Development and Application of Big Data Platform for Garlic Industry Chain

Weijie Chen[1], Guo Feng[1], Chao Zhang[1], Pingzeng Liu[1, *], Wanming Ren[2], Ning Cao[3] and Jianrui Ding[4]

Abstract: In order to effectively solve the problems which affect the stable and healthy development of garlic industry, such as the uncertainty of the planting scale and production data, the influence factors of price fluctuation is difficult to be accurately analyzed, the difficult to predict the trend of price change, the uncertainty of the market concentration, and the difficulty of the short-term price prediction etc. the big data platform of the garlic industry chain has been developed. Combined with a variety of data acquisition technology, the information collection of influencing factors for garlic industry chain is realized. Based on the construction of the big data technology platform, the real-time synchronous acquisition, efficient storage and analysis of the planting, market, storage, processing, export and logistics information in five provinces and seven counties are realized. The application of the big data platform for garlic industry chain has realized the accurate acquisition of garlic planting area, the price and trend of market circulation and the information of export information, analyzed the fluctuation regulation of garlic price, and also realized the short-term precision prediction of garlic price.

Keywords: Garlic industry chain, agricultural big data, planting area, price fluctuation regulation, price prediction.

1 Introduction

The rapid development of new information technologies, such as big data, Internet of things and artificial intelligence, has effectively promoted the pace of modern agricultural construction [Li (2012)]. Agriculture is not only the source of big data, but also the typical area of big data application. Through the research and application of big data collection, storage, analysis and mining technology, large numbers, wide sources and diverse types of data are fully utilized in the field of agricultural products market, and the ability to obtain and use agricultural product market data resources has been continuously improved. It is very necessary to realize more accurate monitoring, analysis and prediction of market economy.

---

[1] Shandong Agricultural University, Tai'an, 271000, China.

[2] Shandong Provincial Agricultural Information Center, Ji'nan, 250013, China.

[3] College of Information Engineering, Qingdao Binhai University, Qingdao, 266000, China.

[4] Department of Computer Science, Utah State University, Logan, USA.

* Corresponding Author: Pingzeng Liu. Email: lpz8565@126.com.

To realize the stable development of the agricultural product market, the United States economic statistician Moore chose 21 most representative indicators from the time series of nearly 1000 statistical indicators, the establishment of a new composed of the leading indicators, and lagging indicator three indicators of climate monitoring system. The system can make use of some macroeconomic indicators to describe the volatility of a country's economic boom. This is an important means of macroeconomic monitoring. This provides a basis for the country to strengthen macroeconomic regulation and control [Trostle (2008)]. Starting from the circulation of the market, Yang Kun et al. [Yang and Wang (2010)] designed agricultural product price information monitoring and forecasting system, and realized the functions of monitoring market price change, analyzing market price trend, and putting forward prediction and warning suggestions. Wang et al. [Wang, Zhao and Zhao (2014)] based on B/S model, using Spring+Struts+Hibernate framework, based on the design idea of system development of open source projects, through the middleware embedded Eviews and Easyfit, a short-term prediction system of agricultural products price was built up. It has realized the short-term prediction of agricultural product price. This strengthened the guiding role of economic analysis results for real production. Li et al. [Li and Liu (2011)] proposed the construction of agricultural information database and combined the radial basis function artificial neural network model to predict the price of agricultural products for a short time. With the emergence of new technologies such as Internet of things, big data and artificial intelligence, the construction of big data in agricultural products has been developing step by step. It also provides technical support for the macro management and decision making ability of the industry and the accurate price prediction. Through the Internet of things, Internet technology and sensor networks for data collection, Ferrández-Pastor Francisco Javier et al., Li Daoliang et al. and Zhang Shirui et al. and so on have achieved the precision management of agricultural products [Ferrández-Pastor, Garcia-Chamizo, Nieto-Hidalgo et al. (2016); Li and Yang (2018); Zhang, Zheng, Shen et al. (2012)]. In the field of data storage, the GFS, GPFS, ADFS distributed file system and non-relational database and spatial data warehouse are applied to the big data storage of agricultural products [Ghemawat, Gobioff and Leung (2003); Mckusick and Quinlan (2009); Zhang, Li, Zou et al. (2014); Han, Mao, Xia et al. (2016)]. In the research and development of big data platform, Monsanto's integrated farmland management system, FarmLogs company's cloud SaaS model farm management, VitalFields company's cloud farm log and Agriculture Consultant and Smart Management company's Yitian farm cloud platform have realized the digitalization and information management of farmland to provide farmers with a full range of farmland services. In order to ensure the healthy and stable development of the agricultural product market, scholars at home and abroad have carried out a variety of research. The research and development of single technology is relatively mature, while the information through and integration of the whole industry chain is less. There is still a problem of the asymmetric information of the whole industry chain and the low integration of big data and agricultural product chain.

In the reality that garlic industry chain is becoming more and more closely linked and industrial data is growing rapidly, with the help of big data technology, the deep integration of big data and the development of garlic industry can be achieved, and the big data platform of the garlic industry chain is researched and developed. To a certain

extent, it can provide data support for data mining, and it can effectively solve the problems of uncertainty of planting scale and yield data, difficult analysis of the influence factors of price fluctuation, difficult to predict the trend of price change, the uncertainty of market concentration and the difficulty of short-term price prediction. To realize the tracking and monitoring of the whole garlic industry chain, complete the data collection and storage, analyze the influence factors of garlic price fluctuation and the characteristics of garlic price fluctuation, accurately predict the trend of garlic price operation, and make the timely forecast service report on garlic price fluctuation is an effective way to guarantee the stable development of garlic industry.

2 Architecture and key technologies

The big data of agriculture is a data set which is produced by the integration of its own characteristics, such as regional, seasonal, diversity, and periodicity and so on. It has a wide variety of sources, diverse types, complex structure and potential value, and it is difficult to apply the usual methods to deal with and analyze the data [Xu, Wang and Li (2015)]. With the support of big data system framework and big data technology, the big data construction of single variety agriculture should be integrated with big data processing processes, such as industrial production, logistics, market trade, purchase and marketing, export, consumption and so on. Big data technology is a collection of data collection, data storage, data processing and data visualization technology [Gong, Li, Chai et al. (2014); Yan and Zhang (2013); Low, Bickson, Gonzalez et al. (2012); Li and Gong (2015)].

