

Comparison of Bigdata Application in User Behavior for Short Video Industry: TikTok and Xiaohongshu

Fangxing Li*

School of Jinan Foreign Language, Jinan, China

*Corresponding author: 100884@yzpc.edu.cn

Abstract. With the progression of innovation, the applications created based on enormous information will gotten to be increasingly copious. Since the computing control of computers and the gigantic sum of information accessible are not easy issue, cleverly applications spoken to by manufactured insights and profound learning will ended up more astute. Within the setting of such industry competition, each brief video stage still must deeply understand the contrasts of its clients in numerous division measurements, so as to realize more precise advertise situating and define more sensible improvement methodologies. This article uses two platforms which analyze user behaviors depending on bigdata. AARRR Model is a technological tool that helps firms to find their problems. Bigdata is a basic section to build up such model. According to the analysis, the two companies built their own big data analysis methods based on this model to help them achieve greater success. These results shed light on guiding further exploration of big data analysis implementation for short video industry.

Keywords: Bigdata; User behavior; TikTok; Xiaohongshu.

1. Introduction

Oracle, which is the biggest supplier of information management software and service around the world, stated that the definition of bigdata is data that contains more noteworthy assortment, arriving in expanding volumes and with more speed [1]. Bigdata develops rapidly as more and more companies are willing to use it. There is data generation in the world. For example, it was big bang of data on 31 December 2016. Volume of data would be doubled on 31 December 2020. Then, data volume will be doubled again on 31 December 2024. The whole of available progressed data at the around the world level created from 150 exabytes in 2005 to 1200 exabytes in 2010. It is expected to expand by 40% each year inside the another few a long time, which is nearly 40 times the much-debated advancement of the world's people [2]. Therefore, a lot of companies use bigdata from population to help them analyze consumer behavior. This essay aims to know how such business use bigdata on consumer behavior via short video industry, especially two famous apps that are TikTok and Xiaohongshu. It is necessary to know about some information about these two firms.

TikTok is a popular social media app that permits clients to make, observe, and share 15-second recordings shot on portable gadgets or webcams. With its personalized nourishes of peculiar brief recordings set to music and sound impacts, the app is eminent for its addictive quality and tall levels of engagement. Novice and proficient makers alike can include impacts (e.g., channels, foundation music, and stickers to their recordings), and can collaborate on substance and make split-screen two-part harmony recordings indeed on the off chance that they're in several areas [3]. As for Xiaohongshu, it is from China. With a mission to "inspire lives", it could be a way of life stage that motivates individuals to find and interface with a extend of assorted ways of life. Millions of clients grandstand their encounters on the stage day by day, from beauty care products and magnificence to design, nourishment, travel, amusement, wellness, and childcare, brought to life outwardly through an assortment of groups counting photographs, content, recordings and livestreaming. The inventive stage coordinating the bona fide substance shared by its community with commerce, rapidly getting to be one of the foremost prevalent goals for making way of life choices [4].

It is meaningful for company to use bigdata to determine the consumer behaviors. Hence, entrepreneurs should know the value of analyzing consumer behavior and advantage of using bigdata. The approach of the enormous information time has changed the way shoppers get item data, and the

data is more adequate and exact. Inside the ordinary promote illustrate, buyers by and large know a certain thing or brand through advancements and need other data reinforce, which is able restrict consumers' sound decision-making. Inside the period of gigantic data, clients can totally get a handle on the thing information through gigantic examination data, significantly get it the thing properties, and determinedly redesign from the situational incorporation of things to long-term consideration [5].

Back to real life, users of TikTok and Xiaohongshu are not hard to find something interesting. Both provide shopping platform rather than only sharing videos. Based on using the softwares, they often can meet something that they have interest or hope to know. There things that refer are from bigdata. Therefore, the motivation of whole paper is to find reasons about such situations. The reminder of the paper is organized as follows. The Sec. 2 will introduce the theory and features of bigdata analyzing consumer behavior. Subsequently, the Sec. 3 and 4 will demonstrate how bigdata is used in TikTok and Xiaohongshu, respectively. Afterwards, comparison about two firms will be clarified in Sec. 5 and limitations and future outlooks will be given accordingly in Sec. 6. Eventually, a brief summary will be given in Sec. 7.

