



# An Overview and Prospects Analysis of Data Mining Technology

Pengpeng Sun

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China

Email: 2732944077@qq.com

**How to cite this paper:** Sun, P.P. (2024) An Overview and Prospects Analysis of Data Mining Technology. *Open Access Library Journal*, 11: e11949.  
<https://doi.org/10.4236/oalib.1111949>

**Received:** July 15, 2024

**Accepted:** August 13, 2024

**Published:** August 16, 2024

Copyright © 2024 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Data mining is the process of extracting useful information and knowledge from mass data. Through statistics, machine learning, pattern recognition and other technologies, data is analyzed and processed to discover potential patterns and laws. This paper provides an in-depth overview of the basic concepts and main technologies of data mining, including data association, data classification, and clustering. The application of data mining in the fields of Internet, finance, medical treatment and environmental meteorology is discussed in detail. This paper also examines the talent requirement and development status of data mining technology, and points out the current challenges, such as data privacy protection, data quality management, model interpretability and usability. This paper aims to help readers gain a comprehensive understanding of the significance of data mining technology and its extensive application prospects.

## Subject Areas

Big Data Search and Mining

## Keywords

Data Mining, Data Analysis, Data Processing

## 1. Fundamental Principles

### 1.1. Definition of Data Mining

Data mining was first proposed at the 11<sup>th</sup> International Joint Conference on Artificial Intelligence in 1989. The concept of data mining originated from machine learning, database systems, pattern recognition and statistics.

The emergence of very large-scale databases makes it necessary to develop automating data collection methods to cope with the large amount of data. The

rapid advances in computer technology, including faster and stronger computability, as well as the emergence and development of parallel architecture, lay a solid foundation for data mining technology. In addition, the realization of high speed data access and the deepening application of statistical methods in the field of data processing further promote the research of data mining technology. The birth of data mining technology is the result of the demand for information processing in the era of big data. The era of big data generates a large amount of data that can be widely used, and there is an urgent need for a technology to transform this data into useful information and knowledge.

Data mining technology is a technology that explores the potential of data by classifying and grouping related data and discovering the patterns and relationships between data. It includes three basic parts: data preparation, exploring data patterns, and data expression. The prerequisite work before launching data mining is to set up the mining model and mining engine so that it can achieve data classification, grouping and patterns discovering as expected. The focus of data mining is on data preprocessing, which is the foundation of the entire data mining application, and its quality directly affects the final result [1].

Data mining can be defined from both technical and commercial perspectives. From a technical perspective, the process of extracting obvious or potentially useful information and knowledge from massive amounts of data is called data mining. From a commercial perspective, data mining is a business information processing technology.

## 1.2. Process of Data Mining

The specific process of data mining technology is mainly divided into five steps:

- 1) Establish the goal of data mining: Firstly, determine the overall goal of the mining task, conduct a preliminary assessment based on the goal, determine the required data type, resource costs, corresponding risks, etc., and finally choose a specific implementation plan based on the evaluation results.

- 2) Data preparation: The first step is data selection. According to the goal determined in the first step, select the source of data mining, and then preprocess the data. This process includes handling missing values, ensuring data consistency, data conversion and data optimization, etc.

- 3) Mathematical modeling: Select a suitable model according to the specific situation. Evaluate the model after building it. Cross-validation can be employed to evaluate the results and optimization algorithms, such as the gradient descent method, can be used to optimize the parameters. Then select the most appropriate model based on the evaluation results.

- 4) Result evaluation: Run preliminary results with the model, and evaluate whether the model can solve the actual problem according to the specific task requirements. If it can solve the problem, choose the model for business deployment. Otherwise, re-optimize the model or choose to rebuild another model.

- 5) Model application: Deploy appropriate models for business purposes.

---

## 2. Data Mining Related Technologies

### 2.1. Data Association Technology

#### 2.1.1. The Concept of Data Association Technology

Data association technology is mainly used to dig out the correlation between valuable data items from large-scale data.

Data association technology mainly includes association rule mining, sequence pattern mining, correlation analysis and other technologies. The first three technologies will be introduced in detail below.

#### 2.1.2. Association Rule Mining

Association analysis is a method of exploring relationships between data items and can be used to explore frequent patterns, association rules, or dependencies. Association analysis can be used to identify sets of items that occur simultaneously in a given data set.

