Group Multilateral Relation Analysis Based on Large Data

Qiang LIU[1, a)], Guo-min ZHOU[1, b)], Guang-xuan CHEN[1, c)] and Yong XU[2, a)]

[1]Big data Key Laboratory of the M.P.S, Zhejiang Police College, Hangzhou 310053, China
[2]Hangzhou Hezhong Data Technology Co., Ltd, Hangzhou 310052, China

E-mail: [a)]Corresponding author: navylq@126.com, [b)]zhouguomin@zjjcxy.cn, c)chenguangxuan@zjjcxy.cn

Abstract. Massive, multi-source, heterogeneous police data and social data brings challenges to the current police work. The existing massive data resources are studied as the research object to excavate the group of multilateral relations by using large data technology for data archiving. The results of the study could provide technical support to police enforcement departments for fighting crime and preventing crime.

1. Introduction

In recent years, with the continuous application of public security information systems, a variety of public security business applications are gradually built and put into practical application, accumulated a large number of valuable data resources [1-2]. Outside of police department, there are also accumulated a wealth of data resources in the information systems from all walks of life. The massive data resources from police data and social data may reflect the attributes, trajectories and trends of the society and the individual, but the information is disorganized and the relationship can not be straightened out [3-4]. Now, ETL technology has been more mature, fusion heterogeneous, multi-source data is no longer a problem.

For large data processing, it is an important subject to dig interpersonal relationships by using massive & multi-source data. You should organize the information into a structured, centralized archival data, according to the classification of the category, and the information with the plane/train trips, hotel accommodation, Internet cafes online data, online shopping consumption, communication, household registration information, case and so on. Then you can use reality and virtual network activities to visualize the direct or indirect relationship among people by the whole, automatic and deep mining analysis.

Based on the existing large data technology represented by Hadoop, this paper uses distributed database technology, Map-Reduce, full-text search and other technical means to organize the massive data resources inside and outside the public into organized and centralized archival data, according to the classification of the main categories, then analyzes the visually interpersonal relationships by using the data described above.

2. Super dimension aggregation method

The information in the public security information system is build up based on a specific needs, for example, the hotel system maily including accommodation records, the driver management system maily including the driver basic information and his car's illegal information. It is difficult to grasp the

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

whole information of people or cars through a single system. However, if we pile up the information purely, a large number of bare data (original file) may be produced, and the data can not form archival data which can be be used effectively. The traditional way is to let the police carry out multi-library, multi-table correlation query in the massive data and spend a lot of time, effort to carry out a variety of complex relationship analysis. So that it's not efficient and unrealistic.

So we can effectively aggregate a large number of scattered bare data for truly combat services, by data archiving and a variety of archives for effective correlation. Data archiving mainly refers to organize the chaotic, scattered data into a structured, centralized archival data, by using the data relevance in the data object. Each file can be analyzed automatically, and all other data with a specific association of the current data object may be summarized into the file of the current data object. As shown in Fig.1 and Fig.2, there are personnel files, vehicle files, case files, virtual identity files, organization files and other files in the aggregative classification data.