2.1 Architecture

According to the information of each link in the industrial chain, in order to meet the requirements of high performance, scalability, high reliability and low energy consumption, the big data processing system is divided into four layers according to the big data processing process, as shown in Fig. 1. The first layer is the physical layer of the agricultural big data system architecture. According to the data resources of industrial chain planting, production, market, supply and marketing, processing and export, combined with big data collection and storage technology, it is divided into pre-production materials, production information, export information and other types of data. The second layer is the information layer of the agricultural big data system architecture. Through a variety of big data acquisition technology to obtain various types of data, data sources of industrial chain elements, and the industrial chain data storage to generate data warehouse. Through data quality analysis and data cleaning, denoising and integration technology, the collected data are classified and stored in the storage system to form a storage mode of continuous accumulation of agricultural data resources. The third level is the data analysis and mining layer of the agricultural big data service system. Through the big data preprocessing technology, big data analysis technology and big data visualization technology, according to the system requirements, the mass data resources can be fully excavated and analyzed. The fourth level is the service layer of the agricultural big data service system. Through the application of agricultural big data, it can achieve the purpose of intelligent production, transparency of industrial management,

information of market management, flexible service and information of scientific research management. This not only makes all the links of the industrial chain information, but also reduces the cost of big data processing in agriculture. It also ensures the reliability of the software, and also reduces the threshold of big data processing.
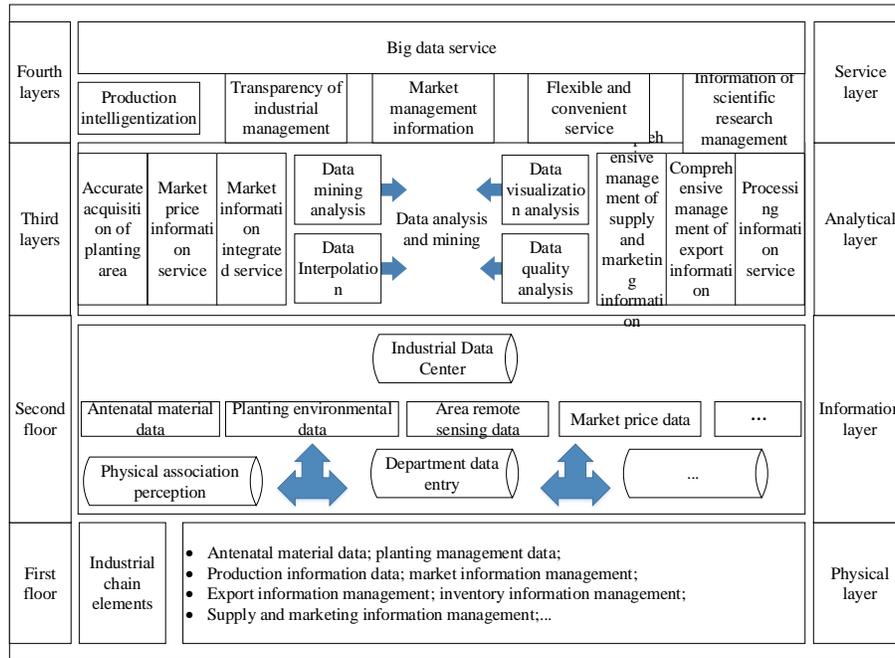
Figure 1: The architecture of agricultural big data system

2.2 Key technology

Using the corresponding big data technology, the agricultural big data service system is built, as shown in Fig. 2.

2.2.1 Data acquisition technology

The research and development of the platform is based on data. Data as the core part of the system is the key element of the whole system. In the process of developing a data centric platform, it is necessary to improve the means of data acquisition. According to the data features of agricultural production, market, storage, processing, export and other links, this paper mainly introduces the data acquisition technology of physical association perception and social network data.

IoT perception is based on intelligent hardware real-time perception of resource dynamic information, and provides scientific and technological means ahead of time for monitoring, operation, management and service. It can monitor all people and things in time, accurately and quickly by deploying sensors such as ecological monitoring, climate monitoring, environmental monitoring, and market monitoring. It has the advantages of universality, compatibility, real-time and intelligence.

RFID technology is a wireless communication technology that can identify specific targets and read and write related data through a radio signal without identifying mechanical or optical contact between a system and a specific target. It has the advantages of fast scanning, small size, diversified shape, anti-pollution ability and durability, repeatability, penetration and barrier free reading, large memory capacity and security.
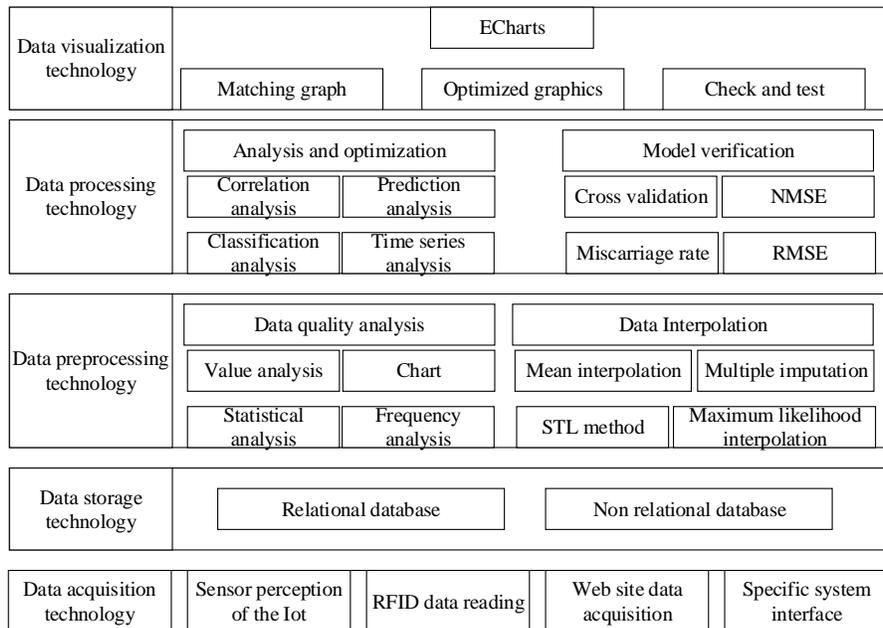
Figure 2: Big data technology

Social network data acquisition is a process of acquiring data by using various soft-ware tools for massive and complex data in the network. Social network data acquisition is divided into internal data acquisition and external data acquisition. Internal data acquisition is based on the needs of companies, schools and government agencies to extract data from their internal databases, resource repositories and data statistics platforms. There are two ways to obtain external data, one is to get external open data sets, such as log on to China Meteorological Website to download some data that is open to the outside world. The second is to use the Python crawler to capture data from the network, such as obtaining the effective price information from the price web site.

