Big Data Analytics on Indian Crop Planning to Increase Agricultural Production

Conference Paper · November 2017

DOI: 10.14257/astl.2017.147.30

5 authors, including:

Ch Chandra Sekhar
Aditya Institute of Technology & Management
10 PUBLICATIONS **6** CITATIONS
SEE PROFILE

Ch Sekhar
Vignan's Institute of Information Technology
27 PUBLICATIONS **9** CITATIONS
SEE PROFILE

Advanced Science and Technology Letters
Vol.147 (SMART DSC-2017), pp.211-216
http://dx.doi.org/10.14257/astl.2017.147.30

Big Data Analytics on Indian Crop Planning to Increase Agricultural Production

Ch. Chandra Sekhar[1], J. UdayKumar[2], B. Kishor Kumar [3] and Ch. Sekhar[4]

[1, 2, 3] Asst. Prof, Dept. of IT, AITAM, Srikakulam, AP, India
[4]Asst. Prof, Dept. of CSE, Vignan's IIT, Vizag, A.P, India
dlaxmics@gmail.com[1], mail2udauykumar@gmail.com [2], kishorbhupathi@gmail.com [3],
sekhar1203@gmail.com

Abstract. To find a hidden pattern in crop sales by applying data clustering on it. By using time or crop type as the clustering factor, to predict the schedule of crop sowing/planting or decide which crop should be sown/planted in the season. If the pattern of such crops are already known, one can set the plan, when to harvest the product, how much the price and how many. To show results, data clustering will be used to analyze the data pattern, crop pattern and crop sales pattern. Besides, it predicts the price of crops in further years or a season which helps farmers to adopt the crop cultivation plan. This paper emphasizes the use of Big Data (BD) which helps the farmers will upgrade the worth of their products via less pesticides to be focused.

Keywords: Big Data Analytics, Hadoop, Crop Sales, Agriculture, ARS, ICT.

1 Introduction

Government is spending a lot on agriculture data. Data is growing much faster than the computation speeds. An example of Big Data is crop sales. Crop sales data will be used to represent the crops data. Since government has actively and continuously collecting crop sales dataset but the size of dataset are considered to be a big data which are a real-world data, which is really a hard problem to analyze it. In order to analyze big data, data mining and statistical techniques can be expanded under parallel and distributed computing platform, also which consumes large amount of storage and computational time on handling large dataset. In conforms to its name, Big Data Analytics turns out as an important research topic. Recently, Big Data get its popularity among data scientists and business persons.

1.1 Origin of the Research Problem

As far Open Government Data (OGD) platform India maintains the data about the crop information [1]. The size of the dataset is very massive, so the traditional data analysis methodologies may not sufficient to predict the crop patterns in the dataset. Here applying Big Data Analytics for crop planning may lead to the increase in the agriculture production.

ISSN: 2287-1233 ASTL

1.2 Hadoop Terminology

In Big Data Analytics, one of the infrastructure technologies is Hadoop. It is affirmative platform to implement a large data analytics task. It comprises the way to find unidentified correlations, hidden patterns, and other essential data to study disseminated dataset of expansive crop deals. If the entire process is done by a single node, it usually gets exhausted and consumes time to analyze crop price and yield information [2]. Hence, the data clustering will be handled under distributed Hadoop environment which serves choice in crop planning by forecast the demand in the market at the earliest [3]. In this paper, a popular Map-Reduce concept utilized clustered file system extensively with Hadoop Distributed File System (HDFS) [4]. The purpose behind the Map-Reduce paradigm is high scalable which executes massively parallel and distributed over a huge number of computing nodes [5]. Therefore, Big Data Analytics is transforming every domain and everything that in the society, including science, healthcare, government, finance, IT, etc. [6]. Big Data Analytics can examine so-called all "5V": volume, variety, velocity, veracity and value. Both organizational and technological innovation required to have impact of Big Data within the agricultural sector [7]. Analytics, the main element, exploits the values from Big Data to invent new models for business and government [8]. The ICT Platform associates the farmers with the buyers of Agri Commodities – viz. Large Retailers, Exporters, Food Processing Units, Mandies. In this paper suggested data mining techniques, the expert can characterize the expansion of farming exercises to fortify different powers in existing agribusiness [9].

