# Application and research trends of big data

**Xi Yan***

Northwood High School, USA

*Corresponding author: jamesyanusa@gmail.com

**Keywords:** Big data; Main technology, Application, Research trends

**Abstract:** With the continuous development of big data technology in the word, its application field has become more and more extensive, which has a decisive significance for the development of social economy. At the same time, big data is still in the stage of continuous improvement, which should be discuss with the needs of research trends. This paper introduced the history and main technology of big data. Also come with the introduction of the application of big data in the healthcare, media and entertainment, transportation fields. Through the research trends of big data, the relevant researches can be considered based on the characteristic of Big data and NLP, and real-time analysis can help to obtain valuable decision-making information.

## 1. Introduction

The term “big data” refers to data that is so large, fast or complex, which is considered as difficult or impossible to process using traditional methods. The action of accessing and storing large amounts of information for analytics has been around a long time, but the concept of big data speeded out in the early 2000s when industry analyst Doug Laney articulated the now-mainstream definition of big data as the three V’s: Volume, Variety, and Velocity. Volume, the main characteristic, is the size of the data that's been collected. The organizations collect data from a variety of sources, including business transactions, smart (IoT) devices, industrial equipment, videos, social media and more.

In the past, storing it would have been a problem, but cheaper storage on platforms like data lakes and Hadoop have eased the burden. Next is Variety, which indeed is what makes big data really, really big. Data scientists and analysts aren’t just limited to collecting data from just one source, but many. And this data can be broken down into three distinct types – structured, semi-structured, and unstructured. Huge volumes of data are pouring in from a variety of different sources, and they are doing so at great speed, giving us our third characteristic of big data: Velocity. The high velocity of data means that there will be more data available on any given day than the day before, and it also means that the velocity of data analysis needs to be just as high. Data professionals today do not gather data over time and then carry out a single analysis at the end of the week, month, or year. In other words, the analysis is live, and the faster the data can be collected and processed, the more valuable it is in both the long and short term.

## 2. The main technology of big data

The main technology of big data is made up of four parts: Data Storage, Data Mining, Data Analytics, and Data Visualization.

### 2.1 Data Storage

Data Storage is the retention of information using technology specifically developed to keep that data and have it as accessible as necessary. There are already some companies, including Microsoft, IBM, Hortonworks, Intel, using data storage such as Hadoop and MongoDB, to help them store and analyse the data they have collected.

  DOI: 10.25236/cseem.2020.174

### 2.2 Data Mining

Next is Data Mining, which is the process of finding anomalies, patterns, and correlations with large data sets to predict outcomes. With a variety of techniques available, data mining allows organizations to continually analyze data and automate both routine and critical decisions without the delay of human judgment. What is more, planning is a critical process within every organization. Data mining facilitates planning and provides managers with reliable forecasts based on past trends and current conditions. The popular ones include RapidMiner, KNIME, Weka, Orange, and more.

### 2.3 Data Analytics

Data Analytics is just as the name suggests, analyzing the data in large amounts, and some companies, like Netflix, Twitter, Yahoo, Spotify, require such technology to assist them for further improvements. Just like Data Mining, companies like KNIME and RapidMiner have also developed through this field.

### 2.4 Data Visualization

Last we have Data Visualization, a graphical representation of information and data, which is more literal than data itself. With this technology, not only professionals, but also normal people can understand and utilize the information gathered. It is usually depicted through charts, graphs, and maps, providing an accessible way to present trends, outliers, and patterns in data. As the “age of Big Data” kicks into high-gear, visualization is an increasingly key tool to make sense of the trillions of rows of data generated every day. Data visualization helps to tell stories by curating data into a form easier to understand, highlighting the trends and outliers. A good visualization tells a story, removing the noise from data and highlighting the useful information. Some popular ones are Toucan Toco, Tableau, Sisense, and more.

## 3. Applications of big data:

In this era where every aspect of our day-to-day life is gadget oriented, there is a huge volume of data that has been emanating from various digital sources.

### 3.1 Big data in Healthcare

Healthcare industry is bound to generate a huge amount of data. Big Data has opened a new path in healthcare. With considerable benefits and opportunities, it has attracted the momentous attention of all the stakeholders in the healthcare industry.

It is also widely used in healthcare, as healthcare facilities and hospitals prefer using large data storage to store and manage patient data. The idea that the healthcare providers can access the data anytime from a secure native data center appeals to most, making data storage a very popular choice.

With a huge space for digital information, Media and Entertainment has moved forward to putting their data such as videos into data storage, and providing it to the society to create more influences. Even further, the data storage has enhanced the information flow in the industry by enabling stakeholders in the transport industry to store more fleet data.

### 3.2 Big data in Media and entertainment

Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs. Today the traditional media content development method has been replaced with myriad media services like pay per view, live streaming and much more. During the process of content delivery, media distributors and providers collect a vast amount of user data. Big data in the media and entertainment industry can help in seeking an in-depth understanding of consumers’ behavior and preferences. Big data has helped many companies in fusing and understanding all the user data from multiple sources, including social media. Also, with the help of analytics solutions, it is now possible

to uncover reasons that drive customers to subscribe and unsubscribe from any particular channel or platform.