2. User Behaviours Analysis on Bigdata

2.1 Theories and features

It is by and large accepted that enormous information examination has five characteristics: colossal sum of information, ultra-fast calculation speed, broadened information sorts, moo esteem thickness and tall data genuineness [6]. All of characteristics of bigdata allow more and more companies use it to help them. Huge amount of data helps managers get fuller picture of every user's behavior. Quick speed of calculation can give firms result fast which save timing cost. Diversified data types make entrepreneurs know more aspects of feeling and behaviors from consumers. It is a good way to improve details of a product. Accurate information provides more possibility for managers to make right decisions. Actually, managers cannot gain useful information from original data because of low value density. In this case, it is necessary to use some methods and models to analyze consumer behaviors via bigdata.

There are direct relationships and indirect relationships between firm's activities and users. The aim behind activity is that make users become performant consumers. The conversion rate depicts the relationship between visits (or clicks) on web site to changes. A change is the method of a potential client getting to be a genuine client [7]. The formula is $\text{Conversion Rate} = (\text{Number of Conversions} \times 100\%) / (\text{Number of Visits})$. Funnel analysis is a capable analytics strategy that appears outwardly the transformation between the foremost vital steps of the client travel [8]. Companies use bigdata to get conversion rate and then managers can use funnel analysis to find which part has the biggest problems and make a plan of improvement. AARRR model is often used on funnel analysis with bigdata. These five capital letters represent five stages of a product as shown in Fig. 1, where the definite meaning is given as follows:

- Acquisition: how users know this product? This is the first and fundamental step for a product. Consumers are bases.
- Activation: how is the user's first experience? This is a process of users finding a highlight feature of a product. Managers should make sure this product has an attraction.
- Retained: will the users back? The aim of retention is that using this product is a habit for users. Turn users into repeat consumers.
- Revenue: how to get more profit? Managers should turn users into paying users and know the difference between users and paying users. The target of developing a new product is profit. Enlarging group of paying users is helpful for income.
- Referral: will the users tell other people this product? In this stage, managers should think whether this product really satisfy users' needs and in which circumstance they will use this product. Finally, it is important to know the reasons behind users sharing this product.



Figure 1. AARRR Model

2.1.1 TikTok

Concurring to analytics location Sensor Tower, TikTok hit 1 billion downloads all-inclusive in February 2019 (barring Android downloads from China.) [9]. This data was three years ago, so there must be more and more downloads in 2022, especially more users appeared in 2020 during COVID-19 because people are limited to be out of door. It means that the data from TikTok is quite huge. Bigdata is useful and important for TikTok to handle problems and improve themselves. Actually, it is not difficult for their users that they can watch the majority of short videos that they are keen to. At the same time, definition of image of TikTok from users is like that ‘what are you playing.’ The using processes in TikTok solve the difficulty that no content to watch successfully. This advantage leads it to success eventually.

Its calculation customizes the suggestions for each client and recordings are played as before long as the app is opened, on the so-called ‘For You’ page, progressing with a straightforward ‘swipe’ [10]. Combined AARRR model with TikTok’s bigdata, they should acquire users firstly. The number of users will be recorded in bigdata. According to collecting users’ habits and interest, bigdata can divided them into different groups, and the processes are illustrated in Table. 1 and Fig. 2 [11].