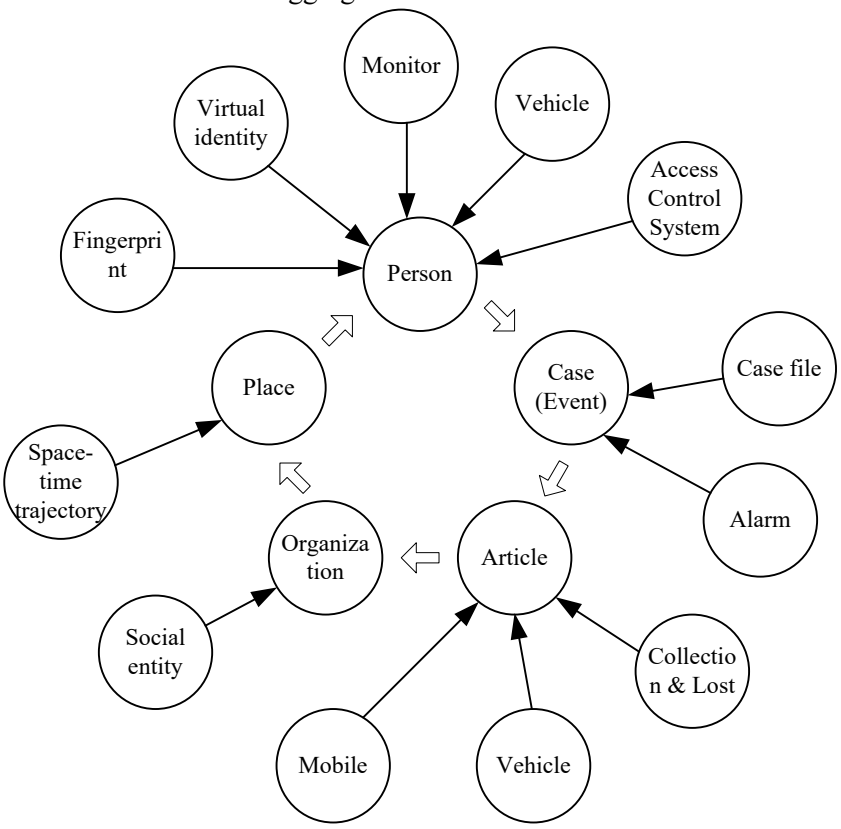

FIGURE 1. The display information of file data (1)

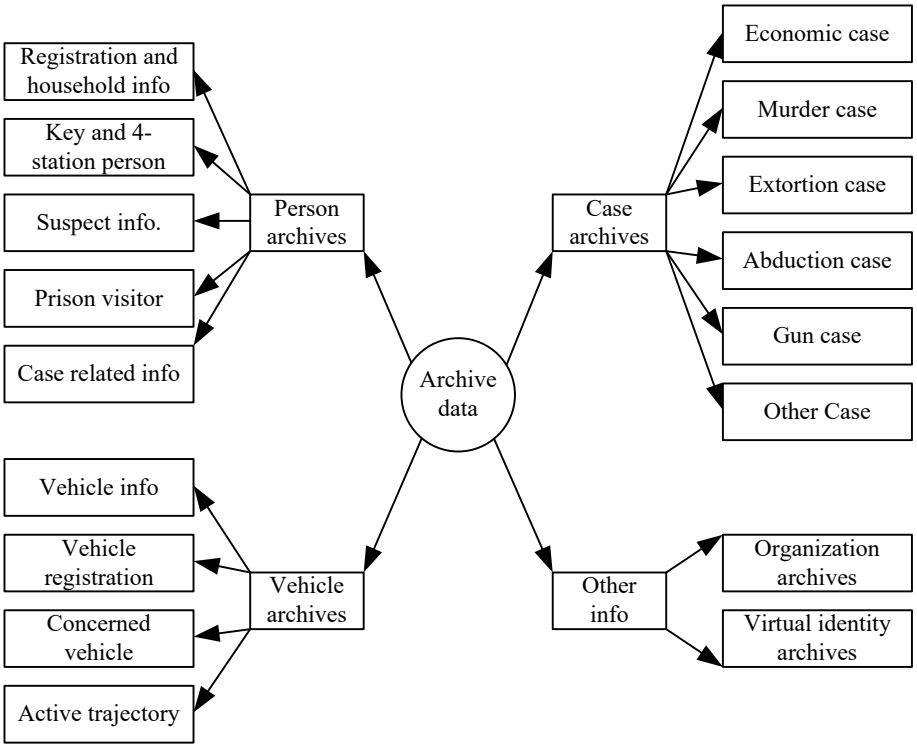

FIGURE 2. The display information of file data (2)

3. Super dimension aggregation method

The data files formed by the hyper-dimension aggregation collation are stored in the NoSQL database. The NoSQL database supports the storage and management for massive structure, semi-structured, unstructured data, also supports search query, analysis and mining, is very suitable for massive data query, full text search, offline analysis, offline transaction computing and other scenes, especially for non-format data query, retrieval [5-7]. The data which is scattered in each information system can be easily analyzed and archived according to NoSQL database, such as personnel, case, article, organization and other data.

First of all, based on the NoSQL database, the public security data is described on the basis of the super dimension aggregation method. Then, you can perform unilateral relational analysis, relationship analysis, or multi-party relationship by using the high-performance batch computing framework MapReduce. Finally, the relationships between the different feature objects can be automatically discovered and displayed graphically, thus visually revealing the interrelationships between objects, including hidden or indirect links.

