The Application of Big Data in the Insurance Industry- with Potential Risks and Possible Solutions

Yude Li[1,*]

[1] *School of Insurance, Shanghai Lixin University of Accounting and Finance, Shanghai, 200120, China*
[*] *Corresponding author. Email: 201230139@stu.lixin.edu.cn*

ABSTRACT

Big data is coming into use across industries. In recent years, the position of insurance in the financial industry is gradually rising. Life insurance and property insurance are two major types of commercial insurance, with the addition of big data technology, diversified applications are also emerging. Some feasible application modes combined with big data in these two insurance types are listed in the paper. Although the technical architecture of big data has become mature, there is still a very broad application and development space due to its flexibility. With the uncertainty of development, the following potential risks are put forward in the article, disclosure of personal privacy, mishandling of information, and Inadequate legal supervision. Some advice is proposed in the paper from the perspective of companies, individuals, and the legal system, and here follows a few examples: strengthen internal self-discipline, enhance self-protection consciousness, introduce a new law, etc.

Keywords: Big data, Insurance Application, Disclosure and mishandling of information, Industry self-discipline, Privacy protection

1. INTRODUCTION

With the development of technology and the economy, people are starting to prepare themselves for insurance in advance. Meanwhile, emerging big data technology also plays an important role in the insurance industry. There are reports showing that in the big data investment insurance industry as early as 2018, the figure will exceed 2.4 billion dollars, and the investment will continue to grow over the next few years. Based on feedback from insurance companies around the world, the benefits brought by big data technologies include increased access to insurance services, a reduction of the workload of policy management, accurate predictions of large loss claims, cost savings in administration and claims processing, accelerated processing of regular business, increased fraud detection rates, etc. [1]. Therefore, it is necessary to combine the traditional insurance model with big data and adapt to the fast-paced market brought by new technology [2].

It's essential to know some of the specific applications of big data in the insurance industry, the risks being brought with it, and some possible solutions that are also critical. This paper takes China as an example. Though there is still a small gap between China's insurance market and that of developed countries, it's still representative since China is the biggest developing country in the world.

PR Newswire reported that the insurance industry is no exception to the trend that big data is rapidly gaining traction from diverse vertical sectors. The different uses of big data in insurance have been found, such as targeted marketing, personalized products, efficient claims processing, proactive fraud detection, etc. The article gives the forecasts for 8 horizontal submarkets, 8 application areas, 9 use cases, 6 regions, and 35 countries [1]. At the same time, Venkatesh discovered that India's insurance industry is going through a rough patch for the redundant practices of the companies and lack of personalization in handling customer needs. Venkatesh thought that Indian insurance companies must adapt to newer technologies (like big data) in the age of technology and speedy transactions [2].

In the area of life insurance, Lavine major researched data mining for selling life insurance in this article. The insurance companies use massive data sets to analyze a person's medical, social, geographic, personal, and financial history, so that can better calculate the odds of future health issues, then make quick and correct decisions. Lavine also mentioned that the new technology of underwriting may help clients who have

been rejected due to illnesses like cancer [3]. To predict the risk level of applicants for believing the risk assessment was a key factor in classifying applicants in the life insurance business, Boodhun and Jayabalan put forward solutions for risk assessment for life insurance companies using predictive analysis. In the model, machine learning algorithms, namely Multiple Linear Regression, Artificial Neural Network, REPTree, and Random Tree classifiers were used on the dataset [4]. In another research, Song et al. proposed the development of a medical big-data mining process that uses the big data that are offered by the open system of Health Insurance Review and Assessment Service. Based on the development of convergence information technology, most parts of human lives have become digitized. With the advanced medical facilities and big-data mining process, efficient and convenient health care becomes possible [5]. In terms of algorithms, El Aboudi and Benhlima in the paper proposed a big data architecture that is extensible based on both stream computing and batch computing, which can further improve the reliability of the healthcare system by real-time warning and accurate prediction of patients' health status. There is a prototype implementation based on spark and MongoDB tools have been built to generate real-time alerts [6]. As for property insurance, mainly in auto insurance, Burgess in the paper discussed the following topics: the science and technologies behind the insurance, actuarial tables, telematics devices, the promises and challenges in insurance technology, customer service improvements, and the Progressive Insurance program named Snapshot. The paper reports the advancements in big data in the insurance sector [7].

