Computers and Electronics in Agriculture 143 (2017) 23–37

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Review

A review on the practice of big data analysis in agriculture

Andreas Kamilaris*, Andreas Kartakoullis, Francesc X. Prenafeta-Boldú

GIRO Joint Research Unit IRTA-UPC, Barcelona, Spain

ARTICLE INFO

Keywords:
Big data analysis
Agriculture
Survey
Smart Farming

ABSTRACT

To tackle the increasing challenges of agricultural production, the complex agricultural ecosystems need to be better understood. This can happen by means of modern digital technologies that monitor continuously the physical environment, producing large quantities of data in an unprecedented pace. The analysis of this (big) data would enable farmers and companies to extract value from it, improving their productivity. Although big data analysis is leading to advances in various industries, it has not yet been widely applied in agriculture. The objective of this paper is to perform a review on current studies and research works in agriculture which employ the recent practice of big data analysis, in order to solve various relevant problems. Thirty-four different studies are presented, examining the problem they address, the proposed solution, tools, algorithms and data used, nature and dimensions of big data employed, scale of use as well as overall impact. Concluding, our review highlights the large opportunities of big data analysis in agriculture towards smarter farming, showing that the availability of hardware and software, techniques and methods for big data analysis, as well as the increasing openness of big data sources, shall encourage more academic research, public sector initiatives and business ventures in the agricultural sector. This practice is still at an early development stage and many barriers need to be overcome.

1. Introduction

Population growth, along with socioeconomic factors have historically been associated to food shortage (Slavin, 2016). In the last 50 years, the world's population has grown from three billion to more than six, creating a high demand for food (Kitzes et al., 2008). As the (Food and Agriculture Organization of the United Nations, 2009) estimates, the global population would increase by more than 30% until 2050, which means that a 70% increase on food production must be achieved. Land degradation and water contamination, climate change, sociocultural development (e.g. dietary preference of meat protein), governmental policies and market fluctuations add uncertainties to food security (Gebbers and Adamchuk, 2010), defined as access to sufficient, safe and nutritious food by all people on the planet. These uncertainties challenge agriculture to improve productivity, lowering at the same time its environmental footprint, which currently accounts for the 20% of the anthropogenic Greenhouses Gas (GHG) emissions (Sayer and Cassman, 2013).

To satisfy these increasing demands, several studies and initiatives have been launched since the 1990s. Advancements in crop growth modeling and yield monitoring (Basso et al., 2001), together with global navigation satellite systems (e.g. GPS) (Aqeel ur et al., 2014) have enabled precise localization of point measurements in the field, so that spatial variability maps can be created (Pierce and N., 1999), a concept known as "*precision agriculture*" (Bell et al., 1995).

Nowadays, agricultural practices are being supported by biotechnology (Rahman et al., 2013) and emerging digital technologies such as remote sensing (Bastiaanssen et al., 2000), cloud computing (Hashem et al., 2015) and Internet of Things (IoT) (Weber and Weber, 2010), leading to the notion of "*smart farming*" (Tyagi, 2016; Babinet Gilles et al., 2015). The deployment of new information and communication technologies (ICT) for field-level crop/farm management extend the precision agriculture concept (Lokers et al., 2016), enhancing the existing tasks of management and decision making by context (Kamilaris et al., 2016), situation and location awareness (Karmas et al., 2016).

Smart farming is important for tackling the challenges of agricultural production in terms of productivity, environmental impact, food security and sustainability. Sustainable agriculture (Senanayake, 1991) is very relevant and directly linked to smart farming (Bongiovanni and Lowenberg-DeBoer, 2004), as it enhances the environmental quality and resource base in which agriculture depends, providing basic human food needs (Pretty, 2008). It can be understood as an ecosystem-based approach to agriculture, which integrates biological, chemical, physical, ecological, economic and social sciences in a comprehensive way, in order to develop safe smart farming practices

* Corresponding author.
E-mail addresses: andreas.kamilaris@irta.cat (A. Kamilaris), andreas.kartakoullis@irta.cat (A. Kartakoullis), francesc.prenafeta@irta.cat (F.X. Prenafeta-Boldú).

http://dx.doi.org/10.1016/j.compag.2017.09.037
Received 24 January 2017; Received in revised form 11 September 2017; Accepted 28 September 2017
0168-1699/

that do not degrade our environment.

To address the challenges of smart farming and sustainable agriculture, as (McQueen et al., 1995) and (Gebbers and Adamchuk, 2010) point out, the complex, multivariate and unpredictable agricultural ecosystems need to be better analyzed and understood. The aforementioned emerging digital technologies contribute to this understanding by monitoring and measuring continuously various aspects of the physical environment (Sonka, 2016), producing large quantities of data in an unprecedented pace (Chi et al., 2016). This implies, as (Hashem et al., 2015) note, the need for large-scale collection, storage, pre-processing, modeling and analysis of huge amounts of data coming from various heterogeneous sources.

Agricultural "*big data*" creates the necessity for large investments in infrastructures for data storage and processing (Nandyala and Kim, 2016; Hashem et al., 2015), which need to operate almost in real-time for some applications (e.g. weather forecasting, monitoring for crops' pests and animals' diseases). Hence, "*big data analysis*" is the term used to describe a new generation of practices (Kempenaar et al., 2016; Sonka, 2016), designed so that farmers and related organizations can extract economic value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis (Waga and Rabah, 2014; Lokers et al., 2016).

