Big Data Analytics in Precision Agriculture: A Survey

Bharath G, Anala M R

Department of CSE, R V College of Engineering, Bengaluru, Karnataka, India

***Abstract*— Farming is undergoing a digital revolution. The result of precision agriculture is the accumulation of large sets of farm data, which must be managed efficiently for use. This paper presents the approach and methodology of precision agriculture in collecting, managing and analysing large data sets using big data analysis tools that contribute to the productivity. A Cloud based IoT and Big Data solution to precision agriculture is discussed that implements M2M architecture to provide end to end solution. Finally we focus on the problems faced by agriculturists in India to carry out modern agricultural practices and also present a survey of new amendments that have been done in the country.**

Keywords: **Big data, database, private data, public data, digital revolution, precision agriculture, M2M Cloud IoT.**

I. INTRODUCTION

The world has witnessed the continuously increasing use of technology in agriculture since the 19th century, when the industry began to plough fields with tractors. It was during this period, the agricultural revolution made a remarkable impact on the productivity as a consequence of vast improvements in technology. The agricultural sector transformed into a much technologically intense and data rich industry with the advent of biotechnology, plant genetics, chemical inputs and guidance systems. Farmers, in developed countries generate and capture huge agricultural data using mobile technology, which are stored and later retrieved by application softwares with the help of the database management softwares. Information is a fundamental and an essential element of any activity in agricultural sector. Information and communication technology (ICT) is extensively utilised in farming to provide indispensible information at right time and at a least expense.

In traditional agriculture, a schedule is predetermined considering factors such as rainfall, suitable weather, etc, and all tasks are performed in order accordingly. Despite their efforts, farmers face difficulty in making proper decisions due to lack of essential information at appropriate time. Thus, it is required to collect real time data on weather, air quality, and soil fertility and even on availability of labour and essentially the capital investment in total and analytical predictions, to make ingenious decisions. This way of farming is known as Precision agriculture.

Big Data has emerged big in integrating various industries across the world among which agriculture too is a part. Agriculture and other agri-businesses require innumerable decisions to be taken based on various influencing factors that are stranded with intricate problems. Big Data is now a driving factor for progress in precision agriculture upon which farmers are relying, with the expectation of maximum yields.

This paper is organised as follows: Section II and section III present the detailed information on Big Data and its application in agriculture while section IV presents the M2M technology achieved using IoT and Cloud Big Data processing. Finally sections V and VI contain a literature survey on use of Big Data applications presented with case studies.

II. BIG DATA – CLASSIFICATION AND CHARACTERISATION

The concept of "Big Data" can be described as large volume of data sets which may be structured, semi-structured or unstructured, characterised based on the degree of organisation. The size of data set generally ranges from tens of terabytes to hundreds of petabytes.

a. Structured data:

Structured data is defined as any set of unprocessed and highly organised data which is stored as a fixed field within a record or file. Relational databases and spreadsheets that organise data into tables are examples of structured data.

b. Unstructured data:

Unstructured data includes all those datasets that cannot be instantly categorised and fit into a relational table, for example, clicked photos and graphics, videos, streaming instrument data, web pages, portable document files, PowerPoint presentations, electronic-mails, entries of a blog, and word processing documents fall into this category of big data.

c. Semi-structured data:

Semi-structured data have some organisational properties such as tags and markers to provide hierarchical structure for records and fields within data. Such type of data does not correspond to regular structures of data model in relational database. Examples are XML documents and NoSQL databases.

Big data contains both structured and unstructured data, however, International Data Corporation (IDC) has estimated that 90% of the big data sets are unstructured. Softwares like Hadoop, Hive, Cassendra, etc, can process databases of both unstructured and structured data that are extremely large, very complex and dynamic.

Characteristics of Big Data:

Big data is well understood by knowing its characteristics which are known as 3V's; volume, variety and velocity.

