

Research Article

Applications of Machine Learning in Public Security Information and Resource Management

Zhihui Wang  and Jinyu Wang

People's Public Security University of China, Beijing, China

Correspondence should be addressed to Zhihui Wang; wangzhihui@ppsuc.edu.cn

Received 3 August 2021; Revised 6 September 2021; Accepted 1 October 2021; Published 12 October 2021

Academic Editor: Mian Ahmad Jan

Copyright © 2021 Zhihui Wang and Jinyu Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The data mining and big data technologies could be of utmost importance to investigate outbound and case datasets in the police records. New findings and useful information may potentially be obtained through data preprocessing and multidimensional modeling. Public security data is a kind of “big data,” having characteristics like large volume, rapid growth, various structures, large-scale storage, low density, and time sensitiveness. In this paper, a police data warehouse is constructed and a public security information analysis system is proposed. The proposed system comprises two modules: (i) case management and (ii) public security information mining. The former is responsible for the collection and processing of case information. The latter preprocesses the data of major cases that have occurred in the past ten years to create a data warehouse. Then, we use the model to create a data warehouse based on needs. By dividing the measurement values and dimensions, the analysis and prediction of criminals’ characteristics and the case environment realize relationships between them. In the process of mining and processing crime data, data mining algorithms can quickly find out the relevant information in the data. Furthermore, the system can find out relevant trends and laws to detect criminal cases faster than other methods. This can reduce the emergence of new crimes and provide a basis for decision-making in the public security department that has practical significance.

1. Introduction

With the accelerated development of modern technologies like communication, computers, and big data (known as high-tech), now other data mining and security technologies are increasingly used in our daily life, thus driving us into an information society [1]. However, illegal crimes have also shown new characteristics. It can be said that modern illegal and criminal acts are already in a period of rapid development. It is not difficult to find that, with the improvement of professionalization and high technology, cybercrimes, high-tech crimes, and so on obviously have the characteristics of the times. New criminal methods and forms of crime are also constantly being updated. In this new situation, the state has put forward two clear goals, namely, (a) request police force from science and technology and (b) revitalize police through science and technological education [2]. At

the end of the last century, under the modern economic and social conditions, to achieve dynamic management and combat crime, the Ministry of Public Security has not hesitated to put forward the slogan “Strong Science and Technology,” cultivate rapid response capabilities, and formulate comprehensive start-up coordination. Under the new situation, our operational strategy requires that we make full use of technology, strengthen the police’s fight against crime, and improve the solution efficiency of public security work [3].

In the past ten years, the development process of the informatization and construction of public security organs is obvious to all, and progress has been made by leaps and bounds. The establishment of a public security informatization network that is all-round and not leaking is “vertical and horizontal to the end” [4]. All police types and businesses have also fully implemented information management. Similarly,

the accumulation of a large amount of business data and some comprehensive applications are also underway. The main advantage of information management lies in the use of advanced computer technology to effectively manage existing information through certain methods, which in turn improve law enforcement efficiency and prevent new illegal behaviors. Finally, it can be very powerful to combat and curb criminal acts. However, at present, the intelligent analysis function is still almost in a blank state, so the massive business data accumulated in the work is limited to simple primary applications such as query, update, and statistics [5]. Hence, how to use the data mining technology to discover and use the hidden regular information behind these data to serve various management works of the department and even provide scientifically valuable evidence for the leadership to make decisions can be seen from this. Moreover, providing the police with some useful technical support and reliable references is an important subject that we will explore to solve these problems.

In this paper, a public security information analysis system, based on the data mining technology, is proposed that consists of two modules: (i) case management module and (ii) public security information mining. The former module is responsible for the collection and processing of case information, and the latter module preprocesses the data of major cases that have occurred in the past ten years to create a data warehouse. By dividing the measurement values and dimensions, the analysis and prediction of the criminals' characteristics and the case environment realizes a detailed analysis of the existing relationship between the environment and the type of case. This paper studies how to use the data mining technology to analyse the information of criminals and find laws and trends of criminal behavior, which is very important for improving (i) the ability of antiterrorism decision-making, (ii) commanding, and (iii) levels of comprehensive application of online information and (iv) strengthening the construction of a modern public security prevention and control system. The principal contributions of our work are as follows:

- (i) We take the outbound and case data in the police records
- (ii) Through multidimensional data modeling and preprocessing (big data), we obtain some useful information from the data
- (iii) A police data warehouse is constructed, and a public security information analysis system is proposed, based on the data mining technology
- (iv) A clustering algorithm divides instances into natural groups and distinguishes the hidden classes in the data instead of using predicted instance classes
- (v) We use the proposed model to create a data warehouse based on the users' needs

The rest of the paper is organized as follows: In Section 2, we briefly discuss the related literature review. In Section 3, we analyse the relevant technologies of public security information systems. In Section 4, we describe the public security police information platform system and result

analysis. Finally, Section 5 illustrates the final thoughts and several directions for future research.

