

Analytics: The Real-World Use of Big Data in Financial Services Studying with Judge System Events

CHEN Si'en¹ (陈思恩), YANG Lüqing^{2*} (杨律青), XU Shouhui³ (徐守辉)

(1. School of Economics, Xiamen University, Xiamen 361005, Fujian, China;

2. School of Software, Xiamen University, Xiamen 361005, Fujian, China;

3. Xiamen Xindeco IOT Ltd., Xiamen 361015, Fujian, China)

© Shanghai Jiaotong University and Springer-Verlag Berlin Heidelberg 2016

Abstract: “Big data” which admittedly means many things to many people is no longer confined to the realm of technology. Today it is a business imperative and is providing solutions to long-standing business challenges for banking and financial markets companies around the world. Financial services firms are leveraging big data to transform their processes, their organizations and the entire industry. Since 2012, the term “big data” has frequently been mentioned and used to describe and define the huge amount of data in the information explosive era and to name related technological development and innovation. As to the police work, the coming of big data era is not only a challenge but also an opportunity. Police agencies should go with the tide of development to start with such aspects as work thinking, top design, public information sharing and application and talent provision so as to promote the new development and progress of police work. This paper expounds the practical effect and significance of police big data application by cases happened in some areas.

Key words: big data, police work, suspect locking, behavior

CLC number: C 81 **Document code:** A

0 Introduction

Big data and Internet analytics are especially promising and differentiating for financial service companies as data is one of arguably their most important assets in financial services. With no physical products to manufacture, data is one of their most important assets arguably. The business of banking and financial management is rife with transactions, conducting hundreds of millions daily, each adding another row to the industry's immense and growing ocean of data. So the question for many of these firms remains how to harvest and leverage this information to gain a competitive advantage.

We found that 71 percent of these banking and financial markets firms reported that the use of information (including big data) and analytics is creating a competitive advantage for their organizations, compared with 63 percent of cross-industry respondents. Innovative banking and financial markets organizations extract value from uncertain data, so the construction of judge system events based on data mining technology

is important.

The data storage in public security department usually has the following features: tera byte level volume storage, varied and interactive systems, very complex data structure, and the rapid growth of the quantity^[1]. If we take the traditional method of data processing, it is definitely a demanding exercise, both time- and cost-consuming. So the big data research and construction in public security is quite necessary^[2]. In recent years, many scientific and efficient big data application platforms emerge as the times require. “Locking suspect in 3 seconds” is no longer a fairy tale.

This paper makes an effort to the parameter identification of the ship response model. Four categories of the response model are considered. Predictions of maneuvers including trained samples by using the identified parameters are compared with the results of free-running model tests. The four categories are consistent with each other. The generalization of the identified model is verified by predicting different untrained maneuvers.

1 Data Filtering and Fast Suspects Locking

There are lots of query requirements in the daily work of law enforcement or security agencies for the

Received date: 2015-08-22

Foundation item: the Major Program of National Social Science Foundation of China (No. 13&ZD148)

***E-mail:** lqyang@xmu.edu.cn

population, crime, immigration, the hotel internet information, vehicle/driver, fugitives, stolen vehicles and other information. But, after many years' accumulation, the data is very large, and may be stored in different regions and systems. It is very difficult for the statistical work^[3]. So, in actual combat operations, the police urgently needs to achieve big data across different areas and departments, and realize the statistical analysis and comprehensive relevance utilization through timely, comprehensive and accurate information support^[4].

2 Study on the Law of Crime

2.1 Analysis of Vehicles on Checkpoint and Their Positions

Here is an example: "two guys who take the same train and live in the same hotel may be accomplices"^[5]. In the past, if policemen wanted to prove this, they needed to piece the different clues together, and it took them many hours' even days to process multi-source data on the traditional architecture. However, on the big data platform, the police can analyze suspects' behavior rules before the fact (such as the Internet record, hotel record, and driving record) and rate their guilt possibilities. This assists the police to narrow the scope of suspects and reduce the workload. Analysis of the vehicles on checkpoint and their position information not only contributes to the supervision of illegal traffic behavior, but also can find the fake-licensed car, car thefts and other illegal acts^[5].