2.2.2 Data storage technology

Data storage is an essential link to ensure the reliability and accuracy of the analysis results. The big data storage system generally includes three aspects of the file sys-tem, the database technology and the programming model.

A file system is a method and data structure used by an operating system to specify files on storage devices or partitions, that is, to organize files on storage devices. Common distributed file systems include GFS, HDFS, Lustre, TFS, and so on. ① Google is a

proprietary and partially open source distributed file system (GFS) developed to meet the needs of the company, which is more suitable for large file storage and reading operations than in writing operations, but GFS has the disadvantages of single point failure and low efficiency in processing small files [Ghemawat, Gobioff and Leung (2003)]. ② The open source community has implemented an open source distributed file system HDFS deployed on Hadoop. HDFS is a file system based on reliable storage of big data on large clusters. It is the storage foundation of distributed computing [Li, Shen, Walter et al. (2017)]. ③ Lustre is a large-scale, secure and reliable cluster file system with high availability. It is developed and maintained by SUN Company. The main purpose of this project is to develop the next generation of cluster file system, which can support more than 10000 nodes and store data in PB data volume. ④ TFS is a distributed file system with high scalability, high availability, high performance and Internet service oriented. It can provide high reliability and high concurrency storage access for external applications, and is widely used in Taobao applications.

The programming model is critical to achieve the application logic and auxiliary data analysis, there are three main programming models: general processing model, stream processing model and graph processing model. The general processing model is used to solve general application problems and is often used in MapReduce [Zhang and Xiao (2015)] and Dryad [Isard, Budiu, Yu et al. (2007)]. The stream processing model uses RAM as the data storage medium to achieve high access and processing rates, S4 [Neumeyer, Robbins, Nair et al. (2010)] and Storm are two distributed stream processing platforms running on the JVM [Li and Gong (2015)]. The graph processing model is used in situations as social network analysis and RDF which can be represented as interrelations of entities.

Database technology is the basic theory and implementation method of the structure, storage, design, management and application of the database, and uses these theories to process, analyze and understand the data in the database. The object of database technology research and management is data, so the specific content of database technology mainly includes: ① Through the unified organization and management of the data, the corresponding database and data warehouse are established according to the specified structure. ② The database management system and data mining system are used to design a data management and data mining application system that can add, modify, delete, process, analyze, understand, report and print the data in the database. ③ Finally, the data processing, analysis and understanding are realized by using the application management system. Among them, relational database and non-relational database are widely applied in database technology.

MySQL is an open source relational database management system. MySQL data-base can start from simple data retrieval, and gradually include some complex con-tent such as connected use, subquery, regular expression and all text based search, stored procedure, cursor, trigger, table constraint and so on, so as to realize the effective storage and management of data. It has the advantages of small size, fast speed, low total cost and open source. It has the advantages of fast speed, reliability and fast adaptability in the process of storing data. Redis is an open source non-relational database written in ANSI C language. Redis is a high-performance key-value data-base and supports data sorting in

various ways. Its advantages are: ① volatile and persistent key storage capability; ② good at processing array types of data; ③ with very fast processing speed; ④ it can quickly process time series data and is easy to handle set operation. ⑤ there are many methods that can be used for atomic processing.

2.2.3 Data preprocessing technology

Data preprocessing refers to processing data before data mining, including cleaning up outliers and correcting erroneous data. The purpose of data preprocessing is to preprocess the dimensionized variables through the data filtering and conversion algorithms, such as classification, segmentation and conversion, and convert the array of original variables into the form that can be used. Data preprocessing can not only improve the efficiency of the system, but also meet the diverse needs of users. According to the purpose of pretreatment and the demand of garlic industry big data service system, data pretreatment is divided into two aspects: data quality analysis and data interpolation.

Data quality analysis usually includes numerical analysis, statistical analysis and frequency analysis techniques. Numerical analysis is based on the need to design and implement some computing methods, and get approximate and accurate results for data quality analysis. Statistical analysis refers to the use of statistical methods and knowledge related to the analysis object, from the quantitative and qualitative analysis of data quality. Frequency analysis refers to the use of statistical methods in a certain range to complete a certain number of data index number or frequency of numerical statistics.

According to the result of data quality analysis, in order to improve the utilization rate of data, it is necessary to complete the interpolation of data according to the requirements. Data interpolation usually includes mean interpolation, multiple interpolation, STL method and maximum likelihood estimation interpolation. The mean interpolation method is a simple and fast method for missing data processing. It fills the missing variable value according to the average value of the variable on all other objects, or the data that will appear the highest frequency or frequency in the same attribute data to compensate for the missing data value. This method has the ad-vantages of low data quality, easy operation and excellent interpolation results. Multiple interpolation is derived from the single interpolation method and is used to fill the missing values of complex data. It constructs m substitution values for each missing value, and then processes the generated m complete data sets by the same data analysis method, obtains the M processing results, and finally obtains the estimated value of the final target variable based on some one. The maximum likelihood estimation interpolation method is used to estimate the unknown parameters through the marginal distribution of observed data. For maximum likelihood parameter estimation, the most commonly used method is expectation maximization. The method is applicable to large samples, and the number of effective samples is sufficient to ensure that the maximum likelihood estimator is asymptotically unbiased and obeys normal distribution. However, this method may fall into local extremum, and the convergence speed is not very fast, and the computation is very complicated.

2.2.4 Data analysis technology

Data analysis is the core and the most valuable part of the field of big data technology. Through data analysis, unknown valuable regulations and results can be revealed, and can help people make more scientific and intelligent decisions. Data analysis technology generally includes data mining and big data analysis technology. In order to realize the data analysis of the garlic industry, the relevant analysis, prediction analysis, time series analysis and machine learning algorithms are introduced in detail in this study.

