

## Research on the Developmental Trend of Data Journalism under the Background and Time of Big Data

Hui Zhi

Wuhan Donghu University, No.1, Tangbeiroad, Jiangxia,  
Wuhan, Hubei, China, 430212

**Abstract**— In this paper, we conduct research on the developmental trend of the data journalism under the current background and the time of big data. Big data is not only a concept, but also a description of a state of society: in the era of the big data, data become important social resources and production data, the news media is no exception. In the time of the data had not been so seriously, the core of the news resources is a reporter on the scene to get first-hand material, is based on the reporter can see, smell, feel the fact description, data is often only a supplementary role. However, in today's era of big data, although the scene is also very important, but based on the various aspects of data mining and analysis and the depth of the formation of information has become more and more important. Our research proposes the novel paradigm for the issues that is meaningful.

**Keywords-** *Big Data, Developmental Trend, Data Journalism, Contemporary Era and Time.*

### Introduction

Today, all walks of life has been more and more to explore the influence of the big data, value and application, in the Internet industry, e-commerce, Internet marketing has been used in such fields as big data. Although the news media will be real data is applied to the normal practice is not much, but in the long run, big data as an important part of information, production mode, the

transmission way of news and information and the general construction of the whole news formats are bound to make a difference. Analysis and explore coping strategies, we can make the news media in the era of big data in the competition to win the initiative [1-3].

According to the literature survey, the functionality of the data could be summarized as the follows. (1) Journalism has a branch in the news called data refers to through the data acquisition, statistics, analysis and create news reports. News and in the era of big data, data needed for the analysis of the data scale is huge, news has been far beyond the traditional data needed for analysis data of the order of the magnitude. (2) In the era of big data, since the huge amounts of data has become an important means of production, so as the main force of the news media, the information transmission industry should fully use the means of production. Strengthen news reports data idea and thinking, to establish a smooth mechanism, to encourage news editorial staff quickly and efficiently massive external data acquisition and processing. (3) In the competitive age of the Internet media, pay more attention to the analysis of user needs and experience, and in accordance with continuously improve communication and communication form is very important, this will bring the media sustainable development power while the huge amounts of data analysis using a core purpose are accurate and user preferences.

Of press and publication of press and publication of basic standards and regulations system of the coordinated development, on the one hand, can effectively promote the news

publishing industry to better adapt to the rapid development of the economic globalization, realize the production of high quality and high efficiency, service and management, on the other hand can effectively promote the development of press and publication industry internationalization process, to gain the initiative in the trade competition. Standard with the press and publication of press and publication laws system in our country is in the context of the planned economic system set up, with the establishment and perfection of the socialist market economic system, as well as the trend of economic globalization intensifying, press and publication of standards and can set up the system of the coordinated development of press and publication laws and regulations system, form the good interactive relationship, for the press and publication industry in the aspects of economy, trade and technology development is very important.

To deal with the mentioned challenges, in this paper, we conduct research on the developmental trend of data journalism under the current background and the time of big data. Big data analysis and application of different from the traditional Internet data analysis, the traditional data analysis has limitations, makes every effort to find "cause-and-effect" relationship, through analysis and big data is the analysis of the entire process, all-round, the whole sample, it can be to objective data reflect the business situation, pay attention to the analysis of "link", and dig the Internet user behavior habits and preferences, obtain the valuable information, to guide the enterprise get the target user, and targeted to optimize the products and services which is the main value in big data.

### The 3 dimensions of data journalism

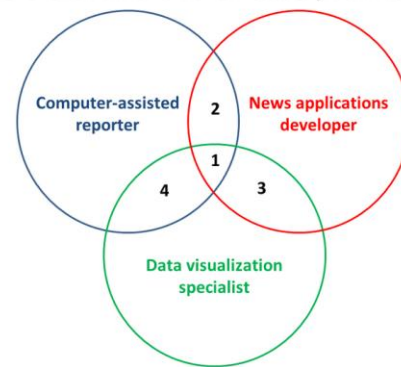


Figure 1. The Primary Components of the Data Journalism

### The Proposed Methodology

#### The Concepts of the Data Mining Techniques.

Data is the source of knowledge. But, having a large amount of data and have many useful knowledge completely are two different things. In the past few years, from the database found that knowledge in this area is developing very quickly. Broad market and the research interest prompted the rapid development of this field. The progress of the computer technology and data collection technology can make people from a wider range and speed of a few years ago unthinkable to collect and store information [4-5].

