A Study of Big Data Application in Agriculture

Jia Chen 1,2,*, Shouwu He 1, Xiaoying Li 1
1 Campus of Nanning, Guilin University of Technology, No.15, Anji Avenue, Nanning, 530001, China
2 Segi University, Faculty of Engineering, Built Environment, and Information Technology, Kota Damansara, Selangor, Malaysia

*9804003@glut.edu.cn

Abstract. Agricultural data is typically characterized by at least 3 characteristics: massive data volume, various data sources, and complex structure. With the popularization of next-generation information technology, precision agriculture has become an emergent development direction in agricultural informatization. This paper proposed a big data application framework for precision agriculture. The model includes data sources, data integration, and data analytics. Use of big data application provides better guidance for precision agricultural.

Keywords: Big Data, Precision Agriculture, Decision Support

1. Introduction

The emergence of next-generation information technology is inseparable from the widespread application of the Internet of Things, cloud computing, and the Artificial Intelligence (AI). These technologies have spawned a large accumulation of data, and we have also entered the era of big data. Big data technology changed the way people think in terms of medical care, transportation, banking and other fields, and it also facilitated agricultural stakeholders to decide how to deal with agricultural information technology. In China, agriculture is a traditional industry with a long history. Since the 1970s, information technology has changed the agriculture, from artificial to intelligent and precise [1].

Precision Agriculture (PA) came to being in 1990s. Its purpose is to use less resource consumption to obtain higher agricultural harvest [2]. Precision agriculture has a wide range of connotations, involving all aspects of natural resources and production processes. Natural resource conditions include land, territory, weather and climate, natural disasters, etc. The production process includes breeding, planting, fertilization, pest control, etc. Nowadays, as sensors have begun to be used extensively, the collection of agricultural data has also been improved from handwritten records or manual digital records to automatic data acquisition. In the whole process of agricultural production, operation and management, sensors are being deployed on a large scale [3]. Sensor data is more accurate, the collection process is faster, and the automated collection process also avoids human error or unexpected termination. Fundamentally, the accumulation of agricultural data has begun to explode.

To implement precision agriculture, how to obtain and accurately process these types of data in time is facing challenges. First of all, agriculture-related organizations have been lack of standardization of agricultural data definition. Second, the existing agricultural information system is generally designed for traditional agricultural service or activity, thus can’t satisfy big data
requirements. Third, the traditional analysis method model cannot be applied to the diversified agricultural production requirements [4]. To promote the production and management of precision agriculture, the methods and tools of agricultural big data application will become the focus in future research.

Big data takes 4v as its main characteristics: volume, speed, variety, and value. The emergence of big data technology provides an effective solution to solve new problems such as data diversity, high data volume, and high speed. This technology is applied to precision agriculture, covering everything from soil conditions, irrigation, and fertilization to real-time monitoring of crops, and extending to all aspects of agricultural production operations and management, providing more comprehensive analysis. Therefore, precision agriculture empowered by big data has become a new direction of agricultural development in the future. Some countries are vigorously supporting the application of big data in the agricultural field. In 2015, the U.S. invested 4.6 billion US dollars to conduct software and big data agricultural application research [5]. In China, the government also formulated relevant policies to promote the function and huge potential of big data in agricultural and rural development.

The purpose of this paper is to design a big data analysis framework for precision agriculture, and to discuss the key aspects of this framework to deal with issues related to agricultural big data.

2. A big data agriculture framework

Precision agriculture is a typical data-driven industry application practice. Through comprehensive analysis of agricultural historical data, geospatial information, sensor network data, and crop knowledge base, a decision support system (DSS) is established to produce accurate quantitative implementation [6]. In the past 50 years, precision agriculture has undergone tremendous development changes. In the early days, mobile communications, map information, weather information, soil fertility and other data were gradually introduced, and information systems made agricultural production more efficient. At present, big data promotes the in-depth aggregation of multi-source data. Through cross-industry and cross-disciplinary data analysis, big data technology is leading precision agriculture into a more in-depth transformation. As a result, agriculture has become more "precise" and "smart".

According to the characteristics of agriculture and the general process of big data processing technology, we decompose the relevant elements of precision agriculture and propose a big data analysis framework for precision agriculture, as shown in Fig.1.

![Fig.1 Big data analytics framework for Precision Agriculture](image-url)
3. Data sources
A large amount of data is required by precision agriculture to serve decision support systems. Usually, data might be structured, semi-structured, and unstructured. Data sources include IoT sensor information, document records, web GIS information, etc.

3.1. Sensor network data
The Internet of Things technology links physical objects in the real world to the digital world. The sensors measure the characteristics of the natural world, quantifies the characterization information and sends it to the information receiving end. For example, sensors deployed in farmland can measure the land's PH value, humidity, precipitation, and temperature in real time. These data records may become the key to system control [7]. The data collected by sensors is the most important part of agricultural big data and the main source of data mining analysis.

