ATLANTIS PRESS

Data Mining in the Big Data Era

Weigao Sun

California State University, College of Engineering and Computer Science, Fullerton, 800 N State College Blvd, Fullerton, CA 92831
swg9527@gmail.com

ABSTRACT

While entering into the era of big data, seeking the secrets hidden in the data has become the main target in every single industry. This paper states the definition of data mining in the era of big data and interprets data mining methods in current popular industries such as sports, music, and health care. This paper also analyzes the future of data mining and some potential applications. The paper finds that in the era of big data, data mining, as a powerful method to seek the secret from the data, has had a significant impact on the fields of sports, health care, and the music industry. Data mining also has a brighter future when considering information security, business management, financial analysis, the education industry, etc. The rise of the big data era provides data mining with a suitable and stable environment, which makes jobs such as Data Analyst so popular.

Keywords: Data mining, big data era, sports, music, health care

1. INTRODUCTION

People's lives are totally changed while computer technology is developing rapidly. Human beings have powerful hardware devices. Electronic products such as mobile phones, computers, tablets are everywhere. With these perfect hardware being used, uncountable users of different age ranges and purposes have been involved. As the most important information transmission, the Internet has played a significant role in providing people with applications and services. The data is transmitted through Internet websites and software. When the information is used and stored every second, a huge amount of data is produced. Data mining, machine learning, and social networking are the main topics related to big data. Data mining is the most important method to find hidden information, which is used in almost every field nowadays. As no relevant studies refer to this field in such a detailed way, this paper will provide an overview of data mining in the big data era specifically in the Sports, Music industry, and Health care, also give a brief look at the future of data mining. It hopes to provide meaningful insights for further study in this field.

2. BIG DATA ERA

With advanced science and technology, whether it is a mobile phone, or a computer can be used by every people. The high-speed network provides extremely fast information exchange speed, and network base stations can be seen everywhere. Therefore, the communication between people is getting closer and human life is getting more and more convenient. Google is the most common search engine which is used every second. It is obvious that covid-19 doesn't stop people from buying groceries or getting an education. Amazon and eBay can meet most of the shopping needs. Online teaching also allows face-to-face education to continue such as Zoom. Every possible action can produce information. With the increasing variety of information, human society has completely entered into the era of big data. When data penetrates into every field today, it becomes an important production factor, which accelerates the progress of people's lives to a certain extent, and also promotes the timeliness of information exchange in different fields.
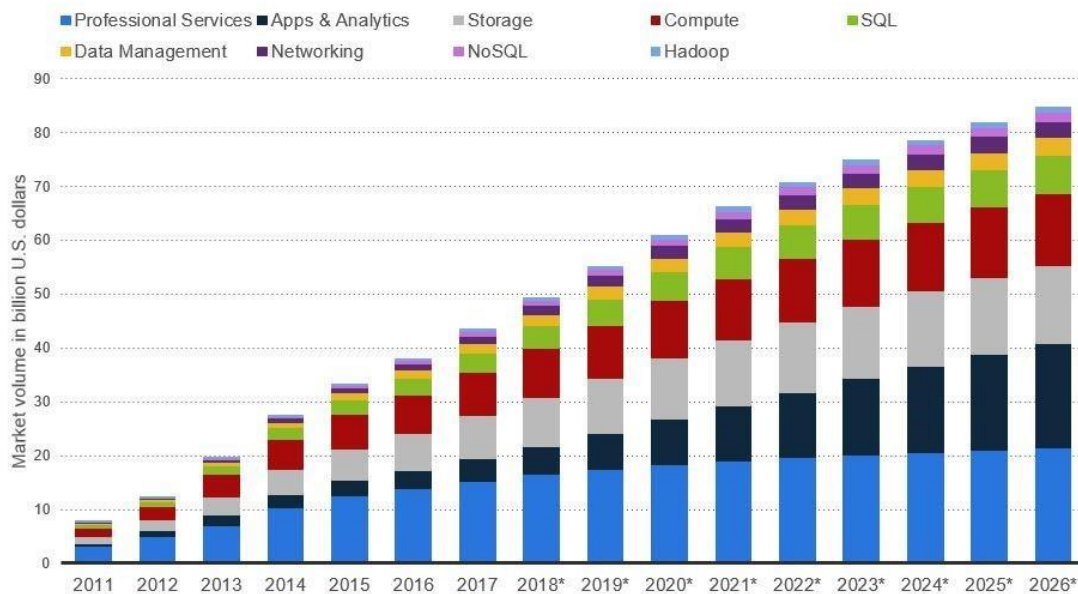
Figure 1 Market value of big data applications [1]