Correlation analysis is the process of describing the close relationship between objective things and using appropriate statistical indicators. Its specific concept refers to the rule of the existence of two or more than two variables in a sense. The purpose is to explore the hidden relationships in data sets. Through correlation analysis, the correlation degree and causal relationship between statistical variables can be analyzed. Commonly used correlation analysis methods include chart correlation analysis, covariance and covariance matrix, correlation coefficient and so on.

Prediction analysis is a scientific cognition activity for future development. The prediction analysis is not simply relying on experience and intuition for simple guess-work. It is a scientific analysis and decision made by combining the theories and methods of a variety of disciplines and through accurate calculation and reference to objective laws. It is a statistical or data mining solution that can be deployed for many other purposes such as prediction, optimization, prediction and simulation. Commonly used methods are exponential smoothing, moving average, and linear regression and so on.

Time series analysis is a statistical method for dynamic data processing. This method uses stochastic process theory and mathematical statistics to study a set of numerical sequences based on time to predict its future development. The main feature of this method is that it is not influenced by other external factors, and the trend of market demand is predicted by the study of time. The commonly used time series analysis methods include AR model, MA model, ARIMA model, grey prediction model, SVM model and so on.

The intelligent prediction method based on machine learning has a unique ad-vantage in forecasting accurate value. Its biggest branch is supervised learning and semi-supervised learning. Among them, dimensionality reduction and clustering are classified as unsupervised learning, and regression and classification are supervised learning. Data in unsupervised learning is not labeled, and data can be processed dimensionality reduction. Data in supervised learning require labels, which are used to predict new data based on existing results. Commonly used methods include neural network, KMeans, decision tree method, regularization method and so on.

In order to achieve more accurate results in the analysis process, it is necessary to further judge the accuracy of the algorithm or model according to the results of cross test, RMSE, misjudged rate and so on. If the accuracy is poor, it is necessary to re select the algorithm and conduct experiments.

2.2.5 Data visualization technology

Data visualization is a way of combining data and visualization technology. The

visualization of big data can not only intuitively display the basic information in a large number of data, but also drive complex data analysis, facilitate the interaction between data and users, and make users more convenient to see the changes in the multiple attributes of the data of the industry objects or events, and provide the convenient operation for the industry [Keim, Qu and Ma (2013)].

With the emergence of some visual software, visual controls and components based on third parties are becoming more and more applied to many fields, increasing the efficiency of work and promoting the development of visualization. The third party control, which is widely used ECharts has many advantages, such as diversified forms, easy operation, easy to call, and can be modified for many times, and has a prominent advantage in the display and analysis of data. The flow chart of the data visualization is shown below.
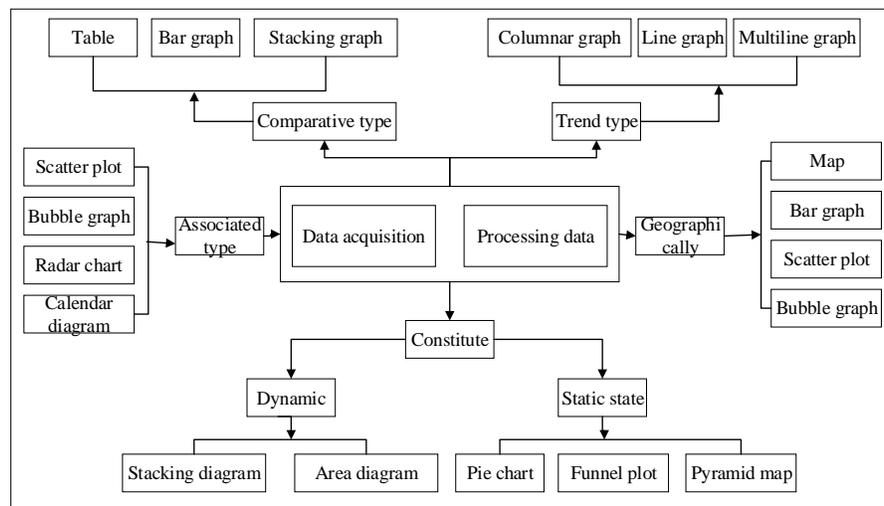
Figure 3: Flow chart of data visualization

3 Platform design

3.1 Overall design

In order to solve the problems such as uncertainty of planting scale and yield data, difficult analysis of the influence factors of price fluctuation, difficult to predict the trend of price change, the uncertainty of market concentration and the difficulty of short-term price prediction, the first premise is to determine the circulation process and the circulation links of the garlic industry chain. As shown in Fig. 4, the main participants in garlic industry chain include garlic growers, distributors, stockbrokers, market brokers, garlic traders, garlic enterprises and consumers. Among them, the farmers hope to be informed of the information services for garlic planting, the data resources needed to be shared by garlic merchants, and the information that consumers want to know about the garlic planting environment, the use of pesticides and fertilizers, the process of circulation processing, and the selection of packaging materials. According to the actual

needs of the participants in each link, the whole industry chain of garlic is divided into 7 links, including planting, purchasing, market, logistics, warehousing, processing and exporting, in combination with the concept of industrial chain management before, during and after production. There are many problems in the circulation process of garlic industry and the numerous participants in each link. Only through a thorough analysis of the characteristics of the whole industry of garlic can the garlic industry chain be managed in an all-round way.
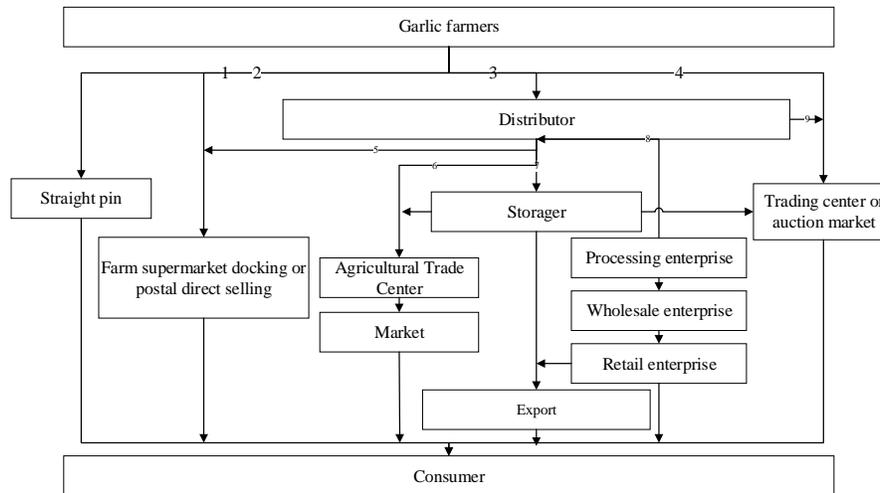
Figure 4: The process of garlic industry

3.1.1 Architecture of platform

This platform is divided into five layers, such as infrastructure layer, data layer, core layer, service layer and user layer, as shown in Fig. 5. The platform covers seven aspects of production, planting, logistics, market, warehousing, processing and export. Through the four subsystems of purchase and sale management, planting management, processing and management, and market management, the data and information of all links are organized effectively to provide services for farmers, enterprises, governments and consumers.

