ORIGINAL RESEARCH PAPER
Volume-6 | Issue-12 | December-2017 | ISSN No 2277 - 8179 | IF : 4.176 | IC Value : 93.98
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

REVIEW OF LITERATURE OF DATA MINING TECHNIQUES FOR CROP YIELD PREDICTION

Computer Science

Shital H. Bhojani Assistant Professor-AAU-Anand– PHD Scholar RK University, Rajkot
Dr. Nirav Bhatt Assistant Professor – RK University, Rajkot – Gujarat

ABSTRACT

Data mining techniques are very popular in the area of agriculture and relatively it's a young research field. Crop yield prediction plays vital role in agriculture sector for decision making. This paper presents a brief study of various data mining techniques used for crop yield prediction by different researchers. Various researches have been done exploring the connections between large-scale climatologically phenomena and crop productivity. Crop prediction methodology is used to predict the suitable crop by sensing various parameters of soil and environment.

KEYWORDS

Crop yield prediction, data mining, classification, clustering, association rule mining, regression

INRODUCTION

Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis. Data mining is a logical process that is used to search through large amount of data in order to find useful data. The goal of data mining technique is to extract patterns that were previously unknown. Once these patterns are found they can further be used to make certain decisions for development. It basically includes Exploration, Pattern identification & Deployment. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining).The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Yield prediction is very difficult task. Data mining functionalities include the discovery of concept/class descriptions, associations and correlations, classification, prediction, clustering, trend analysis, outlier and deviation analysis, and similarity analysis. Data mining is part of our daily lives although we may often unaware of its presence.

LITERATURE SURVEY

There have been a number of research studies undertaken that focus on the importance of using data mining as a supplementary tool in transforming large volumes of agricultural data into meaningful information. Many researchers have been contributed their previous knowledge towards data mining in agriculture. There are many simulations models available for crop productivity predictions. As it depends on economical and environmental parameters so we can not apply these existing models or methods to any other area.

D Ramesh, B Vishnu Vardhan compared statistical model Multiple Linear Regression over the Data Mining Density-based clustering technique. He concluded that comparison of the crop yield prediction can be made with the entire set of existing available data and will be dedicated to suitable approaches for improving the efficiency of the proposed technique [3,7].

M.C.S.Geetha concluded that data mining plays a crucial role for decision making on several issues related to agriculture field. She also discusses on different data mining applications in solving the different agricultural problems [4].

Vrushali Bhuyar focuses on different classifier algorithms used on soil dataset to predict fertility rate. Study shows that among the classifier J48 classifier perform better to predict fertility index [6].

Raorane A.A. and Kulkarni R.V., discussed few data mining techniques in their paper. They concluded that efficient technique can be developed and analyzed using the appropriate data, to solve complex agricultural problems using data mining techniques. Also recommend some of the algorithms and statistical methods that give good results in agriculture growth [8].

Georg Rub, Rudolf Kruse, Martin Schneider, and Peter Wagner mainly focused on Neural Network technique of data mining in their research work. They built and evaluated different networks and substantiated the assumption that the prediction accuracy of the networks rises once more data become available at later stages into the growing season [10-12]. Georg Rub also used Different regression techniques on agricultural yield data and concluded that support vector regression can serve as a better reference model for yield prediction compare to decision tree and neural network [9-12].

S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh attempted to compile the research findings of different researchers who worked on crop productivity data [13].

Dr. D. Ashok Kumar and N. Kannathasan focuses on different data mining techniques we can use in agriculture. Their research survey recommending that a comparison of different data mining techniques could produce an efficient algorithm for soil classification for multiple classes. The benefits of a greater understanding of soils could improve productivity in farming, maintain biodiversity, reduce reliance on fertilizers and create a better integrated soil management system for both the private and public sectors [2].

Anusha A. Shettar and Shanmukhappa A. Angadi used weka tool to implement data mining algorithms on dataset and concluded that: Classification algorithms like J48, LMT, LAD Tree and ID3 gives the better prediction result with respect to yield _class attribute like Poor, Good, Very_Good, and Excellent. This different accuracy result with different classification algorithms [5].

