Research on Fintech Based on Big Data Technology

FENG LI, LINGLING WANG*
School of Management Science and Engineering
Anhui University of Finance and Economics
Bengbu 233030, CHINA

Abstract: - Big data technology has been widely applied in the financial industry. This paper mainly researches on Fintech using big data technology. In the era of big data, it can help enterprises grasp business opportunities in the financial industry, logistics industry, manufacturing industry, medical industry, and other fields and improve enterprise value. Big data technology has been widely applied in the financial industry. In this paper, we introduce many applications of big data technology in banking, securities, insurance and other financial segments.

1 Introduction

Recently, big data technology has a great impact on the financial industry, and the financial industry has great prospects for the practical application of information systems [1-3]. The financial industry has high requirements for the practicality of information system, and has accumulated a large amount of customer transaction data. Therefore, the main information needs of financial industry are customer behavior analysis [4], fraud prevention [5] financial analysis [6] and so on. In the side of customer behavior analysis, the banks conduct marketing, financial product innovation and satisfaction analysis on customers by studying customer behavior data such as credit card swiping, deposit and withdrawal, electronic bank transfer, etc., and send targeted advertising information to customers that may be of interest or favorable information.

As for the prevent fraud, the financial fraud usually takes account fraud and insurance fraud as common cases, which will have a significant impact on the financial order. Fraud can be detected through the behavior patterns of accounts, and big data, predictive analytics and risk segmentation can help companies identify patterns of layered fraud, obtain big data from received claims, and detect fraud early based on predictive analysis. Financial risk analysis and evaluation of financial risks can use many data sources, such as customer managers, mobile banking, telephone banking, etc., as well as data from regulatory and credit rating departments [7]. Under a certain risk analysis model, big data analysis can help financial institutions predict financial risks, including credit risks, market risks and operational risks.

Therefore, in the era of big data, it can help enterprises grasp business opportunities in the financial industry, logistics industry, manufacturing industry, medical industry, and other fields and improve enterprise value. Big data technology has been widely applied in the financial industry [8]. There are many applications of big data technology in banking, securities, insurance and other financial segments.

2 Overview of Big Data

Big data refers to the collection of data that cannot be captured, managed and processed by conventional software tools within a certain period of time. It is a massive, high-growth and diversified information asset that requires a new processing mode to have stronger decision-making ability, insight and discovery ability, and process

optimization ability. Big data covers a wide range of industries, including politics, education, media, medicine, commerce, industry, agriculture, Internet and other aspects in addition to finance. According to the report of McKinsey, an internationally renowned consulting company, information technology, financial insurance, government and wholesale trade have the highest potential in terms of the comprehensive value potential of big data application. In terms of the data volume of each company in the industry, the data volume of information, finance and insurance, computer and electronic equipment, and public utilities is the largest. It can be seen that both the information industry (Internet and telecommunications) and the financial industry are key industries for big data application in terms of investment scale and application potential.

There are many characteristics of big data [9], in this paper, it can be summarized as the following:

(1) *Variety*. Big data includes both traditional structured data represented by transactions, semi-structured data represented by web pages, and unstructured data represented by video and voice information.

(2) *High value*. The core characteristics of big data are massive and multi-source, which means that there should be a large number of databases with strong high-speed processing capacity and relatively complete structure as the support, rather than simply storage and analysis by traditional technical means. Thirdly, the information capacity is huge and large; Useful and effective information can be captured from various sources and turned into valuable resources (including data warehouses). These characteristics of big data can greatly promote economic society and enterprise management efficiency.

(3) *Velocity*. A large amount of information resources with strong high-speed processing capacity are potential and priceless, big data requires rapid data processing with strong timeliness and real-time or quasi-real-time processing. Mining and analyzing large databases can effectively solve these problems and improve the data quality and the extent of application.

(4) *High degree of complexity*. A large number of massive unstructured databases store various and complex useful resources (such as human resource management knowledge, etc.). These types of data are closely related, and this complex relationship is obtained by users after analysis and processing, which is of great value. In addition, big data also has the characteristics of high efficiency and massive information storage.

3 Application of big data in banks

Many Chinese banks have begun to use big data to drive business operations [10].

3.1 Customer portrait

Customer portrait is mainly divided into individual customer portrait and enterprise customer portrait [11]. Personal customer portraits include demographic characteristics, consumption power data, interest data, risk preference, etc. Corporate customer portrait includes production, circulation, operation, finance, sales and customer data, upstream and downstream data of relevant industrial chain. On the basis of customer portraits, banks can effectively carry out precision marketing, including:

(1) Real-time marketing. Real-time marketing is based on the real-time status of customers to carry out marketing, such as the customer's location at that time, the customer's last consumption and other information to carry out marketing targeted (a customer purchases maternity products by credit card, can predict the probability of pregnancy through modeling and recommend pregnant women like the business); Or view life-changing events (changing jobs, changing marital status, moving house, etc.) as marketing opportunities;

(2) Cross marketing. That is, the cross-recommendation of different businesses or products. For example, China Merchants Bank can effectively identify small and micro enterprise customers based on the analysis of customer transaction records, and then implement cross-selling with remote banking.