1.3 Research to Progress and on Pass Solutions to Agricultural Problems

In the ARS (Agricultural Research Service within USDA), studies that measure the effect of environmental conditions on agriculture often give contradictory results at various locations because the factors such as crop, soil, water, weather, climate, and management differences[10]. Therefore, for better to analyze the results, the Big Data approaches could be used as these tools would allow new explorations that enable researchers to:

- strengthen the local best practices over long-term weather conditions;
- introduce, incorporate, and emerge new management practices;
- estimate the results to other locations and agro-climatic conditions;
- combine results over a region; assess future climate change effects; and
- improve theoretical frameworks and analytical tools of agricultural science.

Improvements in ARS computing capacity would help to resolve this issue. An example for this would be integrating weather and precipitation characteristics to forecast variety performance in an arbitrary zone with different climatic conditions.

1.4 Agricultural Supply Chain' Decisions using BD

Agriculture, an industry, probes the various data sources existed in the field of agricultural and analyzes the usage of sources in supply chain improvement. The utilization digital records about farming practices show a significant part in documenting measures to figure out the factors of particular sales. According to this paper's context, the developed country based firms are in a practice to exercise their ability over farmers from developing countries. Thus, foreign companies are more functional in new budding trade crops to incorporate into the universal supply chain.

1.5 A New Age of Agriculture using BD

On this scope, Precision Agriculture (PA) or precision farming has turned into a key pattern in automated nations. For instance, a few tasks like extreme weather conditions, unknown soil types can be performed well by people instead of calculations. Consequently, the information is crunched by algorithms and experts transformed into useful advice to cultivators and their machines to the optimum amount of pesticides, herbicides, fertilizer, and other applications. In this manner, fixed cash crops such as coffee, cotton, tea, and tobacco are extra potential to implement diverse attributes of BD.

2 Data Analytical framework

In this, collect crop sales data using several computers to run data collection program that collect data from internet and store it to a open Agriculture database with two parameters such as date and Crop_Id. Different attributes can be used in the clustering process, such as time (week or month) or crop type. The relationship between some attributes in the data set, such as price and trading volume. The process of submitting, extracting and storing will be repeated as many the numbers of crops in the crop list. In this framework, crop sales will be analyzed by applying data clustering. The data clustering algorithm will run inside of Hadoop platform that provides parallel processing and distributed computing feature to the process. By using decision tree, can predict the grouping of the data based on the pattern that has been discovered previously. To speed up the analytical process, data analytics framework uses Hadoop, Map-Reduce and HDFS to store the input and the results on crop sales data as shown in Fig. 1.

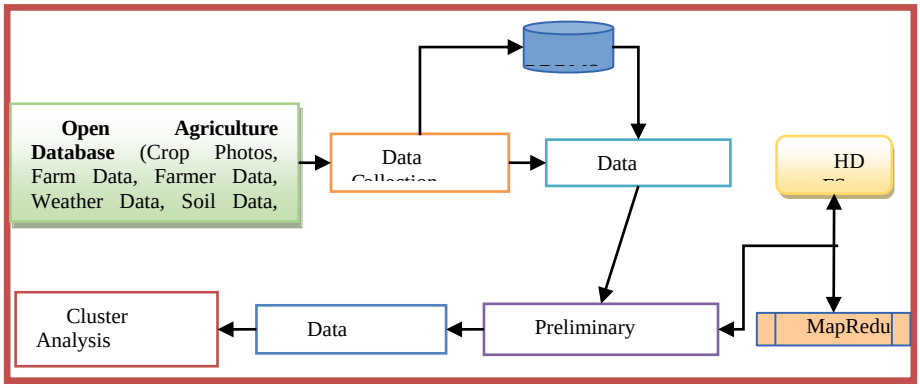

Fig. 1. Data analytics framework on Crop sales data

3 Methodology

A better clustering method is needed to handle the unbalanced data, especially when there are very few transactions in the extreme high price with very low trading volume. The experiments of the data clustering on crop sales data can be divided into several steps as shown in Fig. 2.

