International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

A Survey on Various Data Mining Techniques in Field of Agriculture for Prediction of Crop Yield

Huma Khan[1], Shahista Navaz[2], Dr. S. M. Ghosh[3]

[1]PhD Research Scholar, CVRU, Bilaspur, Chhattisgarh

[2]PhD Research Scholar, CVRU, Bilaspur, Chhattisgarh

[3]Professor, Dept of Computer Science and Engineering, CVRU, Bilaspur, Chhattisgarh

Abstract: ***Agriculture is a key to the economy and infrastructure of India. It plays the significant most strategic role in the progress and financial growth of the nation. In Order to get a proper tab on the agriculture sector of the nation, state wise and geography wise, an appropriate program can be deployed, with the help of this program we can predict the best crop. Crop yield prediction provides information for decision makers to maximize the crop productivity. Data mining technology proved to be a better choice for this purpose and has become an interesting and recent research topic in agriculture to predict the crop yield. This paper presents a brief comparative study of various methods and techniques available that can be used to predict the crop yield in data mining.***

Keywords: K-Means, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Multiple Linear Regression (MLR)

1. Introduction

Data mining is a process of extracting important and useful information from large data sets. It can be used in field of prediction crop yield. Agriculture in India plays a strategic role in the economy of the country. The major drawback for the farmers to get the profit is the condition of global warming that has significant impacts on conditions affecting agriculture. As a result, the rate of suicide among Indian farmers is increasing drastically which causing impact on the agricultural economy. In India, majority of the farmers are not getting the expected crop yield due to several reasons like weather conditions, soil, water and air which leads to insecurity of food. Every farmer is interested in knowing, how much yield he is about expect. Raw data, which is obtained from the history of crop yields, is required for the prediction. These data play an important role in prediction of crop yield. Data Mining is widely applied to agricultural problems. Data Mining is used to analyze large data sets and establish useful classifications and patters in the data sets. The overall goal of the Data Mining process is to extract the information from a data set and transform it into understandable structure for further use.

Different techniques were proposed for mining data over the Years. A detailed and elaborated 10 Data Mining techniques were discussed by the researchers [1]. Wheat yield prediction done by Neural Network's Multilayer Perceptron model [2] by training the network, here sensor input and fertilizers considered as parameters for prediction. Rudolf Kruse et al. [3, 4, 5] used MLPs successfully in their research. In [6] , considered neural network first network with a Multilayered Perceptron and second one with a Radial Basis Function, Support Vector Regression and Decision Regression Tree also considered. Out of these four techniques [6] it showed that the Support Vector Regression technique is the most suitable for this type of problem. Later spatial Auto Correlation technique also improved prediction quality [7].

Data Mining techniques are mainly divided in two groups [8], classification and clustering techniques. Classification techniques are designed for classifying unknown samples using information provided by a set of classified samples. Generally, Neural Networks and Support Vector Machines [9], these two classification techniques learn from training set how to classify unknown samples. Another classification technique, K- Nearest Neighbor [10], does not have any learning phase., The K-Nearest Neighbor uses the information in the training set, but it does not extract any rule for classifying the other.

One of the most used clustering technique is the K-Means algorithm [11]. Given a set of data with unknown classification, the aim is to find a partition of the set in which similar data are grouped in the same cluster. The idea behind the K-Means algorithm is, given a certain partition of the data in K clusters, the centers of the clusters can be computed as the means of all samples belonging to a clusters. One of the disadvantages could be the choice of the parameter K. Another issue that needs attention is the computational cost of the algorithm. There are other Data Mining techniques statistical based techniques, such as Principle Component Analysis(PCA) , Regression Model and Biclustering Techniques [12,13] have some applications in agriculture or agricultural - related fields.

2. Literature Review

Various Data Mining techniques can be applied in the field of agriculture. Some are related to weather conditions or forecasts. For example, the K-Means algorithm is used in performing atmosphere pollution forecast[14] Neighbor (KNN) is applied for simulating daily precipitations and other weather variables [15]. Various changes of the weather scenarios are analyzed using SVMs [16]. Data Mining techniques are also applied to study sound recognition problems. [17] uses SVMs for classification of the sound of birds and other different sounds. In [18] K-Nearest Neighbor

approach was used to analyzed and estimate forest variables for analyzing satellite imagery.