2. Related Work

The method of clustering in public security data has been widely used. For example, Sun and Scanlon [6] have realized the related modeling of the criminal behavior of sexual assault through the method of clustering and self-organizing mapping. Gupta et al. [7] have figured out a way to cluster the investigation reports of crimes to achieve the purpose of unearthing the criminals of the case. The system can even analyse the criminal life of criminals. It not only analyses the correlation between cases and criminals but also realizes the establishment of visual crime classifications through clustering methods, and it can also realize construction by analysing specific criminal gangs. The main basis for the filing of criminals is the duration, severity, frequency of crime, and the nature of the crime. A relatively new comparison method is used to compare the similarity of all criminals, then comprehensively consider the differences in the four aspects to generate a variable distance matrix that can describe the crime career, and finally realize the clustering analysis.

In the past period, because criminal network analysis is a very specific case, it has attracted a lot of attention. Subsequently, the conceptual space method was also proposed. The purpose of this method is to extract the relationship between crimes from the summary of the case and generate a similar network of suspects. Through the weight of cooccurrence, by realizing the calculation of the relative frequency of the suspect appearing in the same case at the same time, the measurement of the strength of the relationship between the two cases can be realized. Through the hierarchical clustering method, the criminal network is divided into many subnetworks, and the interaction mode between the subnetworks is determined by the method of block modeling. Through the related measurement of the centrality, closeness, and proximity of the network, the important members of the criminal group can be found out. "CrimeNetExplorer" is a visualization framework as well as an automated criminal network analysis process [8]. Among them are the main stages related to criminal network analysis, such as structured analysis, criminal network creation, network division, and network visualization.

2.1. Public Security Information System. The research on data mining technology in the field of public security started relatively late. But in recent years, with the advancement of the processing police information, the number of police officers has increased, and police work has begun to use data mining and other related technologies. Kaur et al. [9] suggested the method of establishing a public security data warehouse, discussed how to mine public security data, and realized the relevant analysis of data mining and the overall framework of the public security data warehouse. Hossain et al. [10] defined the relevant attributes of the case, applied

the correlation matrix, and realized the discussion of the relevant methods through the discovery model of the relevant cases. Aiming at some relatively small-scale criminal organizations, Li and Cui [11] have developed a way to explore the relationship between criminal organizations based on social networks. The relationship mining of criminal organizations is mainly to analyse the relationship between the members of the criminal organization and determine the key personnel in the criminal organization.

2.2. Data Mining Technology. Through the related academic research on data mining, the police department's ability to enforce law and combat terrorism will be greatly improved. Based on the specific characteristics of crime and related security tools, data mining technologies mainly include the following: information sharing and collaboration, intelligent text mining, security association mining, classification and clustering, spatial and temporal crime pattern mining, and we analyzed criminal/terrorist networks in six areas [12]. Although the research on the use of data mining technology in police work has been for a long time, there are not many related documents, and there are not many related research results. There are even a lot of relevant data mainly for studying small areas, there is no cooperative relationship between various studies, and the application is mainly realized through some experiments [4].

3. Analysis of Relevant Technologies for the Public Security Information System

3.1. Overview of the Data Mining Technology. Data mining is also called analysing data, which uses semiautomatic or automated tools to mine data and knowledge. The process of data mining is also more complicated. From the database storing large amounts of data, multiple steps, such as practical, previously unknown, and usable knowledge, are mined. Each step constitutes a complete process. Finally, the knowledge is used to make reasonable judgments and scientific decision-making. The main process of data mining is shown in Figure 1.

The general process of data mining technology is described in Figure 1. In the whole data mining, the business object is the basis to be studied, and data mining is carried out around the business object. If researchers can focus on the research business object, dig out the results and verify the accuracy of the results and then data mining can be completed correctly [12]. The contents of each step in the data mining process are as follows.

3.1.1. Analysis of Business Objects. First, it is necessary to clearly know what the goal of data mining is, and then the premise is to clearly analyse the business problems. Although it is impossible to predict the results of data mining, the problems to be studied and analysed are obvious. Therefore, you must be familiar with the business objects, clarify the problems that need to be explored, and plan a general direction.

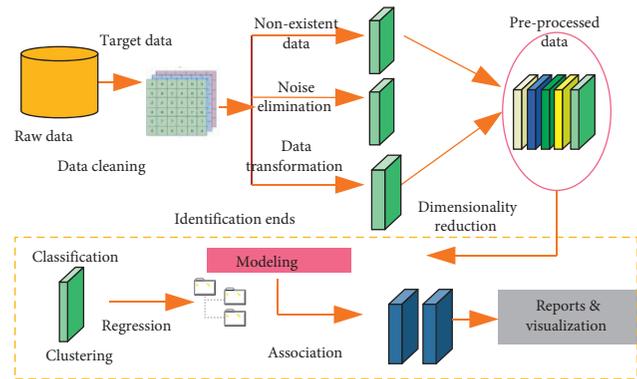


FIGURE 1: Process of data mining.

3.1.2. Data Preparation

- (i) Data selection: collect all the information of related business objects, including internal information and external information, select data suitable for data mining from this information according to needs, and make a good choice on the data.
- (ii) Data preprocessing: after screening the required data, estimate the quality of the data, compress the data set as much as possible, then sample the data, propose the types of mining operations that need to be applied, and make full preparations for the next analysis.
- (iii) The purpose of data conversion is to unify the data types of the source data and the way in which the values exist. For example, convert continuous data into discrete classes and convert Boolean types into integer classes.

