Data Integration for Precision Agriculture - Challenges and Opportunities for the Database community

Luiz Henrique Zambom Santana

Leaf Agriculture, Brazil
luiz@withleaf.io

Abstract. The last years the precision agriculture transformed one of the most ancient activities into a humongous source of data. This can happen by means of sensors that monitor continuously the physical environment (e.g., satellite imagery, high technology machinery, micro weather stations) producing large quantities of data in an unprecedented pace. Although there are many papers describing how to use this data (e.g., in modern Big Data systems, as the input of Machine Learning pipelines), today this is a virtually impossible task without a huge effort conciliation and integration. There are many research opportunities that emerge from this scenario, for instance data accessibility through integration methods, new tools (e.g., visualization, ETL tools), and novel datasets and benchmarks. This is specially interesting in the Brazilian context, our country have more than 800 thousand of hectares of arable land and the agribusiness represents almost 30% of our Gross Domestic Product (GDP). This paper presents the experience of four years of working at Leaf Agriculture, the goal is to list the challenges and opportunities for data integration in the precision agriculture.

Keywords: Data Integration, Precision Agriculture

1 Introduction

Precision agriculture comprises a set of technologies that combines sensors, information systems, enhanced machinery, and informed management to optimize production by accounting for variability and uncertainties within agricultural systems [5]. It provides means to monitor the food production chain and manage both the quantity and quality of agricultural produce.

The evolution of precision agriculture led to the development of many siloed datasets. Although a big part of agribusiness happens in a limited space - inside the farms -, the data is collected by highly specialized systems which leads to many problems on the data integration. This article summarizes the experience collected over the last four years of development inside Leaf Agriculture[1]. This

[1] https://withleaf.io

topic is very important as it tackles economic and social aspects, as well as food safety. Due to it, it is a relevant research issue for the Database community.

The rest of the paper is organized as follows. Section 2 details the Precision agriculture and the state of the art in and data integration for this field. Section 3 concludes the paper with challenges and opportunities of this area.

2 Precision agriculture and data integration

Precision agriculture is the application of technologies and principles to manage spatial and temporal variability associated with all aspects of agricultural production for the purpose of improving crop performance and environmental quality [10]. Precision agriculture generally involves better management of farm inputs such as fertilizers, herbicides, seed, fuel by doing the right management practice at the right place and the right time. The precision agriculture systems deal with different data natures [9], for instance:

- Field data: it comprises polygons and points, the field data includes the boundaries of the space where something is planted and the point data that describes the operations (e.g., planting, harvesting, tillage) that were developed in this area.
- Machinery: it includes tractors, implements, and other vehicles used in the farm activities. Many times these vehicles are autonomous or semi-autonomous, oriented by GPS, and contain dozens of sensors [4]. The main problem with this data is creating a universal identifier that permits the farmer to relate the field operation data with the machinery used in these operations. Also, the maintenance of these machines and the fuel costs are important and must be closely tracked.
- Soil: this type of data is formed by points that describe the soil analysis on properties like calcium and phosphate [1]. It also includes governmental databases (e.g., BDSolos[2]) that describe more static properties of the soil, such as the soil depth and type.
- Imagery: the images returned from satellites (e.g., Sentinel-2[3], Landsat-9[4], Planet[5]) and the images acquired from unmanned aerial vehicle (UAV) [7]. The traditional scouting requires surveys performed by humans in the fields, which is costly, slow, and sparse. These images help the farmers to rapidly identify problems and generate health indexes (e.g., NDVI, EVI[6]) to track their crops.
- Resources identification: having a unique identifier for chemicals, seeds, and crops is also a challenge in this case. These resources vary a lot, for instance, completely different chemicals can have similar market names, the

[2] https://www.bdsolos.cnptia.embrapa.br/(last access 27-05-2022)
[3] https://sentinel.esa.int/web/sentinel/missions/sentinel-2(last access 27-05-2022)
[4] https://landsat.gsfc.nasa.gov/satellites/landsat-9/(last access 27-05-2022)
[5] https://www.planet.com/(last access 27-05-2022)
[6] https://www.usgs.gov/landsat-missions/landsat-enhanced-vegetation-index(last access 27-05-2022)

crop names vary in different languages, the code that represents a seed can easily wrongly inputted by a machine operator.

- Irrigation: water saving is a real challenge, specially in parts of the world like Israel or the northeast of Brazil. However having good data available is specially important [6]. There are some important free datasets available[7], but the access and the model of the data is not common.
- FMIS: For a farmer the data types mentioned before is normally serviced by a FMIS (Farm Management Information System) [3]. It usually consist of a map where the agronomic fields are displayed and a management area where fiscal, economic and planning tasks can be executed.