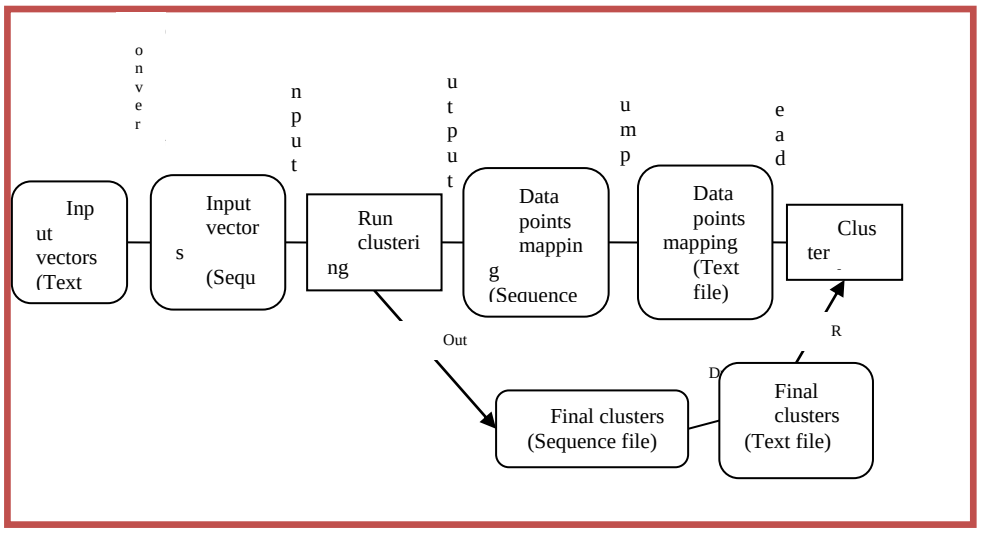

Fig. 2. Crop sales data analysis steps

Create and prepare the input vectors into text file format. The input vectors are converted into sequential file format. The algorithm will be used to do data clustering on it. In the algorithm process and initial clusters will be generated and stored into a directory and final clusters will be stored into a different directory at the end of the

data clustering. In this step, plotting the clusters into graphics can be helpful. The crop sales data domain as the important factor, these data have to be converted to human readable format to make it easier to analyze the result.

4 Results

As per Open Government Data (OGD) Platform India published by National Data Sharing and Accessibility Policy (NDSAP). Analyzed the 20 years crop covered area and production from 1997, the dataset consist of 570 records with attributes of district wise, crop wise, season wise and year wise data. The information is being utilized to revise and investigate crop growing pattern and diversification, high yield production, Agro-climatic region wise performance and crop production contribution to district/state/country.

5 Conclusion

Here people are mostly depending on cultivation rather than jobs because of their illiteracy. Unfortunately, their lack of education reflects on their methods of cultivation. Here our people need a better guidance through technology. Agriculture demonstrator states that who do well do not achieve good results. So, no growth is found in lives of cultivators. In the future, this research will be scaled up in terms of data size and crop variants. Apart from ICT and all, Big Data Analytics is one of the best platforms for Crop Planning to improve agriculture productivity. Big Data Analytics on Crop Planning is very significance work to increase agricultural production and provide the benefits of ICT & other advanced technologies to the common man.

References

1. Govt. of India released under NDSAP, https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics
2. Yang, et. al., "Data Clustering on Taiwan Crop Sales Under Hadoop Platform", Proceedings of the Institute of Industrial Engineers Asian Conference 2013.
3. M. Moorthy, et al., “An Analysis for Big Data and its Technologies”, IJCSET, Vol 4, Issue 12, Dec 2014, pp. 412-418.
4. Anirban Mukherjee, et. al., “Shared Disk Big Data Analytics with Apache Hadoop”, 978-1-4673-2371-0/12©2012 IEEE
5. K. Grolinger, et.al., “Challenges for MapReduce in Big Data”, 2014 IEEE World Congress on Services, June 27-July 2, Alaska, USA.
6. IEEE International Conference by CSI on Big Data on IT in Business, Industry and Government (CSIBIG), March 8-9, 2014, ISBN: 978-1-4799-3063-0.
7. Steve Sonka, “Big Data and the Ag Sector: More than Lots of Numbers”, IFAMA, Volume 17 Issue 1, 2014, pp. 1-20

8. Joseph O. Chan, “An Architecture for Big Data Analytics”, IIMA, Volume 13 Issue 2, 2013, pp. 1-14
9. G. Nasrin Fathima, et al., “Agriculture Crop Pattern Using Data Mining Techniques”, IJARCSSE, Volume 4, Issue 5, May 2014, ISSN: 2277 128X, pp. 781-786.