3.1.3. Data Mining. After getting the converted data, the next step is to mine. According to specific actual requirements, using appropriate algorithms to mine, it will automatically finish the rest of the work and finally wait for the result [13].

3.1.4. Analysis of Results. Analyse the results, evaluate, and draw conclusions. The analysis method generally selects visualization technology and outputs in the form of text or graphics. This is determined by the operation set by the user.

3.1.5. Application Integration. Based on the knowledge of the above steps, it can be found that if the integrated knowledge is integrated into the organizational structure and then put into the relevant business information system, then the business information system can realize application intelligence [14]. The above steps are completed in stages, each stage requires good professional staff in all aspects to achieve, and these personnel can be roughly divided into three categories:

- (1) Proficient business personnel: such personnel are required not only to be quite proficient in business

and analysing business objects but also to put forward business requirements based on the characteristics of each business object and to mine them in a targeted manner so that business problems can be classified as data information.

- (2) Data proficient personnel: this type of personnel can freely apply mathematical knowledge, can analyse business technical knowledge data by themselves, and can effectively extract value data information and transform business requirements into data mining for each step; you can also match the corresponding technology for the operation of each step [15].
- (3) Proficient in data management personnel: this type of personnel is good at mastering data management technology; they can collect various data to form a data warehouse and can effectively use it.

Therefore, it can be known from the above that the process of data mining is a process of collaboration among professionals, and at the same time, it can be known that it is a highly integrated and high-investment technical process. Data mining needs to determine the requirements for data definition and selection of mining algorithms by determining each business object, analyse the technology and business of each stage of data mining, make timely adjustments, and finally make a scientific, clear, and reasonable mining result explanation. At present, the peak of informatization development is data mining, which also reflects the value of the highest application point of informatization.

3.2. Clustering Algorithm. In data mining technology, the clustering algorithm can be regarded as a very practical algorithm. From the current point of view, it is a well-deserved cluster analysis that is widely used in most applications, such as image processing, pattern recognition, data processing, and market research [16]. As a large piece of data mining technology clustering algorithm, it divides instances into natural groups and then distinguishes the hidden classes in the data, instead of using predicted instance classes. Cluster analysis can be used as an independent tool to obtain the approximate distribution of data and analyse each collection of data that combines similar properties. These collections can also be called clusters. Clustering analysis can map out some of the operating mechanisms of some instances in certain fields and create different connections and different relationships between the instances.

In the clustering algorithm, it is mainly necessary to determine the standard to measure the cluster centre and similarity. The clustering algorithm is different from the classification in data mining technology. The classification is the class stored in the known target database. What needs to be done is to extract and classify each record; the difference but similarity is that in the cluster. In the class algorithm, clustering is run without knowing the number of clusters in the target database. Its purpose is to classify all data. Under this classification, standards are specified based on attributes so that these data can be clustered. Minimize between and

maximize between different clusters. In fact, the similarity of most class algorithms in clustering algorithms is related to distance. The reason is that there are quite a lot of data types in the database, so how to measure the distance between two nonnumeric fields? The discussion on this issue is also very intense, and many researchers have proposed similar algorithms. Each cluster obtained by cluster analysis can be treated uniformly in many applications. Clustering analysis algorithms can be classified into hierarchical method, split method, density-based method, and model-based method, and so on.

Generally, N -dimensional “space” is involved when using cluster analysis. This space can be used to solve the measurement problem. When performing data mining on clusters in N -dimensional space, the first thing to do is to survey the distance between the data and the data. Commonly used measurement methods include Minkowski distance, Manhattan distance, and Euclidean distance [17]. The measurement method of Minkowski distance is as follows:

$$s(a, b) = \|X_a - \bar{X}_b\|^{1/r} = \left((x_a^1 - \bar{x}_b^1) + (x_a^2 - \bar{x}_b^2) + \dots + (x_a^n - \bar{x}_b^n) \right)^{1/r}, \quad (1)$$

where $a = x_a^1, x_a^2, \dots, x_a^n$ and $b = x_b^1, x_b^2, \dots, x_b^n$ represent n -dimensional data objects; when the database represents the i^{th} record, there are n fields. Obviously, when using formula (1) to calculate the distance, to make the field meet the requirements, these fields must be processed. Of course, some applications do not make too many requirements for these fields. When r in formula (1) is 1, the distance is also called Manhattan distance. When the value of r is 2, the distance is called Euclidean distance. In the clustering algorithm, the weight of some data depends on the weight and lightness for a specific situation; for example, when it is necessary to implement cluster analysis of lost customers, the time since the last time the customer purchased the product should be given a very heavyweight. The weighted distance is shown in formula (2). Calculation method is as follows:

$$s(a, b) = w \|X_a - \bar{X}_b\|^{1/r} = w_1 \left((x_a^1 - \bar{x}_b^1) + w_2 (x_a^2 - \bar{x}_b^2) + \dots + w_n (x_a^n - \bar{x}_b^n) \right)^{1/r}, \quad (2)$$

where w represents $\|X_a - \bar{X}_b\|^{1/r}$, the weight of the total distance; when it is between 0 and n , the sum of all weights is 1. The clustering methods we generally use often include hierarchical clustering, grid clustering, partition clustering, density clustering, and simulated clustering.