The precision agriculture data integration was subject of studies in many attempts over the years. Both industry and academy are chasing an unique form to the myriad of data available. Among the industry standards, the ADAPT framework[8] is used by many machinery manufacturers to register their operation data. The main problem of this framework is that it is very complex to be used and developed in C# with proprietary libraries. The standard ISO 11783[9] (also called ISOBUS) is a broad standard for machinery information, but the main focus is how the electronic components will operate instead of how humans and computers can extract knowledge and insights of the data in this format. Another example of standard that focus only one vertical of the agriculture, Modus[10] is a XML-based standard to represent soil data. The research community also introduced some attempts of creating data integration. The AGDATABOX [2] is an API that aims to facilitate the input and output of the agronomic data. The idea is to create a unique source of true for this data. Similarly, U-agro [8] propose a unified architecture for farm data, which try to solve the data integration by unifying the source of data in the farm. The main limitation of AGDATABOX and U-agro is to ignore the data already available in the farms.

3 Challenges and opportunities

Agriculture is ultimately an ancient science that's been iterated upon for generations. In the last decades, precision agriculture brought many advances, specially by permitting farm data to be collected. However, the next generation of agricultural systems will need to integrate this data to fully explore its potential. This task presents many challenges and opportunities for database community.

The challenges that are currently faced by the precision agriculture applications are derived not only from the challenging environment where the data collected in farms, but also the following pure computational and research issues can be listed: lack of standards, few professionals (e.g., programmers, DBAs, data scientists) with domain knowledge, huge datasets, different versions of the

[7] https://data.world/datasets/irrigation(last access 27-05-2022)
[8] https://adaptframework.org/(last access 27-05-2022)
[9] https://www.iso.org/standard/57556.html(last access 27-05-2022)
[10] https://www.ispag.org/proceedings/ (last access 27-05-2022)

same data (e.g., many imagery resolutions, micro weather stations in the field competing with climate satellite data), no benchmarks or datasets for development, poorly documented frameworks, intrinsically GIS datasets, and multidimensional data and schema evolution over time [11]. The opportunities are great and includes virtually all areas of database research. For instance, data integration and visualization could bring tailored results for agriculture family farming in the south of Brazil. Employ machine learning and Big Data techniques would permit to predict, estimate, sample or suggest actions for grain growing in the Brazilian center east. Data lakes can store forest imagery forest data for future sustainable analysis. Data mining would be used to find patterns that makes a smarter water usage, specially in Brazilian northwestern region. Blockchain could bring better contracts and more efficiency to agrobusiness as a whole.

In author's opinion - specially in Brazil, where the agriculture is represents a big part of our economy and sustainability is a very important topic to our future - the database community could play an important role by creating the tools, standards and solutions that this field demands.

References

1. Adamchuk, V.I., Hummel, J., Morgan, M., Upadhyaya, S.: On-the-go soil sensors for precision agriculture. Computers and electronics in agriculture 44(1), 71–91 (2004)
2. Bazzi, C.L., Jasse, E.P., Magalhães, P.S.G., Michelon, G.K., de Souza, E.G., Schenatto, K., Sobjak, R.: Agdatabox api–integration of data and software in precision agriculture. SoftwareX 10, 100327 (2019)
3. Carrer, M.J., de Souza Filho, H.M., Batalha, M.O.: Factors influencing the adoption of farm management information systems (fmis) by brazilian citrus farmers. Computers and Electronics in Agriculture 138, 11–19 (2017)
4. Eaton, R., Katupitiya, J., Siew, K.W., Howarth, B.: Autonomous farming: Modelling and control of agricultural machinery in a unified framework. International journal of intelligent systems technologies and applications 8(1-4), 444–457 (2010)
5. Gebbers, R., Adamchuk, V.I.: Precision agriculture and food security. Science 327(5967), 828–831 (2010)
6. Goap, A., Sharma, D., Shukla, A., Krishna, C.R.: An iot based smart irrigation management system using machine learning and open source technologies. Computers and electronics in agriculture 155, 41–49 (2018)
7. Mogili, U.R., Deepak, B.: Review on application of drone systems in precision agriculture. Procedia computer science 133, 502–509 (2018)
8. Morgenstern, M., Alves, R., Battisti, G., Maran, V.: U-agro: Uma arquitetura ubíqua de gerenciamento de atividades na agricultura de precisão. ICCEEg-10 (2015)
9. Mulla, D.J.: Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosystems engineering 114(4), 358–371 (2013)
10. Pierce, F.J., Nowak, P.: Aspects of precision agriculture. In: Advances in agronomy, vol. 67, pp. 1–85. Elsevier (1999)
11. Zhang, N., Wang, M., Wang, N.: Precision agriculture—a worldwide overview. Computers and electronics in agriculture 36(2-3), 113–132 (2002)