Association rules are rules that describe the relationship between item sets. Support, confidence, and lift are three important indicators of association rules. Support indicates the frequency of occurrence of item sets in the entire data set, and confidence indicates the probability that item set B will also appear when item set A appears. Lift indicates the degree of association between two item sets.

Common algorithms for association rule mining include Apriori, FP-Growth and Eclat.

#### 2.1.3. Sequence Pattern Mining

Discovering frequent subsequences in time series data is known as sequence pattern mining, and common algorithms for sequence pattern mining include GSP (Generalized Sequential Pattern) and PrefixSpan.

The GSP algorithm discovers frequent subsequences by expanding the sequence pattern according to the steps. The steps of the GSP algorithm are: generate candidate sequences, scan the database, prune, and repeat.

The PrefixSpan algorithm mines frequent subsequences by recursively projecting the database. The steps of the PrefixSpan algorithm are: database projection, recursive mining.

#### 2.1.4. Correlation Analysis

Correlation analysis is often used to measure the strength and direction of the relationship between variables. Common methods include Pearson correlation coefficient and Spearman rank correlation coefficient.

The Pearson correlation coefficient is mainly used to measure the linear relationship between two continuous variables, and its value ranges from  $-1$  to  $1$ .

The Spearman rank correlation coefficient, on the other hand, is used to measure the monotonic relationship between two variables and is applicable to both continuous and ordinal categorical variables.

### 2.2. Data Classification Technology

Data classification technique is one of the core techniques of data mining, and

the main technical process is to build a classification model by analyzing the existing dataset and using the classification model to classify the new data. The main goal of data classification techniques is to assign the samples in a data set to pre-defined categories, which requires building a suitable model that can accurately classify the data based on the sample characteristics. The model or algorithm used for classification is known as classifier.

Data classification technologies mainly include decision trees, support vector machines, naive Bayes, k-nearest neighbor algorithms, neural networks, etc.

The decision tree model is a type of tree model. It is a non-parametric supervised machine learning method and a commonly used model in the field of data mining and machine learning. Iterative dichotomiser 3 (ID3), C4.5, classification and regression tree (CART), chi-squared automatic interaction detector (CHAID), and quick unbiased efficient statistical tree (QUEST) are common decision tree model algorithms [2].

Support vector machine is a widely used pattern recognition algorithm and a binary classification algorithm based on statistical learning theory. It can map the input raw data to a certain point in high-dimensional space. Since different types of input sample data are clustered at different locations in high-dimensional space, it can realize the classification and recognition of different types of input data by finding the appropriate hyperplane. Thus, when new data are mapped to the same high dimensional space, it can be predicted which category they belong to based on the location of the points they are mapped to [3].

The Naive Bayes algorithm is a simple probabilistic classifier based on Bayes theorem, which assumes that the features that affect classification are independent of each other. Although in reality, the features are likely not completely independent. The Naive Bayes algorithm is based on the Bayesian formula and can calculate the posterior probability of a category given a set of features. The Naive Bayes algorithm uses feature data to calculate the probability of each category, and then selects the category with the highest probability as the final result [4].

The k-Nearest Neighbors (k-NN) algorithm is an instance-based learning method that selects the k-Nearest Neighbors by calculating the distance of a new sample from the samples in the training set and decides the class of the new sample by majority voting.

Neural network is the abbreviation of artificial neural network. It is inspired by the working mode of human brain nervous system. It is essentially a mathematical model. Similar to human brain neural network, artificial neural network is composed of artificial neurons and the connection between neurons. Among these neurons, there are two types of special neurons, one for receiving external information and the other for outputting information. Therefore, neural network can be regarded as a processing system for information from input to output. By inputting information into the neural network, the classification result can be obtained through the output layer.

## 2.3. Data Clustering Technology

Data clustering is a technique for grouping data objects so that data in the same group are similar to each other, while those in different groups are different. The core of data clustering technology is clustering, that is, the data set is divided into a number of disjoint subsets, and the goal of clustering analysis is to maximize intra-cluster similarity and minimize inter-cluster similarity.

Data clustering techniques mainly include K-means clustering, hierarchical clustering, DBSCAN, OPTICS, Gaussian mixture model and other common algorithms.