3.3. Decision Tree Algorithm. Decision tree refers to the use of a tree structure to represent the decision set or the classification of data according to different data characteristics. The law is generated through data and the law is discovered. This is an effective method of supervising and inductive learning. In other words, the decision tree is a tree

structure to represent the decision-making process, which can show the rules of what value should be obtained under certain similar conditions. Under normal circumstances, one event can cause two or more events and get different results. From this feature, the structure of the decision tree is from top to bottom, which is like a flowchart [18]. The top node of the tree is the beginning of the entire decision tree, also called the root node. Each branch of the tree represents a new decision output, and the child nodes on each node represent related attribute tests. We use decision-making related algorithms to pass and derive the number of child nodes.

The key to the problem is how to construct a decision tree model, that is, to build a decision tree. The process is roughly divided into two stages: the first is the tree-building stage, which is also called the recursive process, to obtain a tree; the second is the pruning stage, which aims to reduce the fluctuations caused by the noise in the training set.

3.3.1. Building a Tree. Usually, an important index to measure the quality of a node's split is based on the information gain. If a split has the highest information gain, then this is a split plan. Shannon proposed the theory of information and defined information (i.e., the amount of information) and Entropy as shown in

$$\text{Inf}_i = \log\left(\frac{1}{p_i}\right), \quad (3)$$

$$\text{Ent}_i = \sum_i p_i \log\left(\frac{1}{p_i}\right). \quad (4)$$

The term refers to a weighted average of the information volume of the system, that is, the average information volume of the system, and the information volume is the principle of the information gain index. Because the tree-building algorithm is a recursive process, only one of the splitting methods is needed to discuss and study the specific node N . Suppose that the training set pointing to N is S . This training set implicitly contains m different classes, distinguishing different classes C ($i = 1, 2, \dots, m$). Let S_i be the number of like data in S ; before splitting, the original is

$$E(c_i) = \sum_i p_i \log\left(\frac{1}{p_i}\right), \quad (5)$$

where p_i is any sample of the probability belonging to C_i . Since the information is encoded in binary, a logarithmic function with base 2 is set. According to the above formula, it is easy to get the total number, which is a weighted average.

3.3.2. Pruning. The most used method in pruning is the statistical method, which cuts off some branches that are not edged or even noise. There are many methods of pruning, usually the following two methods:

- (i) *Synchronous Pruning.* When building a tree, if certain requirements are met, such as information gain or effective statistics reaching a preset threshold, the

node stops splitting, and the internal node is regarded as a node on a leaf. Take the class with the highest frequency in the subset as the sign of the leaf node's self, and then store these instance probabilities to distribute the function.

- (ii) *Hysteresis Pruning.* When building a tree, when the independent data in the training set is included in the decision tree and reaches the node if the class label of the training data is different from the class label of the leaf node, then it is called classification error. After the establishment of the tree is completed, the algorithm calculates the probability of each possible error branch passing through each internal node by weighted average and calculates the error rate instead of cutting the node. Because clipping can reduce their error rate, it is necessary to cut all branches under this node. This node is called a leaf. The error rate can be used to verify the test data independently included in the training set data, and the result is a decision tree with a minimized error rate.

4. Results Analysis

Under the protection of the security guarantee system, the public security police information platform comprehensively uses mainstream IT technologies such as server virtualization technology and middleware technology and builds a portal site, web application service layer, and database service layer based on the service architecture (design model), data storage, and backup layer and other four-layer structures. In Figure 2, the portal website is the unified entrance to the public security police information network and is an application system leading to public security information resources. The police have passed unified identity authentication and access control access to information resources with the single sign-on management system (PKI/PM) [18]; the web application service layer provides the web services and application servers required by the police information system. The web server will be based on the specific business needs of the police system. The functional components of the business logic layer are packaged into web services, and the web services provide business logic functions to the presentation layer for users to access; the database service layer provides the operating environment for the database system, and a well-designed database can ensure the stability and reliability of the system. The operation of the data storage and backup layer provides a unified storage, backup, and recovery management functions for the public security information resource database. The data lost in the system can be restored in the shortest time. It is composed of storage area network, cluster technology, dual-system hot backup, storage management software, and other components.

The security assurance system includes equipment security, supporting software security, network system security, application system security, data transmission and reception security, and computer room environment security. The construction of a security protection system must

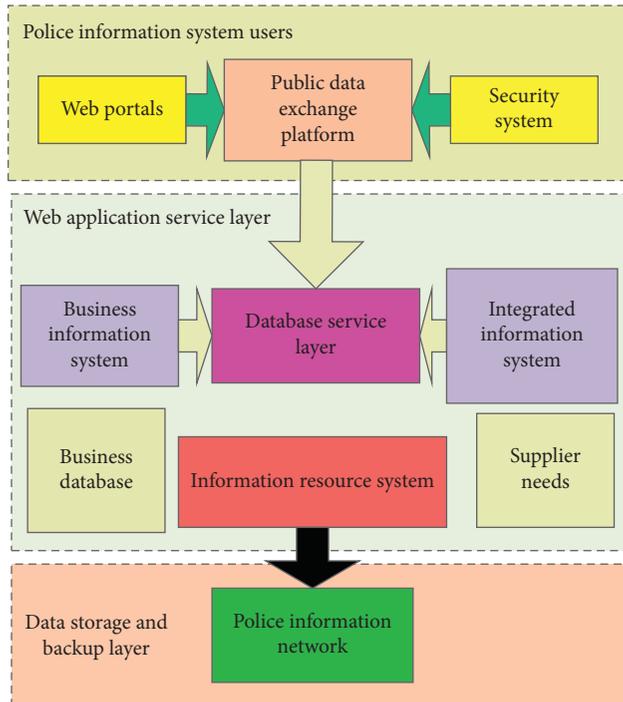


FIGURE 2: Public security police information platform system structure.