3. DATA MINING

3.1. Definition

Data mining is the product of the rapid growth of data. The use of computer data analysis can be traced back to the 1960s. The term "data mining" was first used in the 1990s. It is a powerful technology for extracting specific data or mining useful information from a huge amount of data. It is called Knowledge Discovery in Databases, which refers to KDD. It is an interdisciplinary branch of computer science that includes artificial intelligence, machine learning, statistics, and the cross-application of databases.

3.2. Data Mining Procedure

Data mining has raw analysis steps, management aspects, data preprocessing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating [2].
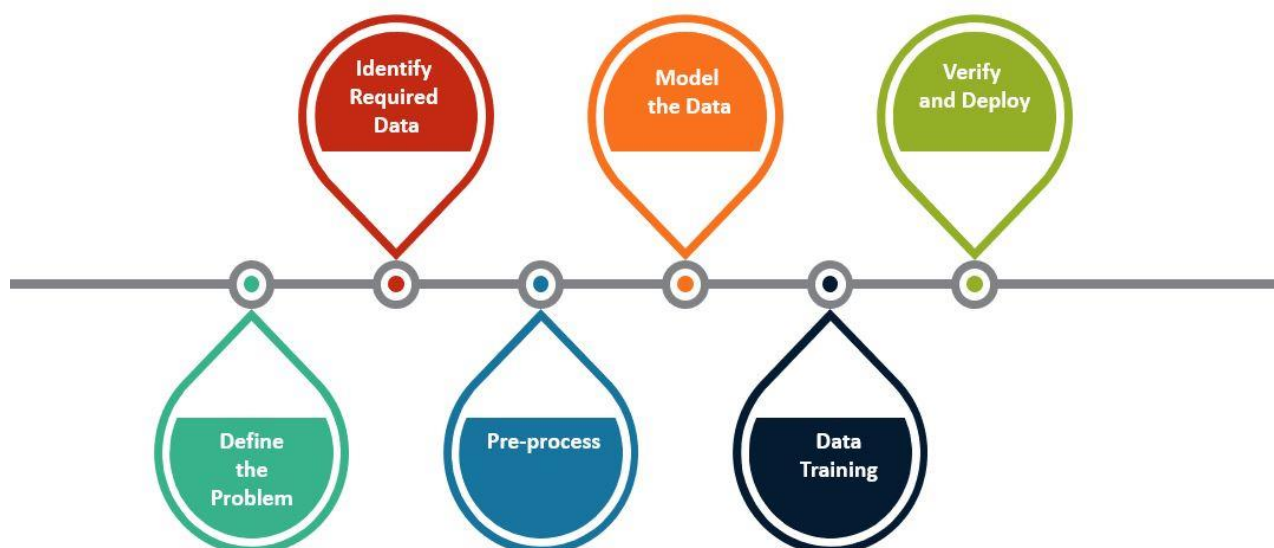

Figure 2 An overview of data mining process [3]

There are usually 3-10 steps when applying data mining, depending on the topic and the situation of the problem, and how they work together with the process. But in general, the process is similar to what is being discussed. Defining the problem is usually the very first step of data mining, which is especially important in order to ensure the right direction. This may not have much connection with the data mining technology itself.

The users will need to identify the business goals, such as what kind of problem to solve or what the final destination is. The users will also need to be certain of the goals, such as how those selected business goals translate into specific data mining project goals [3]. The answers to these questions will help to get the target data. After the first step, find the required data and understand the data. This will give users an idea of how much data preparation and preprocessing may be required. Here comes the hard work of data mining. After finding the required data, go through the process of cleansing and preprocessing. Sometimes, there will be a need for the integration of multiple data sources to prepare the final data. Some of these data sources may even be external to complete some attributes of the data [3]. At this step, users will need to build a mining model to find the result. A mining model is created by applying an algorithm to data, but it is more than an algorithm or a metadata container: it is a set of data, statistics, and patterns that can be applied to new data to generate predictions and make inferences about relationships [4]. Users will need some tools to improve productivity, such as the Data Mining Wizard or SQL Server Management Studio. These tools can help to build the required model and evaluate the results. After evaluating the results from the last step, it tests the model on different sample data sets. The key is to keep iterating the data until the users are satisfied with the consistency of the results. Now verify the final model and plan for deployment. Think about the visualizations needed [3]. It is funny that data mining is also a process for telling a story. Reporting the findings and operationalizing the process are the last jobs.