There is such a case in Beijing. A car owner sold his car to someone else while he installed global positioning system in advance. After the deal, he used mobile phone to track the car' location in real time and took the occasion to steal the car back. The police got the report and searched on all major online trade platforms of second-hand car. They tried all kinds of key words, locked some similar cars and analyzed these car's attribute data one by one. Suddenly, they found that a seller's contact number was exactly the same as the former owner's. This man wanted to change a place and sell the car again! The police checked that day's video camera nearby, and found the suspect. It's easy to finish the evidence taking^[6].

2.2 Video Camera and Rapid Investigation

In an urban environment, airport or mass transit hubs could be thousands of cameras recording the comings and goings of people. It could take hours if not days. Modern concerns about public safety and security include a focus on a range of events from less serious everyday crimes like shoplifting through to personal violent crimes and like homicide through to terrorism ultimately. Minor and major crimes involve people moving about in a known space in identifiable patterns to find weaknesses^[7]. But, owing to the comparison techniques of big data platform, at least 95% of irrelevant images now can be filtered out quickly^[8]. As shown in Fig. 1, the police can track a target individual's movements from location to location, and access all relevant associated recordings.



Fig. 1 Looking for your target in an image album that is pre-sorted

3 Information Mining and Crime Prediction

3.1 Model Comparisons

IBM said that big data can gather a great deal of information and process them, just like "the microscope and magnifier in industrial age"^[9]. It helps us regain a sense of the outside world, via the neural network of big data, matrix and distributed platform^[10]. The flexible application of big data can not only lock suspect quickly, but also prevent crimes.

3.2 Trends and Seasonality of Different Types of Cases

Trends and seasonality are calculated by time series model. The actual value is the superposition of seasonality, trend and noise. The seasonality means the rules of 12 months within one year, and it won't change year and years. The future value can be predicted by seasonality and trends^[11]. Figure 2 shows the statistics of the number of cases across the province. The drill can provide statistical information function, and number of cases around the city district under the view.

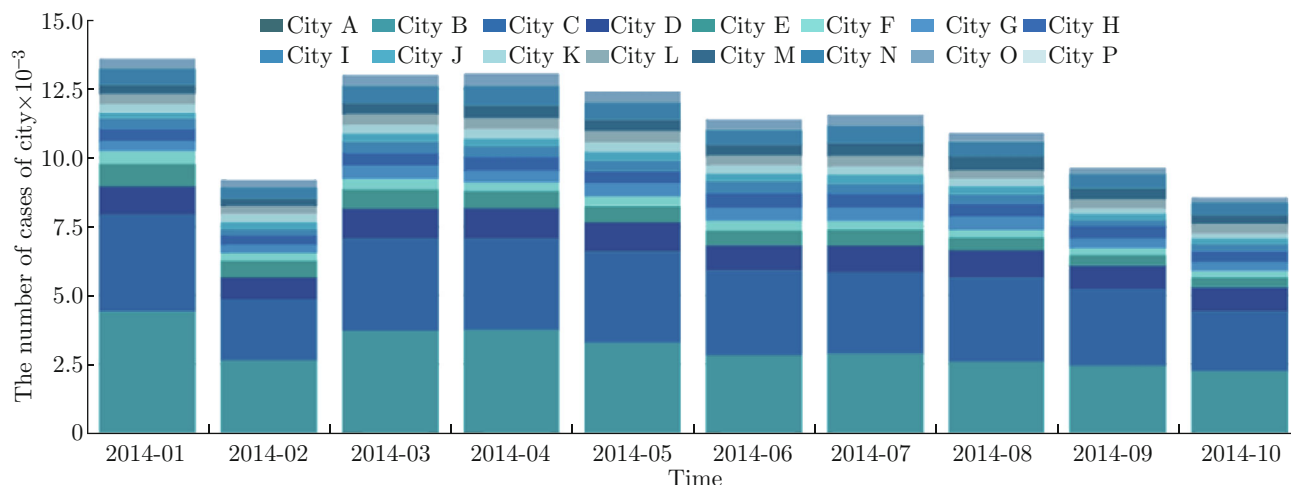


Fig. 2 Case data statistics

3.3 Quantitative Forecast of Different Types of Cases

The blue solid line in Fig. 3 is the actual value. Data is analyzed according to the number of case history. The time series prediction model is established to predict the future number of cases. The blue dotted line is

the predicted monthly value which is just calculated by historical values. And the light blue region and light grey region on the both sides of blue dotted line respectively represent 80% and 90% confidence intervals. According to the algorithm, there will be 80% and 95% actual values fallen into the predicted interval.

