

A Comprehensive Survey on Applications of Big Data in the Financial Industry

Jiuyi Xu^{1,*}

Dalian University of Technology, Dalian, China, 116024

*Corresponding author. Email: joy.xujiuyi@gmail.com

ABSTRACT

Big data (BD) is a general term for data with large volume, multiple sources, fast processing speed, and huge application value. Nowadays, BD has become one of the priority businesses in most industries. However, research documents and reference materials about applications of BD in the financial industry are disperse and it is difficult for scholars to learn and compare them in different businesses. This paper mainly focuses on the overview of applications of BD in the financial industry and introduces the applications of BD in risk control, customer value management, precision marketing and other aspects. This paper also introduces the applications of BD in different specific businesses in the financial business including banking, securities, and insurance. Finally, this paper summarizes application challenges and development trends of the applications of BD in the financial industry and gives some relevant suggestions for both scholars and industries to refer to.

Keywords: Big Data, data analysis, financial industry, business integration, information security

1. INTRODUCTION

Big Data (BD) is born with the information of society, based on the accumulation of massive data, including countless “Data Production – Data Collection – Data Processing – Data Analysis – Information Extraction – New Data Production” circular chains. BD is widely used to help to deal with risk control, customer value management, precision marketing and other aspects in the financial industry. Many financial businesses including banking, securities, and insurance can improve their business capacity and bring more profits by deploying BD environment. However, BD environment is facing many constraints such as inefficient data asset management, difficulties in technological transformation, lack of comprehensive industry standards and policy guarantees, and high pressure on security management. This paper summarizes the applications of BD in the financial industry and some specific application aspects in different financial businesses based on individual and corporate surveys and researches. Meanwhile, this paper points out development and application challenges and trends BD in the financial industry are facing. Finally, this paper proposes some relevant suggestions for the status quo.

2. BIG DATA IN THE FINANCIAL INDUSTRY

2.1. Risk Control

Risk control runs through the entire financial supply chain. Due to the rapid development of the Internet and the sharp expansion of global business, the financial industry is facing increasingly complex market environment which includes the emergence of crowdfunding services (like P2P), the changing financial fraud deceitful tricks, and the stricter requirements for risk control from different regulatory authorities. This requires the financial industry to be able to measure all types of risks in the business environment based on their infrastructures. By implementing BD technology, using massive data inside/outside the industry, the financial industries can improve their control capability of risks and achieve the best balance between risks and returns to ensure the development of the industry.

From business perspective, strategies out of business scenarios are meaningless. The Basel Accord has specified the calculation method of many risks (such as credit risk, market risk, and operational risk) in detail [1]. Analysis models were first and extensively used on

credit risk [2]. All these models are inseparable from BD technology. Institutions cannot simply judge whether a customer is high-risk or low-risk based on single variable if the situation is complicated. This requires experimenters can combine various explicit tags and implicit tags based on different models to implement more accurate analysis. The classification of customer risk level involves customer tag matching, customer behavior analysis and customer clustering. After classification, financial companies can decide what measures should be taken to prevent risks, and what measures should be taken to reduce losses caused by risks.

2.2. Customer Value Management (CVM)

CVM means companies discover the most valuable current and potential customers based on relevant information and improve customer loyalty and retention by satisfying their individual needs for services. Although many financial industries store much customer information, due to the lack of mining and analysis, they cannot gain deeper understanding about customers, and it is difficult for these companies to improve their efficiency of customer management. McKinsey & Company pointed out financial companies can obtain feasible forecasts from data, which are mainly reflected in:

- (1) Analyze the composition of customers and segment them according to their value to the company.
- (2) Discover cross-selling opportunities in specific customer groups based on industry benchmark data.
- (3) The consumption propensity model reveals the possibility of the existing customer base to re-purchase the products to the business staff.
- (4) Identify underutilized channels and create more sales opportunities [3].

BD can help establish a user label system based on customer information such as customer preference, customer occupation, and customer consumption style. Companies can classify customers and use existing data tags and external data tags to portray customers in the system. Financial companies can provide more convincing basis for judgement for customer management and personalized customer service by performing customer clustering and customer behavior analysis.