(3) Personalized recommendation. Banks can provide services or recommend personalized banking products according to customers' preferences. For example, according to customers'

age, asset scale, financial management preferences, etc., banks can accurately locate customers, analyze their potential financial service needs, and then carry out targeted marketing and promotion.

(4) Customer life cycle management. Customer life cycle management includes new customer acquisition, customer loss prevention and customer win back.

3.2 Risk management and control

In the aspect of risk management and control, it includes sme loan risk assessment and fraudulent transaction identification [12-15].

(1) Sme loan risk assessment. Banks can analyze the loan risk by combining the production, circulation, sales, finance and other related information of enterprises with the big data mining method, quantify the credit limit of enterprises, and carry out sme loans more effectively.

(2) Fraudulent transaction identification and anti-money laundering analysis. Bank card cardholders' basic information can be used by basic information, transaction history, customer behavior patterns.

3.3 Operation optimization

(1) Market and channel analysis and optimization. Through big data, banks can monitor the quality of different marketing channels, especially network channels, so as to adjust and optimize cooperation channels. At the same time, it can also analyze which channels are more suitable for promoting which kinds of bank products or services, so as to optimize channel promotion strategies.

(2) Product and service optimization: Banks can transform customer behavior into information flow, analyze customer personality characteristics and risk preference, understand customer habits at a deeper level, and intelligently analyze and predict customer demand, so as to carry out product innovation and service optimization. For example, Industrial Bank currently conducts preliminary analysis on big data, distinguishes high-quality customers through repayment data mining, and provides differentiated financial products and

services according to the difference of customer repayment amount.

(3) Public opinion analysis: the bank can pass the crawler technology, grasping community, BBS and microblog related information about the Banks and banking products and services, and through the technology of natural language processing is negative judgments, particularly timely grasp the Banks and banking products and services of the negative information, find and handle problems in time; Positive information can be summarized and reinforced.

4 Application of big data in insurance industry

In general, the application of big data in the insurance industry can be divided into three aspects: customer segmentation and refined marketing, fraud analysis and refined operation [16] [17].

4.1 Customer segmentation and refined marketing

(1) Customer segmentation and differentiated services. Risk appetite is the key to determining insurance demand. Risk lovers, risk neutral and risk averse have different attitudes towards insurance requirements [18]. Generally speaking, risk averse people have greater insurance needs. In customer segmentation, in addition to risk preference data, customers' occupation, hobbies, habits, family structure, consumption pattern preference data should be combined with machine learning algorithm to classify customers, and provide different products and service strategies for customers after classification.

(2) Customer related sales. Insurance companies can find out the best combination of insurance sales by association rules, and find out the time sequence of insurance purchase in the customer life cycle by using time sequence rules, so as to grasp the opportunity of increasing insurance amount of the insured, and establish resale list and rules of existing insured, so as to promote the sales of insurance policies. In addition to these practices, with big data, the insurance industry can directly target customer needs. Take Taobao freight return

insurance as an example. According to statistics, the freight insurance claim rate of Taobao users is more than 50%, and the profit of this product to the insurance company is only about 5%, but many insurance companies are willing to provide this kind of insurance. After the customer purchases freight insurance, the insurance company can obtain the customer's basic personal information, including mobile phone number and bank account information, and understand the product information purchased by the customer, so as to achieve accurate push. Assuming that the customer bought and returned baby milk powder, we can estimate that the customer has children at home, and can recommend products with higher profit margins such as children's disease insurance and education insurance.

(3) Customer precision marketing. In the field of network marketing, the insurance company all kinds of data can be collected from Internet users, such as the geographical distribution of attribute data, search keywords such as real-time data, shopping behavior, behavior such as data browsing behavior, as well as the interests, relationships, such as social data, can realize the regional orientation, in advertising push demand orientation, preferences orientation, relationship orientation.

4.2 Fraudulent behavior analysis

Based on internal and external transactions and historical data, real-time or quasi-real-time prediction and analysis of fraud and other illegal activities, including medical insurance fraud and abuse analysis and auto insurance fraud analysis [19-21].

(1) Analysis of medical insurance fraud and abuse. Medical insurance fraud and abuse can be divided into two kinds, one is illegal insurance fraud, namely insurance fraud; Another kind is to go to a doctor repeatedly in the limit of insured amount, floating compensation amount, that is, abuse of medical insurance. Insurance companies can use past data to find the most significant factors affecting insurance fraud and the value range of these factors, establish a prediction model, and through the automatic scoring function, quickly classify claims according to the possibility of abuse fraud.

