Research on the Overall Architecture and Application of E-Sports Big Data

Yinbin Liu[1] & Xiaoyue Zhang[1]

[1] School of Management, Shanghai University, Shanghai, China

Correspondence: Xiaoyue Zhang, School of Management, Shanghai University, Shanghai, China. E-mail: xiaoyuezhang111@126.com

Abstract

Big data has a profound impact on the transformation of human society and culture. With the rapid development of E-sports industry in the world, the application research of big data in the field of E-sports has gradually become one of the focuses of academic attention. Based on the method of literature analysis, this paper makes a systematic analysis of 628 articles about big data of E-sports published from 2013 to 2020, and attempts to systematically study the literature from the concepts, characteristics, system architecture and practical applications of E-sports big data, and propose possible future improvements in this field.

Keywords: E-sports, big data, application analysis

1. Introduction

In today's information age, with the deeper integration of the concept of big data with all areas of society, the field of E-sports has also entered the era of big data. As a product of the Internet era, E-sports itself has massive amounts of data, and digitization is a natural gene. Especially in recent years, global E-sports has entered explosive growth and popularization. In 2019, global E-sports revenue exceeded 1 billion U.S. dollars, and the audience exceeded 450 million people (Penguin Intelligence,2019). Behind the huge industry and user scale is the rapid increase in the amount of data in the E-sports field, which generates a large amount of structured and unstructured data. In this context, E-sports big data technology came into being. From event operation to content dissemination, data-driven technology effectively improved industry efficiency.

In recent years, scholars have carried out research on the application of big data in the field of E-sports from different angles. This paper first summarizes the concept and characteristics of E-sports big data. Secondly, from the aspect of system architecture, the framework of E-sports big data system is summarized and divided into four stages: data acquisition, data preprocessing and storage, data analysis and data application. Finally, from the practical application aspect, it summarizes the big data applications of B-end clubs, events, media, sponsors and C-end players in E-sports industry, and puts forward suggestions for the existing problems in the current research.

2. Concepts and Characteristics of E-Sports Big Data

2.1 Concepts

The rapid growth of data and the great value of big data in industrial development make big data widely used from concept to final realization. In 2008, Nature took the lead in publishing a big data special issue "Big data". Research found that the influence of big data has reached various fields from Internet technology to biomedicine (Howe et al., 2008). In 2011, Internet Data Center (IDC) defined big data as: Big data technology describes a new generation of technology and architecture, which is designed to obtain value from large amounts of data through high-speed capture, discovery and analysis (Gantz et al., 2011). In the same year, McKinsey(2011) defined big data as: Data sets whose data volume exceeds the capabilities of traditional data tools to capture, store, manage and analyze.

The current research on the definition of E-sports big data shows a diversified trend, and researchers have proposed different definitions from different angles and levels. From the operational level, Chen et al. (2016) defined it as meeting the precise operation requirements of E-sports games by collecting user behavior data, analyzing business status in real time, and digging in-depth market value. Based on the characteristics of

E-sports big data, industry insider Gao believes that by processing massive and high-growth data, it enables E-sports stakeholders to have more activity decision-making power and information insight (Aggro eSports, 2019). The research report released by JINGDATA (2018) explains the big data of E-sports from the perspective of application, that is, through the analysis of a large amount of data collected in game competitions, useful evidence can be extracted and formed to provide support for E-sports related industries.

2.1 Characteristics

In 2001, Laney (2001) first proposed describing data management from three dimensions, namely the 3V characteristics of big data: volume, velocity and variety. On this basis, the International Data Center also incorporates value into the characteristics system of big data, and IBM (2012) adds that veracity is also an inevitable attribute of big data, which constitutes the 5V characteristics of big data.

From the perspective of acquisition, processing, analysis, and application, E-sports big data has the general basic characteristics of big data, as well as its unique characteristics. According to the summary of the existing research, its specific performance is as follows:

(1) Volume

The initial unit of measurement of big data is at least P (10^3T), E (10^6T) or Z (10^9T) (Xu,2015). As a competitive sport in the new era, E-sports is endowed with massive data due to its industry's special digital attributes. As of the end of August 2020, there are 79,336 games on the top game platform Steam alone, the number of monthly active users has increased by nearly 95 million, and the daily amount of new data has reached the PB magnitude. Meanwhile, the continuous increase in the number of users and the rapid updating of games have accelerated the massive accumulation of data, making the overall output of E-sports big data considerable.