Let {OD} denote a object which contains attributes of various elements, OD={M, T, L, G, OR}. Where M is the personal element, M={m1, m2, m3, …, mi}. T is the time element, T={t1, t2, t3, …, ti}. L is the location element, L={l1, l2, l3, …, li}. G is the articles element, G={g1, g2, g3, …, gi}. OR is the organization element, OR={or1, or2, or3, …, ori}. The principle of social relation self-discovery is shown in Fig.3.

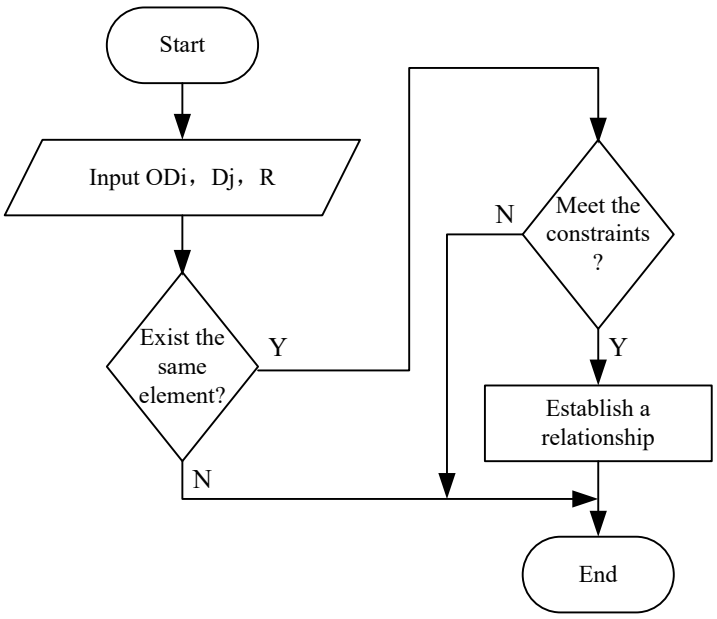

FIGURE 3. Social relation self-discovery

Here, ODi and ODj are two ODs, R is the constraint condition. When the two kinds of data ODi and ODj exist with the same characteristics of the property, the system may compare each other according to the constraints of the alignment. If the result of the comparison satisfies the constraints R, a relationship may be established. Normally, the constraints R are based on public security work experience. The system will automatically tap the social relations of known persons, automatically dig peer, living, Internet, relatives, social ties and other relations. With the deepening of the integration of information resources, a new further social relation may be generated.

After the social relations dig out, the relationship between the various data, and the information represented by the data will form a social relationship network. Based on the six-dimensional theory of space, the platform provides an indirect correlation analysis function through the network analysis technology. The platform could specify the path or use the shortest path, and provide iterative mining function, then a infinite drill analysis of social relations.

4. Conclusions

The massive data resources from police data and social data are used to study the interpersonal relations of a group. The use of super-dimensional aggregation classification methods and social relations self-discovery technology may provide police a new method to find an important clue for investigate and solve a case. The police may easily deal with diversity, complexity and concealment of social relations.

Acknowledgements

This work was financially supported by the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1509219), Science and Technology Program of Ministry of Public Security of China (2015GABJC41) and Scientific Research Fund of Zhejiang Provincial Education Department (Y201533572).

References

[1] Q. Shen, and B. Chen, *Electronic Design Engineering*, Vol. 24 (2016), pp. 25-27.
[2] D. Zhang, K. Y. Qi, Z. Y. Su and J. Zhang, *Information Technology & Standardization*, No. 7 (2015), pp. 18-22.
[3] W. Li, L. Q. Sun, and S. L. Li, *Journal of Chinese People's Public Security University(Science and Technology)*, No. 4 (2013), pp. 20-25.

[4] J. Liu, and L. H. Zhang, *Police Technology*, No. 3 (2015), pp. 4-7.
[5] J. P. Meng, C. Chen, and J. J. Meng, *Microcomputer ITS Application*, Vol. 33 (2014), pp. 19-22.
[6] W. Li, *Information & Communication*, No. 7 (2014), pp. 72-73.
[7] Y. B. Wu, Q. Xue, D. W. Xiang and Y. H. Mai, *Modern Electronics Technique*, Vol. 39 (2016), pp. 44-47.