(2) Auto insurance fraud analysis. Insurance companies can make use of the past fraud events to establish prediction models, claim application classification processing, can largely solve the problem of auto insurance fraud, including auto insurance claim application fraud detection, salesman and garage collusion fraud detection.

4.3 Refined operation

(1) Product optimization and policy individuation. In the past, in the absence of refined data analysis and mining, insurance companies put many people on the same risk level, and the customer's insurance policy did not completely solve the customer's various risk problems. However, insurance companies can use their own data and customers' social network data to solve existing risk control problems, develop personalized policies for customers, obtain more accurate and higher profit margin policy model, and provide personalized solutions for each customer.

(2) Operation analysis. Based on the analysis of internal and external operation, management and interactive data, with the help of big data station, comprehensive statistics and prediction of enterprise operation and management performance. Modeling based on the data of insurance policies and customer interaction, and using the big data platform to quickly analyze and predict the recurrence or new market risks, operational risks, etc.

(3) Selection of agents (insurance sales personnel). According to the performance data, gender, age, working years before joining the company, experience in other insurance companies and the thinking orientation test of agents, the characteristics of the salesmen with the best sales performance are found out and the high-potential salesmen are selected.

5 Application of big data in securities industry

In the era of big data, most securities firms have realized the importance of big data, and the research and application of big data [22]. Compared

with the banking and insurance industry, the application of big data in securities industry started relatively late. At present, the application of big data in the securities industry at home and abroad can be roughly divided into the following three directions:

5.1 Customer relationship management

(1) Customer segmentation [23]. By analyzing the state of the customer's account (type), life cycle, investing time, account value (peak of assets and average assets, trading volumes, commissions, contribution and cost, etc.), trading habits (turnover, market awareness, positions, the average value of shareholding, the average time of shareholding, single deal mean and average daily trading volume, etc.), investment preferences (preference varieties, whether order channels and shen) and return on investment (relative and absolute income, the relative and absolute returns and investment ability, etc.) this year, for customer clustering and segmentation, trading pattern types to find customers, find out the most valuable and earnings potential customer base, and their most in need of service, better resources allocation and policy, improve the service, seize the most valuable customers.

(2) Customer loss prediction. Securities firms can predict the probability of customer churn by modeling the customer's historical trading behavior and churn [24-26].

In 2012, Haitong Securities independently developed "Securities Customer behavior Characteristics Analysis Technology based on Data Mining Algorithm", which is mainly applied in customer depth portrait and user turnover probability prediction based on portrait. The model of customer classification, customer preference and customer loss probability is established by analyzing the massive information of haitong's more than 1 million sample customers and half-year transaction records. The technology's primary intention is to measure the probability of future customer loss through quantitative analysis of customer behavior.

5.2 Investment climate index

In 2012, Guotai Junan launched the "Investment Climate Index for Individual Investors" (referred to as 3I Index) [27], which conveys information such as individual investors' expectations of the market and current risk appetite through a unique perspective [28]. Guotai Junan Research Institute continues to track and monitor a large number of samples of individual investors, and calculates and weighted the comprehensive investment climate index from a series of indicators such as account investment return rate, position holding rate and capital flow [29].

Through in-depth mining and analysis of massive real investment transaction information of individual investors, 3I Index can understand the changes of trading behavior of individual investors, the state and development trend of investment confidence, market expectations and current risk preference. In terms of sample selection, small and medium investors with less than 1 million yuan of capital and more than 5 years of investment are selected. The sample size is as high as 100,000, covering different parts of the country. Therefore, this index is relatively representative. In terms of parameters, investors are optimistic or pessimistic about the market, mainly based on the position rate of small and medium investors, whether additional funds, whether profits. The "3I index" is released once a month. The median value is 100. 100-120 is within the normal range.

6 Conclusion

In the era of big data, it can help enterprises grasp business opportunities in the financial industry, logistics industry, manufacturing industry, medical industry, and other fields and improve enterprise value. Big data technology has been widely applied in the financial industry. The integration and application of financial data and other cross-field data are increasingly strengthened, artificial intelligence is becoming a new direction of financial big data application, and the integration, sharing and opening of financial industry data is becoming a trend, bringing new development

opportunities and great development power to the financial industry

Acknowledgment

References:

[1] Wang, Fatao, et al. "Big data analytics on enterprise credit risk evaluation of e-Business platform." Information Systems and e-Business Management 18.3 (2020): 311-350.

[2] Du, Danyang, Aihua Li, and Lingling Zhang. "Survey on the applications of big data in Chinese real estate enterprise." Procedia Computer Science 30 (2014): 24-33.