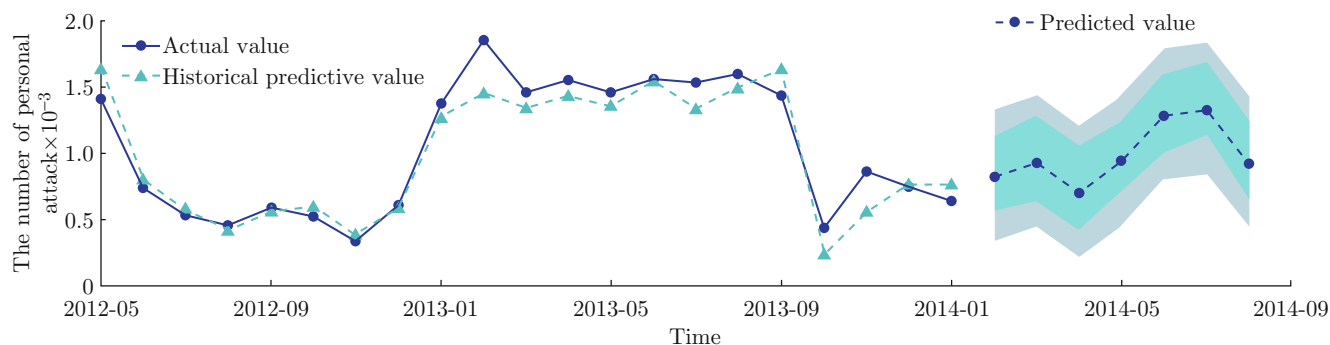


Fig. 3 The time series prediction of target prediction

3.4 Target Prediction

Firstly, police resources can be rationally adjusted and allocated through the prediction of future trends^[12]. Secondly, the policemen can understand event causes and take manual interventions once they find the abnormal value, if the actual value is much more or less than the predicted value. Thirdly, the security services may formulate measures to slow or stop the increase when they find the cases with growth trend. Fourthly, any rule that makes crimes reduced can be promoted to other regions as the experiences. At last, it helps to understand the cause of seasonal regularity and do the manual intervention^[13]. Figure 4 shows personal attack target prediction, where AoP represents analysis of periodiaty, GT represents general trend, and RF represents random factors. Table 1 shows the case

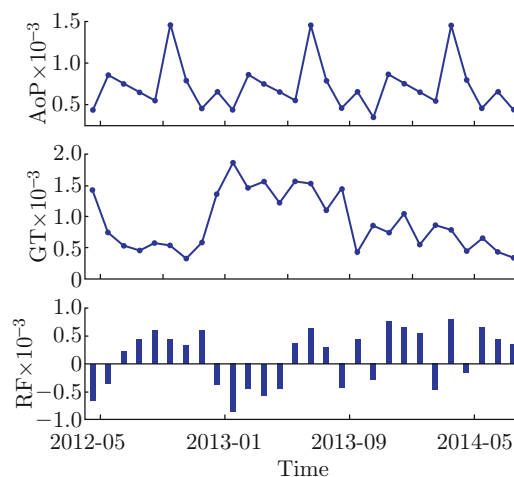


Fig. 4 The analysis of the case decision trend

Table 1 Confidence interval of the case decision trend

Time	Confidence interval				
	95% predicted	80% predicted	The middle value	80% actual	95% actual
2014-12	892	765	654	533	432
2015-01	1 245	952	754	652	435
2015-02	977	765	546	434	321
2015-03	1 135	987	675	454	435
2015-04	876	675	544	435	342
2015-05	1 214	967	675	543	342
2015-06	892	765	654	533	432
2015-07	1 245	952	754	652	435
2015-08	977	765	546	434	321
2015-09	1 135	987	675	454	435
2015-10	876	675	544	435	342
2015-11	1 214	967	675	543	342

confidence intervals.

