

LITERATURE REVIEW AND PROSPECTS OF BIG DATA MARKETING BASED ON ITS RESEARCH FRAMEWORK

HONGWEI WU AND HUI GUAN*

School of Economics and Management
Dalian University

No. 10, Xuefu Street, Jinzhou New District, Dalian 116622, P. R. China
1016490075@qq.com; *Corresponding author: 2316405695@qq.com

Received December 2020; accepted March 2021

ABSTRACT. *In the era of the network economy, big data marketing, as an emerging marketing method, has gradually become the mainstream in the marketing field with its advantages of low cost and high efficiency. In this context, big data marketing has become an important research topic in academia. In accordance with the logical framework of precise customer acquisition, precise positioning and precise marketing, this article sorts out the domestic and foreign literature from the definition, data sources and types, user portraits, and precise marketing of big data marketing, analyzes and summarizes the data sources and data types, and data user portrait label system and method, precise marketing decision. We provide the knowledge framework of big data marketing and put forward the future research prospect.*

Keywords: Big data marketing, User portrait, Precision marketing

1. Introduction. With the advent of the Web3.0 era, more and more Internet users gather in Weibo, Twitter and various social and entertainment platforms. These aggregations leave a large amount of data about users' interests, hobbies, thoughts, and behaviors. These massive structured and semi-structured data make enterprises use big data technology to dig deep into the potential value of users hidden in the data, so as to achieve revenue from precision marketing.

As a new phenomenon, big data marketing has been proved in practice. Compared with the rapid development of practice, the academic circles pay less attention to it, and the related research is relatively few. At present, the research on big data marketing at home and abroad is mostly based on the following three aspects. First, from the perspective of research means, these studies use big data rather than small data sampling to obtain data that truly reflect user cognition, emotion, behavior and response [1]. Second, from the point of view of the purpose of the study, the related research is based on the obtained big data, which makes the empirical research data more real or excavates new variables and enriches the theoretical model [2]. Third, based on massive user generation data, big data technology is used to mine user value and implement precision marketing [3]. However, few scholars have studied the components of big data marketing, and the research on the internal mechanism of big data marketing is even scarce. In order to make people have a comprehensive understanding of big data marketing, it is necessary to carry out new research to construct the research framework of big data marketing, so as to make up for this academic gap, which is the problem that this paper tries to solve.

In this paper, according to the logical framework of accurate positioning and precision marketing, this paper combs the literature on the definition, data source and type, user portrait, precision marketing of big data marketing, summarizes and analyzes the data

source and data type, big data user portrait label system and method, so as to provide a new research perspective for precision marketing decision.

2. The Concept of Big Data Marketing. The word “big data” first appeared in the computer science of the 1980s. Since 2008, authoritative magazines such as “Nature” and “Science” have published “Big Data” special issues, discussing the important value of big data in multiple disciplines. At present, the academic circles have not unified the definition of big data. Japoc et al. [4] define big data as a collection of data beyond the capabilities of traditional databases; Similarly, Chen et al. [5] define big data as a collection of data whose content cannot be collected, stored, managed, and analyzed by traditional database software tools in regular time; Lohr [6] believes that the essence of big data is a collection of processes and business objectives for data storage, processing and access; Armstrong and Green [7] propose that big data needs to be processed using “new” technology. The above four definitions point out that big data is a massive and heterogeneous data set produced in the development of modern networks. However, compared with the first two definitions, the latter two definitions emphasize the corresponding new tools, new technologies and new processes, so that big data can be truly applied to other industries and play the economic and social value contained therein.

Big data marketing concept was first put forward by Larry Webber, who believes that big data marketing is a new marketing method applied to the Internet industry on the basis of massive data and big data analysis technology [8]; Arthur [9] thinks that the big data insight to stimulate user participation and improve marketing effect is big data marketing; McAfee and Brynjolfsson [10] regard big data marketing as an intelligent management activity to obtain information and win competitive advantage.

