Application of Big Data Technology in JD

Ning Shi$^{(\boxtimes)}$ and Huwei Liu

Beijing Wuzi University, Beijing, China
398807172@qq.com, 591878845@qq.com

Abstract. The arrival of the era of big data has brought about changes and impacts on human life, work, and thinking. With the rapid development of the scale and number of e-commerce in China, the e-commerce marketing requires continuous innovation. Big data can tap and utilize the underlying business value behind the data to achieve more precise positioning and marketing. This article analyzes the big data theory and method, discusses the three major challenges of data holding, data processing and data security brought by e-commerce in the era of big data. The era of big data analyzes and accurately updates and changes the target audience. A case study of JD e-commerce company was conducted again to analyze JD's big data platform and the application and practice of marketing based on the platform. Inspired by the case study, we found weaknesses and made suggestions.

Keywords: Big data · JD · Butterfly Festival

1 Research Background

Nowadays, with the widespread use of the Internet and the development of the Internet of Things, information is increasingly large. Behind these data information, important information and commercial value are hidden. Because of technological innovations and applications such as cloud computing, data that was originally difficult to collect for storage and use began to reflect its value and was easily used. With the explosion of information and the breakthrough of modern technology, big data will bring about tremendous changes to life and business. Big data will gradually create more value for humanity. In recent years, big data has been increasingly sought after by various industries.

On the Internet, the use of big data can completely describe the trajectory of each individual's life, and can completely capture the main behavior of each individual or consumer. Big data on the Internet carries the most extensive intentions and needs of consumers. Every search, click, browse, and comment in life is a real life moment that can directly reflect their personality, preferences, and wishes. Using data mining and other technologies to describe the data on

the Internet to describe consumers' habits, etc., can find and use the commercial value behind the data and achieve more precise positioning and marketing.

On January 31, the China Internet Network Information Center (CNNIC) released the 41st "Statistical Report on China's Internet Development" in Beijing. As of December 2017, the number of Internet users in China reached 772 million, and the penetration rate reached 55.8% [1]. Modern people have become more and more dependent on the Internet. The web pages and websites we have browsed are all recorded in the form of data. This is the source of big data.

2   The Concept of Big Data

Big data is one of the hottest words in the world. What is big data?

The Gartner Research Institute believes that after an effective processing model, big data has special values such as better decision-making capabilities, process optimization capabilities, and insight capabilities [2]. Big data has information assets that are diversified in type, have high-speed data growth, and have a quantity of Shanghai characteristics.

Big data has the following four characteristics that summarize big data: Volume (large quantity), Variety (variety), Velocity (high frequency speed), and Value (high value) [3].

Volume refers to the huge amount of data, and the measurement unit of big data cannot be measured in GB at all. It rises from GB to TB, and TB quickly rises to PB.

Variety mainly represents the rich and diversified characteristics of data types. The types of big data include a large amount of unstructured data (pictures, videos, texts, geographic locations, audio files, etc.), different types of data derived from sensors, recorded data of mobile terminals, search browsing data on the Internet, and other real-time online data, etc.

Velocity refers to processing speed. The recording of large amounts of data, etc. imposes higher requirements on the processing technology of big data. It requires faster and more efficient data distribution and discrete technology processing. This is precisely the difference from the traditional nature of data mining technology [4]. The high-speed and high-accuracy data mining processing of massive data can only exert its huge commercial value on the results of processing.

Value refers to the low density of value and high value in business. The 21st century is an era of information. With the use of cloud computing and the production of massive data, the use and use of data in the society will become increasingly dependent in the future. Many industries and companies have been keenly aware of the business opportunities and value. However, due to the inadequacies and limitations of the era and knowledge and technology, the value of big data could not be fully tapped and used.

3 Big Data Application in JD

3.1 JD Introduction

On March 2nd, 2018, JD released the financial statements for the fourth quarter and the full year of fiscal year 2017. The data shows that in 2017, JD net revenue was RMB 362.3 billion (approximately US$55.7 billion), an increase of 40.3% year-on-year; net profit was RMB 116.8 million (approximately US$ 18 million) [5]. JD is currently China's largest B2C e-commerce company, accounting for more than half of China's self-operated e-commerce market share.