Along with the advantages, some problems also appear. Sharma proposed that the insurance industry had dealt with some inherent challenges for a long time. However, the latest developments in big data and AI technologies have made it possible to identify the risks. Sharma also discussed the benefits and the potential risks of using big data and AI technologies. With correct datasets, well-trained AI models, and strict guidelines, comes prosperity to people [8]. Furthermore, Helfand

focused on the essential content about insurance and big data that lawyers need to know and understand. Helfand pointed out that the law is likely to develop in ways that accommodate big data since it's already here, and the insurance lawyers need to find a way to approach it for the present [9]. Mills also raised questions about the accuracy of data with the continued growth in the use of big data in the insurance industry. Nevertheless, the researcher still believed the combination of big data and insurance brings a more efficient and relevant future for both insurers and insureds as long as based on a firm, factual foundation [10].

Some possible solutions have been brought up. For example, Ali et al. speculated the possible situation for big data analytics in health insurance of a strong ethical and regulatory environment and put forward examples of safeguards and participatory mechanisms that should be established [11]. A good ethical network environment is expected.

This paper is major in the application of big data in the insurance industry. Some potential risks of big data are also discussed, and suggestions that might be useful in some ways are proposed in the paper as well. In chapter 2, the specific application of big data technology in life insurance and property insurance is mainly described. The possible risks in aspects like personal privacy, information processing, and the legal system supervision are raised in chapter 3. The paper gives a few feasible solutions from aspects of companies, individuals, and legal regulators in chapter 4.

2. APPLICABLE ASPECTS AND ADVANTAGES

2.1. *Personal insurance*

This part of the paper is about the usage of big data in personal insurance. As shown in figure 1, the premium income of personal insurance in China for the past 10 years as a whole shows a steady upward trend, which means the scale of personal insurance in China is gradually expanding.
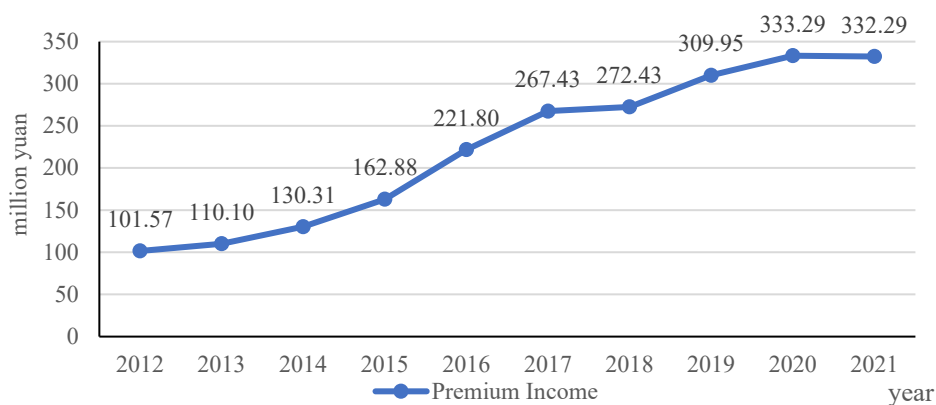

Figure 1. The personal insurance premium income for the past 10 years of National Bureau of Statistics data

Based on the situation, with the addition of big data, there are the following applicable aspects and advantages of big data in personal insurance. There are four major categories included in personal insurance-- life insurance, medical insurance, serious disease insurance, and accident insurance. Medical insurance and serious disease insurance are collectively known as health insurance, and accident insurance is not discussed in this paper for its great, huge randomness.

First here is about the application of big data in life insurance. To underwrite life insurance, it's of great importance to check the insured's status, health status, medical history, and financial status, then compared with underwriting standards, and it will be insured if passes. The problems facing insurance companies today include the underwriting rules being poorly set, low efficiency in obtaining underwriting data, and being highly dependent on the personal qualitative judgment of the underwriter. However, now with the data mining technique, the needed information can be easily analyzed from massive data sets, then calculate the odds of risks that may occur in the future [3]. To better do the risk assessment in life insurance, a model using Multiple Linear Regression,

Artificial Neural Network, REPTree, and Random Tree classifiers are used on the datasets to finish the predictive analysis.[4]. As for health insurance, companies need to conduct rigorous health screening to make critical and effective decisions [5]. To provide better services for health insurance consumers, the participation of big data technology is inevitable. Based on current technology, a big data architecture has been constructed to provide real-time warning and accurate prediction of patients' health conditions, which improves the reliability of health information collection [6]. In general, the use of big data in personal insurance sharply improves efficiency, and data accuracy, and makes more reasonable predictions.

