

# Research and Practice of Big Data in Internet Industry

Honglei Yang

Asia Metropolitan University, Kuala Lumpur, 666100, Malaysia

**Abstract:** In order to better analyze the correlation between big data technology and the internet industry and further promote the healthy development of the internet industry, the research and practice of big data in the internet industry were investigated and analyzed. the current development status of the internet industry and the application of big data technology were investigated. according to the investigation results, the potential risk values in the development process of related industries were calculated, the defense coefficient was standardized, and relevant optimization suggestions were put forward to ensure the healthy development of the industry. Empirical analysis results confirm that the application of big data in the Internet industry can effectively promote the development of modern economy and technology.

**Keywords:** Big data; Internet; Network security; Financial development

## 1. Introduction

Big data is the general name of data generated during the operation of the Internet. Its internal database contains the data resources of the entire Internet, so it can be defined from the perspective of the development of the Internet [1]. Big data is the result of the highly developed Internet and reflects from the side that mankind is entering a more explosive information age. With the development of social informatization, the problem of big data processing brought about by Internet application has become more and more urgent. The commercial value contained in it has become a "cake" for various circles to scramble for. All circles want to seek a new round of innovation and development under this opportunity, and finally improve their position in the market competition and expand their advantages [2]. As a new economic form, internet plus's core lies in integration, which is crucial to the application of big data. Big Data and internet plus, as the most advanced technologies in the development of modern science and technology, have exerted a profound influence on the whole society and have gradually penetrated into every corner of people's production and life, but people's cognition of them is still in the primary stage. In recent years, with the rapid development and wide spread of Internet technology, it has brought great help to the formation of big data environment and has become an essential part of the development of Internet. Judging from the current development of the Internet economy, it is necessary to give full play to the full role of big data in the development of the Internet in order to promote the growth and sustainable development of the Internet economy [3]. At present, the application of big data in the internet plus process has set off a frenzy of scientific and technological development and affected the

development of the whole society. Based on the actual situation, this paper studies the application of big data in the internet plus process and puts forward some practical measures for reference.

## 2. Research on the Application of Big Data in Internet Industry

### 2.1. Survey on application of big data in internet industry

Facing the fierce market competition environment, all walks of life are seeking new development. The application of big data in the internet plus process has greatly improved the survival probability of small and medium-sized enterprises and provided favorable conditions for their development [4]. At the same time, for the traditional industry, it has accumulated abundant data and information in its own development process, while the application of big data has brought about a revolutionary technological change, which has had a significant impact on it. Under such an environmental background, many industries such as transportation, finance, logistics and so on have implemented database construction, which has laid a foundation for building a new mode of commercial development in internet plus. In the view of many traditional businesses, the consumption of consumers is very random due to the limitation of data sources and analysis methods [5]. In response, Internet operating companies have actively established cooperative relations, committed themselves to the development and application of big data, built a full-link tag cloud for the crowd, formed a precision recommendation service module, and based on brand influence, put forward precision marketing strategies, which to a large extent promoted the upgrading of traditional industries to "internet plus" and won the sup-

port and trust of many user groups. The application of big data in the process of "internet plus" has become an important method to upgrade industrial structure, innovate products and services, and optimize user experience. Although there are many differences between big data application and traditional software, traditional software testing methods and strategies can be applied to big data application testing after appropriate adjustments. Traditional software testing cannot fully adapt to big data performance testing. At present, most of the performance testing needs to be completed with special testing tools. Most big data applications are essentially similar to object-oriented software systems [6]. Big data testing is a hot topic for researchers in recent years. The structural characteristics of big data applications determine the dis-

tribution of all kinds of big data faults. To discover, analyze and eliminate faults, many tests are usually required to ensure that big data applications have perfect and good functions, performance, compatibility and security [7]. Big data performance test is a process of information collection and analysis to ensure that website servers can respond to browser requests within a specified parameter range. The data collected during this process are used to predict what load level will deplete system resources. Its main purpose is to develop effective improvement strategies to maintain acceptable system performance. In order to better analyze the current application of big data in the Internet industry, the current application of big data in our country has been sorted out through investigation and research and tabulated as follows:

**Table 1. Survey on application of big data technology in internet industry**

Features	Function	BIM	System	Widely linked cloud	Lubanyun	PKPM	Server	OMMA
Cycle	Design phase	1	1	2	1	0	1	0
	Construction stage	0	0	0	2	1	2	1
	Operational extremes	1	1	0	2	2	1	0
	Three-dimensional visualization	2	1	1	2	1	1	0
	Collision detection	1	1	2	1	2	0	0
	Change detection	0	1	0	2	0	1	1
	Online file storage	1	2	1	0	1	2	1
	Online view viewing	2	0	2	1	2	2	0
Functional characteristics	Online data processing	2	2	1	0	1	2	2
	data sharing	1	1	0	2	1	1	2
	Document identification	0	2	2	1	0	1	1
	File format conversion	1	0	2	2	1	0	1
	Data encryption	0	0	1	0	1	2	1
	Data analysis	2	2	0	2	1	0	1
Cloud deployment	Structural analysis	2	0	2	1	2	0	1
	Private cloud	1	1	2	0	2	0	0
	Public cloud	1	1	1	2	1	2	0
Licensing mode	Hybrid cloud	2	0	0	2	1	0	1
	Free trial	0	1	1	2	1	1	2
	Free open source	0	0	2	1	0	0	2

Based on the above table for further judgment and analysis, combined with the current online business data and the environment of the combination of the Internet research, in the above table, 0, 1, 2 respectively from low to high said its dependence on big data technology. Through the application of big data technology in the above functions, the relevant internet industry and the combination of big data analysis technology, consumers can have a better understanding of the quality of products or services, and can quickly choose satisfactory products from a wide variety of products according to their own needs, which is also the key role of e-commerce development. In this era, personalized recommendation service has become the magic weapon for e-commerce [8]. The so-called personalized recommendation service is the concrete manifestation of the application of big data in the "internet plus" process, and is the service that the e-

commerce platform recommends individuals after analyzing group behaviors. In this process, the website analyzes the user's interest preferences by collecting user behavior data, such as browsing, attention, price comparison and comments, etc. matches the corresponding commodity information, and intelligently recommends it to the user application, which helps the e-commerce providers greatly improve the customer service experience and satisfaction, establishes a good reputation of the website, and makes the e-commerce providers increasingly influential. From this dimension, the application of big data accelerates the transformation process from traditional shops to "internet plus".

**2.2. Risk and defense of big data in internet industry**

Due to the complex environment of big data Internet, in order to solve the risks faced by the Internet industry in