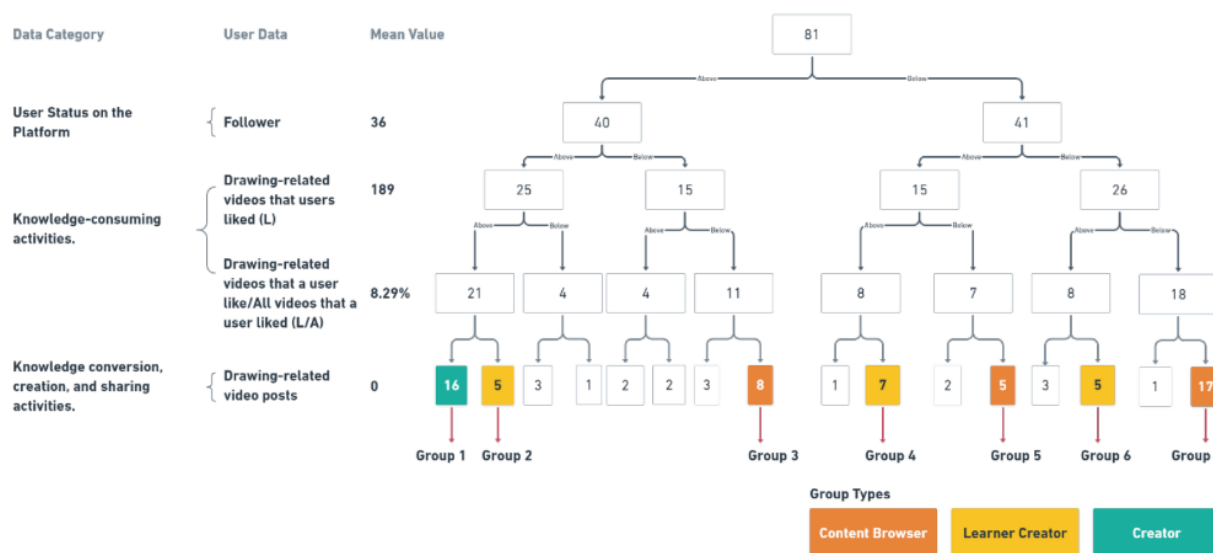


Figure 2. Process of dividing users into different groups

Table 1. Research results

Platform	Content browser	Learner creator	Creator	Total number
TikTok	36.7%	20.7%	19.5%	81

Creator browser is a group of people who do not have interests of specific theme of short videos. Learner creator means users want to learn certain aspects from TikTok short videos. Creator is a symbol that users send videos on platform. After the machine learns the interface of clients from bigdata, it'll thrust comparative recordings at a certain frequency, but it'll not thrust as well numerous

recordings of a certain sort to cause users' stylish weakness. This enters arrange 2 of AARRR show which is actuation. TikTok will physically select a few high-quality substances and thrust it to fans and clients who are fascinated by related labels. After TikTok experts discharge inventive recordings, it will direct conventional clients to mimic through operations. When a certain sort of imaginative video is prevalent, the machine will moreover utilize bigdata to create more suggestions for this sort of video to attract ordinary clients to take an interest. Meanwhile, both TikTok authorities and clients can issue theme challenges to direct clients to form beneath the same theme. The purpose of these activities is to retain users. Then, bigdata can give TikTok percentages of group of people as mentioned in Table. 1 and Fig. 2. Managers in TikTok can make decisions about their firm's future via data. It gives a plenty of sharing alternatives and recognizably prompts clients to spread substance by changing the share icon's color and shape. The numerous sharing alternatives amplify the reach of TikTok recordings, contributing to unparalleled video social spread [12].

2.1.2 Xiaohongshu

Monthly active user (MAU) is a key execution marker (KPI) utilized by social organizing and other companies to check the number of one-of-a-kind clients who visit a location inside the past month [13]. MAU of Xiaohongshu had been 200 million in 2021, which doubled the size in January 2020 (100 million). In 2022, its MAU will rise which is more than two hundred million. This number of users means there is a huge data store for Xiaohongshu, hence it is significant to use bigdata to analyze their user's behaviors in order to provide better services. The tag of this application is 'label my life'. There are two main ways to share life on Xiaohongshu. One is videos, and another is photographs. When users use it for a long time, they will find that they will receive more interesting posts. Notes on the Xiaohongshu platform will be classified by tags. Thus, one can see at a glance, and also recommend other notes with similar attributes according to the notes that users like and favorite, and customize different push content for different users. They also use bigdata to do achieve the goal.

Xiaohongshu takes life substance as the section point, they utilize KOL and KOC to spread the Seeding. It is permitting items to be profoundly coordinates with life. This approach makes clients think that the stage features a higher sense of reference and believe [14]. For Xiaohongshu, the community provides user stickiness to attract traffic for e-commerce. The e-commerce monetizes this part of the traffic and forms a closed loop in the APP. For the algorithm team, there is user data in the community and user behavior data in the e-commerce section. The way to connect the user behaviors on both sides and better understand users is the fundamental starting point of the algorithm.