4 Data Connection and Improving of Disaster Relief System

The application of big data can also improve the relief system for city's public security. On July 21, 2012, it was pouring down in Beijing. A lot of citizens chose microblog to transmit messages. They opened the additional coordinate function, then brought about map locations, and provided convenience for the timely

relief.

In the Ya'an earthquake, the missing people platform launched by microblog and other major internet companies also provided much channel support for relief. Except text and images, a microblog or a WeChat message still includes information such as time and location. National Information Society Agency shows disaster area scenes. It is also an epitome of big data application. Because of the data disconnection between websites and social media, the greatest value has not been developed^[14].

The real-world use of big data in financial services studying with judge system events is SKYNET platform^[15]. The SKYNET programmer collected 55 million cell phone records from Pakistan to identify "interesting" or "suspect" behaviors. Questions are being raised about the "method of identifying terrorist targets based on metadata" because it may identify false positives especially when it comes to the activities of journalists who seek to contact terrorists. SKYNET is a behavior profiling programmer that attempts to identify "interesting travel patterns", including how often a person travels and goes to where. Specifically, the programmer aims at identifying "courier-like travel patterns". It achieves this by analyzing mobile phone metadata that reveals both location and communication data from bulk call records. On the basis of this metadata, SKYNET looks for patterns amongst different people who use phones in the similar ways. SKYNET is outlier detection via machine learning, as shown in Fig. 5.

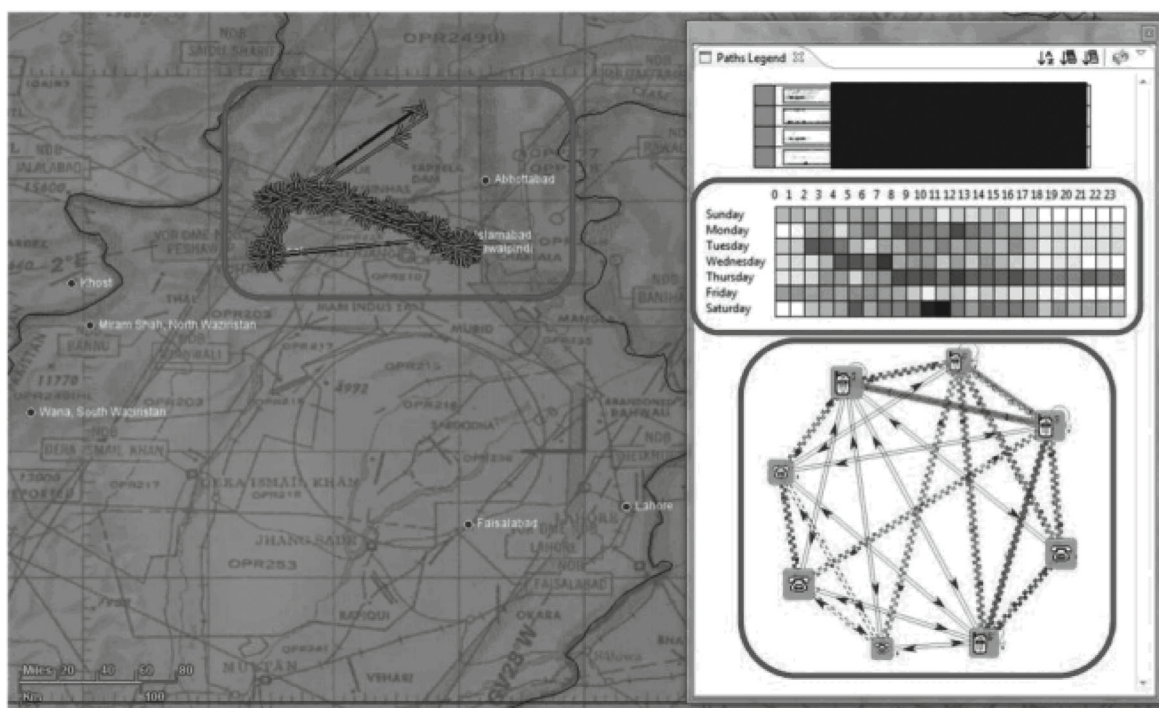


Fig. 5 SKYNET: Courier detection via machine learning

SKYNET uses a cloud computing technology to store and analyze call-data-records (CDRs) from Pakistani Telecoms upload to American National Security Agency Cloud. Analysis of the data includes pattern of life, social network and travel behavior. This is done by using geospatial, extemporal, and pattern-of-life and travel analytics. A mobile phone's international mobile subscriber identification number is a unique identification associated with all mobile phones on a cellular network. It is stored as a 64-bit field and is sent by the phone to the network.