3. The Research Framework of Big Data Marketing. The rise of big data makes marketing must go through leapfrog development. Through the acquisition of network data, data mining can be used to aggregate and classify the data with marketing value to realize accurate marketing, so as to construct a new marketing paradigm in the era of big data. According to the logical process of big data marketing, this paper puts forward three levels of big data marketing research framework: accurate customer acquisition, accurate positioning and accurate marketing.

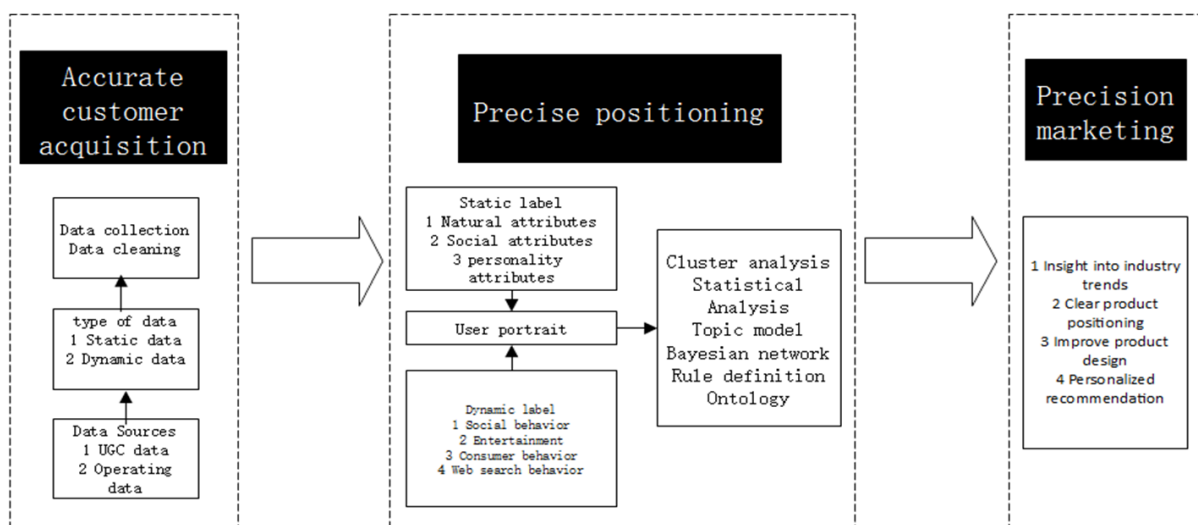


FIGURE 1. Framework diagram of big data marketing research

3.1. Accurate customer acquisition.

3.1.1. *Data source and data type.* Based on the literature, the authors summarize the data sources of big data marketing into two categories. The first category is UGC (user generated content) data, which is the data that users actively publish in social software, including online text data (post, comment) and online picture data. The second category is operational data, specifically including 1) entertainment behaviour data, which is data generated by users when they follow, like, repost, favourite, etc.; 2) consumer behavior data, which is the data generated by users when selecting items, creating orders and commenting feedback; 3) web search data, which is the data generated by users when searching and browsing.

Combined with data sources, data types can be divided into two categories, static data and dynamic data.

TABLE 1. Data sources and data types

Data sources		Website	Type of data	
			Static data	Dynamic data
UGC data	Social behavior data	WeChat, Weibo, Twitter, etc.	Demographic statistics (name, gender, date of birth, place of origin, marriage, education, etc.) Social attributes (family information, company information, communication information, etc.)	Length of visit, frequency of interaction, release of opinions, etc.
	Entertainment behavior data	Douyin, Huya Live, iQiyi, etc.		Like, follow, reprint, collection, etc.
Operational data	Consumer behavior data	Douyin, Huya Live, iQiyi, etc.		Price, quantity, transaction time, etc.
	Web search data	Baidu, Google, Zhihu, etc.		Search, browse, etc.

