

Research on the Application of Big Data in Security Information Collection

Chai Gao

Political College of National Defence University Xi'an, China

3237610382@qq.com

Keywords: Big data; Security; Information collection; Research

Abstract. With the increase of urban population and the increase of mobility, it is significant to strengthen the management of urban personnel information for public security, prevention of key areas and maintenance of stability. Automatic collection of dynamic information, such as data and video, is carried out on the relevant personnel and vehicles in large event venues, resident entrances and exits, and through the reading of the identity card information of the target personnel, and the collision with various types of information data in the background to return the check results. The dynamic information collected will be compared with databases such as fugitive personnel and key personnel and human image comparison systems, actively identify and control dangerous individuals and groups, and control unsafe factors in a timely manner to become the focus of the security work at the scene of control activities. In order to effectively solve many problems in urban public security prevention and control, criminal investigation, and traffic management, in recent years, there have been problems such as the bayonet system and the license plate number retrieval system for personnel and vehicle monitoring, and effectively controlled the security of group activities. "Science and technology security" has become an essential tool for modern policing. Although this kind of system has been implemented for many years, the complete set of information management systems for personnel background information, vehicle control and personnel identification is almost zero so far, resulting in public security personnel in manual investigation of suspects, such as a needle in a haystack, with heavy resistance. Therefore, the large data application platform for large-scale event security comes into being, which effectively provides favorable technical support and guarantee for the safety of large-scale event.

Introduction

The stage is as big as the dream is. This sentence describes the changes that big data has brought to police work. If you use it or not, the data is there, and how to use it well reflects the wisdom and level of the public security organs. Big data applications will be a revolution in information work. One is the dynamic analysis of variable holograms by sampling static analysis. Traditional information collection and analysis is based on traditional sampling statistical methods, and large data has changed our view of statistics. In daily police work, due to limited data collection and processing capabilities, a certain number of samples are generally randomly selected, such as 50 Internet cafes, 50 mobile populations, 50 theft cases, etc., and then static analysis. In the age of big data, instead of relying on sample survey, we can analyze the whole data dynamically, not only avoid the defects of traditional statistical methods, but also be able to get the results and find the problems more quickly. Second, information "separatist" change "unified". In the course of information system construction, avoiding information "island" is always a difficult problem. The public security business involves all aspects of social management. With the development of society, the types of data that need to be collected continue to increase. For example, the previous population system may only need to record the name of the population, age, identity card information, household registration, etc.. With the development of the times, Perhaps based on this information has not been able to find this person, but also need driver's license, license plate, social security situation, mobile phone number, even QQ information, the original system was forced to continue to expand. In the era of big data, all systems of separatism are connected through technical interfaces and dealt with

together. The value of system barriers to invisible information analysis lies in accurate prediction. The analysis of public security information department is guided by actual results. The application of large data makes the information analysis change from causal judgment to correlation analysis. More abundant sources of information, more data, and faster analysis speed, analysis of massive data through a diversified multi-factor scientific forecasting model, making analysis and analysis more accurate, more targeted disposal, not only know "what", It is also possible to find the "why" behind it and make recommendations on "how to do it".

Platform achieves its objectives.

The goal of the platform is divided into two parts, one is data warehouse construction and personnel background review, and the other is to increase technical means to detect and control dangerous vehicles and individuals and groups in a timely manner.

The goal of data warehouse construction and personnel background review is three points: First, the construction of a large data database, The data include the information base of "seven key categories of personnel", "personnel Bank of dangerous goods and explosives", "four personnel bank", "people Bank of released from prison", and the expansion of data from the city to other cities in the autonomous region and other provinces. Extraction of data from other police systems, such as: "hotel industry", "Internet cafes", "aviation", "train booking information", etc., through data warehouse excavation technology, filtering, analysis of data, and the formation of a network of personnel. The second is to solve the data interface and data request service between the system and the service department of the autonomous region, and to achieve information integration with the third police system. The third is to deploy mobile system application terminal, realize mobile personnel background information inquiry, find and control social instability factors in time.