Big data analysis is successfully being used in various industries, such as banking, insurance, online user behavior understanding and personalization, as well as in environmental studies (Waga and Rabah, 2014; Cooper et al., 2013). As (Kim et al., 2014) show, governmental organizations use big data analysis to enhance their ability to serve their citizens addressing national challenges related to economy, health care, job creation, natural disasters and terrorism.

Although big data analysis seems to be successful and popular in many domains, it started being applied to agriculture only recently (Lokers et al., 2016), when stakeholders started to perceive its potential benefits (Bunge, 2014; Sonka, 2016). According to some of the largest agricultural corporations, tailoring advice to farmers based on analyzing big data could increase annual global profits from crops by about US $20 billion (Bunge, 2014).

The motivation for preparing this survey stems from the fact that big data analysis in agriculture is a modern technique with growing popularity, while recent advancements and applications of big data in other domains indicate its large potential (Kim et al., 2014; Cooper et al., 2013). Current relevant surveys (Wolfert et al., 2017; Nandyala and Kim, 2016; Waga and Rabah, 2014; Wu et al., 2016) cover mostly theoretical aspects of this technique (e.g. conceptual framework, socioeconomics, business processes, stakeholders' network) or focus on particular sub-domains such as remote sensing (Chi et al., 2016; Liaghat and Balasundram, 2010; Teke et al., 2013; Ozdogan et al., 2010; Karmas et al., 2014) and geospatial analysis (Karmas et al., 2016). Thus, the main contribution of this survey is that it presents a more focused overview of the particular problems encountered in agriculture, compared to existing surveys, where data analysis is a key aspect and solutions are found inside the big data realm. Our survey highlights the (big) data used, the methods and techniques employed, giving specific insights from a technical perspective on the potential and opportunities of big data analysis, open issues, barriers and ways to overcome them.

2. Methodology

The bibliographic analysis in the domain under study involved three steps: (a) collection of related work, (b) filtering of relevant work, and (c) detailed review and analysis of state of the art related work. In the first step, a keyword-based search for conference papers and articles was performed from the scientific databases IEEE Xplore and ScienceDirect, as well as from the web scientific indexing services Web of Science (Thomson Reuters, 2017) and Google Scholar. As search keywords, we used the following query:

"Big Data" AND ["Precision Agriculture" OR "Smart Farming" OR "Agriculture"]

In this way, we filtered out papers referring to big data but not applied to the agricultural domain. Existing surveys (Wolfert et al., 2017; Nandyala and Kim, 2016; Waga and Rabah, 2014; (Chi et al., 2016; Wu et al., 2016) were also examined for related work. From this effort, 1330 papers were initially identified. Restricting the search for papers in English only, with at least two citations, the initial number of papers was reduced to 232. Number of citations was recorded based on Google Scholar. An exception was made for papers published in 2016–2017, where zero citations were acceptable.

In the second step, we checked these 232 papers whether they *made actual use of big data analysis* in some agricultural application. Use of big data analysis was quantified as satisfying some of its five "V" characteristics (Chi et al., 2016) (see Section 3). We primarily targeted the first three "V"s (i.e. volume, velocity and variety), since dimensions V4 and V5 (i.e. veracity and valorization) were more difficult to quantify. From the 232 papers, only 34 qualified according to our constraints. We were forced to discard (also) a small number of interesting efforts which did not qualify in terms of the data analysis employed and the solutions provided. In the final step, the 34 papers selected from the previous step were analyzed one-by-one, considering the problem they addressed, solution proposed, impact achieved (if measurable), tools, systems and algorithms used, sources of data employed and which "V" dimensions of big data they satisfied.

3. Big data in agriculture

Chi et al. (2016) characterize big data according to the following five dimensions:

- **Volume (V1)**: The size of data collected for analysis.
- **Velocity (V2)**: The time window in which data is useful and relevant. For example, some data should be analyzed in a reasonable time to achieve a given task, e.g. to identify pests (PEAT UG, 2016) and animal diseases (Chedad et al., 2001).
- **Variety (V3)**: Multi-source (e.g. images, videos, remote and field-based sensing data), multi-temporal (e.g. collected on different dates/times), and multi-resolution (e.g. different spatial resolution images) as well as data having different formats, from various sources and disciplines, and from several application domains.
- **Veracity (V4)**: The quality, reliability and potential of the data, as well as its accuracy, reliability and overall confidence.
- **Valorization (V5)**: The ability to propagate knowledge, appreciation and innovation.

Although these five "V"s can describe big data, big data analysis does not need to satisfy all five dimensions (Rodriguez et al., 2017). Big data is generally notorious for being less accurate and stable, usually compromising V4 (veracity). Another relevant "V" could be **visualization**, meaning the need of presenting complex data structures and rich information in an easy-to-understand way (Hashem et al., 2015; Karmas et al., 2016).

According to the above, as (Wolfert et al., 2017) explain, big data is less a matter of data volume than the capacity to search, aggregate, visualize and cross-reference large datasets in reasonable time. It is about the capability to extract information and insights where previously it was economically or technically not feasible to do so (Sonka, 2016).

In the following subsections, the most relevant research efforts, case studies and techniques in terms of solving agricultural problems by using big data analysis are discussed (Section 3.1), together with sources of big data (Section 3.2) and specific techniques employed for big data analysis (Section 3.3).