Yethiraj N G. concluded that there are a growing number of applications of data mining techniques in agriculture and a growing amount of data that are currently available from many resources [15].

Nilesh Dumbre, Omkar Chikane and Gitesh More concluded that we can say that if perfect crop recommendations are given to farmers it will definitely help to increase the crop yield and also in building the economic status of agricultural dependent countries [17].

Rajshekhar Borate, Rahul Ombale and his colleagues found that by using the Data Mining techniques accurate prediction of specified crop yields across different districts will help to farmers of India. They proposed a system which will recommend some inference to the farmers based on the stored data [18].

MATERIALS AND METHODS

Data mining is the process which used to find out useful patterns from large amount of data [19]. Data Mining is a fairly new technology that can facilitate discovery of rules and patterns in sets of data. Typical uses include identifying clusters in data, and inferring sets of rules that describe each category or class in a data set. We need data mining techniques when human expertise does not exist for solving a problem using example data or past experience or also required when humans are unable to explain their expertise over the time. A wide variety of different data mining techniques are available. Generally, each implementation only accepts data in its own individual format, and has its own ways of specifying parameters and output. The main techniques for data mining include Association rules, Classification,

Clustering and Regression.

CLUSTERING

In clustering, the focus is on finding a partition of data records into clusters such that the points within each cluster are close to one another [19]. Clustering groups the data instances into subsets in such a manner that similar instances are assembled together, while dissimilar instances belong to diverse groups. Since the aim of clustering is to find out a new set of categories, the latest groups are of interest in themselves, and their assessment is intrinsic. There is no prior knowledge about data. The different clustering methods are K-mean, k medoids Hierarchical Methods(HM), Partitioning Methods (PM), Density-based Methods(DBM), Model-based Clustering Methods(MBCM), Grid-based Methods and Soft-computing Methods [fuzzy, neural network based], Squared Error—Based Clustering (Vector Quantization), network data and Clustering graph . Hierarchical cluster creates a hierarchy or tree of clusters which mainly useful for data summarization and visualization. Density based cluster method is developed to discover the cluster with arbitrary shape.

ASSOCIATION RULE MINING

Association rule mining technique is used to find out the frequent patterns that appear in data set frequently, such patterns plays an essential role in mining association or correlation among data [19]. In this method, the focus is on finding relationships between the different items in a transactional database. Association rules are used to find out elements that co-occur repeatedly within a dataset consisting of many independent selections of elements (such as purchasing transactions), and to discover rules. It is mainly used for market basket analysis, customer segmentation, store layout, catalog design, and telecommunication alarm prediction, etc...

CLASSIFICATION

Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. It is a process in which a model learns to predict a class label from a set of training data which can then be used to predict discrete class labels on new samples. To maximize the predictive accuracy obtained by the classification model when classifying examples in the test set unseen during training is one of the major goals of classification algorithm. Data mining classification algorithms can follow three different learning approaches: supervised learning, unsupervised learning, or semi-supervised learning [19]. The different classification techniques for discovering knowledge are Rule Based Classifiers, Bayesian Networks (BN), Decision Tree (DT), Nearest Neighbor (NN), Artificial Neural Network (ANN), Support Vector Machine (SVM), Rough Sets, Fuzzy Logic, and Genetic Algorithms.

REGRESSION

It is a statistical measure that can be used to determine the strength of the relationship between one dependent variable and a series of other changing variables known as independent variables (regular attributes). If independent variable contains multiple input attributes like in our research (rainfall, humidity, etc), then it is termed as multiple linear regressions. Linear regression provides a model for the relationship between a scalar variable and one or more explanatory variables. This is done by fitting a linear equation to the observed data. The methods for prediction are Nonlinear Regression (NLR) and Linear Regression (LR).