[3] Khanra, Sayantan, et al. "Big data analytics in healthcare: a systematic literature review." Enterprise Information Systems 14.7 (2020): 878-912.

[4] Khade, Anindita A. "Performing customer behavior analysis using big data analytics." Procedia computer science 79 (2016): 986-992.

[5] Chen, Jidong, et al. "Big data based fraud risk management at Alibaba." The Journal of Finance and Data Science 1.1 (2015): 1-10.

[6] Liu, Yi, Jiawen Peng, and Zhihao Yu. "Big data platform architecture under the background of financial technology: In the insurance industry as an example." Proceedings of the 2018 international conference on big data engineering and technology. 2018.

[7] Bauguess, Scott W. "The role of big data, machine learning, and AI in assessing risks: A regulatory perspective." Machine Learning, and AI in Assessing Risks: A Regulatory Perspective (June 21, 2017). SEC Keynote Address: OpRisk North America (2017).

[8] Bennett, Mike. "The financial industry business ontology: Best practice for big data." Journal of Banking Regulation 14.3 (2013): 255-268.

[9] Sivarajah, Uthayasankar, et al. "Critical analysis of Big Data challenges and analytical methods." Journal of business research 70 (2017): 263-286.

[10] Srivastava, Utkarsh, and Santosh Gopalkrishnan. "Impact of big data analytics on banking sector: Learning for Indian banks." Procedia Computer Science 50 (2015): 643-652.

[11] Li, Li, Fei Liu, and Congbo Li. "Customer satisfaction evaluation method for customized product development using Entropy weight and Analytic Hierarchy Process." Computers & Industrial Engineering 77 (2014): 80-87.

[12] Deumes, Rogier, and W. Robert Knechel. "Economic incentives for voluntary reporting on internal risk management and control systems." Auditing: A Journal of Practice & Theory 27.1 (2008): 35-66.

[13] Power, Michael. "The risk management of everything." The Journal of Risk Finance (2004).

[14] Hopkin, Paul. Fundamentals of risk management: understanding, evaluating and implementing effective risk management. Kogan Page Publishers, 2018.

[15] Mikes, Anette. "Risk management and calculative cultures." Management Accounting Research 20.1 (2009): 18-40.

[16] Liu, Yi, Jiawen Peng, and Zhihao Yu. "Big data platform architecture under the background of financial technology: In the insurance industry as an example." Proceedings of the 2018 international conference on big data engineering and technology. 2018.

[17] Rawat, Seema, et al. "Application of machine learning and data visualization techniques for

decision support in the insurance sector." International Journal of Information Management Data Insights 1.2 (2021): 100012.

[18] Kim, Su-Yeon, et al. "Customer segmentation and strategy development based on customer lifetime value: A case study." Expert systems with applications 31.1 (2006): 101-107.

[19] Harris, Lloyd C. "Fraudulent return proclivity: an empirical analysis." Journal of Retailing 84.4 (2008): 461-476.

[20] Humpherys, Sean L., et al. "Identification of fraudulent financial statements using linguistic credibility analysis." Decision Support Systems 50.3 (2011): 585-594.

[21] Li, Shing-Han, et al. "Identifying the signs of fraudulent accounts using data mining techniques." Computers in Human Behavior 28.3 (2012): 1002-1013.

[22] Choi, Tsan-Ming, Hing Kai Chan, and Xiaohang Yue. "Recent development in big data analytics for business operations and risk management." IEEE transactions on cybernetics 47.1 (2016): 81-92.

[23] Kim, Su-Yeon, et al. "Customer segmentation and strategy development based on customer lifetime value: A case study." Expert systems with applications 31.1 (2006): 101-107.

[24] Khan, Afaq Alam, Sanjay Jamwal, and Mohammad M. Sepehri. "Applying data mining to customer churn prediction in an internet service provider." International Journal of Computer Applications 9.7 (2010): 8-14.

[25] Kumar, Viswanathan, and Werner Reinartz. "Creating enduring customer value." Journal of Marketing 80.6 (2016): 36-68.

[26] Kamakura, Wagner A., et al. "Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction." International Journal of Research in marketing 20.1 (2003): 45-65.

[27] Vaaler, Paul M. "Immigrant remittances and the venture investment environment of developing countries." Journal of International Business Studies 42.9 (2011): 1121-1149.

[28] Wen, Fenghua, et al. "Retail investor attention and stock price crash risk: evidence from China." International Review of Financial Analysis 65 (2019): 101376.

[29] WU, Guo-Hua, Tian-Yin YAO, and Bao-Ping ZHANG. "A Study of the Efficiency of Futures Research Institutes of China." The Journal of Asian Finance, Economics, and Business 7.12 (2020): 555-564.

Sources of funding for research presented in a scientific article or scientific article itself