Some of the Data Mining techniques often used for studying soil characteristics. [19]The K-Means approach is used for classifying soils in combination with GPS-based technologies. [20] uses a K-Means approach to classify soils and plants and Camps Valls et al.[21] uses SVMs to classify crops. Apples are checked using different approaches before sending them to the market.

[22] Used data mining technique to predict rainfall over two techniques and compares yield prediction based on Rainfall between MLR Technique and K-Means. The estimation of average production using MLR Technique was given as 98 % and using K-Means algorithm was given as 96% accuracy.

Fathima et al.[23] used k means and Appriori algorithm, Crop type and Irrigation parameters were considered. Focused on the policies that government could frame by the cropping practices of farmers.

[24] presented the purpose of data mining techniques in the field of agriculture k-means, ID3 algorithms, the k nearest neighbor, support vector machines, artificial neural networks were discussed.

[25]Kaur et al. have used BP neural network and simulated the result using MATLAB. They found suitable data model for achieving high accuracy for price prediction. The prediction is mainly based on only price.

Bendre et al. [26] used Map Reduce and Linear Regression algorithm for weather forecasting. The effective model to improve the accuracy of rainfall forecasting is investigated. The forecasting is done based on only a weather data.

[27] predicted yield using Naïve Bayes, Appriori algorithm. Main focus was on various soil parameters like pH,Nitrogen, moisture etc and comparison accuracy is also presented. Only 77% of accuracy is achieved.

Naïve Bayes, J48, random forests, support vector machines, artificial neural networks were implemented [28]. Climate data and Crop parameters are used for crop yield is prediction. Other parameters like soil are not considered.

Regression Analysis [29], Linear are cited. Described about various environmental factors that influence the crop yield and the relationship among these parameters is also established.

3. Conclusion and Future Work

Data mining is the emerging field of technology can be used for the crop yield prediction. Various techniques of classification and clustering can be used to analyze the data. Biclustering having huge scope in this field and rarely used by the researchers in this field when compared to other data mining techniques. Other than that the multidisciplinary approach of integrating agriculture with soft computing tools can be helpful in forecasting or managing agricultural crops effectively.

References

[1] Wu X, Kumar V, Quilan JR, Ghosh J, Yang Q, Motoda H, McLanchlan GJ, Ng A, Liu B, Yu PS, Zhou Z-H, Steinbach M, Hand DJ, Steinberg D, Top 10 algorithms in data mining. Knowl Inf Syst 14 : 1-37, 2008.
[2] P. Wagner and M. Schneider. Economic benefits of neural network-generated site-specific decision rules for nitrogen fertilization. In J. V. Stafford, editor, Proceedings of the 6th European Conference on Precision Agriculture, 775–782, 2007.
[3] Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner. Estimation of neural network parameters for wheat yield prediction. In Max Bramer, editor, Artificial Intelligence in Theory and Practice II, volume 276 of IFIP International Federation for Information Processing, 109–118. Springer, July 2008.
[4] Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner. Optimizing wheat yield prediction using different topologies of neural networks. In Jos´e Luis Verdegay, Manuel Ojeda-Aciego, and Luis Magdalena, editors, Proceedings of IPMU-08, 576–582. University of M´alaga, June 2008.
[5] Georg Ruß, Rudolf Kruse, Peter Wagner, and Martin Schneider. Data Mining with neural networks for wheat yield prediction. In Petra Perner, editor, Advances in Data Mining (Proc ICDM 2008), 47–56, Berlin, Heidelberg, July 2008. Springer Verlag.
[6] G. Ruß, Data Mining of Agricultural Yield Data: A Comparison of Regression Models, Conference Proceedings, Advances in Data Mining – Applications and Theoretical Aspects, P. Perner (Ed.), Lecture Notes in Artificial Intelligence 6171, Berlin, Heidelberg, 24–37, Springer, 2009.
[7] G. Ruß, A. Brenning, Data Mining in Precision Agriculture: Management of Spatial Information, Conference Proceedings, Computational Intelligence for Knowledge- Based Systems Design, E. H¨ullermeier, R. Kruse, and F. Hoffmann (Eds.), Lecture Notes in Artificial Intelligence 6178, Berlin, Heidelberg, 350–359, Springer, 2010.
[8] A. Mucherino, P. Papajorgji, P.M. Pardalos, A Survey of Data Mining Techniques Applied to Agriculture, Operational Research: An International Journal 9(2), 121–140, 2009.
[9] M. Kovacevic, B. Bajat, B. Gajic, Soil Type Classification and Estimation of Soil Properties using Support Vector Machines, Geoderma 154(3–4), 340–347, 2010.
[10] Cover TM, Hart PE, Nearest Neighbor pattern classification.EEE Trans Info Theory 13(1) : 21-27, 1967.
[11] J. Hartigan, Clustering Algorithms, John Wiles & Sons, New York, 1975.
[12] A. Mucherino, A. Urtubia, Consistent Biclustering and Applications to Agriculture, IbaI Conference Proceedings, Proceedings of the Industrial Conference on Data Mining (ICDM10), Workshop “Data Mining in Agriculture” (DMA10),Berlin, Germany, 105-113, 2010.
[13] A. Mucherino, S. Cafieri, A New Heuristic for Feature Selection by Consistent Biclustering, arXiv e-print, arXiv:1003.3279v1, March 2010.