CONCLUSIONS

In this research survey data mining techniques were discussed in order to estimate crop yield analysis with existing data and their use in data mining by multiple authors. This paper focuses on new research possibilities to the problem of crop yield prediction. *There are a growing number of applications of data mining techniques in agriculture and a growing amount of data that are currently available from many resources. It's a difficult task to predict crop yield due to stochastic rain fall pattern and also variation in temperature. So we can apply different data mining techniques for crop yield predication and can produce an efficient algorithm for crop classification for better decision making for the benefit of mankind.*

REFERENCES

[1] YunousVagh, Jitian Xiao. A Data Mining Perspective of the Dual Effect of Rainfall and Temperature on Wheat Yield. ECU Publications 2012

[2] Dr. D. Ashok Kumar, N. Kannathasan. A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining. IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 1, May 2011,ISSN: 1694-0814

[3] D Ramesh, B Vishnu Vardhan. Analysis of crop yield prediction using data mining techniques.IJRET: International Journal of Research in Engineering and Technology ISSN: 2319-1163 | pISSN: 2321-7308

[4] M.C.S.Geetha. A Survey on Data Mining Techniques in Agriculture. International Journal of Innovative Research in Computerand Communication Engineering(An ISO 3297: 2007 Certified Organization)Vol. 3, Issue 2, February 2015 ISSN(Online): 2320-9801 ISSN (Print): 2320-9798

[5] Anusha A. Shettar, Shanmukhappa A. Angadi. Efficient data mining algorithms for agriculture data. International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 02, Issue 09; September – 2016 [ISSN: 2455-1457]

[6] VrushaliBhuyar. Comparative Analysis of Classification C2Techniques on Soil Data to Predict Fertility Rate for Aurangabad District. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com, Volume 3, Issue 2, March – April 2014 ISSN 2278-6856

[7] D Ramesh , B Vishnu Vardhan. Data Mining Techniques and Applications to Agricultural Yield Data.International Journal of Advanced Research in Computer and Communication EngineeringVol. 2, Issue 9, September 2013, ISSN (Print) : 2319-5940, ISSN (Online) : 2278-1021

[8] Raorane A.A., Kulkarni R.V. Data Mining: An effective tool for yield estimation in the agricultural sector. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com, Volume 1, Issue 2, July – August 2012 ISSN 2278-6856

[9] Georg Ruß. Data Mining of Agricultural Yield Data A Comparison of Regression Models.Otto-von-Guericke-Universit¨at Magdeburg

[10] Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner.Estimation of Neural Network Parameters for Wheat Yield Prediction.IFIP International Federation for Information Processing, Volume 276; Artificial Intelligence and Practice II; Max Bramer; (Boston: Springer), pp. 109–118.

[11] Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner Optimizing Wheat Yield Prediction Using Different Topologies of Neural Networks. L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 576-582,Torremolinos (M alaga), June 22-27, 2008

[12] Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter WagnerData Mining with Neural Networks for Wheat Yield Prediction. P. Perner (Ed.): ICDM 2008, LNAI 5077, pp. 47–56, 2008.Springer-Verlag Berlin Heidelberg 2008

[13] S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh. Data mining Techniques for Predicting Crop Productivity – A review article.IJCST Vol. 2, Issue 1, March 2011, ISSN : 2229-4333(Print)|ISSN:0976-8491(Online)

[14] Ramesh Vamanan ,K.Ramar. Classification of agricultural land soils a data mining approach."International Journal on Computer Science and Engineering (IJCSE), ISSN : 0975-3397 Vol. 3 No. 1 Jan 2011"

[15] YETHIRAJ N G. Applying data mining techniques in the field of agriculture and allied sciences. Vol 01, Issue 02, December 2012 International Journal of Business Intelligent, http://iirpublications.comISSN: 2278-2400, Integrated Intelligent Research (IIR)

[16] Mrs. SwaroopaShastri, Arati K Gunge,Neural Network based Weather Prediction Model towards Ideal Crop Selection. International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 5 Issue 09 September 2016 Page No.17845-17847

[17] NileshDumbre, OmkarChikane, Gitesh More, System for agriculture recommendation using data mining. International Education & Research Journal [IERJ] E-ISSN: 2454-9916 | Volume : 1 | Issue : 5 | Dec 2015

[18] Rajshekhar Borate, Rahul Ombale, SagarAhire, ManojDhawade, Mrs. Prof. R. P. Karande Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in India. International Journal of Novel Research in Computer Science and Software Engineering Vol. 3, Issue 1, pp: (34-37), Month: January-April 2016

[19] Data mining Concepts and Techniques by Jiawei Han and MichelineKamber