1. **Volume** -- It refers to the enormous amount of data generated at every second. It is estimated that about 2.3 trillion gigabytes of data is created every day [1].
2. **Variety**— Data can be in different formats and from different sources. Unlike earlier days, data is now more unstructured, data may be in forms such as photos, sensor data, encrypted packets and so on. Hence it requires advanced technology and tools to store and analyse raw big data.
3. **Velocity**— The speed with which data is collected, stored, analysed and distributed to end users defines the quality of big data management.

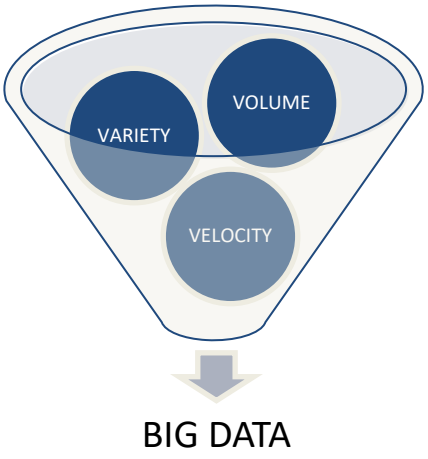

Figure 1: Three characteristics of Big Data

There is one more attribute of big data referred to as Veracity which means uncertainty of data. This character speaks about the extent of accuracy of the collected data.

III. AGRICULTURAL BIG DATA AND BENEFITS

The agricultural data can be classified as private data and public data.

Private Big Data: This data set contains data obtained at the production level and generated by an individual farmer. It mainly includes information regarding ones farmer's field, soil type, irrigation level, yield, livestock, etc, [2].

Public Big Data: At public level, there are funded agencies which collect, maintain and analyse data records. The records may contain data about weather conditions, soil survey, farm program participant records, marketing, etc,[2].

Stages in Big Data process are collection of data, managing aggregated data, and effective usage of processed data

Data Collection: Big data in agriculture involves digital records of farm data. This includes soil moisture content, temperature, weather variability pattern, irrigation facility, financial assistance like insurance and loans schemes, humidity level data, nutrient content data, historical cultivation data of the field and also knowledgeable articles on agriculture written by researchers and innovative agriculture practitioners. Global Positioning System (GPS) and Geographic Information System (GIS) have enabled to quantify the spatial variability within fields. GPS allows collection of geo referenced data and GIS makes spatial analysis and visualisation of interpolated maps [3]. Data is collected via various sensors such as soil moisture sensors, microphone sensor to detect pest using sound detecting technology and detection algorithms, chemical/gaseous sensors to measure gaseous emission from fields (like during ripening of fruits, flower pollination, etc,.) and ultrasonic sensors to detect underground water availability for irrigation [4].

Data Management and Analytics: The primary aim of data management is to make sure that end users get high quality data with easy access. Here are few tools to manage and analyse big data generated from agricultural sector:

1. Hadoop: It serves as a core platform to structure big data. It is a Java platformed programming framework and hence supports processing of large data sets in distributed computing environment [1].
2. MapReduce: MapReduce is programming paradigm supported by Hadoop to handle parallel processing of big data distributed across a large number of computers [1].
3. HDFS: Hadoop Distributed File System, a high performance data access tool used across Hadoop clusters is the primary storage system used by Hadoop applications [1].
4. HBase: It is a column oriented data base management system which runs on top of Hadoop. It is well suited for sparse data sets [1].

Data Use: The end users of the private as well as public big data are farmers and ranchers themselves. The resulting information of big data analysis can be helpful in two ways, as a direct tool contributing to the agricultural productivity and as an indirect tool to aid farmers in making well informed and qualitative on farm decisions. The high resolution geographical maps generated can be used to identify variability of soil in the terrain and the quality measurement values help to decide the type of crop to cultivate based on the nutrient richness of the soil. Historical data on weather pattern are vital in planning different phases of the agricultural activity, especially in places where climatic conditions are unpredictable. Big data applications can be further improvised and enhanced to procure more accurate results and can be made more economical which will entirely transform the sector by appreciating research activity.