In the process of K-means clustering analysis, the number of clusters K is first determined, and then K sets of sample data are randomly selected as the initial cluster centers. The Euclidean distance is used as the standard for measuring similarity, and the square error is used as the clustering criterion function. Then repeat the iteration to minimize the value of the objective function. The K-means algorithm is simple and has high computational efficiency, but it is sensitive to initialization and is prone to falling into local optimality [5].

Hierarchical clustering achieves data clustering by constructing a hierarchical tree (dendrogram). According to different construction methods, it is divided into agglomerative hierarchical clustering and divisive hierarchical clustering. Hierarchical clustering does not require a preset number of categories and has high flexibility, but the computational complexity is high.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based spatial clustering method that discriminates the clustering of data by density. The method measures the density of the space where a point is located in terms of the number of data points in the domain of that point, and continuously searches for neighboring points to classify the data that is dense enough to be classified into clusters. It can classify spatial data with noise into arbitrarily shaped clusters [6].

OPTICS is a density-based non-parametric clustering algorithm. Given a set of point clouds in a certain space, it will output a set of point cloud sequences, which contain clustering information under different density thresholds. Based on this information, the output results can be flexibly adjusted [7].

The Gaussian mixture model is a clustering method based on probability distribution, which fits the data set by mixing multiple Gaussian distributions. The Gaussian mixture model assumes that the data comes from a mixture of multiple Gaussian distributions, each Gaussian distribution corresponds to a cluster, and the probability of each data point belonging to a cluster is known.

## 3. Application Fields of Data Mining Technology

### 3.1. The Internet Field

Data mining technology has a wide range of applications in the Internet field and plays a significant role. It can help enterprises better understand user behavior, optimize services, and improve user satisfaction and commercial bene-

fits.

With the popularization of the Internet, a variety of Internet industries are booming, and with this comes a large number of information processing problems. How to obtain useful data from the complex information is the research goal of data mining technology.

The recommendation system is a typical application of data mining in the Internet field. It recommends personalized content and products to users by analyzing their behaviors and preferences.

Recommendation systems are mainly divided into three types: personalized recommendation, collaborative filtering, and hybrid recommendation.

Personalized recommendations mainly recommend relevant products to users by analyzing user data. For example, e-commerce platforms recommend relevant products by analyzing users' purchase history, browsing history, and shopping cart data. Major video sites recommend videos that users may like based on users' viewing history, search history, and ratings.

Collaborative filtering is divided into user-user collaborative filtering and item-item collaborative filtering. User-based collaborative filtering recommends related content based on the preferences of similar users. For example, if A and B have similar viewing histories, then the content that user A likes may also be liked by user B. Item-item collaborative filtering is relatively simple and recommended based on the similarity between items. For example, if a user has watched movie A, he may also like a similar movie B.

Hybrid recommendation combines content and collaborative filtering to complement each other's strengths and weaknesses for content recommendation. For example, a music player recommends new songs and singers based on the user's listening history and music features.

Optimization of advertisement putting is also an important role of data mining technology in the Internet field. By analyzing users' data through data mining technology, accurate advertisement putting can be achieved to obtain higher advertising effects.

By analyzing users' behavior data, companies build user portraits and deliver personalized ads to users, such as personalized ads in WeChat Moments. Based on users' purchase and search records, relevant ads are delivered accurately.

Real-time data mining can also enable real-time bidding for advertisement putting, adjusting advertisement bids according to the effect of advertisement putting and maximizing the effect of advertisement placement.

### **3.2. The Financial Field**

In the financial field, data mining technology can help financial institutions optimize business processes, reduce risks, and improve customer experience and profitability.

One of the applications of data mining technology in the financial field is credit risk management. Data mining technology builds credit scoring models to assess the credit risk of customers by analyzing their credit records, financial

transactions and demographic information, and these assessment models can help financial institutions to determine loan approvals, credit limits and interest rates.

Data mining technology predicts whether a customer will default through classification models. Based on the customer's transaction behavior and repayment habits, the credit score is dynamically adjusted to promptly reflect changes in credit status.

Fraud detection is an important application of data mining technology in the financial field. Financial fraud is a major risk faced by financial institutions, and data mining technology identifies abnormal behavior and prevents fraud by analyzing transaction data. For example, a rule-based detection system is established, combined with data mining results, to monitor and intercept suspicious transactions in real time.