(2) Velocity

Speed means how fast the data is being generated, and how fast the data must be processed to meet demand. In the era of big data in E-sports, data plays an important role in the professionalism and viewing of the game. Therefore, E-sports events have higher requirements for the speed of data analysis, such as the low-latency real-time data analysis required in the game- Score data, team data, player data, event data, etc.

(3) Multi-dimension

In terms of data types, from the perspective of the audience, E-sports big data involves the data of events, teams and players. From a game perspective, E-sports games need to deal with multi-dimensional data, such as players, heroes, props, lineups, etc. Each dimension requires statistics on damage, team battles, positions, outfits, and game strategies, etc.

(4) Visualization

Visualization is an important characteristic of E-sports data. Data visualization is to convert the original attributes of the data into various visual elements such as symbols, shapes, textures, and finally generate a visual expression form of the data, so that users can understand the connotation characteristics of the data in a visual understanding way(Li et al,2015).Data visualization analysis technology is widely used to analyze player behavior in games and evaluate game performance-game design, playability, detailed characteristics, player experience, etc. (Lan et al,2017).

(5) Heterogeneity

Due to the diversity of sources and variable definitions, e-sports big data has obvious heterogeneity. On the one hand, due to the low degree of data openness of game manufacturers, data analysis enterprises obtain data from various channels and set up their own game databases. On the other hand, it is mainly reflected in the heterogeneity of technology, data analysis tools and models are not the same (Bai et al,2008).

3．E-sports Big Data System Architecture

A big data system is a complex system that provides data processing functions at different stages of the data life cycle (Li &Gong,2015) The core architecture of the entire big data covers four levels: acquisition layer, processing layer, model layer, application layer (Ma et al,2019). Combined with the characteristics of E-sports big data, the system architecture of E-sports big data is divided into four levels: data acquisition, data preprocessing and storage, data analysis, and data application. The details are shown in Figure 1.
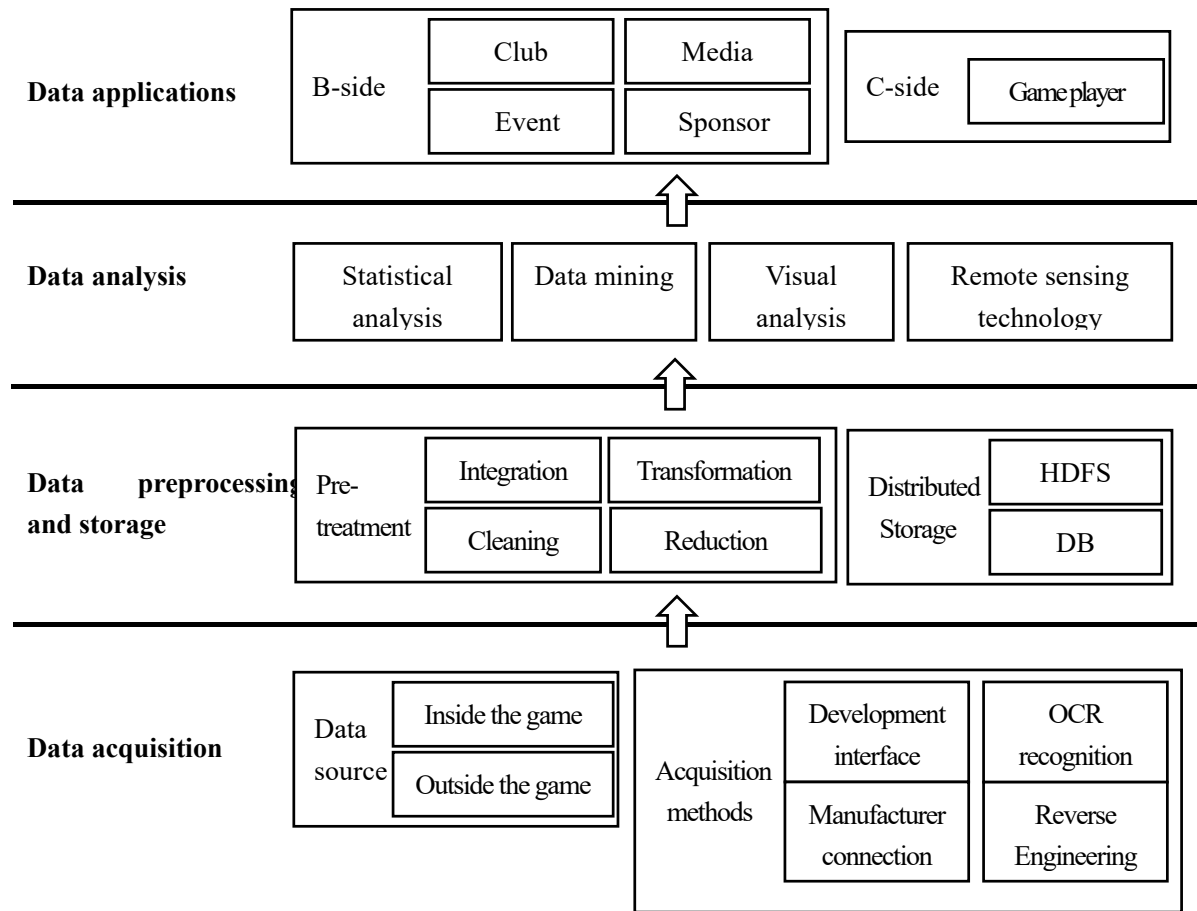
Figure 1. E-sports big data system architecture