4. APPLICATIONS OF DATA MINING

In the era of big data, data mining can be widely used in different fields, such as the financial field, e-commerce, Internet-related products, education, and medicine. By dealing with the data generated by users and target users, the needs can be fulfilled. The direction of future development research can be obtained through these needs, in order to improve the progress in different areas. Three different areas (sports, health care, and music) will be discussed here.

4.1 Sports

The NBA, also known as the National Basketball Association, is the world's most professional basketball league, with a massive fan base and commercial value. The performance of the players and the outcome of the games become crucial factors. Every NBA game will generate a large amount of data. If this data can be effectively used, it can give full play to its potential value [5]. For example, by using data mining, it is possible to predict the outcome of each NBA game. First, record the data from the previous season to get the statistics for the game. The resulting data is then preprocessed to remove

less meaningful data, such as the game when the best player is off because of injury. Analyze the game data of each team and get the characteristic expression that represents the state of each team. Finally, by machine learning the relationship between each game and the team, it is possible to predict future game results. Game results can help team management make important decisions to keep the team's winning percentage and revenue. By using data mining, special training can also be carried out according to the data of the players, and the health of the players can be guaranteed. Each team has a professional data analyst to do this job. These data analysts work with coaches and players to maximize the talents of the players and, of course, like in "Money Ball", identify undervalued players [6]. Dealing with the players' offensive heat map on the court, the hit rate from different angles and ranges, the way of the game-winning shot, the success rate of rebound positions, and analyzing the places that need to focus on training. This can help players strengthen or avoid their weaknesses to make training more efficient. NBA players also use wearable technologies to track their health, avoid injury, and track their fatigue levels [6]. In conclusion, data mining plays an important role in basketball. Data mining can also be applied to most other sports.

4.2 Health Care

The health care system is being modernized with the popularization of medical information technology. Nowadays, the health care system records data on patient files, medical equipment, and medicines into the database system and integrates them with other data such as the situation of patients and the time of consultation. By using data mining technology, it is possible to conduct comprehensive research on the information in the database through the data of patients' medical histories, clinical symptoms, and medical histories of similar patients [8]. Data mining applications can greatly benefit all parties involved in the healthcare industry. For example, data mining can help healthcare insurers detect fraud and abuse, healthcare organizations make customer relationship management decisions, physicians identify effective treatments and best practices, and patients receive better and more affordable healthcare services [7]. The way doctors give advice is also changing. The traditional research model is to use one's own experience to diagnose. In the era of big data, all the information related to human diseases can be inquired about. It will be less persuasive when doctors only rely on their own experience. In other words, data mining can not only provide meaningful diagnostic rules, improve the efficiency of doctors' diagnosis and treatment, and alleviate the contradiction between doctors and patients, but can also effectively warn patients of diseases and recommend treatment measures [8].

4.3 Music Industry

Data analysis in the music industry has been a useful method for a while. When considering the sales of the album, the Billboard rankings, and the total views on YouTube, all the numbers should have meaning, more or less. But with more music consumption becoming digital, there is an ever-increasing amount of information available. Increasingly, numbers are being used to discover potential breakout songs, help advance artists' careers, and even determine the sound of the music itself [9]. It is possible for newly released music or music from years ago to become popular or gain popularity and profit through data mining. By analyzing the purchase data of major music platforms and the number of people who have listened and purchased, the popularity of the song can be predicted, and appropriate advertising can be carried out to maximize benefits. Data mining can be quite powerful, especially in today's pop music, which places more emphasis on appeal and the ability to create value. Alibaba Music has nearly 1 million analysis records and historical user behavior data. Later, the number of artists or songs to be played in the next step will be reduced, and the dark horse will be excluded from the mainstream data. Multiple music platforms will primarily control the pop music trend [10].At the same time, data mining can sum up the music styles and chord trends to get the direction of creating popular music. This is also how new-age musicians can take advantage in order to create the music people like.