2.3. Precision Marketing

Traditionally, marketing is to find and segment target customer groups and then reach individual customer groups. Precision marketing has changed from traditional product-oriented process to customer-oriented process which means evaluating and analyzing customers comprehensively. With the help of

BD, financial institutions can discover the habitual preferences and short-term needs of target customer groups, and then form the customer portrait and find more accurate target customer groups. Financial institutions have much internal/external information with high reliability, and can conduct systematic customer behavior analysis, including product performance, product profitability, customer contribution, customer potential, etc. [4]. Meanwhile, financial companies can also identify potential customers and find cross-selling opportunities more accurately by optimizing predictive models.

Financial institutions can use BD technology to mine, track, and analyze various user data (such as basic personal information data, consumption data, browsing data, and purchase behavior) to improve precision marketing. Based on customer portraits, financial institutions can effectively implement precision marketing, including acquiring customers on the Internet, further segmentation and classification of customers, event marketing and real-time customer marketing, cross-marketing, personalized marketing, and customer life cycle management (including new customer acquisition, customer maintenance, and resuming customer, etc.)

2.4. Others

2.4.1. Product Management

Financial institutions can obtain customer needs by implementing BD technology, so as to make more profitable product settings. Financial institutions can also analyze the characteristics of the product (such as its functions and its life cycle) to make manage product more efficiently [5]. If handling properly, financial institutions can deliver suitable products to customers, which is also good for customer relationship management.

2.4.2. System Management

BD technology can improve the transaction efficiency and simplify the operation and management process for financial institutions. BD can not only help to make front-end business decisions, but also can help to manage back-end information system to improve system management and data utilization. Financial institutions can evaluate the operation of systems to meet the needs of enterprise operation and maintenance, business growth, and performance matching, and ultimately improve the service and management level [6]. System management can provide decision-making foundation for system optimization, upgrade, and expansion based on BD technology.

2.4.3. Process Optimization

BD can help to enhance the internal transparency of financial enterprises and make the flow of information between the upper and lower levels of the enterprise more efficient. Through the application of BD, financial companies can optimize their internal processes and improve their operational efficiency [7]. BD technology can help to promote the exchange and sharing process of cross-level, cross-department, and cross-business information.

3. APPLICATIONS IN FINANCIAL BUSINESS

3.1. Banking Business

3.1.1. Credit Risk Assessment

Traditionally, banks conduct default risk assessments of corporate customers based on static data such as past credit data and transaction data. The lack of forward-looking makes the assessment less valid because not only the historical credit status of the company can affect corporate default, but the overall development status of the company and real-time operating conditions can also be valued. The application of BD technology makes credit risk assessment more credible. The integration of internal and external data is the prerequisite for effective credit risk assessment. In the process of calculating the possibility of customer default, commercial banks should rely on customer-related information the bank has already mastered, as well as information mastered by external institutions including credit history, public evaluations, business management information, income information, and consumption information [8].

3.1.2. Supply Chain Finance

Supply chain finance can be guaranteed by products and accounts capable to help enterprises obtain funds and core companies which have sufficient assets and high credit lines. Banks can form relationship graphs among enterprises in supply chain based on financial data including investments, holdings, and the relationship between shareholders and legal persons, which are conducive to proceed risk control of related companies [9]. The relationship graphs effectively organize fragmented data by establishing association links, making information easier to be understood by humans and machines. Banks can continuously observe changes in business communication among enterprises, gain insights into abnormal interaction through comparison with different baseline data, and provide references for risk control after corporate loans based on these relationship graphs.

3.2. Securities Business

3.2.1. Stock Market Forecasting

The securities industry has its own characteristics, which is different from the general indirectness in the value measurement of products in other industries. The investment and return of customers in the securities industry are showed directly in the form of currency. The presentation and marketing methods of financial products in the securities industry are more professional because of special characteristics of securities industry and stricter regulations. BD can effectively mine more useful information from these quantitative investment data of securities companies and help companies understand market conditions more accurately based on information from data analysis. Securities companies use BD technology to continuously track and monitor many individual investor samples, and to perform statistics on investment yield, holding rate, capital flow and other indicators to estimate the changes in trading behavior status, development trend of investment confidence, anticipation of the market and current risk appetite of individual investors, so as to predict market conditions. The investment model designed by Robert Shiller mainly considers three variables: the cash flow of the investment project plan, the estimated cost of company capital, and the stock market's response to investment (i.e., market sentiment) [10]. BD technology can help to make market sentiment perception possible based on structured and unstructured data used to implement sentiment analysis.