2.2. Property insurance

The topic here is the application of big data in property insurance. The premium income of property insurance has the same tendency to change as personal insurance does, as shown in figure 2. Though the income in 2021 has some decline in the context of the current pandemic, its overall trend is still upward. This also shows that the property insurance market still has a rising trend.
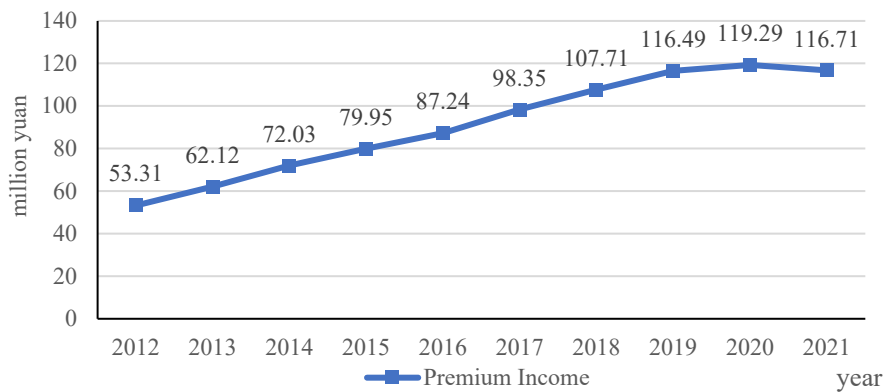

Figure 2. The property insurance premium income for the past 10 years of the National Bureau of Statistics data.

There are many subdivisions of property insurance. However, because of the nature of taking property and its concerned interest as subject-matter insured, the application of some big data technology can cover many kinds of property insurance. This paper takes business property insurance, motor vehicle insurance, and engineering insurance as examples.

The insurance mark of enterprise property insurance is all sorts of fixed assets and mobile assets like plants, machinery and equipment, and raw materials. It will be difficult to determine the loss after all the traditional preparations like checking relevant contents and doing site surveys are completed. But with big data analysis and a built-in sensor can well reduce the cost of time and human resources, and detect risks like unauthorized gas explosions, water line rupture, fire, and so on. The same can be said of home property insurance. The other important part is motor vehicle insurance. Due to the

rapid increase in the number of vehicles, there is a high probability of occurring risks, which brings a higher insurance rate. Traditional motor vehicle insurance pricing is determined by fixed premium rates and internal data. Now with some micromodule equipment like long-range sensors using tracking driving mode, insurance companies have the opportunity to collect a large number of multi-dimensional external data describing the risk of insurance subject matter, such as the road conditions, surroundings, and real-time weather. With more information brought by big data technology, it will be much easier to accurately assess risk and precisely calculate insurance premiums. The method can also be used in cargo transportation insurance and block insurance [7]. The last kind discussed in the paper is engineering insurance. There are the following typical characteristics of engineering insurance. The value of the subject matter insured increases gradually with progress, so the amount insured cannot be determined at the time

of insurance. In addition, because the subject matter is in the construction stage, external risks will increase. Real-time monitoring and recording device for big data applications can be suitable instruments. (i.e., current and voltage online monitoring, real-time temperature monitoring recorder. Such sensors are an important source of big data. The data from the devices can be continuously and automatically provided, to help decision-makers do some predictions.) Such systems help compensate for the loss of materials and manpower. Data results from the devices may assist to anticipate some degree of risk as well.

3. POTENTIAL RISKS

3.1. Disclosure of personal privacy

In the process of popularization and promotion of big data technology, fundamental changes have taken place in the way of data acquisition for the public. At the same time, hidden dangers have been laid for the protection of personal privacy.

In terms of traditional information protection, organizations need user authorization to use information. However, nowadays various tools and systems are employed for getting customer behavior data for big data. That makes the boundaries of personal privacy gradually disappear. When the personal information is collected, it could be passed to the insurer or other 3rd party agencies by such tools for profits. If this personal data is not defended properly, may land in the hands of cybercriminals which means getting exposed to crime. On the other hand, if the insurance companies fail to safeguard the sensitive personal data of customers, bad influences like hefty fines and wrong publicity are brought to the companies and their brand image will be severely impacted [8]. Here is a typical example. Folksam, Sweden's largest insurance company, once it was confirmed that the personal details of nearly a million customers of the company have been leaked to social media sites such as Facebook and Google. Though Folksam has asked partner companies to remove the information, the damage and serious consequences that have been caused remain inevitable. Whether it's an internal leak or an external cyber-attack, the risk of privacy leakage should be a serious concern.

3.2. Mishandling of information

Big data can get a lot of information from the Internet, but the value of this information is difficult to determine immediately.