SKYNET attempts to identify as follows. Who has traveled from Peshawar to Faisalabad or Lahore and back in the past month? Who does the traveler call when he arrives? SKYNET analytic capabilities include mobile phone metadata storage and analysis, pattern-of-life analysis, travel analysis, and social network analysis. SKYNET data sources include mobile phone metadata, global system for mobile communications, and international mobile subscriber identity. In terms of global system for mobile communication metadata, we can measure aspects of each selector's pattern of life, social network, and travel behavior (Fig. 5).

5 Conclusion

In terms of the prospects of big data in the field of public security, data must be online, open, sharing, internet and related^[16]. It is believed that the data standard and sharing mechanism will be more perfect in the future. The data source and the quality will be more reliable. At that time, the big data platform which can process unstructured data (pictures, micro-blogs, and the character's relationships) will have a bigger promotion.

References

- [1] WANG Q, WANG C, LI J, et al. Enabling public verifiability and data dynamic for storage security in cloud computing [C]//*Proceedings of Computer Security, ESORICS 2009*. Saint Malo of France: Springer, 2009: 355-370.
- [2] FORTUNATO S. Community detection in graphs [J]. *Physics Reports*, 2010, **486**(3/4/5): 75-174.
- [3] Cloud Security Alliance. Security guidance for critical areas of focus in cloud computing [EB/OL]. [2012-08-22]. <http://www.cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf>.
- [4] PERLROTH N, RUSLI E M. Security start-ups catch fancy of investor [N]. *The New York Times*, 2012-08-05 (technology section).
- [5] BITTERER A. Hype cycle for business intelligence [R]. Stamford, CT: Gartner, Inc., 2011.
- [6] SHACHAM H, WATERS B. Compact proofs of retrievability [C]//*Proceeding of Asiacrpt 2012*. Melbourne, Australia: Springer, 2011: 90-107.
- [7] BRANTINGHAM P L. Keynote address to the European intelligence and security information conference [R]. Athens, Greece: Computational Criminology, 2011.
- [8] HABEGGER B. Strategic foresight in public policy: Reviewing the experience of the U.K., Singapore, and the Netherlands [J]. *Futures*, 2010, **42**(1): 49-58.
- [9] IBM. IBM's smarter cities challenge: Syracuse [R]. Armonk, NY: IBM, 2011.
- [10] CHEN H, REID E, SINAI J, et al. Terrorism information: Knowledges management and data mining for homeland security [M]. New York: Springer, 2008.
- [11] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: A survey of the state-of-art and possible extensions [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2005, **17**(6): 734-741.
- [12] MCAFEE A, BRYNJOLFSSON E. Big data: The management revolution [J]. *Harvard Business Review*, 2012, **90**(10): 61-68.
- [13] STONEBRAKER M, ABADI D, DEWITT D J, et al. Map reduce and parallel DBMSs: Friends or foes? [J]. *Communications of the ACM*, 2012, **53**(1): 64-71.
- [14] Korea's National Information Society Agency. Evolving world on big data: Global practices [EB/OL]. (2015-11-20) [2012-05-03]. <http://www.koreainformationssociety.com / 2013 / 11 / koreas-national-information-society.html>
- [15] KELLY A R. U.S. government designated prominent Al-Jazeera journalist a 'member of al-Qaida' [EB/OL]. (2015-11-20) [2015-5-8]. <https://firstlook.org/theintercept/2015/05/08/u-s-government-designated-prominent-al-jazeera-journalist-al-qaeda-member-put-watch-list/>
- [16] ZIKOPOULOS P, EATON C, DEROOS D, et al. Understanding big data: Analytics for enterprise-class hadoop and streaming data [M]. New York: McGraw-Hill Osborne Media, 2012.