Using image analysis technology, we can collect and analyze the vehicles of the event venue, resident entrance and exit, and stay, and control the unstable factors in the large-scale activities in time. A face recognition system is built, using static face matching and dynamic face matching, to achieve the support of human image analysis in personnel background review work and real-time comparison of blacklist libraries in active security work. Construction of vehicle identification system, using vehicle, license plate recognition technology, to achieve the venue and the surrounding vehicle identification and vehicle registration information comparison, control the vehicle dynamic information.

Basic Development Technologies

SOA Based Implementation System Design.

Encapsulate the function module into a service and abstract the service interface from it. Each service is self-registered at the start of the service, and its own service is registered on the service bus of the system. The service bus will string all services together, unified management, and provide services to the outside world. When the function modules need to be exchanged, service search and service call are performed through the service bus. In this way, the function modules are loosely coupled and can be extended to new requirements without affecting other modules. The service provider, service agent, and service requester have three operations. That is, publishing, finding, and binding interact.

Adopt B/S, C/S hybrid Architecture.

Reasonable planning of the application structure design of each subsystem, the system adopts the B / S architecture, and the C / S is the auxiliary architecture. The data acquisition control subsystem and mobile law enforcement office system are established by the stability of the data processing of C/S. Using the flexible B/S model to do statistical analysis, comprehensive query, information release and so on.

Database Technology

Data Warehouse.

The public security department must face a large amount of data every day, such as police data(population household registration data, personnel trajectory data), social administrative units, social enterprises and institutions data. These data often have problems such as complex sources, diverse formats, inconsistent, inaccurate, incomplete, and scattered storage, which brings many difficulties to management. Each business system is also independent of each other, thus forming an isolated island of information, data sharing is difficult, business decision-making lacks effective data support, and scientific decision-making is difficult to achieve. Therefore, it is necessary to establish a unified public security data warehouse, integrate all kinds of resources data, and realize centralized management of data. In the data that makes it a collaborative work among business departments, it also becomes a storage of multimedia, document materials, and policies and regulations. It has become a data market data warehouse for decision-making. The following goals must be achieved:

(1) the acquisition and exchange of heterogeneous, multisource and massive data; (2) Provide standardized maintenance management of global data and improve the level of maintenance management of data in compliance with international and national standards and industry standards and norms; (3) Data processing and quality inspection to ensure data validity, consistency and integrity; (4) Achieve the unified storage, unified management, unified retrieval, and unified display of various types of resource data, and achieve the acquisition of environmental protection information on demand; (5) Achieve data sharing and meet the needs of environmental protection departments and other units and individuals for environmental data inquiries and downloads. Data Warehouse(DW) is a relational database that is stored in a specific mode after ETL extraction, conversion, and loading of business data stored in an operational data store to facilitate multidimensional analysis(OLAP) and multi-angle presentation. Its data is based on the basic data and business data of each business system. The data in the data warehouse is detailed, integrated, theme-oriented, and at different times for the purpose of multidimensional analysis needs.

Data warehouse Logical Architecture.

The data warehouse is the data support base of this project. It is composed of source data, data conversion(ETL) and data storage management. Because most of the system's data comes from the outside and is open to external applications. The data warehouse of this project is responsible for the collection, cleaning, conversion and storage of information data from various topics such as household registration, investigation, criminal investigation, communication, and personnel movement trajectory.

Data Types.

Structured data includes relational databases and EXCEL spreadsheets. Unstructured data includes Office documents, text, pictures, XML, HTML, various reports, images, and audio/video information.

Sources of Data.

Data warehouses need to integrate data including dozens of types of information archives such as the resident population, database, personnel database and other subject data(police data: drug users, drug-related personnel, criminals, fugitives, national insurance, anti-evil, pyramid marketing blacklists, counterfeit blacklists, counterfeit money blacklists , false invoice blacklist, personnel network information. Social administrative units, social enterprises and institutions data: bank card blacklists, industry and commerce, taxation, hotel industry, Internet cafes and other information). These data are derived from various business systems, such as: resident population data from the household registration management system, fugitive data from the Ministry of Public Security's online tracking and escape system, drug users, and human intelligence data from the drug control business system. These data are automatically extracted on a regular basis through the "secure access platform".