The truth is, the producers of data cannot determine their value or harm, not to mention the collectors, traders, buyers, and even regulators to determine the value and harm of data in a short time. For example, an individual thinks it doesn't matter if his information is collected, so he doesn't care if anyone does anything with his information. There was a case that happened in August 2018. Beijing Ruizhi Huasheng Technology Co., LTD, illegally steals users' basic information such as what to search online, where to go, and what to buy, and with the information like that to define their targets and advertise for commercial purposes. The company earns 1 million yuan per month for that before being caught by the police. The information may or may not be important, but no one knows until the big data calculates and tells. As big data is generally a combination of discrete massive data, different values and information can be obtained through different models and combination methods. A lot of seemingly irrelevant and harmless information can form important and valuable information through effective combination. As the result, the outflow of big data resources may make it impossible for any party, even the regulators, to assess and anticipate the potential harm. When insurance companies make decisions using the information provided by data derived from big data mining, too much uncertain information increases the difficulty and cost of processing data. In addition, the continued growth in the use of big data raises concerns that not all of it may be accurate [10]. Information not only has to be valuable, but also it has to be clean enough.

3.3. Inadequate legal supervision

The data security environment of China's insurance industry is inadequate. Superficially, the data security management consciousness of insurance institutions is not strong enough, and many of the rules are mere formalities. However, the main problem is the lack of data security management, in both legal systems and regulatory norms.

One of the trickiest aspects of large analytical work is complying with government regulations. Insurance companies' big data analysis may contain sensitive and private information. That is to say, when processing and storing data, it needs to be confirmed that all of the information complies with industry standards or government requirements. Nevertheless, China's existing personal privacy protection law is relatively lagging behind. Though the Act on the Protection of Personal Information has been enacted, there is still not too much attention. Many people only know its existence, but do not realize its importance and significance. With backward consciousness like that, the criminals can't realize the punishment and privacy crime costs, resulting in the endless problems of privacy infringement. According to the Act on the Protection of Personal Information, the punishment measures include but are not limited to warning, fine, and revocation of license. If the circumstances are serious, they shall be sentenced to up to seven years' imprisonment. Despite the strong quantitative punishment measures, there are still dozens of relevant cases in recent years. This is the consequence of inadequate deep-seated legal awareness. Besides, when it comes to the insurance industry, the insurance

law plays an important role in the deliberations of decision-makers. And when it is associated with big data, both the insurance law and the lawyers are facing new problems. For example, when big data analysis becomes a key element in the decision-making process of insurance companies, it may make people suspect that the humanitarian values behind insurance law are being ignored [9]. The relevant insurance laws should be reconsidered.

4. FEASIBLE SOLUTIONS

4.1. Enterprises and companies

There are two aspects for the industries to reduce the risks: industry self-discipline and internal protection.

It's essential for enterprises to strengthen industry self-discipline and consciously safeguard users' privacy rights and interests. When the enterprise website is in a dominant position, it may abuse its own advantages in technology and other aspects to formulate its privacy protection policy so that the standard of privacy protection is blurred and personal privacy is substantially violated. It's where industry self-discipline should play a role. The enterprise itself acts as a manager, should make effective management of industry self-regulation protection of network privacy, and try to formulate the self-discipline standards and regulations in line with national conditions. Each business and its website should ensure that citizens' online privacy is well-protected, and establish a real relationship of mutual trust, to realize the basic protection of data [11]. Enterprises should also establish a mutual supervision mechanism and actively participate in compliance cooperation on personal information protection, change the passive supervision of enterprises into self-regulation mode, and cultivate a benign and safe data flow environment.

Nowadays, data privacy infringement modes are complicated and diversified, which makes the protection of internal data crucial. To improve data security defense capabilities for enterprises, solutions like improving the technical level of data protection, updating technical protections for system vulnerabilities, strengthening database security maintenance, etc. are practicable. Enterprises should also refuse to acquire or sell data with illegal technology tools.

4.2. Individuals

In the era of big data, strengthening personal protection awareness is one of the most basic internal factors to avoid the above risks. In the past, in order to maintain the network environment, the methods of protecting personal privacy tend to prevent users from posting personal information. However, this is obviously not desirable in the reality of real-name networks nowadays. When facing massive amounts of data, it's important to judge what is needed, and get rid of

meaningless data. To raise awareness of personal privacy protection, there are the following aspects to start with. First, it should be necessary to read the privacy statements and assess the risks involved carefully when registering or logging into a website. Remember not to give out too much personal information (i.e., Income level, marital status, home address, health status, work status, etc.) on the website as far as possible, unless it's necessary and inevitable. And never share personal information authorized to third parties easily. After all, there are too many examples (i.e., When users freely disclose personal name, gender, birth date, and mobile phone number, there are the background operators who receive these data last. They can put together complete privacy data according to the user's input, with the possibility of causing telecom fraud and telecom theft.) to prove that infringement and crimes are mainly caused by such leaked information. On the other hand, besides the protection of personal privacy information, it's optimum to take part in civilized network information activities. Do not leak, forward, or disclose the privacy of others, and show them deserved respect.