the development process in the era of big data, the risks of big data in the Internet industry are analyzed and corresponding defense mechanisms are designed to standardize the network data security protocols [8]. Network data security protocol is a kind of handover rule between users. It is established for computers to exchange data safely. Its establishment can effectively resist the risks faced by Internet finance. Because of the openness of the Internet environment, most of the data information is over-disclosed, and there will be some risks in Internet finance, such as credit and loan, which will not guarantee the safety of users' property. Therefore, while the Internet industry is developing, it is necessary to protect the interests of related industries as much as possible and to make data public and transparent. In addition to confidential information, other business situations should be annotated, which is of great significance to the realization of the security of the Internet environment [9]. Therefore, the network security environment defense algorithm is optimized. Set up  $W = \{T_1, T_2\}$  is a set, where  $T_1$  is an attacker;  $T_2$  is the defender;  $S^k = (S_1^k, S_2^k, \dots, S_m^k)$  is the network space where  $T_k$  occurs.  $S_i^1 = \{(s_i^1, h_i^1, \pi_i^1) | 0 < \pi_i^1 < 1, \sum \pi_i^1 = 1\}$  is the attacker's attack behavior,  $s_i^1$  is the attacker  $T_1$ 's malicious act  $s_i$ ;  $h_i^1$  is the attacker  $T_1$ 's target equipment of the attack, This behavior mainly depends on active defense technology and the importance of network environment, and is divided into five attack levels according to the types of equipment and network topology.  $\pi_i^1$  is the attacker  $T_1$ 's The probability commit malicious acts of  $s_i$ .  $S_j^2 = \{(s_j^2, h_j^2, \pi_j^2) | 0 < \pi_j^2 < 1, \sum \pi_j^2 = 1\}$  is the act of defending for a defender.  $s_j^2$  is the defender  $T_2$ 's active defense behavior of  $s_j$ ;  $h_j^2$  is the defender  $T_2$  target equipment for defense, Defense behavior mainly depends on network structure, which is divided into 5 attack levels according to equipment types and network topology.  $\pi_j^2$  is the defender  $T_2$ 's probability of active defense behavior  $s_j$ .  $Q = (Q_i | Q_i = (q_{i1}, q_{i2}, \dots, q_{in}))$  is the space state,  $n$  is the number of network devices,  $q$  is the network defense state within a certain period of time,  $q_{hi} = (host_h, link_{ji})$ ,  $host_h$  is the state that the network equipment obtains the authority within a certain period of time,  $link_{ji}$  initiates an intrusion for network equipment, and the link performance drops [10]. Different states represent different stages of network defense. The mutual transition between states is determined jointly by armor class and attack level. A mathematical model can be constructed according to the behaviors of both defense and attack parties, as shown in Formula (1)

$$q_{ij}^l = W_{ij}^l + \sum_{l=1}^k P_{ij}^l(Q_i | q_i, S_i^1, S_i^2) S_i, P_{ij}^l \geq 0, \sum_{l=1}^k P_{ij}^l < 1 \quad (1)$$

In the formula,  $W_{ij}^l$  is the behavior of both defense and attack in  $Q$  state;  $(S_i^1, S_i^2)$  is the defender's profit;  $q_i$  is the network state after successful attack;  $\sum_{l=1}^k P_{ij}^l(Q_i | q_i, S_i^1, S_i^2) S_i$  is the indirect impact of the attack on the network;  $(Q_i | q_i, S_i^1, S_i^2)$  is the probability of information state transition. In the process of network operation, the network is vulnerable to external attacks, resulting in indirect impact on the network, causing interference factors in the subsequent security assessment and seriously disrupting the assessment results. Therefore, the network entropy countermeasure quantification technology is introduced to weigh the benefits and improve the accuracy of the assessment results. After calculation  $\sum_{l=1}^k P_{ij}^l(Q_i | q_i, S_i^1, S_i^2) S_i$ . It can be seen that its attack behavior has a certain impact on the whole network security structure. The ql entropy difference of the network state after successful attack is:

$$\Delta Z = \sum_i I_i \Delta Z_{host_{i_i}} + \sum_j \theta \Delta Z_{link_{j_j}} \quad (2)$$

$$\Delta Z_{host_{i_i}} = \Delta Z_m$$

In the formula,  $\Delta Z_{host_{i_i}}$  is the sum of entropy differences of all indexes after the host of the whole network device is attacked,  $I$  is the number of devices,  $m$  is the availability index.  $I_i$  is the importance of equipment.  $\Delta Z_{link_{j_j}}$  is the impact of attack on the link; is the link impact factor, which can directly reflect the impact degree of the network transport layer. Taking full account of the interests of both the defender and the attacker, it is necessary to evaluate the quantitative effect of the model confrontation. Through the evaluation effect, the damage caused by the attack behavior to the network active defense can be found. If the attack behavior is greater than the defense behavior, then the attack means adopted by the attacker have successfully invaded the active defense system to check the safe use of the equipment. Use in a given security environment  $(S_i^1, S_i^2)$  Defensor income results as quantitative evaluation criteria. The network security problem is regarded as a multi-stage game between defenders and attackers, and each stage corresponds to a security state. The mixed strategies in the security state can be dealt with in a balanced way by solving, thus obtaining the optimal scheme of confrontation between the two parties in the network security state. The network entropy is introduced to carry out quantitative analysis, and IN is set as each index information and mutual confrontation behavior information of the active defense network driven by big data. Out is the quantita-