Data Integration Concepts.

Data integration includes the collection of basic data such as structured data(derived from the system), Excel data(derived from text), and unstructured data(documents, graphs, etc.), Data

processing, data analysis and data comparison, matching, and checking are carried out to solve the problem of data redundancy and inconsistency of the same data from different businesses, and a set of public security business management databases is also formed. The data structure conforms to the requirements of the Ministry of Public Security. ETL are shorthand for data extraction, conversion, loading, and all processes such as data cleaning, integration, conversion, and loading. It is used to implement the conversion process from business system to operational data storage, operational data storage to data warehouse, and eventually form available data for query and analysis.

Data Integration Services.

The quality of ETL extraction integration data directly affects the final result display. Therefore, when establishing data integration services, we must focus on the following aspects: 1) When data extraction considers different processing methods of different data sources; 2) Well-designed rules will be used to convert and clean the extracted data, deal with redundant and ambiguous data, and make the originally heterogeneous data format unified. 3) When updating operating data storage and operating data warehouses, careful scheduling strategies must be considered; 4) For the data quality problems arising in the ETL process, there must be a guarantee means, that is, a data verification mechanism. Its purpose is to be able to monitor the data quality in the ETL process and generate rewards.

Structured and Unstructured Data Integration Ideas.

For structured data, data acquisition is automatically obtained by using data integration service. By extracting the metadata information of these databases, setting the corresponding relationship between the data source and the data collection and database, the data are extracted to the target database. Data integration can be achieved by timing and manual triggering. The data integration can mainly be based on the principle of metadata information mapping that has been obtained. The data source and the target data structure can be mapped, and the data can be calculated during the mapping process. Data validation based on metadata rules.

Determine whether the user is performing a single table extraction or all data extraction. Data extraction is tabulated. Loop each source data table to get the field mapping relationship between the source and the target. Reinsert data into the target table according to certain rules. According to the checkout rule of metadata extraction: Data extraction takes a certain area of a table as the smallest extraction unit. If it fails, data can be rolled back in the smallest unit to maintain data integrity. For the processing of NULL and empty data, unified conversion is performed, the processing is empty, and an empty record dimension is added to the project and domain dimensions. Take the -1 number for the unknown dimension, add the -1 number to the project and domain dimensions and describe it as "unknown". Data extraction integrity is the main, but if there are no fields in the original business table, data import is not allowed, and the system will give the corresponding log. Error configuring field, table data is not allowed to import. When a table fails, the program supports single table import processing. Add a table ID field to the target table to determine that the data comes from a table(easy to locate data). The "empty record" data and "unknown record" data are configured in the dimension configuration table.

Excel data integration: By setting the correspondence between the Excel file and the data table, the Excel data is read through the MS API, and then added to the data, the database, and the business and technical verification of the data. There is no information system data at the lower level and the data reported by enterprises and institutions adopts the file import method. For Excel data integration, the system provides data entry template, the system can realize the dynamic management of template.

Unstructured data integration: Such file data can be extracted from metadata, and full-text retrieval can be carried out. After the metadata is extracted, the files are stored in the database and a full-text search index is established.

Scan graphics class data integration: This type of data integration sets specific name rules based on specific files and is automatically added to the data collection library. Due to the large size of the scanned file, it needs to be compressed. Through the compression of a patented format, up to 1 % of JPEG content can be compressed.

Face Recognition

Suspects at Large are Quickly Identified.

The remote face retrieval based on wireless communication can be realized by using the police face recognition client. In the course of on-site patrols, when the police officers on duty find suspicious personnel carrying false identity cards, they can take face photos on the spot through specific applications installed on the police link and send them to the face retrieval platform through the 3G network. The platform will return the identification results, personnel photo list, etc. to the police pass to facilitate the identification of suspicious personnel on the spot of the civilian police on duty. It can support the concurrent amount of 200 search applications.