[14] Jorquera H, Perez R, Cipriano A, Acuna G Short term forecasting of air pollution episodes. In: Zannetti P (eds) Environmental modeling 4. WIT Press, UK, 2001.
[15] Rajagopalan B, Lall U, A K-Nearest Neighbor simulator for daily precipitation and other weather variables. Wat Res Res 35(10) : 3089–3101, 1999.
[16] Tripathi S, Srinivas VV, Nanjundiah RS Downscaling of precipitation for climate change scenarios: a Support Vector Machine approach. J Hydrol 330:621–640, 2006.
[17] Fagerlund S Bird species recognition using Support Vector Machines. EURASIP J Adv Signal Processing, Article ID 8637, p 8, 2007.
[18] Holmgren P, Thuresson T Satellite remote sensing for forestry planning: a review. Scand J For Res 13(1):90–110, 1998.
[19] Verheyen K, Adriaens D, Hermy M, Deckers S High-resolution continuous soil classification using morphological soil profile descriptions. Geoderma 101:31–48, 2001.
[20] Meyer GE, Neto JC, Jones DD, Hindman TW Intensified fuzzy clusters for classifying plant, soil, and residue regions of interest from color images. Comput Electronics Agric 42:161–180, 2004.
[21] Camps-Valls G, Gomez-Chova L, Calpe-Maravilla J, Soria-Olivas E, Martin-Guerrero JD, Moreno J Support Vector achines for crop classification using hyperspectral data. Lect Notes Comp Sci 2652:134–141 , 2003.
[22] D Ramesh1 , B Vishnu Vardhan "Data Mining Techniques and Applications to Agricultural Yield Data" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013
[23] Fathima, G.N., Geetha, R., "Agriculture Crop Pattern Using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Engineering, Vol. 4, Issue 5, pp.781-786, 2014
[24] Veenadhari**,** S., Misra, B., Singh, C.D., "Machine learning approach for forecasting crop yield based on climatic parameters", International Conference on Computer Communication and Informatics, pp.1-5, 2014.
[25] Kaur, M., Gulati, H., Kundra, H., "Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications", International Journal of Computer Applications, Vol. 99– No.12, pp.1-3, 2014.
[26] Bendre, M. R., Thool, R.C., Thool, V. R., "Big Data in Precision Agriculture : Weather Forecasting for Future Farming", 1st International Conference on Next Generation Computing Technologies, pp.744-750, 2015.
[27] Hemageetha, N., "A survey on application of data mining techniques to analyze the soil for agricultural purpose", 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp.3112-3117, 2016.
[28] Sujatha, R., Isakki, P., "A study on crop yield forecasting using classification techniques", International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE), pp.1-4, 2016.
[29] Sellam,V., Poovammal, E., "Prediction of Crop Yield using Regression Analysis", Indian Journal of Science and Technology, Vol. 9(38), pp.1-5, 2016