tive effect value of attack behavior prediction and countermeasure. Set the initialization value vector to:

$$H^0 = (h_1^0, h_2^0, \dots, h_k^0) = (0, 0, \dots, 0) \quad (3)$$

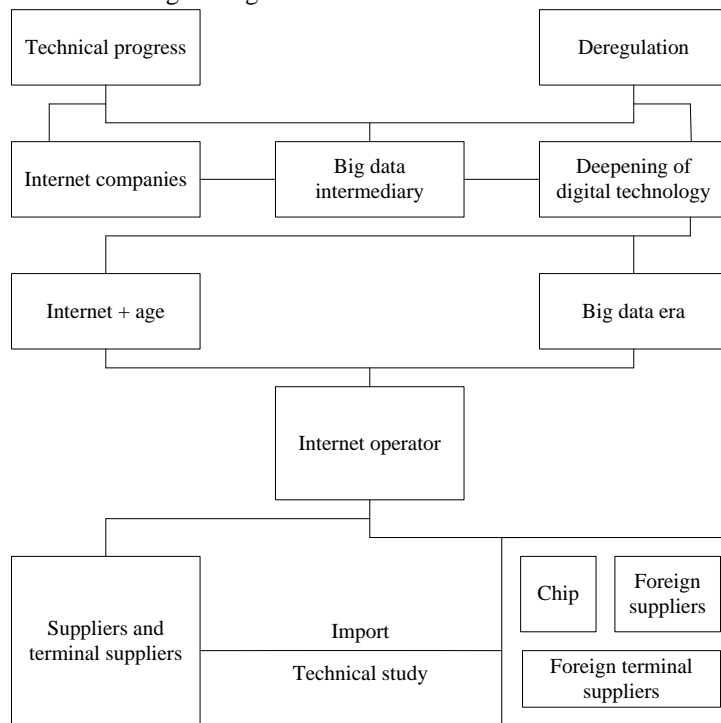
Based on the above algorithm, the entropy difference of each stage is calculated. Combined with the mathematical model, the value vector is updated, and the mixed equilibrium is calculated to obtain  $(Q_i | q_i, S_i^1, S_i^2) S_i$  and state probability, Analyzing the security situation of the active defense network, thus realizing the security evaluation of the active defense network. This paper analyzes the principle of network active defense, constructs an evaluation mathematical model, introduces the network entropy countermeasure quantification technology to weigh the benefits, reduces the influence of interference factors, improves the accuracy of evaluation results, and uses the defender's benefit results as a quantitative evaluation standard to effectively evaluate and defend the security of active Internet information data.

**2.3. Optimization of practice methods of big data in internet industry**

After the defense of data security in the Internet industry is realized, the current application of big data technology in the Internet is further analyzed and relevant optimization suggestions are put forward. Due to the huge market demand space, the flat development of production and sales has become more and more obvious, making small enterprises begin to transform into large-scale social production, which in turn leads to the lengthening of the

industrial chain and shortening of the life cycle of small enterprises. However, according to relevant research results, under the current market economy conditions, personalization and scale often have certain contradictions. For large enterprises, the cost problem caused by personalized production cannot be ignored, while small enterprises have market problems in personalized production. However, C2B based on big data solves the above problems well.

C2B of Big Data can meet the development needs of different industries and fields by providing big data services to manufacturers for a fee. Therefore, this paper analyzes the Internet sharing thinking based on big data and its actual application. Its elements include: the foundation of the development of Internet sharing thinking: cloud computing, social network, big data, search technology, mobile Internet, Internet of Things technology, which is not only the biggest driving force of future innovation in information technology economics, but also an important condition for the integration of Internet industry and economic industry. Enterprises in the "internet plus" era are called the economic leaders of the new world of the Internet. Their innovative theories on economy have been unanimously recognized by the business community, thus further promoting the transformation of the economic model of information technology. According to the above theoretical analysis, the current application framework of internet plus information technology based on big data is obtained, as shown in Figure 1.