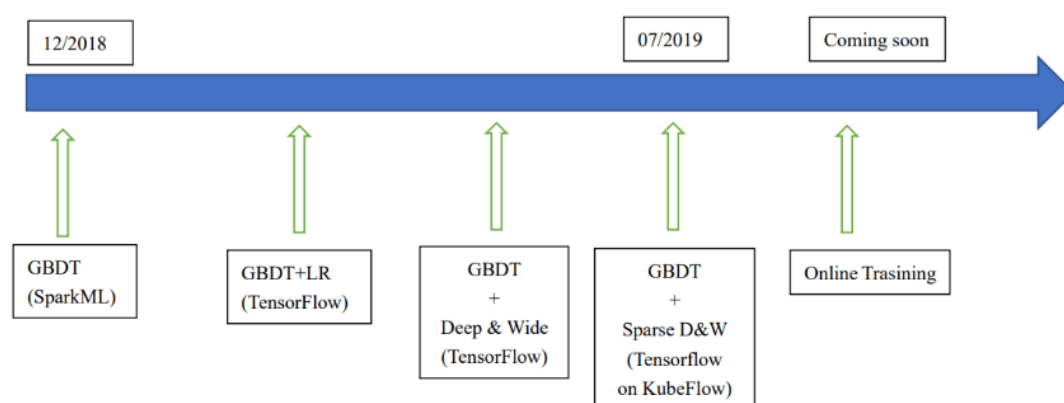


Figure 3. Model of predicting

Seen from Fig. 3, different algorithms are implemented to analyze users in Xiaohongshu via bigdata. They are used to do GBDT model in December 2018. With the advancement, the proposal expectation demonstrate of Xiaohongshu has advanced to the show of GBDT+Sparse D&W. There are 9 fundamental forecast assignments for bigdata, counting tap, stow away, like, fav, comment, share and take after. Tap is the biggest show of Xiaohongshu, which produces almost 500 million

tests a day for demonstrate preparing. It should be noted that distribution in the GBDT model has a lot of user behavior statistics, which generate some static information and dynamic features to describe users or notes. Describe users through user profiles and demographic information, such as static information such as gender and age. Notes are divided into two dimensions: author and content, such as author scores, note quality, tags, and topics. Dynamic features are few, but very important. Dynamic features include user feedback such as whether users click on browsing and searching, whether they have in-depth behaviors, and the like. These interactive data have a real-time pipeline from offline directly into the online model, and online will use these data to predict the click rate and other indicators of interaction quality, and then make recommendations based on the invisible classification of users and notes. Xiaohongshu builds a bridge and arrange for clients, which permits them to discover like-minded individuals rapidly and precisely. In this manner, it disposes of the sense of forlornness of clients, who may not respect themselves as confined islands [15].

2.2 Comparison

Although these two companies use bigdata to analyze their users' behavior, they still have some differences. TikTok's traffic algorithm is almost the most complex among all traffic platforms, and of course it has the largest traffic. TikTok is a typical "tag" to "tag" platform. The platform will disassemble the concerns into about 150 tags for users according to usual browsing preferences, and which videos they can browse are determined to a certain extent by your user tags. If the browsing preferences change, the user tags will also change, and the swiped videos will also change with the tags. The platform will form a creator label based on the content creators publish. The number of labels is also 150. If the content they publish changes, the creator label will also change. After the creator publishes the video, the video will match similar user tags according to the creator's tag. This is the "tag"-to-"tag" traffic algorithm that talked about above. After the short video is matched to the user, the data performance of the video will be used to measure whether the video is worthy of further recommendation. TikTok's recommendation for a single video will assess 5 key data:

- completion rate. Completion rate = viewing time / work time. The higher the completion rate, the more attractive the work is to watch. The qualified line of the market is usually around 15%-20%, and the completion rate of more than 40%-50% is already very good. To find a way to increase the completion rate, the usual way is to set suspense at the beginning or guide to open the comment area to prolong the viewing time.

- Like rate. Like rate = likes/plays. The higher the number of likes, the higher the number of recommendations. The like rate of the first wave of recommendations must reach at least 3%-5%, which means that for every 100 views, at least 3-5 likes are required.

- Message rate. Message rate = message volume / playback volume. The data level of the comment rate has a lot to do with the video type. It is not easy to use the average data to measure it, but it is certain that the better the comment rate performance, the higher the weighted recommendation. Therefore, creators can take the initiative to guide comments in the video or in the copywriting and comment areas to increase the rate of comments.

- Forwarding rate. Forwarding rate = forwarding volume / playback volume. The forwarding rate has little effect on the video that is still circulating in the primary traffic pool, but if you want to break through the traffic level, the forwarding rate is a key indicator.

- Conversion rate. Conversion rate = followers / views. The ratio of road-to-fans, and the new fan rate brought by a single video is also the key data to impact the advanced traffic pool.