3.1.2. *Data acquisition and data cleaning.* Common data acquisition methods are mainly divided into web crawler and platform database. Web crawler is to use crawler tools to obtain user explicit data, such as basic information and user generated content, from user submitted registration information, Weibo text and so on. Platform database is through log mining technology, from the Web log and background database to obtain the technology. Compared with explicit data, implicit data can reflect user characteristics more truthfully and objectively.

There are a lot of low-value or even worthless data in the data set. If these data are not processed and analyzed directly, the analysis results will inevitably be affected. Therefore, data cleaning must be performed before analyzing these data. The main methods are text deduplication, denoising processing, mechanical compression, short sentence removal and word segmentation processing [11].

3.2. Precise positioning.

3.2.1. *Construction of user portrait labeling system.* User portraits are data modeling of users in the real world. Teixeira et al. [12] think that user portrait is a formal description of user based on specific business needs. Therefore, the user portrait is both from reality and higher than reality. So, to build a scientific and reasonable user portrait, we must first determine the user label. Based on the existing literature research, this paper constructs the user label system from the static and dynamic angles according to the data source and data type of big data marketing summarized above.

Static labels mainly show some stable user attributes, including natural attributes, social attributes and personality attributes. In [13], in order to identify the relationship between population characteristics and behavior differences, the user portrait is constructed from two aspects: user basic information and behavior preference; in Zhou et al.'s work [14], to verify whether the user's personality factor has a positive impact on the user's behavior, the personality factor is modeled as a dimension of the user's portrait to describe the user's characteristics.

Dynamic tags are mainly used to express the dynamic characteristics of users, including social behavior, entertainment behavior, consumer behavior and web search behavior. Wan et al. [15] explore the influence of Weibo information emotion type on user's willingness to share from two aspects of social behavior and entertainment behavior through empirical research; Yu and Wang [16] use the user's social behavior and web search behavior in addition to the natural and social attributes of users when constructing library user portrait models.

TABLE 2. User portrait tags

Label type	Attributes	Variable
Static label	Natural attributes	Name, gender, age, etc.
	Social attributes	Education, occupation, income, marital status, etc.
	Personality attributes	Personality, temperament, ideals, interests, etc.
Dynamic label	Social behavior	Length of visit, frequency of interaction, release of opinions, etc.
	Entertainment behavior	Like, follow, reprint, collection, etc.
	Consumer behavior	Category, price, quantity, transaction time, etc.
	Web search behavior	Search, browse, etc.

3.2.2. *User portrait tag extraction technology.* At present, user label extraction techniques mainly include clustering analysis, statistical analysis, topic model, Bayesian network, rule definition and ontology.

Clustering analysis: Clustering analysis is to divide users into several categories according to user characteristic data, to find the group connection between users, and to make the organization master the group characteristics and attributes of users. Marine-Roig and Clavé [17] use cluster analysis, blog and comments hidden information according to the keyword meaningful classification; Huang et al. [18] use K-means clustering analysis to divide mobile users from four dimensions (consumption ability, call volume, network request volume and displacement) into regular call type, instant Internet type, home saving type and random high consumption type.

Statistical analysis: Statistical analysis is to use statistical tools to make statistics on indicators with strong correlation with user behavior characteristics, mining user behavior characteristics, interest preferences, and so on, and then to achieve the extraction of user labels. Li [19] uses the method of statistical analysis to analyze the interest, social behavior and entertainment behavior of App users in order to obtain the correlation between user attributes and App types; Chen et al. [20] based on the user network social process generate (social relations, user behavior) and publish various data for statistical analysis, and then build user portrait.

Theme model: The theme model is used to detect hidden text information and extract relevant user preferences from the external performance of the user. Guo et al. [11] to explore the factors affecting hotel user satisfaction, use LDA theme model to quickly

identify various topics from a large number of comments; Ma et al. [21], on the basis of mining the user attributes of social network, the user attributes are predicted by the theme model.