proceed from all aspects, closely combining organization, strategy, operation, and technology to form an all-round and integrated security protection system, so that the security of the police information platform can be truly guaranteed.

This system uses data mining technology to establish a data mart. The data comes from the information system of various functional departments of the public security organs and the criminal public security information analysis system. Based on the characteristics of the public security business, we use the very powerful SQL Server 2012 to implement the data processing process [19]. The SQL Server 2012 structure level type and new aspects, functions, and improvements can achieve multiple levels even in the same dimension. The establishment demonstrates multiple face-to-face dimensions, improves the ability to analyse multidimensional data sets, and improves the effect of multidimensional analysis. Microsoft SQL Server 2012 is an engine that provides a relational database, which can realize the storage of related case information and the establishment of a data warehouse. Microsoft SQL Server 2012 Integration Services development kit can load data, convert data, extract data, and other functions. Microsoft SQL Server also has the function of analysing data. It puts the data in the data warehouse through a series of processing and then puts it into a multidimensional cube so that analysts can easily analyse and query the data. The software system is implemented using Java programming language, MyEclipse is used for development tools, and SQL Server 2012 is used for storage database.

4.1. Data Preprocessing. The role of ETL (Extraction-Transformation-Loading), that is, extraction, transformation, and loading, is to select and extract data from messy,

inconsistent, and distributed data sources (including plane data, relational data, and logical data) to temporary layer, adjust dirty data, clean lengthy, convert format, integrate data, and finally load and unload the processed results in the target data warehouse. ETL processed result set provides a basic guarantee for online processing and data mining, whether ETL Success is related to the success or failure of the entire project, and it is also the most critical part of the project. According to the current experience of building data warehouses, when mining data warehouses, the entire process of ETL generally takes up most of the time, and when the amount of data is large, it may take longer. Therefore, we should pay more attention to it.

The entire process of ETL takes a long time and is very complicated, so the process needs to be managed. Management includes a series of operations such as ETL scheduling, error handling, management, and logging. At present, most ETL tools manage the process. To ensure high-efficiency operation, ETL generally runs in an automated manner in the background, so reasonable planning must be made. If the process fails, then we must intervene artificially. Therefore, both management and scheduling are particularly important in the entire ETL process. The Microsoft SQL Server 2012 Integration Services (SSIS) used in this paper is a relatively common modern ETL software that can produce a platform such as the main solution for high-performance data warehouses (including data warehouse conversion, extraction, and loading (ETL)) [20].

4.2. Detection of Crime Distribution in Public Security Information City. The type and quantity of urban crimes reflect the state of urban social security to some extent. When the social security situation occurs or is about to change, the distribution type and value of urban crime will also change accordingly. To conduct a comprehensive analysis of the urban social security situation and judge the current social security situation and changes, it is first necessary to pre-process the original data, extract valuable information from it, and classify redundant data and meaningless statistics. Eliminate and obtain the category that accords with the objective law, then find the characteristic that can describe the city's comprehensive public security situation in the complicated urban crime statistics, and analyse and extract the inherent characteristics implied by the crime data. This is the basis and premise of the relevant research in this thesis, and its quality is directly related to the effect and accuracy of the judgment of the social security situation.

This paper designs the following experiments to analyse the urban crime case data: (1) analyse and calculate the distribution characteristics of the annual case data of a certain type of case; (2) calculate the statistical distribution characteristics of the monthly case data; (3) calculate the statistical distribution characteristics of monthly cumulative case data; (4) comparatively analyse the data characteristics.

Since the number of days in 12 months is different, this paper takes 30 days as a standard month. Part of the original data is shown in Figure 3. In the experiment, the distribution characteristics of theft cases within 30 days were calculated

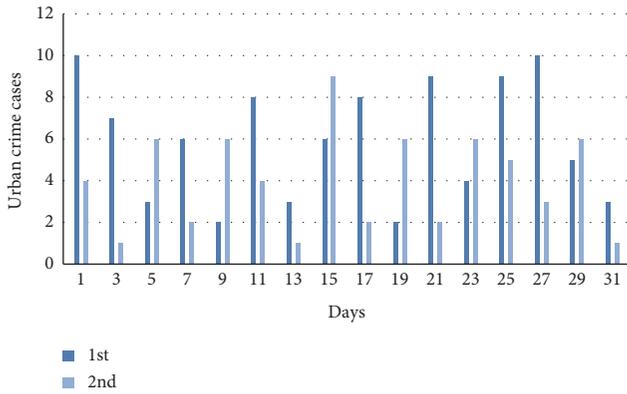


FIGURE 3: Number of crime cases in 1st and 2nd quarters.

by sliding with 15 days as a unit from 30 days. The experimental results are shown in Figure 4. In the experiment, it was calculated every 30 days (the last month was set to 36 days). The distribution characteristics of the accumulated data of theft cases and the experimental results are shown in Figure 5.