3.1 Data Acquisition

E-sports data mainly includes in-game data (business regular data, player behavior data, project operation data) and out-of-game data (forum media data). Compared with traditional sports, E-sports has great advantages in game data acquisition, and can provide online information about events, players and teams (Makarov et al,2017). At present, there are four main ways to collect E-sports data: (1) Development interface. For example, Steam platform open DOTA Data, Blue Hole officially provides PUBG interface. (2) Game manufacturers connection. Game manufacturers can choose third-party data service providers to carry out cooperation. (3) OCR recognition. Intercept event video and live broadcast screens to identify data and events. (4) Reverse engineering. Including client-based request protocol, local-based metadata file, client-based code (Yu,2019).

3.2 Data Preprocessing and Storage

Due to the wide range of E-sports data sources, the original data has large differences in definition and structure, and has defects such as disorder, repeatability, and incompleteness (Salvador et al., 2000). Data preprocessing helps to improve data quality, and enhance the speed and accuracy of subsequent data analysis. The big data preprocessing process mainly covers data integration, data cleaning, data transformation and data simplification (She et al,2019). The pre-processed E-sports data is generally stored in HDFS, DB, HBase and other systems. As a distributed and scalable big data storage system, it can well support data random and real-time read-write access, and is suitable for the massive data magnitude of esports (Zhou & Wang, 2018).

3.3 Data Analysis

Big data analysis is the process of applying descriptive, predictive and prescriptive models to data to answer specific questions or discover new insights (Zeng, 2017). The traditional analysis methods of E-sports big data mainly include statistical analysis (descriptive statistics, inferential statistics) and data mining (regression analysis, cluster analysis, association analysis, etc.). Due to the increasing complexity of E-sports big data, visual analysis,

Internet of Things equipment, AI technology, remote sensing technology and other technologies are gradually incorporated into the analysis system of E-sports big data.

3.4 Data Application

Existing E-sports big data is mainly used in the industry. On the one hand, based on the needs of in-game players, it provides various basic user data for them to improve players' competitive level. On the other hand, it focuses on marketing and operations, including game communication effects, user level distribution, and advertising effects. The service objects can be divided into five main users: game players, clubs, events, media, and sponsors.

4. Analysis of the Application Status of E-Sports Big Data

At present, the application of E-sports big data is in the initial stage of exploration, and its research has not yet formed a system. Aiming at the application objects of E-sports big data, this article conducts an inductive analysis from five levels of E-sports B-side clubs, events, media, sponsors and C-side players.

4.1 Club Application Research

The club data application mainly includes team win rate prediction, player technical analysis, opponent analysis, tactical formulation, etc. Shim et al. (2011) performed a regression analysis on the team data after the "Halo 3" game, and the results showed that the team performance can be effectively predicted based on the player's ability and winning rate. In designing player techniques and tactics, Bauckhage et al. (2014) used spatial clustering techniques to evaluate athletes' behavior, analyze their abilities, and help them choose suitable game map areas. Ontanón et al. (2013) proposed semi-Markov models to predict the location information of opponents in "StarCraft", so as to formulate targeted combat plans for the team.

4.2 Event Application Research

The data application on the event side is based on the interpretation of game information and data, so as to make scientific decisions on event communication and star player creation, so that the audience can have a more intuitive event experience.

Block et al. (2018) designed an analysis tool based on Dota 2 event –"Echo", which can statistically analyze the performance data of athletes, and transform the dynamic data points into a graphic form easy to understand for presentation, so as to improve the audience's watching experience.