### 3.3 Big data in Transportation

Health data analytics involves the extrapolation of actionable insights from sets of patient data, collected from sets of patient data, collected from records in the past. The more precisely transportation companies know how their routes are being used, the more cost efficiently they can deploy staff and trains. Big data analytics help the public transportation sector to predict passenger volumes as precisely as possible. This can be used to plan operating procedures more efficiently and affordably. Moreover, transportation companies can increase customer retention by doing away with short trains during times of peak passenger volume and increasing the frequency of trains in a service-oriented way, for example.

## 4. Research trends of big data

### 4.1 Big data and NLP

Natural Language Processing is a based on deep learning that enables computers to acquire meaning from inputs given by users. In the context of bots, it assesses the intent of the input from the users and then creates responses based on contextual analysis similar to a human being.

The modality of textual data has been somewhat under-represented in big data research thus far. This is despite the fact that large amounts of data are stored in unstructured textual format. This topic has emerged in several areas in parallel in recent years: information retrieval and search engines, text mining, machine learning, web-derived corpus linguistics, digital libraries, high performance and parallel computing. Common of all these areas is some or all of the main parts of the NLP pipeline: collection, cleaning, annotation, indexing, storage, retrieval and analysis of voluminous quantities of naturally occurring language data from the web or large-scale national and international digitization initiatives.

In this context, numerous issues should be considered including those linked to the four Vs of big data: (1) Volume: is having more data for training and testing NLP techniques always better? (2) Variety: are all types of data available on a sufficiently large scale? (3) Velocity: how are parallel methods best applied to carry out NLP on a large scale? (4) Variability: how does inconsistent data impact on the accuracy of NLP techniques?

### 4.2 Big data and Real time analysis

Real-time big data analytics means that the data is processed as it arrives and produces a result without exceeding a time period based on the user’s decision-making, or it can be seen as an instant analytical system triggered by an action.

Although in general organizations value managing data in real time, not all the companies go for real-time big data analytics. The reasons could vary with everyone: the lack of technical support, insufficient funds, fear of the associated challenges, or failure of overall team management. However, the value of real time analysis can be depicted with the people who implemented real time. Real time analysis can provide better performance by analyzing big data in real time. It takes in the big data as people perform actions on certain tasks and processes the data in a short amount of time behind. After processing, real-time data finds its way to a real-time dashboard or turns into either a notification or a system’s action. The result will then go back to the action performers and return what they seek for. If thoroughly planned and properly implemented, real-time big data analytics definitely can become a great advantage. The result will then go back to the action performers and return what they seek for. If thoroughly planned and properly implemented, real-time big data analytics definitely can become a great advantage. Taking into account how different the interpretations of real-time can be, it’s important to have a clear understanding of the society’s requirements to the analytical system.

As mentioned earlier, the current related research in various aspects cannot perfectly solve the core problems of universality of technical support, low-cost real-time big data analytics and the balance of

society's requirement with analytical system. So there are still a lot of challenging tasks waiting for us to study.

## 5. Conclusion

Overall, big data has been effectively used within a variety of fields, improving the quality of life for people with big data analysis and application of big data. People are able to track and monitor their every action and thus gain a better understanding of their own lives. Even though there are consequences like the privacy of people are leaked, like how the society is becoming transparent for people, and controlled by the industries with the data, but the benefits are stronger than the disadvantages that big data brings.

## References


[1] Doug Laney. “3-D Data Management: Controlling Data Volume, Velocity and Variety.” *META Group*.01(2001):949

[2] A, Aisling O’Driscoll, J. D. B, and R. D. S. B. "'Big data', Hadoop and cloud computing in genomics." *Journal of Biomedical Informatics* 46. 5(2013):774-781.

[3] Fu, Zhe. "An Overview of Research Applications on Big Data." *Information and Computers:Theory Edition* 9(2014):91-92.

[4] Purcell, Bernice. "The emergence of 'big data' technology and analytics." *Journal of Technology Research* (2013).

[5] C. Coronel, S. Morris, P. Rob. Database Systems: Implementation, and Management (10th Ed.) [M]. Boston: Cengage Learning, 2013.

[6] Naik, Amrita, and L. Samant. "Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime." *Procedia Computer Science* 85(2016):662-668.

[7] Hartama, Dedy. "Analisa Visualisasi Data Akademik Menggunakan Tableau Big Data." *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)* 3(2018):46.

[8] Wyber, Rosemary, Vaillancourt, Samuel, Perry, William, Mannava, Priya, Folaranmi, Temitope, Celi, Leo Anthony. “Big data in global health: improving health in low- and middle-income countries.” *Bulletin of the World Health Organization*93(2015):203-208

[9] Smith Michael D., Telang Rahul. *Streaming, Sharing, Stealing:Big Data and the Future of Entertainment*. Cambridge: The MIT Press:2016.

[10] Qi Shi, Mohamed Abdel-Aty. “Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. ”*Transportation Research Part C* 58(2015):380-394

[11] Wang, Y. K. "Azure cloud platform for real-time big data analysis of Twitter tweet keywords." *Computer Programming Skills and Maintenance* 000.012 (2015):68-72.

[12] Jin, Xiaolong, et al. "Significance and Challenges of Big Data Research." *Big Data Research* 2.2(2015).

[13] Fang, Hua, et al. "A survey of big data research." *IEEE Network* 29.5(2015):6.

[14] Schroeder, and R. "Big Data and the brave new world of social media research." *Big Data & Society* 1.2(2014):1-11.

[15] Yu, and Shui. "Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data." *IEEE* Access 4(2017):2751-2763.