Rapid Identification of Persons who Refuse to Tell the Truth in Criminal Investigations.

In the process of criminal investigation and investigation, when the arrested suspect refuses to confess his true identity or needs to verify the identity of the suspect, he can use camera equipment to extract the suspect's photo and enter the photo into the face retrieval system. The face retrieval system will retrieve the face in the corresponding face database(in the escape library, temporary population database, etc.) and return a similar face list from the end of the similarity to provide convenience for the investigators to confirm the identity of the suspect.

Quick Confirmation of People in Network Photos such as QQ and Weibo during the Case.

In public security emergencies, if the public takes photos of criminals who are committing crimes and upload them to the Internet, the case handlers can input the images collected on the Internet into the face retrieval system to realize the rapid identification of criminals. To make sure he gets caught first. Drunken, psychotic, dementia, quick identification.

Identification of Priority and Priority Persons in an Emergency during Stabilization Period.

During the period of maintaining stability, rapid identification of the key or primary personnel in an emergency is carried out to help master information such as their social identity, social relations, and interest relations. The case is quickly handled and the case is reported to the news media at the first time.

Analysis of the Likelihood of Merging Cases.

The face recognition system can always make collision comparisons between two different face databases, find people who have reached a certain degree of understanding, and find clues through manual confirmation. According to the face photos of the suspects arrested by the city public security department and the fugitive personnel library in the CCIC library, the 7 types of key personnel library, the local key personnel library, and the temporary population library, face retrieval can be compared. Or the N: N method of collision comparison of personnel information with the National fugitive database(CCIC library) helps the civilian police to quickly filter out whether the suspect is involved in other cases.

Vehicle Surveillance Identification

Connect with the visual vehicle monitoring and identification system, access the vehicle monitoring system through the API, push the vehicle blacklist to the service end for detection and identification. The vehicle monitoring and identification system mainly uses high-definition camera technology combined with vehicle recognition and license plate recognition algorithm to achieve the capture, modeling and storage of license plates entering and leaving vehicles, real-time blacklist comparison and linkage retrieval with large data security platforms.

Terminal Application Technology

The bayonet.

The personnel identity check module will be based on the C/S architecture design and use the network as a public security line or offline stand-alone version to facilitate the complexity of the on-

site environment and ensure that identity checking can also be achieved in the event of interference with the network. Through the integrated video camera technology, second-generation identity card information reading technology, document scanning and identification technology and fingerprint information recording technology, the system realizes the rapid registration of the people who traffic, and synchronously collects the identity information and activity trajectory of the participants.

Records of personnel movements. The personnel movement track records are the security check and entry registration of personnel entering the event site. This function allows for the registration of activities for those who hold identity documents and those who do not hold identity documents. Through the integrated video camera technology, second-generation identity card information reading technology, document scanning and identification technology and fingerprint information recording technology, the system realizes the rapid registration of the people who traffic, and synchronously collects the identity information and activity trajectory of the participants. Identity card holders can quickly register for admission by reading ID card information. The system also provides activity recording functions for non-ID card holders, such as: children who do not apply for ID cards, foreigners who hold passports, etc.. Their movements can be recorded by taking photographs, scanning their documents and recording fingerprints.

Identity information reading. Identity information reading function is mainly accomplished by the combination of support hardware and software algorithm. The system obtains relevant information through the integration of video camera, identity card, passport information reader, document scanner and fingerprint recorder, and the information is read through the information recognition algorithm.

Comparison of personnel identity information. When the entrance registration is completed, the system compares the entrance information with the background information base of the personnel to obtain the results of the person's manual verification and the results of the system automatic comparison. The system automatically uses the accurate comparison method to match the name and ID number to prevent the occurrence of the same name leading to the result of misjudgment.

Offline data. In order to reduce the amount of communication data on the mobile end and prevent the delay of data transmission when the network signal is not good, the mobile end needs to be able to select the corresponding data to be packaged into offline data packets according to the requirements of the user and downloaded to the mobile terminal for use.

