International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 7
July 2017

Data Mining Techniques Used In Agriculture

Dr. Devesh Katiyar, Department of Computer Science,
Dr. Vinodani Katiyar, Department of Information Technoloyg,
Dr. Shakuntala Misra National Rehabilitation University, Lucknow

Abstract: *We have come a long way where agriculture has evolved from the traditional ways of farming and processing into a high-tech business. Data mininga technique of examining large pre-existing databases in order to generate new information also pays a vital role in the field of Agriculture crop yield analysis. Data mining proves to be fertile ground for future innovations in agricultural statistics.*

Key Words: Data Mining, Agriculture

INTRODUCTION

An interdisciplinary subfield of computer science was evolved in 1960's and initially statisticians used the term data fishing and later in 1990's the term Data Mining was used to describe it. The overall goal of data mining is to extract information from the data sets and to transform it into understandable form involving intersection of artificial Intelligence, Machine Learning, statistics and database systems.In other word we can say data mining to be the analysis step of knowledge discovery in database. In other words, we can say that data mining is mining knowledge from data.

Data mining techniques are used to find patterns, classify records, and extract information from large data sets. These techniques, often used in the private sector for market research, fraud detection, and customer relationship management, can also be used by statistical agencies to analyze their large survey datasets. While large datasets are common in many statistical agencies, data mining techniques have not been widely used to improve the production of official statistics. With large datasets, information is often hidden, but data mining techniques can be used to distill or uncover it. Various techniques can be used to classify data into subsets, predict outcomes based on the data, cluster records into like subgroups, or assign propensity scores for some measure to records.

DATA MINING TECHNIQUES

Data mining techniques are mainly divided in two groups,classification and clustering techniques. Classificationtechniques are designed for classifying unknown samplesusing information provided by a set of classified samples.This set is usually referred to as a training set as it is used totrain the classification technique how to perform itsclassification. Generally, Neural Networks and Support Vector Machines, these two classificationtechniques learn from training set how to classify unknownsamples.

Another classification technique, K- Nearest Neighbor,does not have any learning phase, because it uses the training set every time a classification must be performed. A training set is known, and it is used to classify samples of unknown classification. The basic assumption in the KNearestNeighbor algorithm is that similar samples shouldhave similar classification. The parameter K shows thenumber of similar known samples used for assigning a classification to an unknown sample. The K-NearestNeighbor uses the information in the training set, but it does not extract any rule for classifying the other.

In the event a training set not available, there is no previous knowledge about the data to classify. In this case, clustering techniques can be used to split a set of unknown samples into clusters. One of the most used clustering technique is the K-Means algorithm. Given a set of data with unknown classification, the aim is to find a partition of these in which similar data are grouped in the same cluster. The parameter K plays an important role as it specifies the number of clusters in which the data must be partitioned. 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 clusters. The center of the cluster can be considered as the representative of the cluster, because the center is quite close to all samples in the cluster, and therefore it is similar to all of them. There are some disadvantages in using K-Means method. 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 Modeland BiclusteringTechniques have some applicationsin agriculture or agricultural - related field

APPLICATION OF DATA MINING IN AGRICULTURE

Acquiring knowledge from empirical data often turns out to be tedious and error prone therefore to simplify and smoothen the process many rules and techniques were developed.DSSAT, CROPSYST and GLEAMS are amongst several models which are used for the simulation of soil dynamics. These models are able to simulate the dynamics by applying some soil parameters which are needed to the specific requirement. LL is the lower limit of plant water availability; DUL is the drained upper limit; PESW is the plant extractable soil water are some of the most frequent used parameters.

The K-Means approach is used for classifying soil in combination with GPS based technologies. A very specify and relevant context for reference can be where and independent component analysis technique was used for mining spatio temporal data was applied to find patterns in weather at North Atlantic Oscillation(NAO). Applied data warehousing and Online Analytical Processing(OLAP) technologies are of uttermost importance in evaluating agricultural data. A data warehouse provides aflexible yet efficient and reliable storage structure for vast amount of data while OLAPtechniques provide mechanisms for ad hoc and in depth analysis of this data. Traditional analytical tools and database techniques may not succeed here due their rigid nature. Techniques used in their work are equally applicable at any geographic location provided that related data is available.

Studies had shown that how data mining integrated agricultural data including pestscoating, pesticide usage and meteorological recordings is useful for optimization of pesticide usage. Unsupervised clustering of the data was performed first through Recursive Noise Removal (RNR). These clusters reveal interesting patterns of farmer practices along with pesticide usage dynamics and hence help identify the reasons for this pesticide abuse mechanism of performing the mapping from nominal to numeric values(actually ranking)based on the transmittance as well as the statistical properties of the plants were proposed. Spectral analysis (using chemical means) is a tedious and time-consuming process, thus difficult to repeat, each and every time, for classification of (numerically) unclassified agricultural varieties. Supporting statistical method was also proposed based on linear regression curve fitting using normalized nominal attributes. Subsequently a rank is assigned to the variety based on its R2 value and slope of the plot.