Through data processing and calculation, the chart shown in Figure 6 can be obtained, which can be compared with the daily statistical value, the statistical analysis value in units of 30 days, and the incremental statistical analysis value every 30 days. Figure 6 shows the total number of thefts in a city, as of records from the year 2019, and the mean values that are mapped to a Poisson distribution. Our investigation and observations from the datasets are discussed in the following paragraphs.

From the data analysis in the figure, we obtained the following observations and findings:

- (i) Overall, the average Poisson distribution of crime incident data is relatively stable, fluctuating slightly around 5.12. In terms of the incident data of a certain type of case, the Poisson distribution better describes its mathematical characteristics, the distribution characteristics of different periods remain relatively stable and are affected by the crime incident data for a period, and the performance is a small fluctuation. Therefore, the Poisson distribution of urban crime statistics is used to describe its internal law, and the mean of the Poisson distribution over a certain period is used to quantitatively describe its characteristics.
- (ii) The change in the mean value of the Poisson distribution describing the crime distribution is not sensitive to the statistics at a specific moment (e.g., there were 18 high-incidence cases at time 12, and the mean value of the Poisson distribution was basically unaffected), which shows that the characteristic data is not sensitive to occasional disturbances and changes accordingly only when the signal lasts for a period. Therefore, using the mean value as a feature can better prevent the occasional disturbance statistic from interfering with the feature, and the mean value can better reflect the overall trend of the statistic.

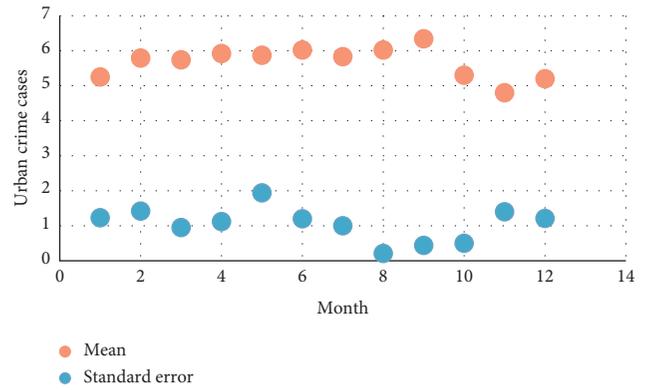


FIGURE 4: Average distribution of theft cases per 30 days in a city in 2019.

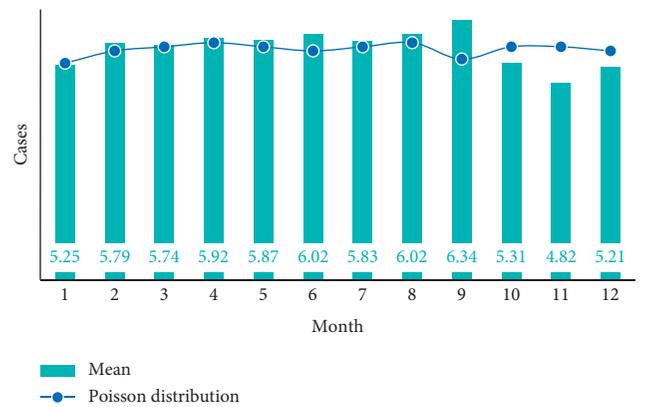


FIGURE 5: The average Poisson distribution of cumulative thefts per 30 days in a city in 2019.

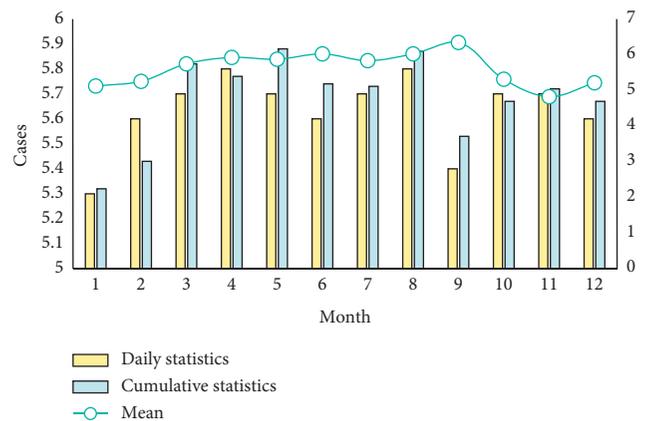


FIGURE 6: The number of thefts in a city in 2019 and the mean value of the Poisson distribution.

