and using database query. Electronic forensics measures come from traditional forensics methods, which are just injected with the elements of the times and closely combined with the era of big data.

The emergence and use of new electronic forensics measures have brought substantial convenience to the development of data-driven investigation mode. The combination of investigation work and mass line is the investigation policy established at the beginning of China's establishment. Applying human flesh search to investigation work is a new attempt of investigation work relying on the mass principle in the era of big data. Social networking site investigation is a method of investigation and evidence collection often used by investigation organs in handling cases. Social networking sites are full of a lot of personal information, including a lot of personal privacy information. In real life, the perpetrator extends the devil's paw of crime to social networks, and online fraud information can be seen everywhere. Data investigation on social networking sites is an effective evidence collection measure. The target of network data retrieval and storage includes electronic data acquisition and replication. It needs to be clear that network monitoring is a technical investigation measure that has gone through strict examination and approval procedures. In the investigation practice, the network monitoring measures mainly include Trojan horse program, springboard technology, network attack, network trap IP search and positioning technology.

#### **IV. PRECAUTIONS FOR APPLYING DATA-DRIVEN INVESTIGATION MODE**

##### **4.1 Adhere to the integration and complementarity of multiple investigation methods**

Strictly speaking, the investigation mode of data-driven investigation mode is the integration of traditional investigation means and big data technology. Take the investigation measure of bottom-up queuing as an example. Bottom-up queuing in the era of big data should first implement data-oriented accurate scheduling, and then implement the traditional manual scheduling when the accurate scheduling has no actual effect. It is a comprehensive application of data scheduling and manual scheduling. The data application platform can process and analyze the data and dig out the data information such as the individual characteristics, behavior mode, interpersonal network and activity law of the criminal subject. However, if such data information is to be used as criminal evidence, it must be mutually confirmed with the investigation resources obtained from the actual investigation activities. In other words, the investigation method of data-driven investigation mode should be implemented in physical space and real

space. The data processing of the data application platform is based on the correlation between the data. The result is a probability deduction. Its strength lies in the prediction of the development of things, and its weakness lies in the identification of deterministic and unique facts.

#### 4.2 Balance the relationship between investigation efficiency and citizens' right to privacy

Data driven investigation mode can improve the application ability of data information and reduce the cost of data information processing, but the protection of citizens' privacy will be ignored in the process of data information collection, mining and analysis. Citizens are under omni-directional, multi-dimensional and panoramic information control, and personal privacy is at stake with the improvement of big data application ability and the expansion of application fields. The data-driven investigation mode not only improves the investigation efficiency, but also brings the risk of violating citizens' privacy. In the era of big data, information acquisition methods are faster and more efficient. Citizens' personal information is scattered in all corners of society. The personal information data collected by investigation organs are more comprehensive and more aggressive than traditional methods. In order to ensure the smooth development of data-driven investigation and maintain the balance between investigation efficiency and citizens' right to privacy, it is imperative to establish the principle of proportion with the lowest degree of damage to citizens' right to privacy. We should formulate specific rules for the collection and use of data and information related to citizens' privacy, and standardize the behavior of investigative organs in collecting and using citizens' personal information.

#### 4.3 Rationally understand the application of big data in investigation

In the era of big data, digital ecology has become a trend. It is undeniable that the intelligence and efficiency of big data have brought convenience to investigation activities. However, big data is not perfect, and it also has something that can not be done. In other words, we should rationally understand the application of big data in investigation. First, recognize the shortcomings of big data application in investigation. The strength of big data lies in its grasp of the development trend of things and its ability to make overall and macro analysis. The weakness lies in the identification of certainty and uniqueness of things. Second, recognize that big data is not omnipotent, and there is something that can not be done. For some crimes, such as impulsive crime or passionate crime, the psychological activities and motive of the perpetrator at the time of committing the crime are areas that cannot be explored through big data technology. In addition, big data also has technical bottlenecks, which makes it impossible to effectively analyze and process all types of data resources.

## V. CONCLUSION

The new, intelligent and dynamic characteristics of criminal crime in the new era show that China's traditional criminal investigation model can no longer meet the needs of criminal investigation in the era of big data, so it is urgent to change in time. Data driven investigation mode, a modern investigation mode dominated by data, can not only meet the needs of combating crime in the new era, but also improve the efficiency of criminal investigation and better serve criminal investigation and crime prevention. The emergence of electronic forensics measures and the development of big data technology provide dual

assistance for the smooth operation of data-driven investigation mode. It should be noted that traditional criminal investigation methods still have practical significance, and big data technology is not the master key. In the investigation practice, we should strive to realize the integration and complementarity of traditional investigation means and big data technology, and build a more perfect modern investigation mode.

## REFERENCES

- [1] Jiang Wei, Fang Binxing, Tian Zhihong. Network Security Evaluation and Optimal Active Defense Based on Attack Defense Game Model. *Acta Computer Sinica*, 2009, 32 (004): 817-827
- [2] Miao Yongqing. Stochastic Model Method and Evaluation Technology of Network Security. *China Science and Technology Investment*, 2017, 4: 314
- [3] Yi Hua Zhou, Wei Min Shi, Wei Ma. Research on Computer Network Security Teaching Mode for Postgraduates Under the Background of New Engineering. *Innovation and Practice of Teaching Methods*, 2020, 3 (14): 169
- [4] Bao Xiuguo, Hu Mingzeng, Zhang Hongli. Two Quantitative Analysis Methods for Survivability of Network Security Management Systems. *Acta Communication Sinica*, 2004, 25 (9): 34-41
- [5] Yang Yi, Bian Yuan, Zhang Tianqiao. Network Security Situation Awareness Based on Machine Learning. *Computer Science and Application*, 2020, 10 (12): 8
- [6] Li Zhiyong. Hierarchical Network Security Threat Situation Quantitative Assessment Method. *Communication World*, 2016, 23: 70-70
- [7] Hu Wenji, Xu Mingwei. Analysis of Secure Routing Protocols for Wireless Sensor Networks. *Journal of Beijing University of Posts and Telecommunications*, 2006, 29 (s1): 107-111
- [8] Wei Yonglian, Yi Feng, Feng Dengguo, Yong W, Yifeng L. Network Security Situation Assessment Model Based on Information Fusion. *Computer Research and Development*, 2009, 46 (3): 353-362
- [9] Xu Guoguang, Li Tao, Wang Yifeng. A Network Security Real-time Risk Detection Method Based on Artificial Immune. *Computer Engineering*, 2005, 31 (12): 945-949
- [10] Li Weiming, Lei Jie, Dong Jing. an Optimized Real-time Network Security Risk Quantification Method. *Acta Computa Sinica*, 2009 (04): 793-804