**Figure 1. "Internet plus" data application framework**

As can be seen from fig. 1, the main contents of the framework include: internet economics is an important theoretical basis for the mutual integration of the internet industry and the economic industry, and it is also an important role in the process of integration. In order to better ensure the application and development of big data in the Internet industry, the big data processing level is optimized, including:

**Big data processing layer:** this layer uses HDFS database and Hbase database to manage structured and unstructured data related to e-commerce. The data mainly comes from the e-commerce network management side and the billing side, including: core network management data, detailed list data, network optimization platform data, complaint data, user information table, etc. After pre-treatment and algorithm processing, these data are stored in Hbase according to standard data format.

**Big Data Management:** Based on Hadoop management platform, specific data preprocessing scripts and algorithm models are established to realize analysis and management of user value and user perceived data. Pre-processing of data mainly includes data processing and noise data processing. E-commerce personalized recommendation system mainly uses big data analysis algorithm model, including analytic hierarchy process and clustering threshold method, to analyze the value and perception of users.

**Big Data Business Layer:** Combines and manages user value and user perceived business, analyzes all dimensions affecting e-commerce user value and perceived business, and finds out their correlation. For example, sorting out the relationship between user value and income, terminal, business and package, sorting out the relationship between user perception and decline, etc.

**Big Data Display Layer:** This layer displays the results of data analysis with charts, assists in the planning of personalized recommendation system for e-commerce, and focuses on geographic display of user values and perceptions and output of relevant charts.

Therefore, people can realize the effective application of big data technology in the Internet industry.

### 3. Empirical Analysis

In order to verify the practical effect of big data in the Internet industry, analysis and investigation were carried out, and the relevant data collected were classified and numbered, namely: T0, T1, T2, T3, T4, T5, T6; The types of data are: Big Data in Internet Economy, Big Data in Internet Agriculture, Big Data in Internet Trade, Big Data in Internet Industry, Big Data in Internet Transportation; The stability of traditional big data automatic classification and processing system and cloud computing-based big data automatic classification and processing system are analyzed respectively.

Table 2. Automatic classification of big data based on internet

Experimental serial number	Accuracy of data classification/%	Data application efficiency/%	Forecast value of data classification
T0	91.17	92.34	92.15
T1	92.54	94.13	92.54
T2	93.52	96.35	93.69
T3	94.74	97.20	94.72
T4	95.56	98.34	95.89
T5	96.13	98.52	96.38
T6	97.52	99.05	97.89

According to the above table, the application efficiency of big data in the Internet industry is extremely high. Based on the above parameters, further analysis is made to collect and analyze the economic and quantitative growth trend of the Internet industry in China in recent years, comparing the development trend of the Internet industry in the traditional environment with the development value of related Internet industry based on big data technology and the development trend of the Internet industry in the big data environment, and drawing a graph as follows.

As can be seen from the above figure, with the development of Internet technology in the past ten years, the overall level of the Internet industry that abandons big data technology is obviously lower than the development trend of the Internet industry that combines big data technology. Moreover, the growth of the number and

economic benefits of Internet industry based on big data technology has been significantly improved. Among them, from 2000 to 2005, the economic benefits of Internet industry were the best. From 2005 to 2010, Internet industry based on big data has mushroomed. From 2010 to 2015, with the maturity and progress of big data technology, the number and benefits of Internet industry have shown an obvious trend of improvement. From 2015 to today, Internet industry based on big data has tended to be stable. Therefore, it can be confirmed that big data technology plays an important role in promoting the development of the Internet industry, and the application of big data technology can effectively promote the healthy development of the Internet industry.