Regarding the marketing of specific events, Glory of Kings launched the official event APP- "Life of Kings", which analyzes game data to configure tasks and rewards, attract online users to offline, and achieve precise marketing for different groups of people.

4.3 Media Application Research

Data analysis on the media side uses big data to count some milestones or interesting data and information, and form widely disseminated content for dissemination. For example, PentaQ, an E-sports data service provider, is not only committed to industry data analysis, but also more in-depth content areas, providing readers with in-depth interviews and character columns. Global comprehensive data service Sportradar has reached a cooperation with ESL (Electronic Sports League) and media companies to develop a dynamic application specifically for E-sports data, so as to bring more exclusive content to the market through data (Li,2016).

4.4 Sponsor Application Research

The data analysis on the sponsor side is mainly reflected in the in-depth interpretation of E-sports event audience portrait and team performance through data, so as to select the appropriate brand image and release channels for the sponsor brands. Yang Chunhui et al. (2018) analyzed the sponsors of The League Of Legends World Championship and believed that the analysis of the characteristics of E-sports event users, such as age, gender and consumption ability, would help the sponsors to carry out accurate sponsorship. Based on the event history analysis (EHA) method, Jensen et al. (2018) used longitudinal data analysis to explore the risk factors and average sponsorship duration of E-sports sponsorship, and provide reference for the marketing decisions of E-sports sponsors.

4.5 Game Player Application Research

The C-side application of E-sports data is mainly for E-sports players. On the one hand, it links to the game API port, obtains and analyzes player data, and provides players with tactics or guidance. On the other hand, the player data is used to score the player's win rate, habitual heroes, game style, etc., so that players can better understand their game level and style.

In terms of tactics and guidance, El-Nasr et al. (2016) proposed to use game telemetry to obtain user behavior data, and analyze the effects of the data through related models to discover the advantages and disadvantages of players and provide them with improved tactical design.

In the study of player preference, Churchill et al. (2016) analyzed the player's preferred game types and roles by mining user viewing data and chat records on the Twitch platform to provide players with recommendation information.

5. Discussion

Although big data has become a hot research topic in the field of E-sports, there are still many basic problems in the research of E-sports big data that have not yet been resolved, which are worthy of further discussion:

(1) Promote the establishment of a unified data semantic standard. Semantics is the interpretation and logical representation of data in the field of E-sports. E-sports data sources are complex and diverse. In addition to open data interfaces, game manufacturers choose to establish cooperation with different data service providers and open data sources, resulting in relatively isolated E-sports data information with different definitions and measurement standards. In the future, researchers can work with industry and government agencies to jointly promote the establishment of standards.

(2) The data analysis method needs to be further consistent with the era of E-sports, and the research on visual analysis is still inadequate. The visualization of E-sports data results is the focus of current research and is also a technical difficulty. However, the work of visualization in E-sports is still in its infancy. Among the literature reviewed, there is still no research review related to visualization.

(3) Promote the practical application of E-sports big data research, break through the basic level of analysis. Most current researches still focus on the personal characteristics of players and are limited to basic applications within the industry. With reference to the development law of sports big data, big data can gradually reveal the relationship between the various elements of E-sports development, and in-depth management and execution. How to draw lessons from the development experience of sports big data still needs to be discussed.

6. Summary

This article reviews the basic concepts of E-sports big data. Based on the basic characteristics of big data, it summarizes the six characteristics of E-sports big data, and emphasizes that the architecture of E-sports big data system is composed of four stages: data acquisition, data preprocessing and storage, data analysis and data application. In addition, from an application perspective, this article introduces the actual application of big data E-sports industry B-side (club, event, media, sponsor) and C-side players. At present, the research and application of E-sports big data is still in the stage of exploration and development. The research system is not perfect, and some technologies are not mature. With the continuous improvement of E-sports industry and the increasing attention of the society and academia, the development prospect of E-sports big data will be broader.

References

Aggro eSports. (2019). The E-sports Big Data Era is Coming-China Entertainment Think Tank and Qihuang E-sports Deepen E-sports Big Data. Retrieved from https://zhuanlan.zhihu.com/p/70556323

Bai, Y. F., Wang, L. F., & Bao, W. B. (2008). Integrated technology and heterogeneous analysis of vast amount biological data. *Agriculture Network Information*, (10), 138-141.

