November-December 2019
ISSN: 0193-4120 Page No. 5383 - 5388

Agricultural Area Detection Using Data Mining

*B. Bharath Kumar, Mary Subaja, D. Mahalakshmi

**UG Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India*
Associate Professor, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India
bharathbobba@gmail.com, marysubajachristo.sse@saveetha.com, mahalakshmid.sse@saveetha.com

Article Info
Volume 81
Page Number: 5383 - 5388
Publication Issue:
November-December 2019

Article History
Article Received: *5 March 2019*
Revised: *18 May 2019*
Accepted: *24 September 2019*
Publication: *26 December 2019*

Abstract
The data mining concept has become one of the most essential things to process huge amounts of data and get the work done in various fields. Initially, manual computations are done as a part of the data analysis. But, with the advent of computers, the computation and analysis processes have become a bit easier than earlier. Along with the increase in technology, the data to be stored and analyzed has also increased at higher rates. In general, the government officials make a note of the actual area present in a village/town/city/state/country and they'll be assigned the task of updating this data at regular intervals of time. To complete this task, they generally conduct surveys and make a note of different categories of land that is present in that particular area. After the collection of this data, the data belonging to all the areas present in a particular state are combined to get an overall idea of the land belonging to different categories in that state. In the same way, the data belonging to all states are brought together and analyzed to understand the total categories of land present in that particular country.
As we all know that Agriculture is one of the key sectors that is making human life possible to survive on this planet, there comes the need to have the statistics of this agricultural land. It is hard to process this agricultural area data in normal methods. Data Mining stands as the best choice to complete this task. Thus, the time series mining approach is followed for agricultural area detection.

Keywords: *agriculture, analysis, data mining, hdfs, hive, sqoop.*

1. Introduction

India is one of the five largest countries in the world that produces a large variety of agricultural products. According to the statistics obtained from the survey conducted in 2018, it is said that agriculture employed 50% of the Indian workforce and contributed 17-18% to the country's Gross Domestic Product. The people from almost all the parts of the world consider this as a work that they are closely attached to, for it makes their survival possible. Agricultural lands and practices can be vastly observed in rural areas rather than in the towns.

November-December 2019
ISSN: 0193-4120 Page No. 5383 - 5388

While the empty lands in the towns are being used to build some sources of entertainment or shelter, the empty lands present in the rural areas are being converted and used as the agricultural lands. Even barren lands are being converted into agricultural lands by following some scientific methods. Thus there comes the need to have an updated database regarding the agricultural lands and the crops that are being harvested.

By this process of having an updated database regarding the agricultural area present in the country and throughout the world, the farmers can be helped in one way or the other. How the farmers get benefited through this agricultural area detection is that the importance and necessity of growing the crops that are most needed by the people in a country can be taught to the farmers to get the desired food products. This way, the income generation for farmers can also be increased. Thus, both the tasks i.e., the hunger-quenching task of the people present in the country or the world and the income generation improvement for the farmers gets done.

Data Mining offers several ways to analyze and make use of the data that we have. But, we need to choose specific methods from them based on our requirements. For instance, the estimation of the political result is generally done using the sentimental analysis. In the sentimental analysis, we gather the data obtained from the polls and the posts from the social media sites. The data obtained from these polls and posts are analyzed, positive polls and positive posts related to a specific political party are separated and their score is obtained. In the same way, the score of other parties is obtained and the final score is calculated by comparing all these values. One of the best methods to perform the agricultural area detection is the time series mining approach.

2. Literature Review

Debanshu Ratha et al performed Harvest observing at spatiotemporal scales permits a superior comprehension of harvest elements and yield evaluation. In this unique circumstance, crop biophysical parameters. These attributes have indicated great potential for the evaluation of harvest development. Specifically, manufactured gap radar (SAR) imaging has drawn significant consideration for rural applications because of its capacity to screen taking all things together climatic conditions and their affectability to dielectric and geometric properties.

Renmin Zhang et al performed Millimeter-wave (mm Wave) transmissions experience the ill effects of serious constriction. To battle the undesired way misfortune, it is basic to set up directional transmissions with huge scale clusters for mm Wave interchanges. With an expanding number of reception apparatuses, be that as it may, it is restrictive to receive conventional numerous information different yield (MIMO) channel estimation innovation because of overwhelming overhead. Writers in proposed a half and half computerized/simple structure, where the advanced precoding was planned dependent on proportionate baseband channel through essentially diminished measurement & simple bar shaping was worked via looking through code words in a predefined codebook.

Jianxiao Liu et al proposed Quality administrative systems (GRN) could be induced from articulation profiles and communications between administrative targets. It can clarify the new capacity of quality and the instrument of the existence procedure, and give significant

data to medicate structure or medicinal related fields. As of late, the earlier information is utilized to improve the learning effectiveness and precision of Bayesian system strategy, for example, utilizing the co citation in PubMed and schematic comparability in Gene Ontology. Altogether, the current research work for the most part utilizes the hunt scoring based Bayesian system learning technique. This technique frequently utilizes the neighborhood or arbitrary pursuit methodology, and it is a combinatorial blast issue with the expansion of the hub number, bringing about quite a while to construct the system. Contrasted and the pursuit scoring based Bayesian

System learning strategy, the learning productivity of reliance investigation based technique is moderately high, and it can get the worldwide ideal arrangement. The three-stage reliance examination calculation (TPDA) is a generally utilized reliance investigation based Bayesian system structure learning strategy.