Benefits of Private Big Data:

1. Production benefits [2] – Farmers achieve maximum yields by investing less on farms inputs and are not subject to high risk of incurring loss.

2. Environmental benefits [2] - Benefits of farm environment includes improved soil quality, water availability. Reduced inputs, pesticides, fertilisers, water and energy often result in off farm environmental benefits.

3. Business [2] - The rise in the use of big data and technological advances are making way for the expansion of business opportunities and establishments of new ones around agriculture.

Benefits of Public Big Data:

1. Data Management Authority [2] – Statutes and government organisations govern the public data and set standards for quality. Statistically, pubic data is found more reliable than various private data because of the common standards set. Hence, they are viewed trusted and authoritative.
2. Privacy and Security [2] – Public data is kept confidential by the managing authorities. Often agencies anonymise data to protect individual's identity.
3. Equal Access [2] – Technology along with collecting, storing and analysing data presents transparency. The access to internet and open data allows equal use of public data.

IV. M2M MODEL OF CLOUD IOT BIG DATA PROCESSING

Internet of Things is seen as the ultimate futuristic technology to connect intelligent devices located at different remote locations that reduce human efforts in many areas, where agriculture is not an exception. Research in progress is advancing the Machine to Machine (M2M) epitome of communication towards cloud computing and Internet of things by creating innovative platforms that are scalable and enable smart, virtualised and secured services [5].

RTUs (Remote Telemetry Units) are used to transmit vital farm parameters such as temperature, precipitation, wind speed and leaf wetness, as measured by smart sensors, with the help of IoT and its sub domains. IoT cloud provides a platform to process the Big Data thus obtained, very conveniently and allows to be visualised as web applications. The platform can provide forecasts in detail, alerts that require instant action and notifications of diseases, along with treatment recommendations [6].

To implement Cloud IoT architecture, a distributed operating system, SlapOS (Simple Language for Accounting and Provisioning) [7], which is based on master-slave architecture, is being used. SlapOS provides IaaS, PaaS and SaaS for various cloud applications.

Processing of big data, which includes RDBMS complemented by other DMS such as NoSQL and systems based on query and searching, can be performed by cloud softwares, for example Exaleed [8] is considered as one of the leading search based application platform for business and government [9]. Exalead CloudView is capable of transforming data in any format collected from any source into structured and informative blocks of data for businesses that can be easily searched and queried by innovative, information access intensive applications.

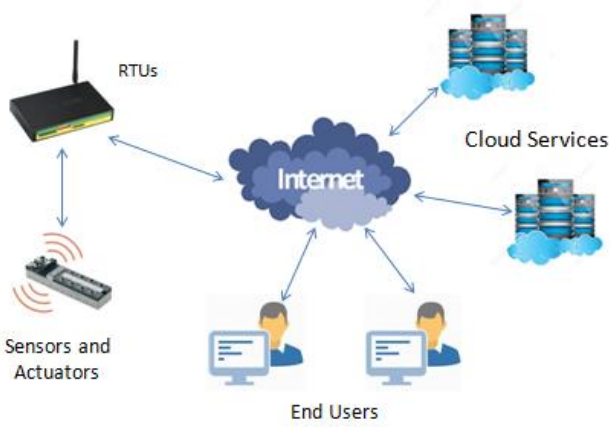

Fig-2: General Architecture of M2M Model

A monitoring station (RTU with sensors) will be situated at the middle of the field to measure different farm parameters.

Plant disease detection is one major application, where the system measures the heat accumulated during the culture or the pressure index and recommend a proper disease management process through big data processing carried out on cloud platform. A similar study was conducted in Grape vine field in Romania [9], where pressure index values of a disease called Powdery mildew that was a serious threat throughout the season, were measured in an attempt to test the appropriateness of the solution provided by the system at different levels of the pressure index found. It was observed that the system efficiently issued first treatment recommendation in the early stage of the disease. Such a disease management can be as well performed on all types of crops to take care of them at right time.