4.3. Regulators

First of all, some suggestions are put forward from the perspective of legislation in this part. With the emergence and popularization of the Internet and data, network privacy is not a completely new type of privacy, but an extension of privacy in the network. However, the right to privacy is not an independent right of personality in the legislation of privacy protection in China but has been included in the scope of reputation right to protect. Therefore, there is no complete system for the protection of privacy or network privacy. It is necessary to stipulate that privacy is an independent right of personality, and the network privacy right is protected as a special type of privacy right, which makes it more reasonable for legal protection. The protection scope of network privacy should include but not be limited to personal data, personal property information, personal email address, network operating traces, and other acts that cause infringements of Internet privacy. Publicity on relevant existing laws and legislation is equally important.

Secondly, the government should define its own position, give full play to its own functions, and effectively supervise the industry's self-regulation protection of network privacy right. As an internal leadership force, the government focuses on guidance, cultivation, and regulation, instead of intervening from the outside. Scientific, economic, interactive, and operational management should be expected to achieve legal self-discipline, creating a solid foundation and good conditions for the future development of a healthy network environment.

5. CONCLUSIONS

This paper shows the benefits of big data used in insurance industries. In life insurance, big data technology helps to improve underwriting efficiency, make risk predictions, enhance the personal experience, etc. In property insurance, superior applications include but are not limited to monitoring property data in real-time, reducing loss cost, risk prediction, insurance assessment, etc. Potential risks and suggestions for ideas are following proposed. Enterprises and companies should strengthen industry self-discipline and internal protection to reduce the risk of disclosure and adverse use of customer personal information. Individuals should consciously protect personal information from being used by malicious organizations. As for the regulators, some necessary measures like introducing new protection laws should be considered. A perfect system and environment can never be achieved by the efforts of one person or a certain organization. No matter what it is, only by joining hands and making efforts together can we achieve the desired results.

REFERENCES

[1] PR Newswire, Big Data in the Insurance Industry: 2018 - 2030 - Opportunities, Challenges, Strategies & Forecasts, PR Newswire US, 2018.

[2] S. Venkatesh, Big Data - Can it make a big impact in the Insurance sector, Journal of the Insurance Institute of India, vol.6, no.4, 2019, pp.92–97.

[3] A. Lavine, The Big Data Dive for Life Insurance, REP, vol.38, no.2, 2014, pp.36.

[4] N. Boodhun, M. Jayabalan, Risk prediction in life insurance industry using supervised learning algorithms, Complex & Intelligent Systems, vol.4, 2018, pp.145–154. DOI: https://doi.org/10.1007/s40747-018-0072-1

[5] C.W. Song, H. Jung, K. Chung, Development of a medical big-data mining process using topic modeling, Cluster Computing, vol.22, no.1, 2019, pp.1949–1958. DOI: https://doi.org/10.1007/s10586-017-0942-0

[6] N. El Aboudi, L. Benhlima, Big Data Management for Healthcare Systems: Architecture, Requirements, and Implementation, Advances in Bioinformatics, vol.2018, 2018. DOI: https://doi.org/10.1155/2018/4059018

[7] R. Burgess, Insurance meets technology: Advances in big data promise personalization and challenges for carriers, Indiana Lawyer, vol.28, no.26, 2018, pp.23.

[8] Y.K. Sharma, Practical Utility of Artificial Intelligence and Big Data in the Insurance Industry, Journal of the Insurance Institute of India, vol.9, no.2, 2021, pp.70–81.

[9] R.D. Helfand, Big Data and Insurance: What Lawyers Need to Know and Understand, Journal of Internet Law, vol.21, no.3, 2017, pp.1–35.

[10] H. Mills, Bad Big Data: The continued growth in the use of big data in the insurance industry raises concerns that not all of it may be accurate, Best's Review, vol.118, no.6, 2017, pp.20.

[11] C.W.L. Ho, J. Ali, K. Caalsc, Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance, Bulletin of the World Health Organization, vol.98, no.4, 2020, pp. 263–269. DOI: https://doi.org/10.2471/BLT.19.234732