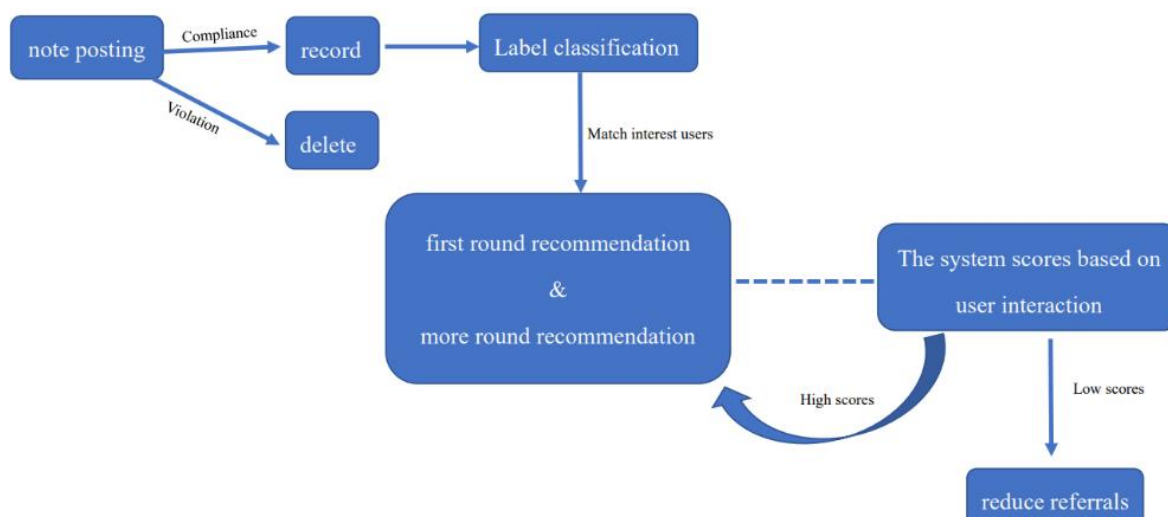


Figure 4. Traffic distribution model

The Fig. 4 exhibits the Traffic distribution model. The algorithm of Xiaohongshu is similar to that of TikTok, and it is also a traffic algorithm of "tag" to "tag". The difference is that based on different user habits, TikTok focuses more on active recommendations, while Xiaohongshu focuses more on search recommendations. Based on the platform positioning of Xiaohongshu, more than 65% of the traffic comes from search, so the search traffic algorithm is more refined, i.e., focus on the logic of the search traffic algorithm. The matching of search results and needs is mainly the matching degree of core keywords and query. The particular substance shown within the look comes about is to discover the data that can best meet wants of clients. The keywords in the title of a note can be described as the top priority, and the official also clearly reminds to fill in the title will have more likes. It can be seen that the title is an important option used by Xiaohongshu to identify content attributes. To make the notes more visible, the most basic work is to optimize the title.

3. Limitations and future outlooks

Since its birth, enormous information has continuously been looked for, and the concept of enormous information has been profoundly established within the hearts of the individuals. In any case, the advancement of enormous information has not however come to the level of total common sense that individuals envision. In spite of the fact that huge information is in a hot organize presently, behind the boom, enormous information itself still has a few exceptionally genuine issues. In terms of the case that these issues will not be an enormous issue within the utilize of huge information now. It may be a covered-up peril that ruins the advancement of huge information.

If enormous information needs to create, the preface is to have sufficient industry information to back it. Huge information was initially born in high-tech businesses such as the Web. Since of the disposition of these businesses, these businesses can get significant client information exceptionally well. Subsequently, within the application of huge information, high-tech businesses such as the Internet have accomplished the most excellent advancement. In conventional user-oriented offline businesses, the information of clients in these businesses cannot be collected well, and the collection of information in these businesses slacks distant behind desires of the huge information period. In a few extraordinary cases, clients in certain businesses are in a single frame, and such client information cannot be well organized and summarized.

When it comes to individual security assurance of huge information, this is often unquestionably an awfully vital point. Since bigdata itself must collect an expansive number of client information, the current information collection is all information gotten through legitimate channels. In any case, for numerous clients, they are continuously stressed that their data will be spilled, and clients cannot anticipate their information data from being spilled, since within the data age, everybody continuously

communicates through the Web. After the user's information data is spilled, they can't know the results caused by the spill, and the results are exceptionally genuine. Subsequently, the improvement of enormous information should fathom the problem of individual protection. On the one hand, it cannot be utilized without confinements, and everybody has the proper to know their individual security and the correct to deny. On the other hand, it is vital to discover a secure and solid way to share individual security information, hence enormous information can create. This is also the best way for both individuals and big data to win together. As a cutting-edge new technology, big data technology has become more and more mature in use today, but more efforts are needed in the follow-up maintenance and protection of user information. After all, as long as the cutting-edge technology of big data breaks through the industry predicament, the benefits brought by it are huge.