In 2017, JD established a strategy for unbounded retail. At the beginning of this year, JD made major adjustments to its organizational structure and established three major business groups, namely the Fast Consumer Business Group, the Electronic Entertainment Group, and the Lifestyle Business Group. At the cooperation conference of the consumer products division of the Fast Forward Business Group, JD Supermarkets announced the establishment of a "brand value community" together with its brand name. They upgraded their past cooperation to a "symbiotic cooperation" model and evolved toward the integration of "people, goods, and field" driven by big data and artificial intelligence. And to achieve data sharing, category construction, user win-win situation, scene symbiosis, and truly realize the symbiosis with the brand business, cooperation and win-win.

3.2 JD Big Data Platform and Application Framework

JD big data platform is divided into four major modules, namely customer modules, tool product modules, model modules, and technology modules. As shown in Fig. 1.

In the customer big data module, the source of customer big data mainly includes the general consumer, JD partners, and JD internal customer data. One of the partners with JD is data from merchants, data from suppliers, and data from other partners, such as Tencent. The data from the JD Group mainly comes from four parts: customer data from JD Mall, data from JD Finance, data from international business and patting network. From JD customer big data module, it can be seen that JD's acquisition of big data comes mainly from external data on the one hand, and JD's internal data on the other hand.

In the tool product module, it mainly includes the data compass of professional analysis such as recommendation system, search engine, JD hui. There is also JD independent innovation R&D dispatching platform, data integration development platform, data knowledge management platform, JD analyst platform, data mining platform, and data quality monitoring platform. The data compass can be divided into seller version, industry version and so on. Through the recommendation system, search engine, JD, etc., the data of ordinary consumers in the customer big data module is processed and applied.

The model module mainly includes user portrait, sales forecast, credit model, merchant rating, distribution grid and wind control model. The user's portrait is
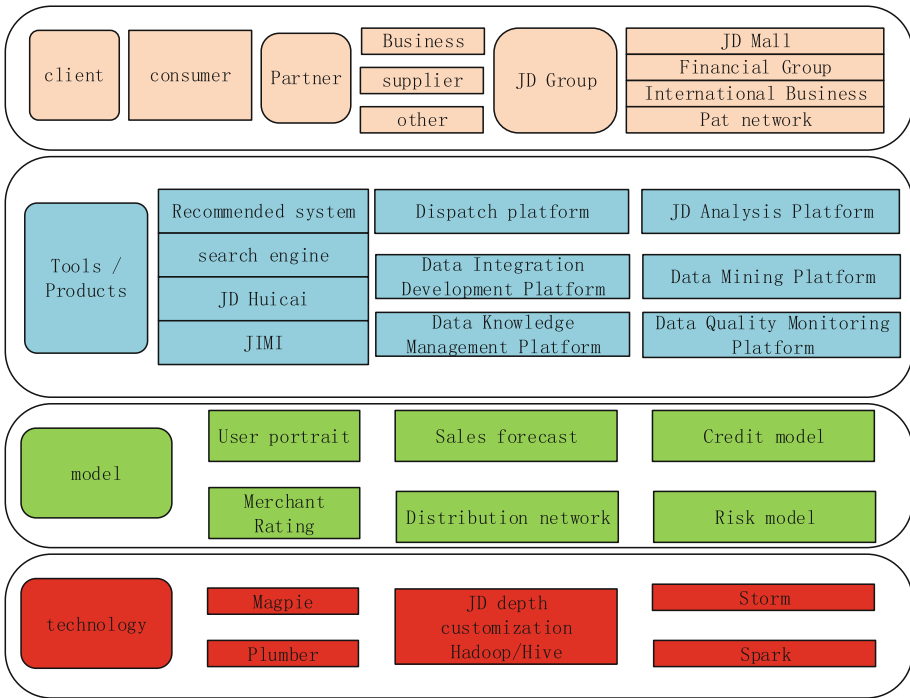
Fig. 1. JD big data platform four modules map.