Different sensor manufacturers have their respective methods when defining data collection specifications, which may lead to data inconsistencies. Ignoring the inconsistency may lead to errors and mistakes in decision-making. For example, the collected rainfall can be measured by hour or day, and the results obtained are completely different. In order to ensure the consistency of sensor data, we have given a pre-defined demonstration of the specification of data collection. Table 1 shows examples of data collection specifications for rainfall, light, and soil sensors, and defines data types, measurement units, and collection frequency. The air temperature sensor measures the temperature 10 times per hour, in degrees Celsius (°C), stored as a floating-point number format. The soil moisture sensor records in percentage. The specifications defined in this table eliminate data differences caused by different collection strategies.

| Data Acquisition | Data type | Unit | Sample Rate |
|------------------|-----------|------|-------------|
| Rainfall         | Float     | mm   | 10samples/hour |
| Water Evaporation| Float     | mm   | 10samples/hour |
| Air Temperature  | Float     | °C   | 10samples/hour |
| Sunlight Time    | Float     | Hour | 10samples/hour |
| Sunlight Intense | Float     | w/m2 | 10samples/hour |
| Soil moisture    | Float     | %    | 10samples/hour |
| Soil PH          | Float     | N/A  | 10samples/hour |

Tab.1 Date sample rules

3.2. Historical data
The information on agricultural production activities has been accumulated over a long period of time to form historical data. The historical data involves logs of farmland, seeding, fertilization, insecticide, harvesting, storage, breeding, etc., as well as statistics from industries and institutions. Historical data also includes cross-cutting data related to agriculture, some unfinished historical data and books. In the information age, there are many options for collecting agricultural documents. Internet offers videos, audios, pictures, and text on the website. Some authoritative agricultural websites provide a lot of policy information, market information, agricultural technology information, and agricultural economic information.

3.3. Spatial data
Among the many data sources of agricultural big data, spatial data has a special status. Satellite remote sensing data, map data, resource survey data, agricultural zoning data, etc. all contain a large amount of spatial information. In the current actual development of precision agriculture, the application of
GIS technology can effectively achieve the development goals of precision agriculture. GIS helps to organize and manage agricultural spatial data. Using maps as a platform to display agricultural statistical information and real-time data in a visual manner is an important means to improve the readability of big data analysis.

4. Data integration

4.1. Data quality
In actual data analysis procedures, high data quality is essential to success. Errors may occur during data creation, data transfer and manual data entry. Different systems have different methods and formats for collecting and managing data. Some are stored in text format, and some are stored in XML files, table files or databases. Different data formats and standards serve their respective systems and applications, but this creates new problems in the information sharing and exchange phase. In order to solve and avoid such troubles, data standards must be defined first.

Zhiqing L, Bo G, et al (2018) discussed standardized agricultural data standards [8]. As shown in Table 2, the crop information standards for agricultural species include naming, aliases, icons in the information system, suitable land types for planting, planting time, depth, etc. The table provides a reference for establishing a standardized information database. Furthermore, in terms of data exchange, through standardized data interfaces and protocols, exchange and expansion with third-party application systems can be realized.

| Attributes        | Description/Norm                                      |
|-------------------|-------------------------------------------------------|
| Crop Name         | The Latin name of the crop                            |
| Crop Nicknames    | Other names, local names                              |
| Icon              | An icon to identify crop                              |
| Field             | The field where the plant grows on                     |
| Soil Type         | Type of suitable soil                                  |
| Sowing Date       | Best season to sow                                    |
| Sowing Depth      | How deep to sow(in CM)                                |
| Row Spacing       | Distance between rows(in CM)                          |
| Plant Spacing     | Distance between plants(in CM)                        |
| Pest Control      | Common types of pests                                 |
| Fertilizing       | Fertilizers needed                                    |

Tab.2 Crop information

4.2. Data integration
Data integration is to centralize data from different sources and heterogeneous data in order to provide a complete data foundation for data analysis. Usually, extracting data from different databases or data tables is to form a subject data warehouse relating to a certain field.

Big data technology provides a new solution for processing large amounts of heterogeneous data and unstructured data [9]. Hadoop is a distributed system infrastructure of the Apache open source organization. The core component, HDFS (Hadoop Distributed File System), provides unified storage support for massive, structured, and unstructured data.

ETL (Extract-Transform-Load) tools can be used to integrate multiple source data and construct data warehouses to form thematic databases for various purposes [10]. Typical ETL activities include Extraction, Transformation, and Loading of data. Software tools are developed to take data out of a variety of sources, carry out necessary transformation and cleansing, and put into data marts for application analysis. Srividya K. Bansal (2014) designed an ETL framework to achieve data fusion.
and data release [11]. The ETL tools helps dealing with data warehouse analysis, and will be widely used in big data applications.