Taotao Lai et al has used this system so as to expand the strength of the multi-model fitting strategies, we propose a compelling iterative plan that contains a productive novel worldwide voracious hunt procedure to create exact speculations near the genuine models. Besides, a common data theoretic technique is utilized to intertwine the speculations to discover genuine models. We show that the novel worldwide eager pursuit system can successfully decrease the blunder in parameter estimations from hypothetical and exact avocations. In the wake of playing out the worldwide voracious inquiry methodology, more than one speculation may compare to a similar model occasion. Accordingly, a combination system is directed to intertwine the speculations comparing to a similar model case, by utilizing common data hypothesis. In the wake of directing the combination methodology, the quantity of held speculations might be not as much as that of the genuine model occurrences. Thus, the ordinary insatiable examining methodology is utilized to produce speculations for the staying model occurrences. The worldwide ravenous pursuit technique, the combination procedure, and the traditional eager testing methodology are performed iteratively until a palatable arrangement is accomplished.

Junhu Ruanet al has found that Digital physical frameworks can interface the digital world with the physical world, which are accounted for in different areas, for example, car, compound forms, human services, assembling and transportation. The present advancement of Internet of Things (IoT) is accelerating the development of more digital physical frameworks. In particular, increasingly more IoT gadgets are sent in horticulture, creating different agrarian digital physical frameworks. IoT is perceived as one promising apparatus to accomplish accuracy farming and further successfully diminish the serious weight brought to the world by the developing populace. In the most recent writing, expanding consideration has been invested on improving horticulture yields by IoT frameworks and systems. To manage the measurement catastrophe issue looked by most traditional mining calculations, we join granulation methods and hereditary calculation (GA) with help vector machine (SVM) to propose a micro GA-SVM. In the incorporated indicator, three granulation strategies, that is, Min-Median-Max granulation, Quartile-Median granulation and Fuzzy granulation, are acquainted with separate huge information in agrarian digital physical frameworks into little scale granules, and GA is utilized to locate the ideal estimations of SVM punishment attribute and part attribute from the diminished micro elements.

November-December 2019
ISSN: 0193-4120 Page No. 5383 - 5388

3. Proposed Approach

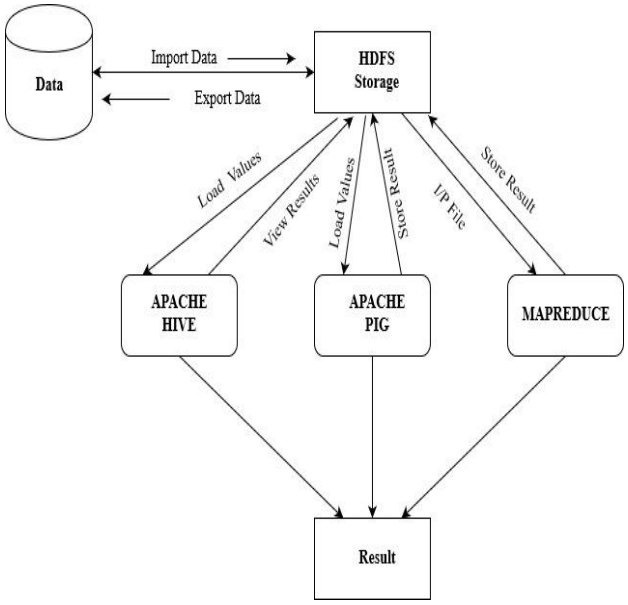

Figure 1: System Architecture

The above picture depicts the overall processes that are involved in detecting the agricultural area. Initially, the data present in the database is imported into the hdfs storage. From this hdfs storage, the APACHE HIVE and APACHE PIG loads the values present in it. The other functionality present in the apache hive is that we can view the results. The other functionality present in the apache pig is that we can acquire the results and store them in the hdfs storage. In the case of map reduce, the input file will be acquired from the hdfs storage and stores the obtained result in it. Finally, the results obtained after processing the data apache hive, apache pig and map reduce are store in a different module and are used for the analysis and detection of the required area value.

MySQL: This is used to manage and make use of the data. In the current project, the data related to agricultural area needs to be gathered and processed, however this can only help us in the data gathering task, this can be of great use in maintaining that data as partitions. The overall area data is obtained, from which we need to separate the agricultural area data. This makes use of the RDMS, which stands for Relational Database Management System. As the name implies, this Relational Database Management System uses relations to store and manage the stored data.

Sqoop: This is an interface application that is used to move data between the applications such as MySQL and the Hadoop. The database present in the Mysql regarding the agricultural area needs to be moved to the Hadoop Distributed File System using the Sqoop. The agriculture related data can be moved into Hadoop Distributed File System or Hive from Mysql, which in turn creates the java classes. Previously, stream of information from the relational database management system is sent to the Hadoop distributed file system by making use of the feature namely ‘send out’ device. Initially, the metadata of the tables containing the actual data is prepared and then the Sqoop performs this send out operation.