Risk management is another important application of data mining technology in the financial field. Data mining technology identifies and predicts various risks in the financial market by analyzing historical and real-time data to help financial institutions develop effective risk management strategies.

In value-at-risk analysis, historical simulation methods, Monte Carlo simulation methods, etc. are used to estimate the risk level of investment portfolios. It simulates the performance of assets under extreme market conditions and evaluates the risk resistance of the investment portfolio.

### 3.3. The Medical Field

The application of data mining technology in the medical field greatly improves the quality and efficiency of medical services and promotes the development of medical research. In the construction of hospital informatization, data mining technology has powerful data processing capability, which can efficiently process huge medical data [8]. By analyzing and mining a large amount of medical data, medical institutions can discover hidden laws and knowledge, and provide scientific basis for disease prevention, diagnosis, treatment and management of diseases.

Data mining technology can be used for disease prediction and diagnosis. By analyzing a patient's historical health data, living habits, genetic information, etc., data mining technology can predict the probability of certain diseases and help doctors take preventive measures in advance. It can also help doctors to provide auxiliary decision support in diagnosing diseases and improve the accuracy and efficiency of diagnosis.

Data mining technology plays an important role in drug research and development, accelerating the discovery and development of new drugs, utilizing association rules and classification models to analyze disease data from compound libraries, screening potential drug candidate compounds, discovering new indications for existing drugs, and improving the efficiency and success rate of drug development.

Data mining technology analyzes real-time health data to monitor the health

status of patients and provide timely intervention measures. It analyzes health data collected by wearable devices to monitor the health status of patients and provide personalized health management suggestions. It uses anomaly detection algorithms to identify abnormal changes in health data and provides timely warnings of potential health problems.

For example, by analyzing the blood glucose data of diabetic patients, abnormal fluctuations can be discovered in time.

Since data mining can quickly detect the potential patterns and connections between data, and can be presented more intuitively using visualization software, it is ideally suitable for processing the large amount of laboratory test data and imaging data generated during the diagnosis and treatment of treatment, and then integrates the medical care records and other medical records generated by patients in the process of long-term treatment to identify potential diabetes risks and complications, providing the basis for healthcare professionals to implement appropriate intervention plans, thereby achieving good diagnosis and treatment results [9].

### **3.4. The Environmental Meteorology Field**

Data mining technology has important applications in the fields of environmental meteorology field, helping scientists and decision makers to better understand natural phenomena, predict environmental changes, and develop effective management and response measures.

Short-term weather forecasting requires high-precision data and complex analysis techniques. Data mining technology can provide more accurate weather forecasts by analyzing a large amount of meteorological data. It uses time series models to analyze historical meteorological data to predict future weather changes [10]. For example, by analyzing past temperature, humidity, wind speed and other data, it can predict the weather conditions in the next few days [11].

Data mining technology can help monitor and protect ecosystems, analyze ecological data, and assess the health of ecosystems.

## **4. Talent Demand and Development of Data Mining Technology**

### **4.1. Talent Demand of Data Mining Technology**

With the arrival of the big data era, data mining technology is increasingly used in various industries, driving the demand for data mining professionals. Data mining technology talents need to have a variety of skills, including data analysis, programming, mathematical statistics, etc., as well as good business understanding and communication skills.

First of all, technical skills. Data mining technology talents need to master the basic principles and methods of data analysis and statistics, be able to process and analyze large amounts of data, and extract useful information and patterns. They should be familiar with commonly used statistical methods and tools in

statistics, such as regression analysis, variance analysis, hypothesis testing, etc. They should be proficient in using data analysis tools such as R and Python, and master relevant statistical and data analysis libraries.

The second is programming and algorithms. Data mining technology talents need to have a solid programming foundation, be familiar with common programming languages and data mining algorithms, master Python, R, Java, C++ and other commonly used programming languages, be able to write efficient data processing and analyzing code, be familiar with commonly used data mining algorithms, such as decision trees, random forests, support vector machines, cluster analysis, association rules, neural networks, etc., and able to choose appropriate algorithmic models according to specific problems.

Database and big data technologies are also required for data mining technology talents, who need to have the ability to process large-scale data and be familiar with data management and big data processing technologies.