Bayesian network: Bayesian network maps user behavior features to Bayesian network, predicts user characteristics by user behavior habits, and then quantifies user characteristics. Wang [22] to speculate the user's interest, maps the user behavior feature to the conversion weight between the nodes in the Bayesian network, and thus obtains the user interest label; in [23], based on the user online review data and hotel performance, a Bayesian network model is established to explore the relationship between the two.

Rule definition: Rule definition is based on research needs and data characteristics, combined with specific business rules, to customize user tags, and calculate relevant indicators. In Tahar and Samir's work [24], in order to clarify the user's retrieval preference and enhance the effectiveness of personalized retrieval, on the basis of calculating the weight of the user's retrieval words and labels, the correlation and correlation degree of the two are calculated by the rule definition; Chen and Hu [25] use the weight calculation rules of four labels, social attribute, interest, social behavior and information search behavior, to extract user labels when constructing UGC website user portraits.

Ontology: Ontology is based on ontology technology to describe users by using structured information and relational information defined in ontology. Tao et al. [26] based on the user's natural attributes, social attributes, and personality attribute data, through ontological methods, construct medical user portraits to achieve personalized recommendations; Liu et al. [27] extract user labels from user's natural attributes, social attributes and consumption behavior data, and construct tourist user portraits by ontology.

3.3. Precision marketing. Big data precision marketing is based on the premise of big data technology to obtain, clean and analyze information, through the construction of user portraits, visual form to display user behavior habits, consumption habits and other information, in case analysts can deeply analyze the value hidden in user data. Through mining the data in the data, really understand the needs of users, really achieve accurate marketing.

Insight into the industry situation: Meng et al. [28] construct the power user portrait model, describe the behavior characteristics of the power user group, provide intellectual support for decision-making, and realize accurate marketing; in [1], through the theme analysis and multiple regression analysis of the restaurant evaluation on the AllergyEats.com website, five groups of topics that have the greatest impact on the restaurant evaluation are identified to improve the service.

Clear product positioning: Wang et al. [29] construct the product positioning model based on the character model, which can objectively analyze and obtain the user's needs from many angles and make clear the product positioning; Nasraoui et al. [30] build a dynamic and evolvable portrait of user behavior based on user web logs to provide basis and guidance for website positioning.

Improving product design: Li [31] based on the characteristics of female headset to build a user portrait, designs a sibyl brand of female headset to meet the needs of users; Rao [32] based on communication user data, makes a comprehensive use of cluster analysis, analysis of variance and other methods to build retained and lost user, in order to explore the factors of user loss, so as to improve products and services.

Personalized recommendations: In [33], based on the data of library user behavior, the model of recommending library resources from the perspective of single user and multi-user is constructed, and the accurate recommendation of library resources for users in big data environment is realized; in [34], based on the analysis of the characteristics of college students' information needs, a personalized information recommendation system for college students is constructed.

4. Big Data Marketing Research Prospect. First of all, as an important analysis tool of big data marketing, user portrait can display user label more intuitively and vividly. However, the existing construction of user portrait is based on a specific field, not on the whole field. Therefore, based on different domain data, it is a direction that can be studied in the future to construct a more comprehensive and detailed user portrait that can be applied to various fields.

Secondly, the value of big data assisted business analysis and decision-making has been demonstrated in the practice of European and American countries. However, the domestic research on big data still stays at the level of data acquisition, cleaning, mining technology, and has not yet risen to the big data driven marketing decision paradigm. Therefore, only from the management point of view, explore how to use big data to assist in the formulation of organizational strategy, implementation strategy, personnel deployment and other practical issues, can really play the value of big data [35].

Thirdly, because big data technology has high technical threshold and use cost, only a few enterprises can use it, and these enterprises are mostly concentrated in the fields of telecommunications, finance and so on, which have higher entry threshold. At the same time, the database established by well-known enterprises lacks a unified big data information sharing platform isolated from each other. This makes enterprises in big data precision marketing data island problem. And only the big data in circulation can objectively and accurately describe the whole picture of customers and give play to their value in marketing. Therefore, how to open, circulate and share big data is a problem that needs to be explored and solved in the future. In addition, the personal privacy, ethics, security and property rights brought about by the acquisition and use of big data need special attention.