Data mining is a combination of a variety of the professional and technical application, it used the methods and techniques include: computer technology, statistical analysis method, all kinds of the algorithms, cooperation, agents, and the distributed target management technology, etc., therefore, the data mining method and technology development, development and application is the main field of the data mining research. According to the literature review, the state-of-the-art mining algorithm can be organized as the follows. (1) Data preparation stage. Knowledge discovery process of object data is raw data that is not suitable for knowledge in the data mining, need to carry on the corresponding processing, such as the choice of data, such as purification. Make its generation

process data, after the transformation, including the interaction between the values of the discrete data and continuous data conversion, data worth packet classification, such as the calculation of the combination between data items, ready for the back of the data mining the correct data. (2) Data mining stage. Data mining stage is the most critical step in knowledge discovery is a technical difficulty. This stage is also according to the goal of knowledge discovery, choose corresponding algorithm, analyze data, search or produce the corresponding model and specific data sets. (3) Explain the evaluation phase. After data dig according to the model, some may be inaccurate or untrue, so need to evaluate the income pattern,

determine the effective and useful patterns, and knowledge is obtained. Evaluation can be based on user experience for many years that can also use the data directly verify its accuracy.

The ultimate goal of data mining is to discover people imperceptible, implicit model. In general, the most convenient to understand these patterns is a statistical model. Second is on the outside of the data detection, for large-scale data set of pattern recognition, classification or clustering. Finally from the theoretical and the computational solution in most database management systems that exist in the multidimensional space and a lot of the problem of data processing.



Figure 2. The Keywords of the Data Mining and Corresponding Principles

**The Journalism Development Path.** The people's livelihood news in people-oriented thoughts as the basis, in the civilian angle of view and narrative of humanities concern and ordinary life, survival, life and livelihood of the people form a kind of content such as news. Its generation and development conform to the trend of the times and more in line with the new requirements of socialist harmonious society, to a certain extent, enriched the people's spiritual and cultural life in today's development and transformation of the social background, the news development breakthrough in how to improve the people's livelihood that is worthy of attention and research [6].

The emergence of new media makes way people access to information and consumption pattern, a major change in the appearance of

mobile phones, the Internet makes people to be able to understand the news and information anytime and anywhere, traditional journalism and communication system also becomes a bottom-up dissemination system dominated by the public. Under the new media basic environment of TV news development faces many obstacles and the challenges could be summarized as the follows. (1) TV news interaction is poor with update speed slow, news quality is lower than the quality of the new media. The negative impact of television news more direct, as the first media and society of the negative impact of television news is more direct, once the information appears error the scope and degree of social negative impact will be far more than other media. (2) A variety of media are stripping TV news audience and the television

news ratings down. Basically can convey various personalized information to each person's hand, this makes the new media has the great scope of the audience right from the start, and the new media characteristics: interactive and across time and space makes the news information no longer suffered from a single publisher control with constraints as the spread of news information become more freedom.

**The Big Data Characteristics.** Big data itself is a relatively abstract concept, we understand from the surface of large data, but as the application is more and more extensive, deeper and deeper research on big data, can be found that big data is not only large in quantity scale, but also includes the data structure is quite complicated, the connection degree between data and the data was very high. So Gartner gives such the definition: big data is need to deal with the new model that can have better decision-making, insight found mass force and the process optimization ability, as high rate of growth and diversification of information assets [7-8].

The general applications of the big data applications could be generally summarized as the follows. (1) Streaming data tuple usually with a label or the rest of the time sequence attribute. Therefore, the same streaming data is often by

sequential processing. The arrival of data sequence is unpredictable, however, due to the dynamic change of time and environment, there is no guarantee that the replay before the data flow and data flow in the order of the data element consistency. This leads to data do not match the physical sequence and logic sequence. Moreover, the data source is not controlled by the receiving system and the generation of data is real-time and unpredictable. (2) Data streams often contain the fault element, spam, etc. So the streaming data processing system must have good fault tolerance and heterogeneous data analysis ability that can complete dynamic cleaning, format of data processing, etc. (3) Using batch data mining the appropriate model, as it is concluded that the specific meaning, wise decisions, and ultimately to make effective measures to achieve the business goal is to the big data batch priority. Big data after the first batch processing system is suitable for the storage of computing, real-time demand is not high, at the same time, accuracy and completeness of the data is more important. In the following figure three, we demonstrate the principles.