a combination of experienced business staff and engineers after analyzing a large amount of data, and describes the consumer characteristics of a user group or a single customer with similar attributes and characteristics. Through the analysis of the user, the attributes of the user's portraits are tagged, and the users in the same tag can use the recommendation system to promote the products in a unified manner. User portraits are the most important part of achieving accurate marketing. The rating of the store and the results of the system are displayed on the front page of the business rating system and the back office system. Users and operators can quickly and easily find what they need, and both parties share information. The user clearly understands the performance and level of service, merchandise and timeliness of the stores in front of him. Rich and clear quantitative information provides efficient and objective data support for his decision to enter stores and shopping. The result of the rating is a report card submitted by the operator. The transcripts detail the good and bad shops, help operators find their own strengths and weaknesses, and promote their continued efforts to learn. This will not only improve the operational dead ends, but also attract more users and improve operational results through better rating results, creating a virtuous circle. The sales forecast is JD forecasting of the monthly, or even the entire year's sales of JD after analyzing the situation and data of warehousing and distribution, so as to facilitate the preparation and planning of

cargo warehousing and marketing of JD in the next month or year. Risk control is through the analysis of user behavior and words and actions to identify the malicious user, then the corresponding shield and reduce the risk of the mall, which is the application of big data in risk control.

In the technology module, technologies such as Magpie, Plumber, Storm, Spark, and JD's deeply customized version of Hadoop are included [6]. The technology module in the big data platform is the basic framework of the big data platform, and JD has many innovations in this technology. Through in-depth study of Hadoop distributed technology, many innovations are realized in the localization deployment process, such as:

(1) Directional development, support for index creation based on HDFS;
(2) Directional transformation, combining small files at the Map level, saving resource consumption;
(3) Support dynamic allocation of resource queues;
(4) Control of the task runtime and HDFS read and write support through parameters;
(5) Self-developed calculation functions applicable to JD Big Data Service Platform;
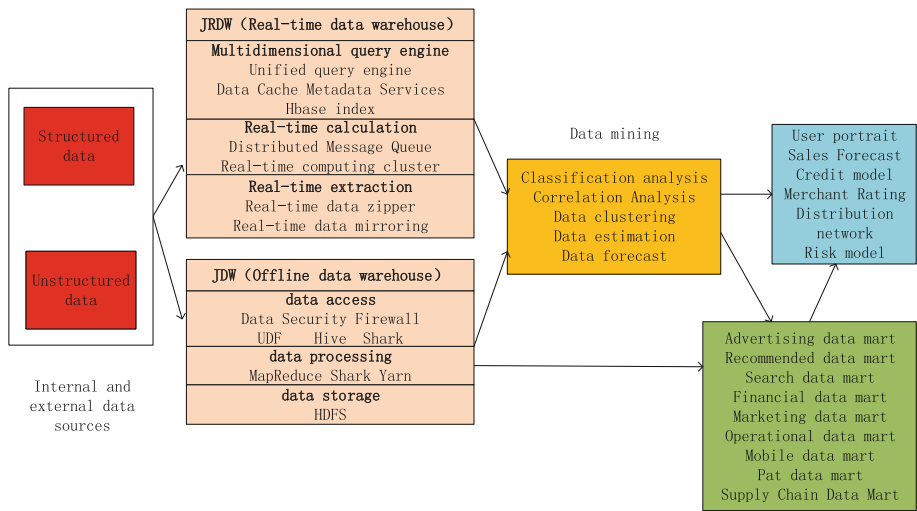(6) Optimize and improve Hadoop storage performance, support more types of data storage formats, etc.

Fig. 2. JD big data platform implementation mechanism.

In addition to the four major modules of JD Big Data, the implementation mechanism and basic architecture of JD Big Data Platform are shown in Fig. 2. First, collect internal data of the company. The data is mainly divided into

structured data and unstructured data. Due to the characteristics of big data to be processed in real time, JD has a large number of users who generate a large amount of data online every day. In order to obtain real-time data processing, maximum value can be obtained for timely processing and recommendation. Therefore, the data is divided into real-time data warehouses and ideal data warehouses to extract, calculate, and query real-time data. For historical data or offline data, it is also necessary to use HDFS for data storage, and use Map Reduce, spark, Yarn and other technologies for data processing, and then perform data access. In order to better serve enterprises and departments, JD has established a data mart, which is a line-oriented data production environment based on JDW. Provides application services for each line, including dozens of departments such as advertising, recommendation, search, finance, marketing, operations, BDA, mobile, pat, and supply chain, and thousands of users provide data services. Compared with the general data warehouse, data marts are more professional and easy to use. In fact, whether it is a real-time data warehouse or an offline data warehouse or a data mart, all the data are pre-processed and analyzed. Data mining is mainly achieved through classification analysis, correlation analysis, data clustering analysis, data estimation, and data prediction analysis. It faces big data. JD conducts sampling through distributed computing, iterates through memory calculations, and abstracts and simplifies common processes to lower the threshold. After a series of data mining, the data is returned to the data mart for use, or stored directly. Faced with the needs of different business departments, there are currently six business models such as user portraits, sales forecasts, estimation models, credit models, business ratings, distribution networks, and risk models. The corresponding business units can then call up to promote current business.