5. Conclusion. This paper summarizes the research results of big data marketing at home and abroad in recent years, and constructs the research framework of big data marketing on the basis of combing the concepts of big data and big data marketing. Combined with the existing research results, the operation mechanism and related technology are expounded in detail. In a word, at present, the research of big data marketing is still in the initial stage, whether it is the concept definition of the basic level, the research paradigm, or the application direction of the application level, the data reliability and so on, all need to be solved. The related research of big data marketing still has a long way to go.

Acknowledgment. This work is partially supported by my mentor, Professor Guan Hui. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

- [1] H. Wen, E. Park, C.-W. Tao, B. Chae, X. Li and J. Kwon, Exploring user-generated content related to dining experiences of consumers with food allergies, *International Journal of Hospitality Management*, vol.85, 2020.
- [2] P. Panchapakesan and J. Ahn, Guest satisfaction & dissatisfaction in luxury hotels: An application of big data, *International Journal of Hospitality Management*, vol.84, 2020.
- [3] Q. Liu, K. Niu, Z. He and X. He, Microblog user interest modeling based on feature propagation, *2013 6th International Symposium on Computational Intelligence and Design (ISCID)*, Hangzhou, China, 2013.
- [4] L. Japac, F. Kreuter, M. Berg, P. Biemer, P. Decker, C. Lampe, J. Lane, C. O'Neil and A. Usher, Big data in survey research: AAPOR task force report, *Public Opinion Quarterly*, vol.9, no.4, pp.839-880, 2015.
- [5] M. Chen, S. Mao and Y. Liu, Big data: A survey, *Mobile Networks & Applications*, vol.19, no.2, pp.171-209, 2014.