Data mining technology talents need to have a solid mathematical foundation and master relevant theoretical knowledge in order to understand and apply complex algorithms and models. Understand linear algebra knowledge such as matrix operations, eigenvalue decomposition, singular value decomposition, etc. Master probability theory knowledge such as probability distribution, random variables, Bayesian theory, as well as mathematical statistics methods such as sampling, estimation, hypothesis testing, etc. Understand common optimization methods and algorithms, such as gradient descent method and Lagrange multiplier method.

Data mining technology talents need to have good business understanding and be able to apply data analysis results to actual business to provide support for decision-making. Understand the business processes and characteristics of the industry served, and be able to analyze and mine data according to industry needs. Have the ability to solve actual business problems, be able to discover problems through data analysis, propose solutions, and evaluate the effectiveness of solutions.

Data mining technology talents have a broad development prospect in the era of big data. With the increasing demand for data analysis and mining in various industries, the demand for data mining technology talents will continue to grow. In order to meet this demand, educational institutions and enterprises need to work together to provide systematic theoretical knowledge and practical training to cultivate professionals talents with multifaceted skills. By improving the training quality of data mining technology talents, enterprises can better utilize data mining technology to improve business efficiency and decision-making capabilities, providing strong support for enterprise development.

## **4.2. Development of Data Mining Technology Development**

### **4.2.1. Current Status of Data Mining Technology Development**

Data mining technology has developed rapidly in the past decades and has become one of the key technologies in the information age. With the advancement

of big data, artificial intelligence and computing power, data mining technology has been more and more widely used in various industries.

Technological progress and innovation. The development of big data technology has greatly contributed to the advancement of data mining technology. The maturity of big data processing frameworks and tools such as Hadoop and Spark has made it possible to process massive amounts of data.

Hadoop ecosystem: Hadoop's MapReduce framework and HDFS distributed file system provide reliable big data storage and processing capabilities and support large-scale parallel computing. HDFS (Hadoop Distributed File System) allows the storage and management of large-scale data sets, while MapReduce supports the execution of distributed computing tasks.

Spark: As a rising star of Hadoop, Spark provides faster in-memory computing capabilities and rich APIs, supporting a variety of application scenarios such as batch processing, stream processing, and graph computing. Compared with MapReduce, Spark has a faster processing speed and supports complex data analysis tasks such as machine learning and graph computing.

The development of machine learning and deep learning technologies has provided powerful algorithm and model support for data mining. In particular, the breakthrough progress of deep learning in the fields of image, speech, natural language processing has greatly expanded the application scope of data mining.

The widespread use of deep learning frameworks such as TensorFlow, PyTorch, and Keras has simplified the development and application of complex models. TensorFlow and PyTorch provide powerful tools for research and industrial applications, while Keras is known for its simplicity and ease of use, making it suitable for quickly building and testing deep learning models.

Innovative algorithms such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory Networks (LSTM) provide more flexible and efficient tools for data mining. These algorithms perform well in processing images, time series data, and natural language processing tasks, greatly enhancing the effectiveness of data mining.

The convergence of cloud computing and edge computing provides powerful computing resources and flexible data processing capabilities, making large-scale data mining more efficient and convenient.

Cloud computing AWS, Google Cloud, Azure and other cloud service providers provide powerful computing and storage capabilities to support large-scale data mining tasks. Cloud computing enables enterprises to expand computing resources on demand, process massive amounts of data, and support various data mining applications.

The development of edge computing technology, driven by the growth of the Internet of Things (IoT) and smart devices, has made it possible to process and analyze data at the device end, reducing the latency and cost of data transmission. Edge computing is suitable for applications that require real-time response, such as autonomous driving and industrial IoT.

Data mining technology shows strong application potential and broad devel-

opment prospects in the era of big data. With the continuous progress of technology and the expansion of application fields, data mining technology will play an increasingly important role in various fields such as business, finance, healthcare, environment and so on.

#### **4.2.2. Challenges of Data Mining Technology**

Although data mining technology has demonstrated great potential and wide application in many fields, it still faces many challenges in practice.

**Data quality issues.** Data quality is the basis for successful data mining. Real-world data usually have various quality issues such as missing values, inconsistencies and duplicate data, which can affect the accuracy and reliability of data mining results.