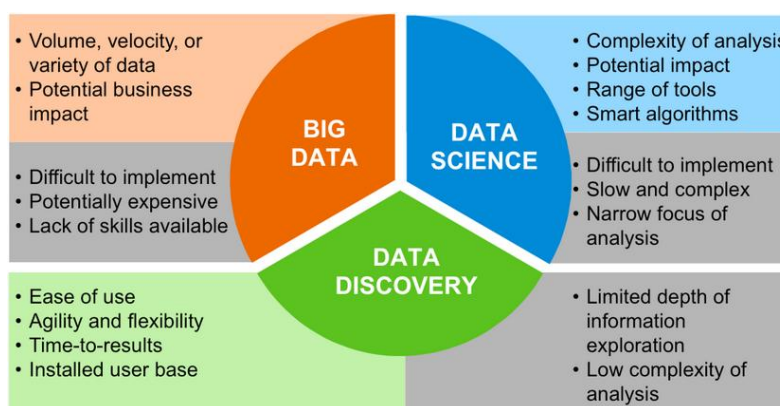


Figure 3. The General Features of the Big Data Scenarios

**The Developmental Trend of Data Journalism.** Big data is a huge commercial value, but also changed the dissemination means and way of life. Under this background, the media

market needs more professional, to fill the industry focus the demand of the market or blank. Relative mass, focus is one of the few, but it is most accurate newspaper industry, the core of the

target user stick to the target user, also makes the transformation and upgrading of the road to be "not lose a reader".

Statistics tell us that the sample selection, the more the more results we will get close to the truth. As the data source, scope and efficiency of data acquisition are related to the data are accurate and precise for newspaper industry, on the one hand, can set up their fixed information site, to collect the required information according to certain time period and on the other hand, with the help of the government, organization of the existing data collection points and information. Data is boring, but the extensive and profound, infinite mystery. Comb through the data contrast, traced back, can find a lot of can't see the surface of the "invisible treasure". Data analysis is to go deep into the data, to reveal the value behind the data. Newspaper industry to play to their talents and professional advantages, carding industry data collected, the understanding of the important data are accurate and in place, reveals the development of market, the trend of the industry and industry law of development, give enlightenment on the industry and at the same time, also can invite the industry experts and scholars to participate in to increase the data analysis of the breadth and openness.

To sum up, press and publication standards for the construction of the system and the system of the press and publication laws and the basic regulations is the press and publication industry healthy and sustainable development of the important basic work, press and publication of standards and the press and publishing regulations system with the goal of consistency makes building good news publishing standard and interactive relationship of press and publication laws and regulations system possible. But the press and publishing standard with the press and publishing regulations system to realize the coordinated development that also depends on its overall environment conditions mature, and the formation of the basic conditions and

perfect also calls for the joint efforts of all parties.

## Conclusion

In this paper, we conduct research on the developmental trend of the data journalism under the current background and the time of big data. Since the "big data" is domestic news spread among hot topic. In contrast, is about the "news" data research is very lack. The latter is a response to the current global journalism development change in the emerging field of the big data era, and is regarded as the future trend of the development of journalism. This article attempts through to domestic and international news of the data for the interpretation of literature, in combination with the mainstream news media practice data of classic case analysis, and interpretation of the concept of "news" data connotation and basic characteristics, sorting out data and accurate news, computer aided news, data visualization, the relationship between the data show the latest achievements of news practice. In the future, more of the corresponding research will be conducted.

## References

- [1] Coddington, Mark. "Clarifying Journalism's Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting." *Digital Journalism* 3.3 (2015): 331-348.
- [2] Lin, Chao Chen. "Where is the limit of big data? A case study of journalism practices pertaining to datasets of e-Government in Taiwan." (2014).
- [3] Gynnild, Astrid. "Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets." *Journalism* 15.6 (2014): 713-730.

- [4] Verbeke, Mathias, et al. "Data mining for computational journalism." status: published (2013).
- [5] Boyd, Danah, and Kate Crawford. "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon." *Information, communication & society* 15.5 (2012): 662-679.
- [6] Picard, Robert G. "Deficient tutelage: Challenges of contemporary journalism education." *New Directions in Journalism Education* (2014): 4.
- [7] Neville, Tina M., and Deborah B. Henry. "Evaluating Scholarly Book Publishers—A Case Study in the Field of Journalism." *The Journal of Academic Librarianship* 40.3 (2014): 379-387.
- [8] Blom, Robin, Lucinda D. Davenport, and Brian J. Bowe. "Reputation Cycles The Value of Accreditation for Undergraduate Journalism Programs." *Journalism & Mass Communication Educator* 67.4 (2012): 392-406.