Bauckhage, C., Sifa, R., & Drachen, A. (2014). Beyond heatmaps: Spatio-temporal clustering using behavior-based partitioning of game levels. *2014 IEEE Conference on Computational Intelligence and Games*, 1-8. https://doi.org/10.1109/cig.2014.6932865

Block, F., Hodge, V., Hobson, S. (2018). Narrative bytes: Data-driven content production in esports. *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video,* 29-41. https://doi.org/10.1145/3210825.3210833

Chen, J., Su, Y., & Tang, Y. (2016). Research and Practice on Real-time Big Data Computing Framework for Mobile Games. *Mobile Communications*, (5),79-84.

Churchill, B. C. B., & Xu, W. (2016). The modem nation: A first study on Twitch. tv social structure and player/game relationships.2016 IEEE international conferences on big data and cloud computing (BDCloud), 223-228. https://doi.org/10.1109/bdcloud-socialcom-sustaincom.2016.43

El-Nasr, M. S., Drachen, A., & Canossa, A. (2016). Game analytics. Springer London Limited.

Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC iView*, 1–12.

Howe, D., Costanzo, M., & Fey, P. (2008). Big data: The future of biocuration. *Nature*, (455), 47-50.

IBM. (2012). What is big data. Retrieved from http://www-01.ibm.com/software/data/bigdata/

Jensen, J. A., & Turner, B. A. (2018). Event history analysis of longitudinal data: a methodological application to sport sponsorship. *Sport in Society,* (21), 376-393. https://doi.org/10.4324/9780429487521-13

JINGDATA. (2018). 2018 Esports Industry Report.

Lan, J., & Wu, Y. C. (2017). Visual analysis of gameplay data. *Journal of Computer Aided Design & Computer Graphics*, (4).

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, (6), 70.

Li, Q., Xu, P., & Qu, H. M. (2015). FPSSeer: visual analysis of game frame rate data. *Proceedings of IEEE Conference on Visual Analytics Science and Technology* (pp. 73-80). Los Alamitos: IEEE Computer Society Press. https://doi.org/10.1109/vast.2015.7347633

Li, W. (2016). Sportradar US executive: how E-sports data is driving the industry. *China Electronic Athletics*, (11),78-79.

Li, X. L., & Gong, H. G. (2015). A survey on big data systems. *Scientia Sinica Informationis*, (1), 1-44.

Ma, L. M., Shi, D., & Gao, Z. Y. (2019). Big data technology and its industry applications: A study based on conceptual framework of railway. *Journal of Beijing Jiaotong University (Social Sciences Edition),* (3), 58-67.

Makarov, I., Savostyanov, D., & Litvyakov, B. (2017). Predicting Winning Team and Probabilistic Ratings in "Dota 2" and "Counter-Strike: Global Offensive" Video Games. https://doi.org/10.1007/978-3-319-73013-4_17

Manyika, J., Chui, M., & Brown, B. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. San Francis-co: Mckinsey Global Institute.

Ontanón, S., Synnaeve, G., & Uriarte, A. (2013). A survey of real-time strategy game AI research and competition in StarCraft. *IEEE Transactions on Computational Intelligence and AI in games,* (4), 293-311. https://doi.org/10.1109/tciaig.2013.2286295

Penguin Intelligence. (2019). The World and China: 2019 E-sports Industry Development Report.

Salvador, G., Julián, L., & Herrera, F. (2000). Data Preprocessing in Data Mining. *Computer Science, 72*.

She, C. Q., Zhang, Z. S., & Liu, P. (2019). Overview of the application of big data analysis technology in new energy vehicle industry: Based on operating big data of new energy vehicle. *Journal of Mechanical Engineering,* (20), 3-16.

Shim, K. J., Hsu, K. W., & Damania, S. (2011). An exploratory study of player and team performance in multiplayer first-person-shooter games. *2011 IEEE Third International Conference on Privacy,* 617-620. https://doi.org/10.1109/passat/socialcom.2011.155

Xu, G. Q. (2015). The development of archives industry under the background of big data. *China Archives,* (6), 70-71.

Yang, C. H., Li, Z. W., & Zhang, T. (2018). An empirical research on commercial sponsorship of E-sports-taking S7 global finals as an example. *Liaoning Sport Science and Technology,* (6), 41-44.

Yu, Y. Y. (2019). *Application of artificial intelligence and big data system in E-sports data processing platform.* Retrieved from https://max.book118.com/html/2019/0707/8055056131002033.shtm.

Zeng, Z. L. (2017). Big Data Analytics: Direction, Methods and Tools. *Information Studies: Theory & Application*, (1), 5-9.

Zhou, L., & Wang, L. L. (2018). Analysis of the source and acquisition mode of E-sports game data. *Electronics World*, (6), 144-144.

Copyrights