**Data integration issues.** Data integration involves bringing together data from different sources and formats to form a unified data set to support comprehensive analysis and mining. This process often faces multiple challenges.

**Privacy protection.** With the development of data mining technology, privacy protection has become an important issue. Especially when processing data involving sensitive personal information, it is a key challenge to protect personal privacy while ensuring the effectiveness of data mining.

**Data security** refers to protecting data from unauthorized access, tampering, and destruction. With the widespread application of data mining technology, data security issues are becoming more and more important.

**Model complexity,** with the development of data mining technology, the models and algorithms used are becoming more and more complex. Although complex models can improve the accuracy of analysis, they also bring difficulties in understanding and interpretation.

**Model Interpretation** refers to the ability to understand and explain the predictions of a model. Especially in fields that require high credibility, such as finance and medicine, model interpretability is particularly important.

**Computing resource requirements.** Large-scale data mining requires powerful computing resource support, especially when dealing with massive data and complex models, the demand for computing resources is more prominent.

**Real-Time Requirements.** In many application scenarios, such as financial transactions, intelligent monitoring, and real-time recommendations, data mining requires real-time processing and response capabilities.

## **5. Conclusion**

Although data mining technology has shown great potential in a number of fields, it still faces challenges in various aspects, such as data quality, privacy protection, model complexity, computational resources and talent demand. Through continuous technological innovation, interdisciplinary integration and talent cultivation, these challenges can be effectively addressed and the further development and wide application of data mining technology can be promoted. The future of data mining technology is full of opportunities. With the continu-

ous advancement of technology and the expansion of application fields, it will play an increasingly important role in all aspects of society and economy.

### Conflicts of Interest

The author declares no conflicts of interest.

### References

- [1] Chen, Z.Q. and Wu, H.Q. (2024) Application of Data Mining Technology in Network Security. *Cyberspace Security*, **15**, 121-125.
- [2] Shen, F.L.Z. and Wang, R.P. (2024) Application of Decision Tree Model in Clinical Research Data Analysis. *Shanghai Medicine*, **45**, 14-18.
- [3] Zeng, Q.T., Chen, G.H. and Li, W.X. (2024) Rapid Detection and Classification of Steel by Laser Induced Breakdown Spectroscopy Based on Particle Swarm-Support Vector Machine Algorithm. *Spectroscopy and Spectral Analysis*, **44**, 1559-1565.
- [4] Li, T., Sun, Y.Y. and Li, X.L. (2024) Research on Auxiliary Diagnosis of Diabetes Based on Machine Learning Classification Algorithm. *Computer Knowledge and Technology*, **20**, 27-29. <https://doi.org/10.14004/j.cnki.ckt.2024.0489>
- [5] Zhang, Y., Xu, Y.M. and Zhang, Y. (2021) A Multivariate Linear Regression Prediction Model for Substation Line Loss Rate Based on a New K-Means Clustering Algorithm. *Journal of Electric Power Science and Technology*, **36**, 179-186. <https://doi.org/10.19781/j.issn.1673-9140.2021.05.022>
- [6] Huang, J. and Yang, L.Q. (2024) A Robust AdaBoost Regression Model Based on Improved DBSCAN Algorithm. *Journal of Hefei University (Comprehensive Edition)*, **41**, 1-9.
- [7] Pan, Q., Lin, Q.X. and Liu, Z.Y. (2022) Thunderstorm Cell Identification Method Based on Radar Data Based on OPTICS Clustering Algorithm. *Meteorological Science and Technology*, **50**, 623-629. <https://doi.org/10.19517/j.1671-6345.20210375>
- [8] Wu, Q. (2024) Analysis on the Application of Data Mining Technology in Hospital Informatization. *Big Data Era*, No. 3, 48-51.
- [9] Liu, T.N., Liu, J.L., Huang, J.W., et al. (2024) Application Progress of Data Mining Technology in Diabetes Management. *Journal of Jinan University (Natural Science and Medicine Edition)*, **45**, 11-20.
- [10] Viktor, M-S. (2013) Big Data: A Revolution That Will Transform How We Live, Work, and Think.
- [11] Shen, W.H. (2016) Re-Analysis of Meteorological Big Data and Its Application. *China Informationization*, No. 1, 85-96.