Hive: Apache Hive is simply called as the Hive. This is a data warehouse project built on top of the apache Hadoop to perform the data query and analysis. This gives a Structured Query Language like interface to manage and make use of data that is stored in various databases and file systems that integrate with the Hadoop. Initially, the agriculture data stockroom intended for querying and analysis is just an organized data that is stored in the tables. Hive sorts out agricultural data tables into allotments. This is a method of transforming the bulk data into related parts dependent on the analysis results of the divided sections.

PIG: Apache Pig is shortly called as Pig. This is a high level platform for making programs that run on Apache Hadoop. The language used in this is the Pig Latin. So, the

developers use pig latin language to compose the contents and process them using the intelligent mode using the grunt shell. Each of those contents is transformed to the map and reduce assignments.

Map Reduce: Map Reduce is the system which helps us in composing applications to process huge amounts of agriculture data. The two main functions of this system are the Map and Reduce. This executes the tasks in three phases namely the map organize, mix arrange and finally the decrease organize. The main functionality of Map is to process the information. The information record is sent to the mapper work line by line. The main function of Reduce is to process the data obtained from the mapper.

4. Conclusion

The efficient functionality of this method depends on the values that are used during the analysis process. In some cases, the values present in the datasets that are provided for analysis might contain missing values. These missing values affect the overall result. To avoid this, the missing values are to be filled with some approximate values. The overall result consists of the area including the inhabited lands and the barren lands. The overall data obtained in the analysis is considered and the agricultural area data is separated.

Initially, the agricultural area data are considered together. Further, the categories of crops are considered and then the statistics related to those categories are separated. The main objective of this project is fulfilled right after gathering the agricultural area data. But, the sub-tasks involved in this analysis process includes finding the type of agricultural practices that are being followed in a certain country and the continent. These results will also help manage import and export activities throughout the world. The agricultural products that are in stock which are more than required can be exported to the areas that require them.

The older method used to carry out this agricultural area detection process is using the Hadoop. With the advancement in technology, the Hadoop system has become less efficient in processing this data and in yielding efficient results. Thus, the efficient spark tool is used to process the data regarding the area occupancy and usage to obtain the best results.

References

[1] Y. Kim and J. J. van Zyl, “A time-series approach to estimate soil moisture using polarimetric radar data,” IEEE Trans. Geosci. Remote Sens., vol. 47, no. 8, pp. 2519–2527, Aug. 2009.

[2] F. Canisius et al., “Tracking crop phenological development using multitemporalpolarimetric Radarsat-2 data,” Remote Sens. Environ., vol. 210,pp. 508–518, Jun. 2018.

[3] R. Zhang, H. Zhang, and W. Xu, “Fast beam alignment algorithm for multi-user mmWave communications,” Electron. Lett., vol. 54, no. 25, pp. 1456–1458, Dec. 2018.

[4] L. Liang, W. Xu, and X. Dong, “Low-complexity hybrid precoding in massive multiuser MIMO systems,” IEEE Wireless Commun.Lett., vol. 3, no. 6, pp. 653–656, Dec. 2014.

[5] V. Tresp, “A Bayesian committee machine,” Neural Comput., vol. 12,no. 11, pp. 2719–2741, 2000.

[6] Y. Wang, T. Joshi, X. S. Zhang, et al., “Inferring gene regulatory networks from multiple microarray datasets,” Bioinformatics, vol.22,no. 19, pp.2413-2420, 2006.

[7] F. Petralia, P. Wang, J. Yang, et al., “Integrative random forest for gene regulatory network inference,” Bioinformatics, vol. 31, no. 12, pp.197-205, 2015.

[8] I. Tsamardinos, L. E. Brown and C. F. Aliferis, “The max-min hillclimbing Bayesian

network structure learning algorithm," Machine learning, vol. 65, no. 1, pp.31-78, 2006.

[9] J. Fu, Y. Cheng, J. Linghu, et al., "RNA sequencing reveals the complex regulatory network in the maize kernel," Nature communications, vol. 4, no. 4, pp.2832, 2013.

[10] Y. X. R. Wang and H. Huang, "Review on statistical methods for gene network reconstruction using expression data," Journal of Theoretical Biology, vol.362, pp.53-61, 2014.

[11] DebanshuRatha et al, "A Generalized Volume Scattering Model-Based Vegetation Index from Polari metric SAR Data"

[12] Renmin Zhang et al, "A Generalized Volume Scattering Model-Based Vegetation Index from Polari metric SAR Data"

[13] Jianxiao Liu et al, "Gene Regulatory Relationship Mining Using Improved Three-Phase Dependency Analysis Approach"

[14] Taotao Lai et al, "Efficient Robust Model Fitting for Multi structure Data Using Global Greedy Search"

[15] JunhuRuan et al, "A Granular GA-SVM Predictor for Big Data in Agricultural Cyber-Physical Systems"