User layer	Growers	Enterprise	Consumer	Government

Service layer

Web page data visualization

Planting agricultural information	Market price	Historical price	Inventory quantity	Garlic processing
Farming operations	Market trend	Real time price	Inventory	Garlic export
Garlic plucking	Market volume	Price prediction	Physical distribution management	

Core layer

| Data synchronization | Framework service | Model optimization | Privilege priority | Process priority |

Data Layer

Data access technology and operation

| Land information database | Market information database | Garlic information database | Other databases |

Batch processing of data acquisition; Real time processing of data acquisition

| Platform internal data | Relevant department data | Internet data |

Infrastructure

| Data base | Network equipment | Data processing software | Other equipment |

Figure 5: The overall architecture chart of the big data platform for garlic industry chain

3.1.2 Key technology

The key factors affecting the garlic industry, such as planting area, export volume and capital speculation, are complex and space-time variable. Only the analysis of sample data cannot explain the variation of garlic industry well, and cannot meet the needs of the big data construction of single variety. The garlic industry has the characteristics of big data capacity, diversity, structural and unstructured, high redundancy and so on. The research and development of garlic industry chain big data plat-form technology framework which has the characteristics of distribution, parallel and high efficiency and support for the whole life cycle of big data engineering. As shown in Fig. 6. The platform technology architecture uses the mashup model data structure, which is based on the big data mix and mashup architecture. The data ware-house is mainly responsible for the traditional structured data business processing, and the big data platform is responsible for storage and processing of unstructured data and semi-structured data. Realize low-cost data storage and efficient data mining [Yur'Evich and Vasil'Evich (2013); Zhang (2015); Wang, Liu, Zhang et al. (2016)].
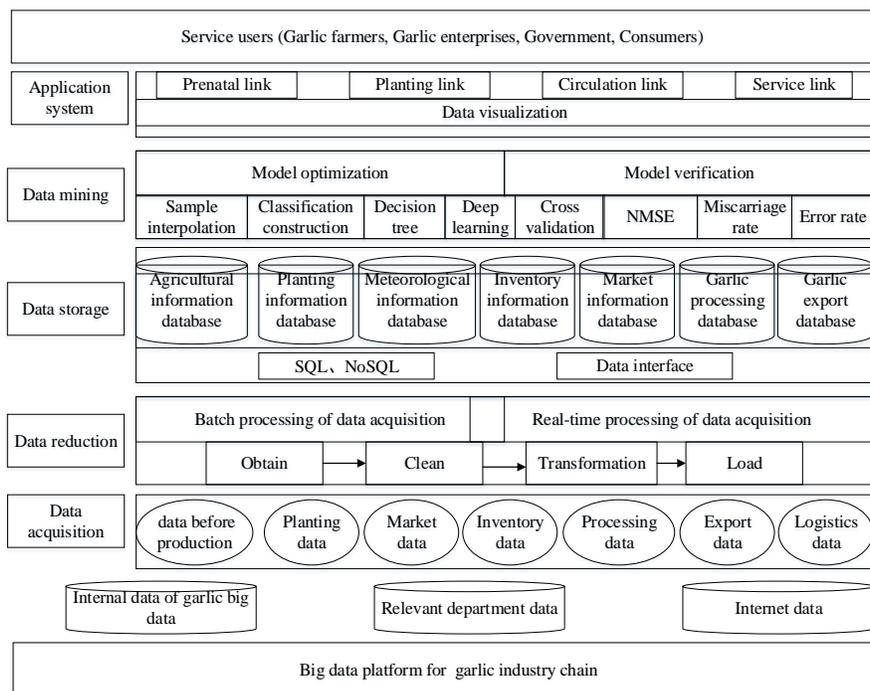
Figure 6: Technical framework of garlic big data platform

3.2 Function module

According to the overall framework of the platform, the functional modules of the platform are determined by the analysis of the standard operation and circulation management regulations in the actual production process of the industry, as shown in Fig. 7. Big data platform of garlic industry chain is divided into eight functional modules, including prenatal management, planting management, wholesale market management, price prediction management, inventory management, supply and demand management, export management and user management.

Among them, antenatal management is the preparatory link before planting garlic, mainly to complete the input of garlic prenatal management information, including planting information management module, planting site basic information management, planting site responsible person information management. The planting information management module, including the basic information management of planting area, production, geographical location and so on. Antenatal material information management module, production input information management of agricultural and artificial production before production. Garlic planting management covering the entire production process from planting to harvesting, including various planting field planting garlic varieties and planting date, planting area, sowing quantity, irrigation time, irrigation date, irrigation, farming operation records, growth of each growth period link monitoring index, garlic picking and garlic harvest cultivation information management. It has realized the information management of garlic production and planting, and provided data support for