The above is the module and implementation mechanism of JD Big Data Platform. JD's big data platform can be used in many different departments and businesses, such as supply chain, pricing and risk control.

3.3   JD Big Data Application and Practice

JD has built its own precise marketing framework based on the big data platform, as shown in Fig. 3 below:

In JD big data precision marketing framework, collecting user data through log data, transaction-related data and non-transaction-related data left by users on the Internet has become the bottom of the precision marketing framework. User behavior modeling is then performed, such as user attribute recognition, user interest modeling, user relationship modeling, user life cycle, and user credit model [7]. Above these users behavior modeling, the user is portrayed to achieve user marketing value and user risk rating assessment, and is provided as the underlying data to various marketing systems. Finally, after knowing the portrait of the user, it uses EDM, SMS, APPPUSH, and product packaging advertisements to promote and achieve accurate marketing. Based on precision marketing on big data platforms, marketing methods are cross-platform, cross-terminal, and
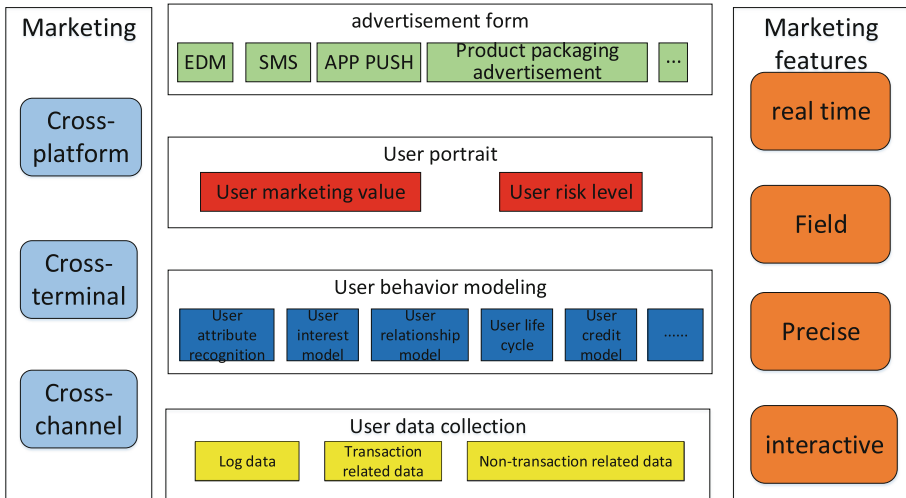
Fig. 3. Marketing structure.

cross-channel, with real-time, on-site, accurate, and interactive marketing features. Big data precision marketing is a dynamic and cyclical process. Continuous precision marketing is implemented before, during, and after implementation [8].

Among them, the user portrait is the most important foundation and part of big data precision marketing [9]. After the user's portrait is an analysis of a large amount of data, the data is used to describe the individual user as a basis or prediction of sales or product recommendation, which is also a solution for converting data into commercial value. Users' long-term behaviors such as page views and shopping are all multi-dimensional and huge data measured in terabytes, and these data constitute the basic data of user portraits. The data of user portraits can be roughly divided into three major parts. One is the basic demographic attributes, including natural attributes and social attributes such as age, gender, occupation, income, and life stage. The other is behavioral preference. This data mainly comes from the user's behavior when browsing the page to infer its preferences, such as the personal attention field. The last one is the data purchased by the user, the purchased goods, unit price, return rate, and evaluation feedback. The user's basic data and behavioral preferences are excavated with data and labeled with different labels to characterize the user's personal characteristics. In addition, by analyzing the crowd of the same tag, or by analyzing the user behavior preference characteristics in different dimensions of different business scenarios, the data is directly converted into the user image. A user may be portrayed by a 3000+ tag to match between the product and the user. The prediction of the user's future behavior leads to the user's behavior so as to obtain a huge business opportunity.