- [6] S. Lohr, *Data-ism: The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else*, HarperCollins Publishers, New York, NY, USA, 2015.
- [7] J. S. Armstrong and K. C. Green, *Guidelines for Science: Evidence and Checklists*, <http://dx.doi.org/10.2139/ssrn.3055874>, 2017.
- [8] A. Amado et al., Research trends on big data in marketing: A text mining and topic modeling based literature analysis, *European Research on Management & Business Economics*, vol.24, no.1, pp.1-7, 2018.
- [9] L. Arthur, *Big Data Marketing: Engage Your Customers More Effectively and Drive Value*, Wiley, 2013.
- [10] A. McAfee and E. Brynjolfsson, Big data: The management revolution (cover story), *Harvard Business Review*, vol.90, no.10, 2012.
- [11] Y. Guo, S. J. Barnes and Q. Jia, Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation, *Tourism Management*, vol.59, pp.467-483, 2017.
- [12] C. J. V. Teixeira, J. S. Pinto and J. A. Martins, User profiles in organizational environments, *Campus Wide Information Systems*, vol.25, no.3, pp.329-332, 2015.
- [13] A. Ma, *Design and Realization of Accurate Selling System Based on User Behavior Analysis*, Master Thesis, Nanjing University of Posts and Telecommunications, 2013.
- [14] M. Zhou, F. Wang, T. Zimmerman, H. Yang, E. M. Haber and L. Gou, Computational discovery of personal traits from social multimedia, *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, San Jose, CA, USA, 2013.
- [15] X. Wan, T. Yang, Y. Yuan and H. Sheng, The influence of the type of microblog information emotion on users' sharing willingness, *Information Science*, vol.37, no.8, pp.97-101,107, 2019.
- [16] X. Yu and Y. Wang, Research on the construction of library precision service model based on user portrait, *Library and Information Service*, no.22, pp.41-48, 2019.
- [17] E. Marine-Roig and S. A. Clavé, Tourism analytics with massive user-generated content: A case study of Barcelona, *Journal of Destination Marketing & Management*, vol.4, no.3, pp.162-172, 2015.
- [18] W. Huang, S. Xu, J. Wu and J. Wang, The profile construction of the mobile user, *Journal of Modern Information*, no.10, p.61, 2016.
- [19] Y. Li, *A Practical Study on the Statistical Methods of User Portraits in the Background of Big Data*, Master Thesis, Capital University of Economics and Business, 2016.
- [20] D. Chen, Q. Zhang, G. Chen, C. Fan and Q. Gao, Forum user profiling by incorporating user behavior and social network connections, in *Cognitive Computing – ICCV 2018. ICCV 2018. Lecture Notes in Computer Science*, J. Xiao, Z. Mao, T. Suzumura and L. Zhang (eds.), Cham, Springer, 2018.
- [21] C. Ma, C. Zhu, Y. Fu, H. Zhu, G. Liu and E. Chen, Social user profiling: A social-aware topic modeling perspective, in *Database Systems for Advanced Applications. DASFAA 2017. Lecture Notes in Computer Science*, S. Candan, L. Chen, T. Pedersen, L. Chang and W. Hua (eds.), Cham, Springer, 2017.
- [22] Q. Wang, Research on user interest model construction based on Bayes network, *Wireless Internet Technology*, no.12, pp.101-102, 2016.
- [23] J. Zhang, Z. Zhang and Y. Yang, The power of expert identity: How website-recognized expert reviews influence travelers' online rating behavior, *Tourism Management*, vol.55, pp.15-24, 2016.
- [24] R. Tahar and K. Samir, A geo-social user profile for a personalized information retrieval, *International Conference on Information Management & Engineering*, Oran, Algeria, 2016.
- [25] Z. Chen and Z. Hu, User portrait study on UGC website, *Computer Systems & Applications*, vol.26, no.1, pp.24-30, 2017.
- [26] X. Tao, Y. Li and N. Zhong, A personalized ontology model for web information gathering, *IEEE Transactions on Knowledge & Data Engineering*, vol.23, no.4, pp.496-511, 2014.
- [27] S. Liu, J. Sun, Y. Su and Y. Zhang, Literature review of Persona at home and abroad, *Information Studies: Theory & Application*, vol.41, no.11, pp.155-160, 2018.
- [28] W. Meng, X. Wu, J. Li, J. Wang, Y. Du, Y. Liang, Y. Wang, M. Xu and X. Lin, Power user portraits based on big data technology, *Telecommunications Science*, no.S1, pp.23-28, 2017.
- [29] Y. Wang, S. Yu, L. Lu et al., Study on capturing of user requirements based on Persona, *Computer Engineering and Applications*, vol.51, no.14, pp.247-252, 2015.
- [30] O. Nasraoui, M. Soliman, E. Saka, A. Badia and R. Germain, A web usage mining framework for mining evolving user profiles in dynamic web sites, *IEEE Transactions on Knowledge & Data Engineering*, vol.20, no.2, pp.202-215, 2008.
- [31] N. Li, The product design practice on the use of the Persona method to identify user needs: Taking the design of sibyle female head headset as an example, *Art & Design*, no.6, 2017.
- [32] X. Rao, *Based on the Reservation and Loss of User Persona Method Improve the Effect of User Research*, Master Thesis, Central China Normal University, 2017.

- [33] Q. Wang and F. Zhao, Design and analysis of library resource recommendation model based on user profile, *Journal of Modern Information*, no.3, pp.105-109,137, 2018.
- [34] Y. Yang and C. Yang, New media information recommendation strategy for personalized needs of college students, *The Party Building and Ideological Education in Schools*, no.24, pp.74-75, 2019.
- [35] H. Wu and H. Guan, The research on the training mode of the big data marketing in the background of the new business subject, *Proc. of the 2nd Asia-Pacific Social Science and Modern Education Conference (SSME2019)*, pp.12-16, 2019.