the follow-up data analysis and price prediction. The wholesale market management records the origin, destination, market volume, market transaction price, logistics time and information management of commonly used customers of circulating garlic in wholesale market. Price prediction management, system administrators according to different regions, different varieties of garlic historical information and real-time quotes information, use the price prediction model to predict the garlic price data and store it into the database in real time. Garlic inventory management information includes vessel shell, storage variety, storage price, storage source, shipment volume, shipment variety, shipping price and shipping direction. Supply and demand management, including the management of the supply and demand of garlic, the quantity and the trend of supply and demand. Export management, including the management of exports of garlic, export volume and export price information. User management includes information management of farmers, garlic enterprises, consumers, and government personnel in the system platform. At the same time, in order to standardize system information and information management in all links, user rights are set up in this system. System administrators have the highest permissions of system management, while farmers and garlic enterprises have some rights of system management. Government employees have the right to get system data and monitoring system.
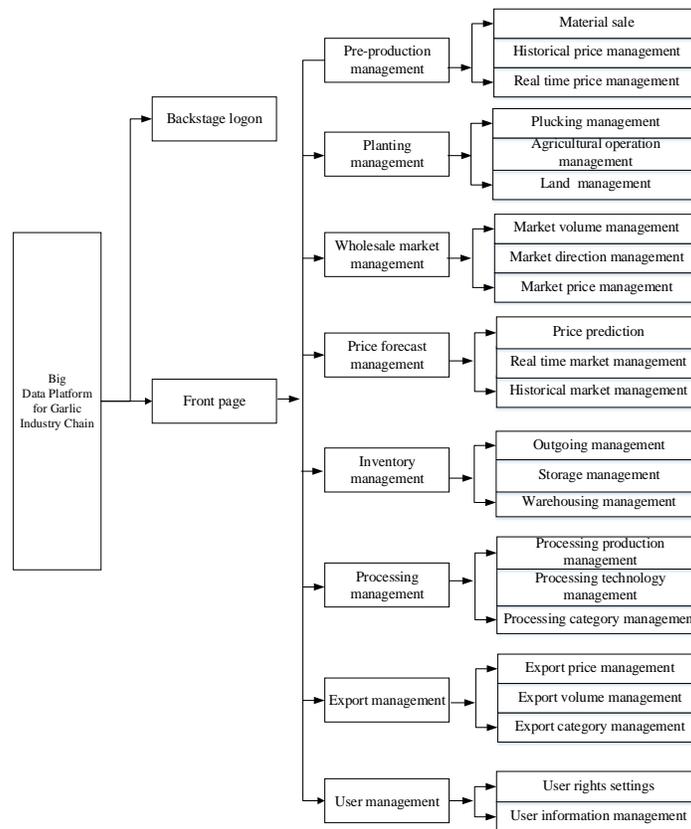
Figure 7: Function module for garlic industry chain

4 Garlic big data platform

4.1 Functional module division

The big data platform of garlic industry chain adopts the framework of platform + system, which realizes the collection of all the data of garlic planting, circulation and sales, which lays the foundation for data mining and analysis. At first, the four subsystems of purchase and marketing management, planting management, processing management and market circulation management are constructed on the basis of big data platform, which realizes the information function of industrial chain.

4.1.1 Purchase and marketing management subsystem

The purchase and marketing management module consists of two parts: demand information management and supply information management. Demand information management mainly through the system interface to publish demand variety, quantity, demand person contact address, demand person contact phone and related transaction type. Supply information management mainly through the system inter-face to release the supply of variety, quantity, supplier related contact and transaction types and other information. On the one hand, the electronic flow instead of the real logistics, breaks through the limitation of time and space, and reduces the inter-mediate links, so that the trading activities can be carried out at any time and place. On the other hand, the supply and demand information submitted by the registered users can further obtain the information of the volume of garlic trade and trade trend, which plays a key role in the information transparency of the garlic industry chain.

4.1.2 Planting management subsystem

Planting management module mainly has the function of planting management and the management function of growth period. Have access to the household registration and garlic prices, by planting management functions of real-time recording of the planting area, planting time, planting and harvesting category, farming operation records and other production information, also can through the monitoring data during the growth period of record management function of each link and growth period. Through this module, it can not only analyze the supply factors that affect the price change of garlic, but also realize the information management of garlic production, and provide data support for the prediction of garlic production and price.

4.1.3 Processing management subsystem

Processing management system mainly covers the processing products, processing technology, processing companies and processing standards, data processing companies have access to the system's real-time through recording processing information, and combined with other circulation information, for the selection of processing category, enhance enterprise brand competition, adjusting the production structure of enterprises to provide data services.

4.1.4 Market management subsystem

The market management subsystem mainly covers information such as market, logistics and export. With the support of big data technology, the comprehensive acquisition of multilink and multi type data, deep mining analysis, real-time dynamic monitoring and warning and decision service are completed, and the analysis results are classified into the market management subsystem interface. This realizes the function of information inquiry, data export and dissemination of information, and completes the process of data service in the market management of garlic industry chain.

4.2 Function module implementation

The big data platform of garlic industry chain mainly completed the functions of precision planting management, circulation monitoring and management, analysis of garlic price law and accurate short-term prediction of garlic price. Precision planting management includes the accurate acquisition of the planting area. The circulation monitoring and management relies on data acquisition technology to collect the price information of garlic in the market in real time. Using data analysis technology to explore the price law of garlic, and realize the short-term precise prediction of garlic prices, so as to provide services for the early warning of garlic price fluctuation.

4.2.1 Precision planting management

(1) Accurate acquisition of implants.

The factors affecting the huge fluctuation of garlic price include planting area, production, natural disaster, speculation and so on, among which the planting area is the most basic and most important data of the industry. In order to solve the problem that the sources of garlic planting area are different, the big data platform of garlic industry chain has realized the precision management of garlic planting area in China. Fig. 8 shows the main distribution area of garlic in China, and the generation of distribution map lays the foundation for obtaining accurate garlic planting area and determining the storage price of garlic.

Figure 8: Garlic regional distribution in China

(2) Precision acquired acreage

Garlic cultivation is mostly covered with plastic film mulching. In the early stage of garlic growth, the planting area of garlic was measured by means of remote sensing

image of planting area by means of geostatistical analysis tools in ArcGIS. Fig. 9 shows a remote sensing image of the main producing area of garlic in Shandong province. The image is classified and processed, which can effectively and accurately obtain the planting area of garlic in the production area.
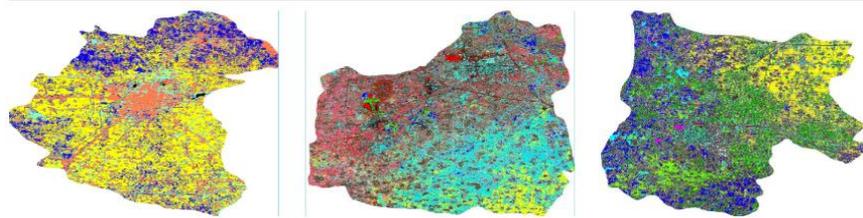
Figure 9: Remote sensing in Shandong province