Logically speaking, JD portraits of users are mainly based on the data mining and analysis of massive data, forming a benchmark learning method or

criterion, and iterating continuously through machine learning methods until the best solution is finally formed, and the program is promoted [10]. However, in actual applications, user portraits are not simply relying on model building. The final step is to provide front-line personnel with business experience description portraits for proofreading and supplementation. This is a combination of the model and the actual market. The user's portrait will be based on the actual application scenarios of the user's portrait model and business experience to give different weights to show priorities. The user's portrait is declared successful after the business person takes the user's portrait and repeatedly verifies the modification in reality. The engineer will promote this portrait to the JD station application. Extracting and summarizing the user's portrait from the basic data, after being verified by the actual situation, it is extended to the application of the whole station. This continuous and rapid iterative approach enables JD to face up to changes and produce a large number of products with excellent performance. The user image is shown in Fig. 4.
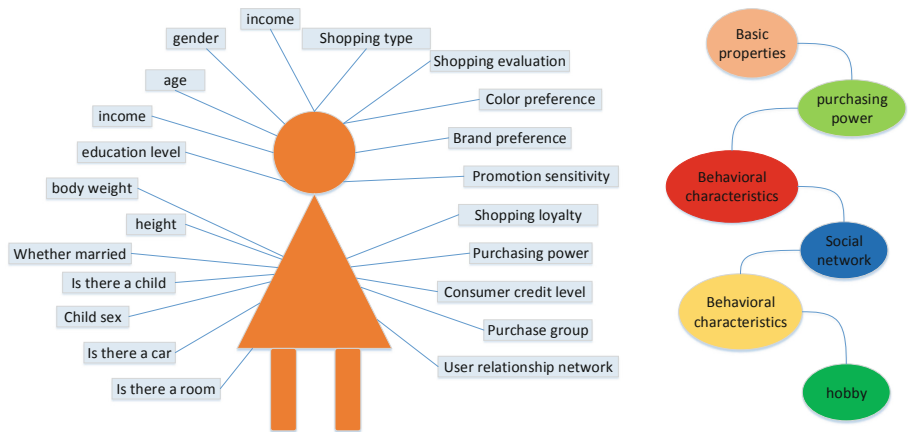

Fig. 4. User portraits.

Through the big data platform's acquisition and analysis of big data, JD will give each user a different label. Each user will have a lot of labels and images, and these labels will be dynamically updated according to the user's browsing and trading behavior. In this way, the user can be recommended for the product that the user is interested in, and the recommendation effect can be obtained from the user's feedback after the transaction, and the tag is updated again.

After applying big data, JD can get a variety of consumer shopping habits and preferences, find similar people and potential customers, and forecast and launch popular products in the market. Compared with traditional e-commerce providers, JD advantages are obvious, and it also reflects the extremely high value of big data technology. According to JD big user data [11], over the past

five years, the number of female online shopping users has grown by more than 200%, making JD more focused on women's shopping habits. In the past five years, the transaction volume of the JD platform has grown by a factor of 10, from RMB 86.9 billion in 2012 to RMB 939.2 billion this year. In addition to the success of brand building and the purchasing power of users, big data has played a big role.

Through the analysis of big data, corresponding accurate marketing strategies and personalized services are provided. Take JD Beauty-Butterfly Festival data as an example to analyze.

According to the latest JD Big Data, the sales volume of JD Beauty Cosmetics Co., Ltd. increased by 102% year-on-year.. In 2011–2017, the penetration rate of online beauty and makeup consumption continued to increase, and the online shopping habits of products have been formed. The proportion of online sales of major cosmetics groups has increased substantially.

The consumption growth index for each tier-one city in 2016 (the index reflects the relative level of each tier-level region and does not reflect the absolute value of growth) shows that in 2016, the "value-added consumption" in the Five-line area region grew fastest, reaching 203% year-on-year. The year-on-year growth of the six-wire region and the year-on-year growth of the fourth-line region were as high as 186% and 160% respectively. In contrast, the first-tier regions, second-tier regions and third-tier regions maintained rapid growth in 2016 compared with the same period in 2015, but the growth was only 100%, 118% and 139% respectively.

In 2017, consumers' skin care and beauty awareness increased significantly. In terms of specific categories, as consumption upgrades continue to ferment, from January to April 2017, from the perspective of year-on-year sales growth, the number of make-ups occupied four, and the "beauty of the eyebrows" category was highly favored. It is well known that the outbreak of makeup is closely related to the degree of economic development.