V. LITERATURE SURVEY: BIG DATA ANALYTICS IN U.S AGRICULTURE

Annually, farmers in US generate about $375 billion from crops [10]. Almost all new farm equipments are equipped with sensors using which terabytes of data is collected in each stage from ploughing to harvesting and marketing their produce. Agriculturists say that 80% of their farm data reside on tractors and it is reported that they have witnessed 15% reduction of cost on inputs [10]. Adoption of precision agriculture has increased yields by 13% [10]. The extent of using modern technology in US has reached a significant level that farmers choose whether to use data themselves, share it locally or upload it to the cloud.

John Deere, a leading manufacturer of agricultural products and tools in US, has always been a pioneering industry to introduce advanced technology in farming machinery. The eponymous founder of the esteemed company, John Deere, in

the early 19th century personally designed steel ploughs and hence transmuted the lives of the itinerant people to settle in a place by earning their livelihood through agriculture. It has initiated several service oriented programs to avail farmers the best use of crowd sourced data and real time monitoring of huge data pooled over cloud platforms by thousands of farmer across the continent or globe [10].

Myjohndeere.com is a web portal for authenticated farmers to collect and analyse the data collected from their own machinery from anywhere. It also allows sharing and using the aggregated data from other users around the world [10].

Emerging companies such as Farmers Business Network count Google Ventures as an investor and have made collecting, gathering, and analyzing data from farms across nations as their primary business [10].

Though, agriculture is a low margin business, effective big data management has made a significant impact on profitability in US.

VI. PRECISION AGRICULTURE AND BIG DATA MANAGEMENT IN INDIA

Agriculture sector is the backbone of Indian economy. In spite of it, its contribution to the country's total GDP is only about 17.9% as declared by the Planning Commission of India, in 2014. Around 70% of the rural households rely on agriculture for their living. India stands second in the world as a producer of agricultural products and has marked 7.68% of the total global agricultural output.

Water is very much essential in agriculture. Though India is the most irrigated country after China, only one-third of the cropped area is under irrigation. India is in the tropical region, hence has a very uncertain and unreliable monsoons with which we cannot expect sustained progress in agriculture. Agricultural lands have been used for cultivation for more than thousand years that has led to the depletion of soil nutrients and hence the fertility. Manures and fertilisers play an important role in nourishing soil but farmers lack the knowledge on how to use chemical fertilisers and in what amount it is required to replenish the soil fertility. Another problem in most of the developing countries, including India is the lack of use of newer tools and mechanism. Even today, a significant number of farmers use traditional and simple hand held tools like wooden ploughs driven by ox, sickles, etc,. The major concern in India is the inadequate help provided by the government to face challenges and risks.

Big data and its applications could become a single solution to 3/4th of the existing problems. With this perspective in mind, in June 2016, ICAR - National Academy of Agricultural Research Management (NAARM) had organised a ten-day training programme with an objective of training the participants in Big Data Analytics and application in Agriculture, in Hyderabad [11]. The Indian Council of Agricultural Research (ICAR) is a sovereign organisation controlled by the Department of Agricultural Research and Education (DARE), under the Ministry of Agriculture and Farmers Welfare. Earlier it was known as Imperial Council of Agricultural Research, when it was formed in 1929 as a society registered under the act of Societies Registration, 1860 [11]. This very initiative is a head start to implement Big Data applications in agriculture in India.

Precision agriculture was introduced in India by Tata Groups as Tata Kisan Kendra (TKK) in the first place. The project of precision farming by TTK aims at delivering end to end solution to farmers [12]. TKK plays a vital role in integrating the modern farming techniques and traditional practices to increase productivity, thus enabling the farmers to obtain proper market value for their produce. Figure-2 shows the practices followed by TTK.