- (iii) The variation range of the historical cumulative statistics of the Poisson distribution is much lower than that of the monthly statistics [21]. This shows that as the time range increases, the Poisson distribution tends to a certain steady state. However, within a certain period, its distribution

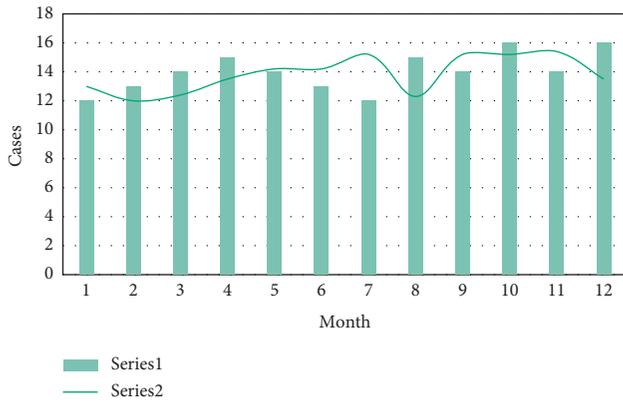


FIGURE 7: Estimated average value of theft cases in 2008 in 30 cycles.

characteristics have changed significantly. This puts forward requirements for how to determine the period for calculating the eigenvalues of the Poisson distribution.

4.3. Experimental Detection of Feature Vector of Decision Tree Algorithm. The contradiction between the new requirements of public security information construction and the traditional business data separation model is increasingly prominent under the background of “intelligence information dominates police affairs.” Building a comprehensive application platform for police information that collects various types of police, various businesses, and various types of data has become one of the new goals of public security information construction. In this general environment, various types of data are highly concentrated and highly integrated, which provides a good data foundation for public security information analysis and research and judgment. Compared with the previous comprehensive analysis and intelligence research and judgment, the allocation of research and design costs has changed a lot [22].

Unification, data transmission security, centralized collaboration of distributed data, and other important contents in the past information analysis and judgment almost no longer need to be considered under the new application platform, and most of its main energy and resources can be transferred to the analysis and processing of information; it provides great convenience for large-scale data processing and large-scale information analysis. According to the foregoing, the following experiments are carried out on the statistics of the high incidence of urban crime cases: (1) Starting from January 1, 2019, at intervals of 5 days, the mean of the Poisson distribution fitting is performed on the 30-day case data. (2) Starting from January 1, 2019, with 30 days as the base, 10 days of data are added for each statistics, and then the mean value of Poisson distribution fitting is performed.

The experimental results are shown in Figure 7, combining the experimental results of the previous experiment (for theft case statistics), it can be found that the average data of the crime statistics distribution basically stabilized after

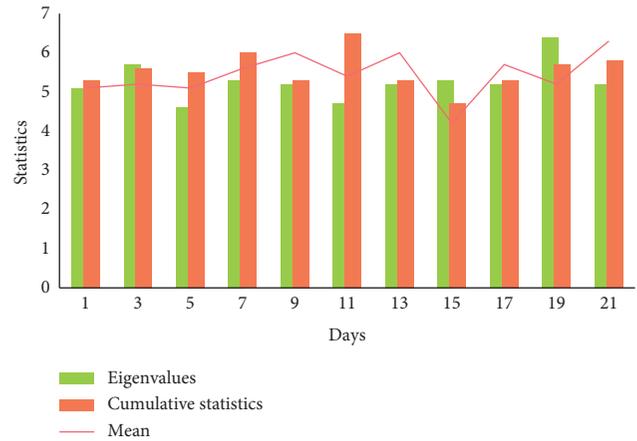


FIGURE 8: Three types of statistical mean comparison chart.

about 30 days after the start of the statistics, and after 60 days, the average was approximately a straight line. At this point, it can be considered that the impact of historical data on current data is small enough.

A comparison chart of statistical feature quantities of theft cases designed according to the above algorithm is shown in Figure 8. Among them, series 1 is historical cumulative data, series 2 is 30-day periodic data, and series 3 is data considering exponentially weighted decay. Obviously, most of the status of series 3 data is between series 1 and series 2; that is, series 3 is affected by current statistical values and historical statistical values when changing. Furthermore, we observed that series 3 better reflects the influence of historical data and current data on characteristic data.

5. Conclusions and Future Research Directions

The public security information analysis and mining system is an auxiliary analysis system, which can meet the relevant needs of public security and police affairs, and is a related expansion based on the data mining of case information. This paper starts from the application of data mining technology in police work, combining public security business and rich experience in case handling, to research and design a more common data mining system. The successful establishment of this system means that the investigative thinking of case-handling personnel is broader, and the case can be handled with higher efficiency than the manual system. The practice has proved that, through the technology of data mining, the relevant crime data in the public security information database can be processed and mined, and some related information contained in the data can be quickly found, trends and laws can be found, and cases can be solved as soon as possible. It can effectively reduce the probability of new crimes. It can provide relevant decision-making and basis for police work, which has very important practical significance [11, 23]. The research in this paper is the application of data mining technology in public security and the establishment of a case analysis mining system. The practice has proved that this design and

implementation method is very effective and can meet the requirements of the police in handling cases.