4.2.2 Control and management of circulation

It is an effective way to solve the information asymmetry of the garlic industry that dynamic real-time monitoring affects the change of the main factors of the price fluctuation in the agricultural products market, the study and analysis of the regulation of price fluctuation and the accurate and timely prediction of the fluctuation trend of the market price of agricultural products. The big data platform of the garlic industry chain uses data acquisition technology such as artificial collection and web crawler to obtain real time price and export information from government platforms and authoritative websites that publish garlic price information. And through the advanced model and the effective data mining analysis method to analyze, judge and predict, give a forecast and make an early warning scheme for the possible trend of the interaction. Fig. 10 and Fig. 11 show the garlic market circulation monitoring and management interface.
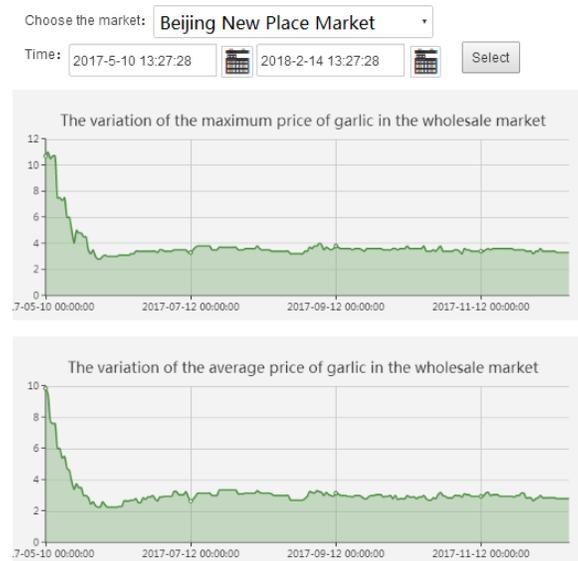
Figure 10: Monitoring and management of garlic market circulation
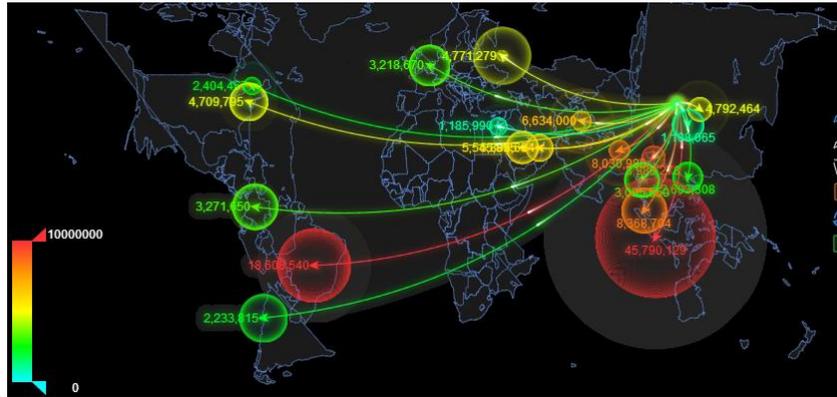
Figure 11: Monitoring and management of export trend

4.2.3 Price regulation and forecast

The price of garlic fluctuates more frequently, which has a great influence on the daily life of the residents. On the basis of the price data obtained in real time monitoring, the big data platform uses big data analysis technology to analyze the fluctuation law of garlic price and realize the short-term and accurate prediction of the price. Mastering the overall history of garlic prices and future fluctuations is extremely important for stabilizing the development of garlic industry.
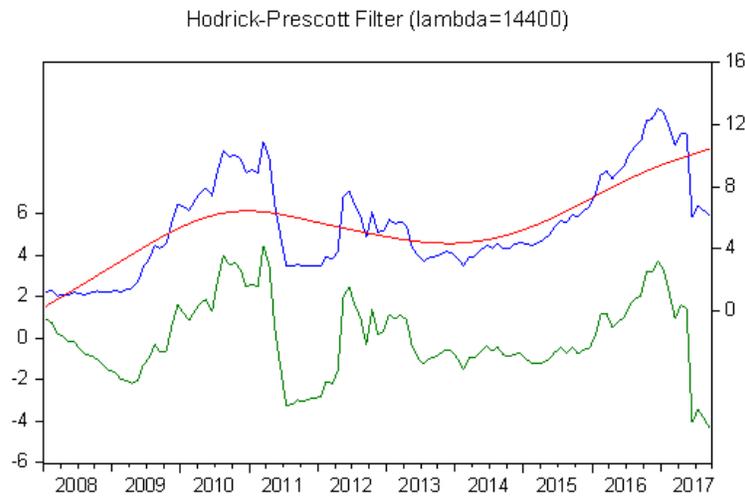
Figure 12: The price fluctuating characteristics of garlic
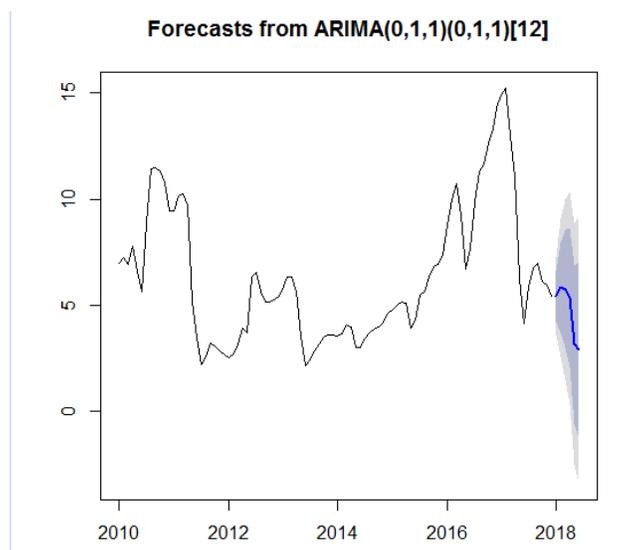
Figure 13: The price prediction of garlic

5 Conclusion

In order to solve the problem that the data of garlic planting scale and yield are not clear , the influence factors of price fluctuation are difficult to be analyzed, the trend of price change is difficult to predict, the degree of market concentration is not clear, and the short-term price prediction is difficult and so on, according to the service concept of big data of agricultural products, according to the construction target of "Big data + Garlic industry" and the support of big data technology, a big data plat-form for garlic industry chain is developed.