5. THE FUTURE OF DATA MINING AND POTENTIAL APPLICATIONS

Considering information security, data mining will become an opportunity for its development. As the ubiquity of data reduces the security of its own information, various technologies and products that serve information security have become the primary issue in the field of information security. Therefore, how to secure the safety of the data industry chain is of great significance to the development of information security. Scientists have proposed various emerging technologies from existing levels. For example, from the perspective of data processing, there is a distributed processing method called MapReduce, and the more well-known application tools are Hadoop and DISCO. In addition to how to extract valuable information and deal with the underlying structured technical support, data mining algorithms and machine learning algorithms are all indispensable. In terms of business management and industrial services, data mining will become a turning point for many industries, such as enterprises and service organizations. Data mining technology will not only bring economic benefits to enterprise management but also change the management model. In each field, the data analyst and data-mining technology will become more popular.

Financial Analysis: The banking, financial, and insurance industries rely on high-quality and reliable data. User data can be used for a variety of purposes in the loan market, such as predicting loan payments and determining credit ratings. Data mining methods make such tasks more manageable and easier to access [11].

With the rapid growth of computers and the internet, the security of computer systems and data is the most concerning problem. By using data mining algorithms that have better efficiency and a lower mistake rate, it is not hard to create a new model for an Intrusion Detection System.

The Retail Industry: After the pandemic, e-commerce takes over the retail market. Data mining can help retailers find the best-selling strategies while it is convenient to get the data on sales, views, returns, and customer service.

Education: As the demand for elite education goes up worldwide, different institutions are looking for innovative solutions to cater to the rising needs [11]. With data mining, it is possible to predict the student's and their parents' behaviors.

6. CONCLUSION

Through the analysis above, the paper finds that in the era of big data, data mining, as a powerful method to seek the secret from the data, has a significant impact on the fields of sports, health care, and the music industry. Data mining also has a brighter future when considering information security, business management, financial analysis, the education industry, etc. The rise of the big data era provides data mining with a suitable and stable environment, which makes jobs such as Data Analyst so popular. Artificial intelligence technology can also bring data mining more possibilities, not only in efficiency but also in accuracy. Besides, mastering the secrets in the data is the key to mastering success. Since this is an overview of data mining technology in the big data era, no detailed empirical research conducted here. Thus, there are many other related topics and specific knowledge that can be provided in future research.

REFERENCES

[1] Columbus, Louis. "10 Charts That Will Change Your Perspective Of Big Data's Growth." Forbes, 23 May 2018, www.forbes.com/sites/louiscolumbus/2018/05/23/10-charts-that-will-change-your-perspective-of-big-datas-growth/?sh=19c8f17d2926.

[2] CosmoLearning. (n.d.). Topics: Data mining. CosmoLearning. from https://cosmolearning.org/topics/data-mining/

[3] Dontha, Ramesh. "Data Mining Steps – Digital Transformation for Professionals." DigitalPro, digitaltransformationpro.com/data-mining-steps. Accessed 14 May 2021.

[4] Minewiskan, M. Mining models (Analysis Services - data mining). 2022. Microsoft Docs. from https://docs.microsoft.com/en-us/analysis-services/data-mining/mining-models-analysis-services-data-mining?view=asallproducts-allversions

[5] Jiang, X., & Feng, B. Data Mining in NBA. 2017, from https://bitdm.github.io/2017/projects/p09/proposal/

[6] Randerson112358. 2019, May 22. How the NBA uses Data & Analytics. Medium. Retrieved February 27, 2022, from https://randerson112358.medium.com/how-the-nba-uses-data-analytics-6eac3c43a096

[7] Data Mining research in the internet filed--Computer Knowledge and Technology, No.39 2019. (n.d.) https://www.cnki.com.cn/Article/CJFDTotal-DNZS201936005.htm

[8] Koh, H. C., & Tan, G. (n.d.). Data Mining Applications in Healthcare. Original Contributions.https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.92.3184&rep=rep1&type=pdf

[9] Setaro, S. How data is making hits and changing the music industry. 2021, October 4.Complex. from https://www.complex.com/music/2019/09/data-changing-music-industry

[10] Gorbunova, I. B., & Kiseleva, Y. N. Music computer in teaching the "Listening To Music" course. Propósitos y Representaciones, 2020, 8(2). https://doi.org/10.20511/pyr2020.v8n2.522

[11] Sharma, Rohit. "12 Most Useful Data Mining Applications of 2021." UpGrad Blog, 3 Apr. 2021, www.upgrad.com/blog/12-most-useful-data-mining-applications-of-2020.