4. Conclusion

In conclusion, bigdata plays a significant role on analysis of user behaviors for these two companies. Their functions and values can be realized via Internet which makes them have huge data base. Actually, general models (e.g., AARRR model) need to be adjusted to implement in each company. So, they have different ways to handle this bigdata, but AARRR model is a basic knowledge to help them. There are some limitations of bigdata (e.g., privacy of users). At the same time, these two companies are just a symbol of whole short video industry, but they cannot represent whole industry. There are other firms to use better way to handle bigdata and then attract more users. In addition, short video industry cannot represent all industry. Some industries have no condition to collect and build up bigdata. Many industries need bigdata to analyze consumer behaviors rather than user behaviors. However, two methods of these two companies can provide an experience for other companies. They also make managers know the importance of right analysis of bigdata and user behaviors. Mastering the traffic algorithms of several platforms allows users to obtain as much traffic as possible. Bigdata has steadily entered into different businesses from the wilderness of science. All through the nation and overseas, huge information has shaped a mechanical scale and has risen to the national vital level. Enormous information innovation and applications have appeared an in-depth improvement slant. Huge data-oriented cloud computing innovations and enormous information computing systems have been persistently presented, modern enormous information mining strategies and calculations have developed in expansive numbers. In addition, modern enormous information models and unused commerce designs have risen one after another. On this basis, conventional businesses have started to utilize enormous information to realize change and overhauling. Overall, these results offer a guideline for implementation of bigdata techniques for corporations.

References

- [1] Sagioglu S, Sinanc D. Big data: A review. 2013 international conference on collaboration technologies and systems (CTS). IEEE, 2013: 42 - 47.
- [2] United Nations Global Pulse. Big Data for Development: Opportunities & Challenges. 2012.
- [3] TikTok. Investopedia, Retrieved from: <https://www.investopedia.com/what-is-tiktok-4588933>
- [4] Pamarathna R. Impact of xiaohongshu on its user based and society: a review. IRE Journal, 2019.
- [5] Zhang C., and Tan T. The Impact of Big Data Analysis on Consumer Behavior. Journal of Physics Conference Series, 2020, 1544: 012165.
- [6] Hofacker C F, Malthouse E C, Sultan F. Big data and consumer behavior: Imminent opportunities. Journal of consumer marketing, 2016.
- [7] McDowell W C, Wilson R C, Kile Jr C O. An examination of retail website design and conversion rate. Journal of Business Research, 2016, 69 (11): 4837 - 4842.
- [8] Copas J, Shi J Q. Meta-analysis, funnel plots and sensitivity analysis. Biostatistics, 2000, 1 (3): 247 - 262.
- [9] Bresnick E. Intensified Play: Cinematic study of TikTok mobile app. University of Southern California, 2019, 4 (4): 1 - 12.

- [10] Vázquez-Herrero, J., Negreira-Rey, M. C., and López-García, X. Let's dance the news! How the news media are adapting to the logic of TikTok. 2020.
- [11] Jung, H., and Q. Zhou. Learning and Sharing Creative Skills with Short Videos: A Case Study of User Behavior in TikTok and Bilibili. International Association of Societies of Design Research (IASDR) 2019.
- [12] Zulli, D., and Zulli, D. J. Extending the Internet meme: Conceptualizing technological mimesis and imitation publics on the TikTok platform. *New Media & Society*, 2020, 146144482098360.
- [13] Metcalfe B. Metcalfe's law after 40 years of ethernet. *Computer*, 2013, 46 (12): 26 - 31.
- [14] Chen Chuyue, et al. Based on COVID-19. Analysis of The Strategic Model of E-commerce Enterprises—
—a Case Study of Xiaohongshu (RED). *Proceedings of 3rd International Conference on Innovations in Economic Management and Social Science (IEMSS 2021)*. Ed., 2021, 251 - 258.
- [15] Lian K, Chen Z, Zhang H. From the Perspective of Feminism: Market Positioning of Xiaohongshu. 2021 5th International Seminar on Education, Management and Social Sciences (ISEMSS 2021). Atlantis Press, 2021: 190 - 194.