Users can make offline data packets according to the jurisdiction area, or they can choose to make customized methods such as the list of key monitoring personnel, personnel types, and optional methods for special inspections. The data Baoye can screen personnel information, such as a lightweight data packet that contains only basic personnel information and review results, or a heavyweight data packet that contains all personnel background data. Offline packets can support multiple packet forms at the same time.

APP Mobile Police.

Network Communications. The APP mobile end program in the project will run on the Internet, and the Internet can not be directly connected with the public security information network. It needs to be used as a data transfer between the networks through the public security bureau's networking information exchange platform. The Internet access link of the public security bureau introduces the Internet information into the public security information network in a one-way way, thus ensuring the security and confidentiality of the public security four-level network.

The APP server program will be deployed in two steps. The information forwarding program will be deployed in the application service area of the Internet information exchange platform. The information data sent by the mobile APP program will be forwarded to the application server in the public security information communication network.

Offline data. In order to reduce the amount of communication data on the mobile end and prevent the delay of data transmission when the network signal is not good, the mobile end needs to be able to select the corresponding data to be packaged into offline data packets according to the requirements of the user and downloaded to the mobile terminal for use.

Users can make offline data packets according to the jurisdiction area, and they can also choose to make customized methods such as the list of key monitoring personnel, personnel types, and self-selected methods for special inspection work. The data Baoye can screen personnel information, such as a lightweight data packet that contains only basic personnel information and review results, or a heavyweight data packet that contains all personnel background data. Offline packets can support multiple packet forms at the same time.

Positioning and tracking. The mobile APP application can realize GPS based map positioning at the mobile end and automatically return the positioning information to the background terminal program by data exchange with the "PGIS platform". Zhongtun, the monitoring personnel can master the specific location information of each terminal in real time on the big screen, and realize centralized scheduling.

Identification of personnel. Through the mobile communication terminal, using mobile phones to take photos, enter identity card information and other means, obtain personnel information and check identity with real-time or offline data, so as to achieve personnel identity check. If there is a problem with the identity of the personnel, send a prompt message, and no record of the prompt through the check.

Personnel trajectory record. The reconciliation of personnel information can also be accompanied by the taking of photographs of personnel through video cameras, recording of personnel trajectory information(GPS positioning data, time). The identity check process is recorded in the database and can be queried.

Personal Status Mobile Review. The background examination results of the inquiry personnel by name, ID number, etc., as well as the details of the background information of the person, the information is displayed according to different contents.

Acknowledgment

At the beginning of the era of big data, the research results of large data from all walks of life are beginning to show results. It is important to use smart data in the integrated construction of security, defense, management, and control. The application of big data will be a necessary way to crack down on illegal crimes and comprehensive social management.

Based on the technology of large data market construction, face capture, intelligent analysis terminal, this paper mainly uses various terminals to collect human images, identity information, and vehicle information as basic data sources. Security for related major events. This paper combs the application in the work of public security and police, its significance and academic value.

References

- [1]A Mantelero .Big Data: I Rischi della Concentrazione del Potere Informativo Digitale e Gli Strumenti di Controllo (Big Data: The Risks of the Concentration of Power Over Information and the Possible Remedies), *Diritto Dellinformazione E Dellinformatica*, 2012.
- [2] M Kayaalp . Patient Privacy in the Era of Big Data. *Balkan Medical Journal*,2017.
- [3]T Glenn, S Monteith. Privacy in the Digital World: Medical and Health Data Outside of HIPAA Protections. *Current Psychiatry Reports*,2014.
- [4]DN Jutla, P Bodorik, S Ali .Engineering Privacy for Big Data Apps with the Unified Modeling Language. *IEEE International Congress on Big Data*,2013.
- [5]NY Lee, BH Wu.Privacy Protection Technology and Access Control Mechanism for Medical Big Data. *Iiai International Congress on Advanced Applied Informatics*,2017.
- [6]T. R. Gopalakrishnan Nair, R. Selvarani. Defect proneness estimation and feedback approach for software design quality improvement[J]. *Information and Software Technology*.2011 (3).