This type of decision is made by an experiment, but an important question is whether, overall, there is a need to design and craft secure machine learning algorithms that way which can balance three aspects that are performance overhead, security optimization, and performance generalization. Other clustering techniques along with decision support systems should be used to further investigate the validity of our findings [22, 24]. Despite the use of machine learning, simple methods like regression analysis and their impact on the proposed system will be of great importance. Deep learning methods including LSTM, CNN, and GCN would also be considered as potential future work [11]. Finally, other related datasets should be used to generalize the findings of our research.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] F. Hussain, S. A. Hassan, R. Hussain, and E. Hossain, "Machine learning for resource management in cellular and IoT networks: potentials, current solutions, and open challenges," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1251–1275, 2020.
- [2] Y. Liu, F. R. Yu, X. Li, H. Ji, and V. C. M. Leung, "Blockchain and machine learning for communications and networking systems," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1392–1431, 2020.
- [3] P. S. Bithas, E. T. Michailidis, N. Nomikos, D. Vouyioukas, and A. G. Kanatas, "A survey on machine-learning techniques for UAV-based communications," *Sensors*, vol. 19, no. 23, Article ID 5170, 2019.
- [4] M. Wang, Y. Cui, X. Wang, S. Xiao, and J. Jiang, "Machine learning for networking: workflow, advances and opportunities," *IEEE Network*, vol. 32, no. 2, pp. 92–99, 2017.
- [5] U. Challita, A. Ferdowsi, M. Chen, and W. Saad, "Machine learning for wireless connectivity and security of cellular-connected UAVs," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 28–35, 2019.
- [6] A. Y. Sun and B. R. Scanlon, "How can big data and machine learning benefit environment and water management: a survey of methods, applications, and future directions," *Environmental Research Letters*, vol. 14, no. 7, pp. 73–75, 2019.
- [7] R. Gupta, S. Tanwar, S. Tyagi, and N. Kumar, "Machine learning models for secure data analytics: a taxonomy and threat model," *Computer Communications*, vol. 153, pp. 406–440, 2020.
- [8] J. Li, Z. Zhao, and R. Li, "Machine learning-based IDS for software-defined 5G network," *IET Networks*, vol. 7, no. 2, pp. 53–60, 2018.
- [9] P. Kaur, M. Sharma, and M. Mittal, "Big data and machine learning based secure healthcare framework," *Procedia computer science*, vol. 132, pp. 1049–1059, 2018.
- [10] E. Hossain, I. Khan, F. Un-Noor, S. S. Sikander, and M. S. H. Sunny, "Application of big data and machine learning in smart grid, and associated security concerns: a review," *IEEE Access*, vol. 7, pp. 13960–13988, 2019.
- [11] C. Li and J. Cui, "Intelligent sports training system based on artificial intelligence and big data," *Mobile Information Systems*, vol. 2021, Article ID 9929650, 11 pages, 2021.
- [12] Y. Wu and F. Meng, "Categorizing security for security management and information resource management," *Journal of Strategic Security*, vol. 11, no. 4, pp. 72–84, 2018.
- [13] Y. Wu, "Developing a taxonomic framework of security methods for security management and information resource management," *Journal of Strategic Security*, vol. 13, no. 2, pp. 64–77, 2020.
- [14] Y. Wu and F. Meng, "Creating a security methods taxonomy for information resource management," *Digital Libraries at the Crossroads of Digital Information for the Future*, vol. 1, pp. 293–297, 2019.
- [15] W. B. Daoud, M. S. Obaidat, A. Meddeb-Makhlouf, F. Zarai, and K. F. Hsiao, "TACRM: trust access control and resource management mechanism in fog computing," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, pp. 1–18, 2019.
- [16] S. Hashemi and M. Zarei, "Internet of Things backdoors: resource management issues, security challenges, and detection methods," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 2, Article ID 4142, 2021.
- [17] B. Cooper, J. Wang, T. Bartram, and F. L. Cooke, "Well-being-oriented human resource management practices and employee performance in the Chinese banking sector: the role of social climate and resilience," *Human Resource Management*, vol. 58, no. 1, pp. 85–97, 2019.
- [18] D. Wang, N. Zhao, B. Song, P. Lin, and F. R. Yu, "Resource management for secure computation offloading in software-defined cyber-physical systems," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 9294–9304, 2021.
- [19] H. Zhu, "Research on human resource recommendation algorithm based on machine learning," *Scientific Programming*, vol. 2021, Article ID 8387277, 10 pages, 2021.
- [20] S. S. Gill and R. Buyya, "SECURE: self-protection approach in cloud resource management," *IEEE Cloud Computing*, vol. 5, no. 1, pp. 60–72, 2018.
- [21] A. Keshavarzi, A. T. Haghghat, and M. Bohlouli, "Adaptive resource management and provisioning in the cloud computing: a survey of definitions, standards and research roadmaps," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 11, no. 9, pp. 4280–4300, 2017.
- [22] M. Zakarya and L. Gillam, "Modelling resource heterogeneities in cloud simulations and quantifying their accuracy," *Simulation Modelling Practice and Theory*, vol. 94, pp. 43–65, 2019.
- [23] X. Yu, Q. Hu, H. Li, J. Du, J. Gao, and L. Sun, "Cross-domain recommendation based on latent factor Alignment," *Neural Computing & Applications*, vol. 12, 2021.
- [24] T. V. Pai and P. S. Aithal, "A review on security issues and challenges in cloud computing model of resource management," *International Journal of Engineering Research and Modern Education (IJERME)*, vol. 2, no. 1, pp. 65–70, 2017.