5. Decision support model
The purpose of the big data system is to provide decision support for precision agriculture big data processing provides new opportunities for intelligent computing. Deep learning, artificial neural networks, machine learning and other algorithms are favored by decision logic. Agricultural production activities change with time and with events. Therefore, fixed decision logic coded for a particular application cannot adapt to changing needs. The actual production process often faces a series of different decision support [12]. Professional agricultural websites contain a lot of agricultural knowledge. Using web crawlers, effective information from these web pages can be obtained and used to build domain knowledge bases. Decision-making models for precision agriculture should then be trained under algorithms like machine learning, image processing, and artificial neural networks.

In order to enable the decision model to meet the fickle requirements of users, we put forward a novel mechanism for decision support model. We first build a module library containing various decision support functions. We also introduce a reconfigurable user intervention interface among the data processing chain. The user intervention interface enables users to select relevant data from data marts, and select the proper decision module from the module library. As Fig.2 shows, to make a fertilizing plan, we form the data marts by extracting crop, weather, soil, pest, and real-time image. In the computation phase, users choose a fertilization plan among many decision modules. The model library contains a set of reasoning and prediction models. All the models have also been well trained under machine learning algorithms. Training data is provided by knowledge bases in plant nutrition science, soil science, and fertilizer science.

6. Conclusion & Future research

6.1. Conclusion
Big data is an emerging trend in agricultural applications. The ultimate goal of establishing big data system for precision agriculture is to promote the optimization of agricultural economy by providing more effective support and tools. Driven by agricultural data, we designed big data analytic framework for precision agriculture. We also discussed the details of the data processing flow. The use of this big data method has certain guiding significance for the integration, mining and use of large amounts of data generated in agricultural production and applications.
6.2. Future research

Our next step is to apply the framework to a real-world example. We are now working on the practice of banana water-fertilizer program in Guangxi, China. This project plans to use web information and historical data to establish a banana water and fertilizer decision model by using neural network algorithms. In this plan, sensors are deployed to collect and monitor the nutrient content in the soil, weather conditions, crop growth, diseases, pests, accordingly to design banana fertilizer and irrigation demand plans. In the actual banana planting process, based on the real-time data from the farmland, our framework generates optimal plans and formulae for irrigation and fertilization.

Further research is needed to investigate how the support decision module is affected by data quality and to provide guidelines for precision agriculture practice.

Acknowledgment

This work was supported by the Guangxi Education Department Basic Ability Promotion Project for Young and Middle-aged Teachers in Colleges and Universities, under No. 2018KY0240 and 2019KY0270.

References

[1] Janssen SJ, Porter CH, Moore AD, Athanasiadis IN, Foster I, Jones JW and Antle JM 2017 Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology Agricultural Systems 155 pp 200-212

[2] Shibusawa S 1998 Precision farming and terra-mechanics Fifth ISTVS Asia-Pacific regional conf. pp 22

[3] Goap A, Sharma D, Shukla AK and Krishna CR 2018 An IoT based smart irrigation management system using Machine learning and open source technologies Computers and electronics in agriculture 155 pp 41-49

[4] Luoisaburwood T 2016 Agriculture Technology Investment Storms to $4.6bn in 2015 as Global Investors Take Note. AgFunderNews (Retrieved Jul 1, 2019) from https://agfundernews.com/agriculture-technology-investment-storms-to-4-6bn-in-2015-as-global-investors-take-note 5380.html

[5] Tan L 2016 Cloud-based decision support and automation for precision agriculture in orchards IFAC-PapersOnLine 49(16) pp 330-335

[6] Kendall H, Naughton P, Clark B, Taylor J, Li Z, Zhao C and Frewer L J 2017 Precision Agriculture in China: Exploring Awareness, Understanding, Attitudes and Perceptions of Agricultural Experts and End-Users in China Advances in Animal Biosciences 8(2) pp703-707

[7] Mohanraj I, Ashokumar K and Naren J 2016 Field Monitoring and Automation Using IOT in Agriculture Domain Procedia Computer Science 93 pp 931-939

[8] Zhiqing L, Bo G, Pingting C, Yaling W and Xingcheng C 2018 A study of Tri-Agri cloud platform design Jiangsu Agricultural Science 46(17) pp 209-212

[9] Yuanzhu Z, Huatan K, Yina L and Ying Kei T 2018 Unlocking the power of big data in new product development. Annals of Operations Research 270(1-2) pp 577-595

[10] Vassiliadis P, Simitsis A and Skiadopoulos S 2002 Conceptual modeling for ETL processes Proc. of the 5th ACM int. workshop on Data Warehousing and OLAP pp 14-21

[11] Bansal SK 2014 Towards a semantic extract-transform-load (ETL) framework for big data integration 2014 IEEE International Congress on Big Data pp 522-529

[12] Kamilaris A, Kartakoullis A and Prenafeta-Boldú FX 2017 A review on the practice of big data analysis in agriculture Computers and Electronics in Agriculture 143 pp 23-37