3.2.2. Robo-advisor

Robo-advisor is one of the new innovations in the application of BD to match the diversified needs of customers in securities business in recent years. The robo-advisor can provide online investment advisory services, which use different models to analyze personalized data such as customer risk preferences and transaction behaviors and then provide customers best customized investment advice. Robo-advisor can automate the process of data analysis and behavior suggestions based on BD technology, and then provide more retail customers with customized services [11].

3.3. Insurance Business

3.3.1. Fraud Identification

The amount of compensation directly affects the profits of an insurance company, and more and more insurance companies start to normalize the process of compensation by implementing BD technology. Insurance fraud will produce many abnormal compensations, and this behavior seriously damages the profits of insurance companies. Insurance companies

need to launch many special investigations which often takes months or years to identify suspicious insurance fraud behavior which cannot stop insurance fraud in time and cannot regain lost benefits in most cases. Insurance companies can build fraud identification models by analyzing data from investigations of fraud incidents occurred in recent years and then these models can help identify fraud patterns and significantly improve the accuracy and timeliness of fraud identification [12].

3.3.2. Risk Pricing

The insurance companies usually price insurance amount based on the result of risk judgment. The application of BD makes risk judgments for different groups more efficient and accurate [13]. For example, based on the driving data of drivers (including driving frequency, driving speed, and sudden braking and sudden acceleration frequency) collected by monitoring devices, the personal behavior data of drivers (including disputes frequency in social media, daily behavior, and personality) collected from social media, and health data of drivers collected from the medical system, the insurance company can achieve more accurate risk judgment of customers and make insurance business more profitable.

4. CHALLENGES AND TRENDS

4.1. Application Challenges

4.1.1. Data Asset Management

Low-quality data, few data acquisition method and decentralized data analysis systems make data asset management in the financial industry more difficult and less efficient. Missing data, data duplication, data errors and inconsistent data formats has always been found in financial data. The financial industry has relatively limited data sources, and institutions need to collect more external data to help to process data analysis. The data standardization in the financial industry is still not up to standard and data are always scattered in various data systems.

4.1.2. Applied Technology and Business

The current architecture of system financial institutions used to process data is relatively complex, involving varieties of system platforms and it is difficult for financial companies to implement large-scale BD technology transformation. Meanwhile, the implementation of BD technology for the financial industry is still in the experimental process and relatively few valid and complete cases and solutions can be referred.

4.1.3. Industry Standards and Regulations

The relevant standards and regulations for financial BD are currently still incomplete. Compared with BD in other industries, financial BD involves more personal data analysis and the requirements for information protection are more stringent. Meanwhile, data from different financial institutions are still in different formats and cannot be utilized efficiently because of lack of effective integration and coordination which causes few cross-domain and cross-enterprise data applications in the market.

4.2. Development Trends

4.2.1. Competitiveness Criteria

BD technology capability is becoming the core element of the competitiveness of financial industry. The demand for BD in the financial industry is business-driven. companies eagerly aim to make marketing and risk identification more accurate, business decisions more targeted, products more attractive to gain more profits from the market by implementing BD technology.

4.2.2. Data Integration and Sharing

The financial industry has gradually realized the commercial value brought by data sharing. In the future, the data circulation of business market will be more robust. Financial institutions will be able to obtain data more conveniently from other industries such as telecommunications, e-commerce, and medical care. Through data integration and sharing, financial institutions can achieve more precise marketing and risk control. At the same time, data integration allows the financial industry to launch more scenario-based financial products and conduct deeper integration business with other industries.