With the rise of "Men Color" era, men are also becoming one of the main consumers of beauty makeup. According to the comparative analysis of 2015–2016 big data, in the aspect of skin care beauty, the per capita consumption gap between male users and female users has decreased from 26.6 RMB in 2014 to 13.7 RMB.

According to the "Fine Beauty User Map", Guangdong, Beijing, and Jiangsu are the largest gathering places for young people (ages 19–25) who are keen on value-for-money consumption, ranking the top three. In addition, big data also shows that white-collar workers have become the main force of the "Appearance Association", and over 29% of users come from white-collar workers and general staff. It is worth noting that in 2016, both male and female users, facial skin care products accounted for more than 50%.

The big data analysis based on user evaluation found that the cost-effectiveness and service have become the most important factor for online beauty consumers, and the proportion of perception evaluation is over 70%. In addition, presenting a beauty care product is becoming a way of expressing

family care. The role of the wife is very important. Among the top five keywords, "wife likes" and "wife bought" occupy the top two.

4 Conclusion

In the context of big data, consumers can be divided into multiple groups. For each group, they can provide targeted services. The increase in the amount of data such as consumer behavior can provide a certain basis for e-commerce users to accurately screen and consume. Through the big data window, e-commerce companies can carry out more personalized and intelligent ad-push services, and can create more accurate business models. In addition, e-commerce through big data, can also better increase the viscosity of users to better carry out new product development and service, and ultimately reduce operating costs.

E-commerce providers can better drive and operate data. For internal links, companies can use data to analyze and evaluate them, and then use these data views for management. Taking JD as an example, it integrates its resources such as JD Mall, JD, and JD Logistics, and then uses these data resources to accurately target customer groups and consumer behavior. At the same time, it also elevates the competition for e-commerce from a simple price war to a differentiated competition. At present, JD has formed a data platform product, such as Data Cube and Quantum Hengdao, which includes operations analysis of stores, product analysis and buyer behavior analysis, industry analysis, financial analysis, and supply chain analysis.

In the context of big data, data assets have become a consensus and have become a relatively core industrial trend. In the future of enterprise competition, it is a competition of scale and activity. The economic benefits and effects of data generation have become more important for companies, and thus have spawned various data-related businesses. Data assets are a kind of capital manifestation of the current Internet ubiquity. Its role in the Internet is not only reflected in the service itself, but also has a certain degree of financial value. The data function not only reflects the use value in the product, but also can grow into a kind of real value. Currently, leaders of data assets, such as Apple, Google, etc., are using different forms of software development to better collect different types of data and better play the commercial value of big data. This new type of integration will eventually form a four-in-one pan-Internet company with "terminal + application + platform + data", which can better survive in the fierce big data competition.

References

1. China Internet Information Center, October 2017. http://tech.sina.com.cn/i/2018-01-31/doc-ifyrcsrv9714983.shtml. Accessed 31 Jan 2018
2. Li, B.W.: E-commerce Big Data. Publishing House of Electronics Industry, Beijing (2014)

3. Ma, J.G.: The concept, characteristics and application of big data. Natl. Def. Sci. Technol. **10**(17), 10–17 (2013)
4. Meng, X.: Big data management: concepts, techniques, and challenges. Comput. Res. Dev., 146–169 (2013)
5. Speedway. http://finance.jrj.com.cn/tech/2018/03/02184824186090.shtml. Accessed 02 Mar 2018
6. JD R&D System: JD Technology Decryption. Publishing House of Electronics Industry, Beijing (2014)
7. Chen, X.: Discussion on business model of big data business. E-commerce, 16–17 (2013)
8. Wang, X.: Application of big data technology in precision marketing. Inf. Commun. Technol., 21–26 (2014)
9. Huang, S.: The deconstruction and reconstruction of marketing system under the background of "big data". Mod. Commun. (J. Commun. Univ. China), 13–20 (2012)
10. Zhang, D.: The opportunities and challenges of electronic commerce development in the era of big data. Foreign Trade, 85–86 (2014)
11. China Economic Net. http://finance.ce.cn/rolling/201803/02/t20180302_28318714.shtml. Accessed 02 Mar 2018