### 4. Concluding Remarks

Big data has penetrated into people's lives. In short, the emergence of big data has brought more vitality and vigor to the development of Internet economy, and has also promoted the development of Internet economy to a certain extent. From a long-term perspective, integrating big

data into the development of the Internet economy is a necessary trend of the times and will also become one of the important driving forces for future economic development.

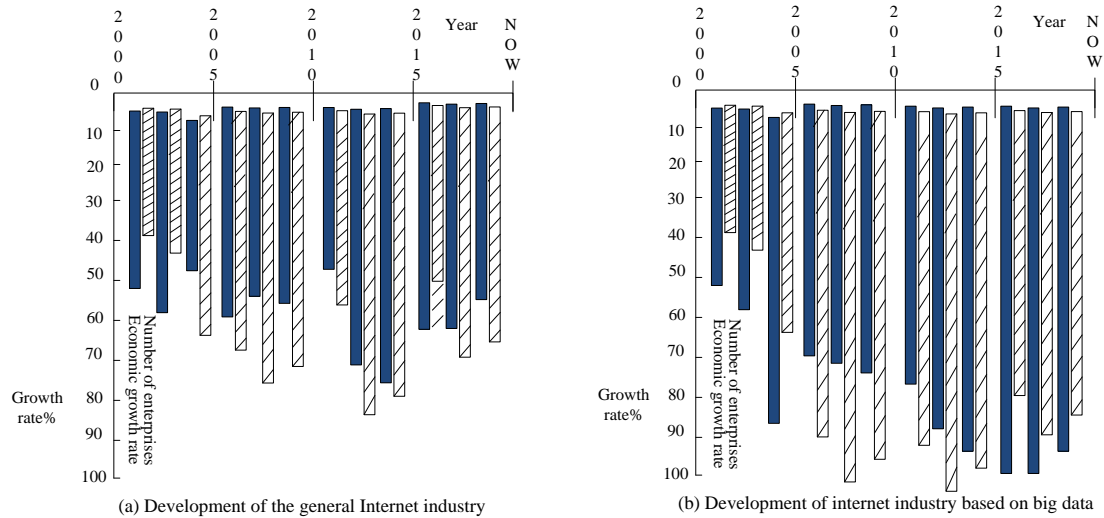


Figure 2. Survey results of big data internet industry development

References

[1] Ding Cun Honghu, Zhao Yuan. Research on the Importance of big data to internet microcredit companies-taking "ali small loan" as an example. Heilongjiang Education (Theory and Practice). 2017, 25(z1), 26-27.

[2] Zhou Xin. Research on evaluation index system and method for high quality development of manufacturing enterprises based on internet big data-taking listed enterprises in chongqing as an example. Proceedings of the 6th National Statistical Modeling Contest in 2018,13 (01), 43-44.

[3] Hua Qingwei. Application and research of computer big data in internet learning. Electroacoustic Technology. 2018, 42, 397(3), 80-82.

[4] He Zhirong. Research on the application model and value of big data based on internet finance. China's Circulation Economy. 2017, 31(5), 39-46.

[5] Zhang Jiwang. Application and discussion of big data technology in "internet plus" era. Electronic Technology and Software Engineering. 2017, 36(13), 157-157.

[6] Ding Linchong, Zhang Jing. Research on big data technology and its application in communication industry. Information and Computer (Theoretical Edition). 2017, (14), 116-117.

[7] Che Yulong. Research on the new marketing format of "internet plus publishing house" in big data era. News Research Guide. 2017, 8(17), 241-241.

[8] Liu Zhijin. Discussion on application of big data in internet maintenance of unicom. Digital Technology and Application. 2018, 36, 332(2), 49+195.

[9] Li Bing. "Innovation and practice of accounting theory in internet plus" big data environment. Science and Technology Economic Journal. 2018, 26, 637(11), 129+134.

[10] Zhu Wei, Qing Liu, Cheng Yan, et al. Path analysis of promoting industrial development by using internet and big data-taking hubei province as an example. Scientific and Technological Progress and Countermeasures. 2017, 34(24), 64-70.