4.2.3. Data Security

Financial data has the characteristics of unlimited replication and flowability in addition to the basic characteristics of BD. These characteristics bring challenges to data security management. The number of malicious attacks on the Internet has increased significantly and incidents of organization data being stolen have merged one after another. Sometimes, the emergence of information leakage may leak nearly all the data assets in the organization at one time. If data are leaked, it will spread rapidly, and then cause even more serious data tampering and fraud behaviors occur. Meanwhile, the leakage of financial data will expose much personal basic information and consumption information, which will bring much inconvenience for customers. Stricter regulations and complete standards

for the data security management of financial institutions needs to be proposed.

5. CONCLUSION

BD can help solve different problems occur in risk control, CVM, precision marketing, and other aspects in financial industry. Moreover, BD can be widely used in different specific businesses of the banking, securities, and insurance business. However, financial BD need to make improvements in data asset management, solutions with increasing business applications, and industry standards and security regulations. In the future, the level of the application of BD will become one of the essential competitiveness criteria. Data integration and sharing will become the trend and data security will get more and more attentions. To make BD bring more profits for the financial industry, there are some suggestions for challenges and trends:

- The application level of BD technology and promote data integration, technology integration, and system integration should be improved.
- Both the technical department and the business department should participate in the process of finding and solving problems based on BD simultaneously.
- Top-level formal financial BD technology standards and regulations should be strengthened as soon as possible.

AUTHORS' CONTRIBUTIONS

This paper is independently completed by Jiuyi Xu.

ACKNOWLEDGMENTS

Prof. Forts has offered me valuable suggestions in the academic studies. Moreover, teachers and their assistants from Cetus Talk helped me solve many problems during the writing of this thesis. Finally, my beloved family members who have always been supporting me should be nominated in the acknowledgements part.

REFERENCES

- [1] Jackson, P., *Bank Capital Standards: The New Basel Accord*, Bank of England Quarterly Bulletin, Spring, 2001.
- [2] Moradi, S., Mokhatab Rafiei, F., *A Dynamic Credit Risk Assessment model with Data Mining Techniques: Evidence from Iranian Banks*, *Financ Innov* 5, 15, 2019.
- [3] McKinsey & Company, *Retail Banking in Asia: Actionable insights for new opportunities*, 2015.
- [4] Jianhui Zhang and Junxuan Zhu, *Research Intelligent Precision Marketing of E-commerce Based on the Big Data*, *Journal of Management and Strategy*, Sciedu Press, vol. 5(1), pages 33-38, February, 2014.
- [5] Li, J., Tao, F., Cheng, Y., et al., *Big Data in Product Lifecycle Management*, *Int J Adv Manuf Technol* 2015: 81, 667-684.
- [6] S. Valeev and N. Kondratyeva, *Large Scale System Management Based on Markov Decision Process and Big Data Concept*, 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), 2016, pp.1-4.
- [7] Qin, S. Joe., *Process Data Analytics in the Era of Big Data*, *AICHE Journal*, Vol. 60, No. 9, 09, 2014, p. 3092-3100.
- [8] Wang, F., Ding, L., Yu, H., et al., *Big Data Analytics on Enterprise Credit Risk Evaluation of E-business Platform*, *Inf Syst E-Bus Manage* 18, 2020, 311-350.
- [9] Yu, Wantao, Wong, Chee Yew, Chavez, Roberto and Jacobs, Mark A., *Integrating Big Data Analytics into Supply Chain Finance: The Roles of Information Processing and Data-driven Culture*, *International Journal of Production Economics*, Elsevier, vol. 236, 2021.
- [10] Shiller, Robert J., *Speculative Asset Prices*, *American Economic Review*, 104(6): 1486-1517.
- [11] M. Day, T. Cheng and J. Li, *AI Robo-Advisor with Big Data Analytics for Financial Services*, 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2018, pp. 1027-1031.
- [12] B. K. Jha, G. G. Sivasankari and K. R. Venugopal, *Fraud Detection and Prevention by Using Big Data Analytics*, 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 267-274.
- [13] Rick Swedloff, *Risk Classification's Big Data Evolution*, 21 Conn. Ins. L.J. 339 (2014-2015).