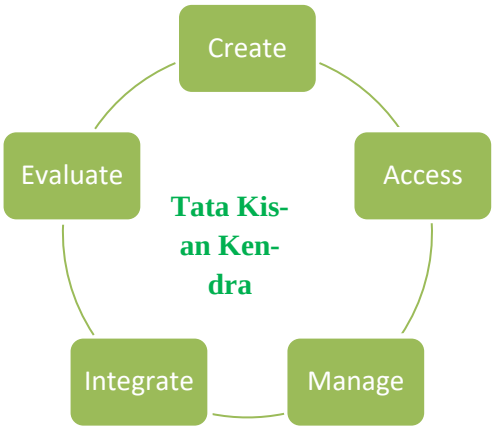

Figure-3: TTK's Precision Farming Practices

TKK conducts workshops on regular basis to provide suggestions to farmers and also evaluates their performance which is critical for analysing the methods adopted. Most importantly TTK welcomes intellects and promotes research and development to invent much powerful and productive tools and innovative ideas [12]. Application of Big data technology in agriculture can be extended to a greater distance by the collaborative assistance of IT industries and support by the government.

VII. CONCLUSION

In this paper we have made an attempt to bring out the significance of the role played by big data analytics in precision agriculture which has radically changed the field of agriculture.

The paper discuses the development and use of M2M remote telemetry as an entailment of Big Data processing on Cloud infrastructure which could be a promising solution for numerous problems, such as disease prevention and control as described, faced during crop production.

This paper briefly describes the current state of research on big data applications in agriculture and the present scenario of agricultural sector in India.

As our future work we foresee to implement the proposed model and big data processing to add an aid to agriculture in India.

REFERENCES

[1]. www.datasciencecentral.com

[2]. Megan Stubbs. “Big Data in U.S. Agriculture”. Congressional Research Service, pp. 2-8. January 6, 2016

[3]. S.Shibusawa and C.Haché. “Data Collection and Analysis methods for Data from Field Experiments”. Encyclopedia of Life Support Systems. ISBN : 978-1-84826-583-7

[4]. Chandra Sukanya Nandyala and Haeng-Kon Kim. “Big and Meta Data Management for U-Agriculture Mobile Services”. International Journal of Software Engineering and Its Applications. Vol. 10, No. 2 (2016), pp. 257-270

[5]. F. Bonomi, R. Milito, J. Zhu, and S. Addepalli. "Fog computing and its role in the internet of things." In Proceedings of the first edition of the MCC workshop on Mobile cloud computing, pp. 13-16. ACM, 2012.

[6]. G. Suciu, S. Halunga, O. Fratu, A. Vasilescu, and V. Suciu. "Study for renewable energy telemetry using a decentralized cloud M2M system." In IEEE Wireless Personal Multimedia Communications (WPMC), 16th International Symposium on, pp. 1-5, 2013

[7]. W. Saad, H. Abbes, M. Jemni, and C. Cérin. "Designing and implementing a cloud-hosted saas for data movement and sharing with slapos." International Journal of Big Data Intelligence 1, no. 1, pp. 18- 35, 2014.

[8]. EXALEAD – Dassault Systemes, www.3ds.com/productsservices/Exalead

[9]. George Suciu, Alexandru Vulpe, Octavian Fratu. “M2M Remote Telemetry and Cloud IoT Big Data Processing in Viticulture”. R mania International Symposium on, pp. 1-5, 2013

[10]. www.businessinsider.com/big-data-and-farming

[11]. www.icar.org.in

[12]. Vivek Talwar, Nrupesh Mastakar and B. Bowonder. “ICT Platform for Enhancing Agricultural Productivity: The case study of Tata Kisan Kendra”, International Journal of Services Technology and Management (IJSTM), Vol. 6, No. 3/4/5, pp. 1-15,2005