(1) The platform combines the status of garlic production and management in five major producing areas in Shandong, Jiangsu, Hebei, Henan and Yunnan provinces. The role and function of farmers, enterprises and government in the garlic industry chain are fully considered, and the key control points of planting, management, sales, storage, processing, logistics and export of garlic industry chain are also considered. This realizes the tracking and monitoring of the economic operation of garlic industrial chain.

(2) The platform implements the precision management of planting area function based on big data analysis. On the basis of data accumulation, the platform accurately located the national garlic planting area and got the planting area, which lays the foundation for the accurate analysis of garlic production.

(3) On the basis of full analysis of the factors affecting the price of garlic, the platform realizes the real-time monitoring of the market circulation and export links, the study of the regulation of price fluctuation and the short-term precision prediction of the price. This provides a scientific reference for the government to regulate the price of garlic. It has enhanced the scientificity, foresight and effectiveness of the business decision-making of the garlic merchants, and provided theoretical guidance for the stable and healthy development of the garlic market.

References

Ferrández-Pastor, F. J.; García-Chamizo, J. M.; Nieto-Hidalgo, M.; Mora-Pascual, J.; Mora-Martínez, J. (2016): Developing ubiquitous sensor network platform using internet of things: application in precision agriculture. *Sensors*, vol. 16, no. 7, pp. 1141.

Ghemawat, S.; Gobioff, H.; Leung, S. (2003): File and storage systems: the google file system. *Acm Symposium on Operating Systems Principles Bolton Landing*, vol. 37, pp. 29-43.

Ghemawat, S.; Gobioff, H.; Leung, S. T. (2003): The google file system. *ACM Sigops Operating Systems Review*, vol. 37, no. 5, pp. 29-43.

Gong, X. Y.; Li, B. H.; Chai, X. D.; Gu, M. (2014): Survey on big data platform technology. *Journal of System Simulation*, vol. 26, no. 3, pp. 489-496.

Han, J. Q.; Mao, K. B.; Xia, L.; Liu, Q.; Ying, M. A. et al. (2016): Research on agricultural big data based on spatial data warehouse. *Journal of Agricultural Science & Technology*, no. 5, pp. 17-24

Isard, M.; Budiu, M.; Yu, Y.; Birrell, A.; Fetterly, D. (2007): Dryad: distributed data-parallel programs from sequential building blocks. *ACM SIGOPS Operating Systems Review*, vol. 41, no. 3, pp. 59-72.

Keim, D.; Qu, H.; Ma, K. L. (2013): Big-data visualization. *IEEE Computer Graphics & Applications*, vol. 33, no. 4, pp. 20-21.

Li, D. L.; Yang, H. (2018): State-of-the-art review for internet of things in agriculture. *Transactions of the Chinese Society for Agricultural Machinery*, vol. 49, no. 1, pp. 1-20.

Li, G. (2012): Research status and scientific thinking of big data. *Bulletin of Chinese Academy of Sciences*, no. 6, pp. 647-657.

Li, X. L.; Gong, H. G. (2015): A survey on big data systems. *Scientia Sinica Informationis*, vol. 45, no. 1, pp. 1-44.

Li, Y.; Liu, J. (2011): Design and implementation of produce price forecast system. *Hubei Agricultural Sciences*, vol. 50, no. 14, pp. 2976-2978.

Li, Z. Z.; Shen, H. Y.; Walter, L.; Jeffrey, D. (2017): An exploration of designing a hybrid scale-up/out hadoop architecture based on performance measurements. *IEEE Transactions on Parallel\s&\sdistributed Systems*, vol. 28, no. 2, pp. 386-400.

Low, Y.; Bickson, D.; Gonzalez, J.; Guestrin, C.; Kyrola, A. et al. (2012): Distributed graphlab: a framework for machine learning and data mining in the cloud. *Proceedings of the Vldb Endowment*, vol. 5, no. 8, pp. 716-727.

Mckusick, M. K.; Quinlan, S. (2009): GFS: Evolution on Fast-forward. *ACM*, vol. 7, no. 7, pp. 10-20.

Neumeyer, L.; Robbins, B.; Nair, A.; Kesari, A. (2010): S4: Distributed stream computing platform. *IEEE International Conference on Data Mining Workshops*, pp. 170-177.

Trostle, R. (2008): Global agricultural supply and demand: factors contributing to the recent increase in food commodity prices. *Electronic Outlook Report from the Economic Research Service*.

Wang, S.; Liu, P.; Zhang, Z.; Zhang, Y.; Song, C. et al. (2016): Development of management methods for "bohai sea granary" data. *Journal of Chinese Agricultural Mechanization*, vol. 37, no. 3, pp. 270-275.

Xu, S. W.; Wang, D. J.; Li, Z. M. (2015): Application research on big data promote agricultural modernization. *Scientia Agricultura Sinica*, vol. 48, no. 17, pp. 3429-3438.

Yan, X. F.; Zhang, D. X. (2013): Big data research. *Computer Technology & Development*, vol. 23, no. 4, pp. 168-172.

Yang, K.; Wang, Z. (2010): Design and implementation of agricultural prices monitoring and forecasting system. *Agriculture Network Information*, no. 11, pp. 37-39.

Yur'Evich, G. E.; Vasil'Evich, G. V. (2013): Analytical review of data visualization methods in application to big data. *Journal of Electrical & Computer Engineering*, no. 4, pp. 1-7.

Zhang, H. R.; Li, Z. L.; Zou, T. F.; Wei, X. Y.; Yang, G. C. (2014): Overview of agriculture big data research. *Computer Science*, vol. 41, no. S2, pp. 387-392.

Zhang, S. R.; Zheng, W. G.; Shen, C. J.; Xing, Z. (2012): Agricultural product price information collection terminal of embedded portable wireless. *Computer Engineering & Design*, vol. 33, no. 2, pp. 514-518.

Zhang, Y. (2015): Design of data center of rdb platform and big data platform mashup. *Computer Technology & Development*, vol. 25, no. 5, pp. 172-178.

Zhang, Y.; Xiao, L. (2015): Communication-efficient distributed optimization of self-concordant empirical loss. *Mathematics*, pp. 1-46.

Zhao, A. P.; Wang, X. D.; Xiao, J. K. (2014): Design and implementation of agricultural product prices short-term forecasting system. *Agriculture Network Information*, no. 4, pp. 5-10.