Effect of pesticides on humans can’t be directly checked because of the poisonous nature of pesticides, therefore the usage of pesticides on cotton crop has been taken into consideration for the purpose. The COFClustering Tool cannot only be used forpesticide data, but also possesses the flexibility to deal with any numeric data. Spatial data mining methods to extract interesting and regular knowledge from large spatial databases of agriculture were studied aiming at discerning trends in agriculture production with reference to the availability of inputs. The predicted and real vs. Counter graph illustrated how closely the poly analyst prediction follows the actual value of the attribute over the range of the dataset. Applying the data mining techniques to agriculture the target for different food grains can be achieved. Their study demonstrated the scope for application of spatial mining tools for a utility study and analysis. The specific application of Polyanalyst gave a clear scope for evaluation and comparison of predicted and real values. Influence of climatic factors on major kharifand rabi crops production in Bhopal District ofMadhya Pradesh State was studied. Thefindings of the study revealed that the decision tree analysis indicated that the productivity of soybean crop was mostly influenced by Relative humidity followed by rainfall and temperature. Relative humidity and Evaporation. For Wheat crop the analysis indicated that the productivity is mostly influenced by Temperature followed by Relative humidity and Rainfall. The findings of decision tree were confirmed from Bayesian classification. The decision tree in the study area fast to execute and much to be desired as representations of knowledge interpretations. The rules formed from the decision tree are helpful in identifying the conditions responsible for the high or low cropproductivity.Powdery Mildew of Mango a devastating disease of mango was predicted using Decision Tree induction, Rough Sets (RS) and hybridized Rough Set based Decision Tree Induction (RDT) in comparison with the standard Logistic Regression (LR) method. The induction algorithms shown better performance over logistic regression.

A web based expert information system based on ID3 algorithm were studied in which an expert system provides advisory services to Tomato growers regarding pests, diseases and their control measures. The web based system has also provision for the growers to interact with other growers on the management practices of tomato crop cultivation.

Fruit defects are often recorded (for a multitude of reasons, sometimes for insurance reasons when exporting fruit overseas). It may be done manually or through computer vision (detecting surface defects when grading fruit). Spray diaries are a legal requirement in many countries and at the very least record the date of spray and the product name. It is known that spraying can have affect different fruit defects for different fruit. Fungicidal sprays are often used to prevent rots from being expressed on fruit. It is also known that some sprays can cause rusting on apples. Currently much of this knowledge comes anecdotally, however some efforts have been in regards to the use of data mining in horticulture.

Wine is widely produced all around the world. The fermentation process of the wine is very important, because it can impact the productivity of wine-related industries and also the quality of wine. If the fermentation could be categorized and predicted at the early stages of the process, it could be altered in order to guarantee a regular and smooth fermentation. Fermentations are nowadays studied by using different techniques, such as, for example, the k-means algorithm,[4] and a technique for classification based on the concept of blustering.[5] Note that these works are different from the ones where a classification of different kinds of wine is performed. See the wiki page wine for more details.

The detection of animal's diseases in farms can impact positively the productivity of the farm, because sick animals can cause contaminations. Moreover, the early detection of the diseases can allow the farmer to cure the animal as soon as the disease appears. Sounds issued by pigs can be analyzed for the detection of diseases. In particular, their coughs can be studied, because they indicate their sickness. A computational system is under development which is able to monitor pig sounds by microphones installed in the farm, and which is also able to discriminate among the different sounds that can be detected.[6]

Before going to market, apples are checked and the ones showing some defects are removed. However, there are also invisible defects that can spoil the apple flavor and look. An example of invisible defect is the water core. This is an internal apple disorder that can affect the longevity of the fruit. Apples with slight or mild water cores are sweeter, but apples with moderate to severe degree of water core cannot be stored for any length of time. Moreover, a few fruits with severe water core could spoil a whole batch of apples. For this reason, a computational system is under study which takes X-ray photographs of the fruit while they run on conveyor belts, and which is also able to analyses (by data mining techniques) the taken pictures and estimate the probability that the fruit contains water cores.

Recent studies by agriculture researchers in Pakistan (one of the top four cotton producers of the world) showed that attempts of cotton crop yield maximization through pro-pesticide state policies have led to a dangerously high pesticide use. These studies have reported a negative correlation between pesticide use and crop yield in Pakistan. Hence excessive use (or abuse) of pesticides is harming the farmers with adverse financial, environmental and social impacts. By data mining the cotton Pest Scouting data along with the meteorological recordings it was shown that how pesticide use can be optimized (reduced). Clustering of data revealed interesting patterns of farmer practices along with pesticide use dynamics and hence help identify the reasons for this pesticide abuse.

SOME OF THE DATA MINING SOFTWARES USED

WEKA: - Weka supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection.

Data Detective: -The powerful yet easy to use data mining platform and the crime analysis software of choice for the Dutch police.

DataLab (more focus on marketing):- A complete and powerful data mining tool with a unique data exploration process, with

a focus on marketing and interoperability with SAS.

GhostMiner :- Complete data mining suite, including k-nearest neighbors, neural nets, decision tree, neurofuzzy, SVM, PCA, clustering, and visualization.

WITNESS Miner: -A graphical data mining tool with decision trees, clustering, discretisation, feature subset selection, and more.

Advanced Miner Professional: - Provides a wide range of tools for data transformations, Data Mining models, data analysis and reporting.

Angoss Knowledge Studio: - A comprehensive suite of data mining and predictive modeling tools; interoperability with SAS and other major statistical tools.

GainSmarts: - Uses predictive modeling technology that can analyze past purchase, demographic, and lifestyle data, to predict the likelihood of response and develop an understanding of consumer characteristics.

XLMiner: -Data Mining Add-In For Excel.

CONCLUSION

As we have been continuously striving for the better result in the agriculture field. This new deviation of Data Mining in the field of agriculture would prove to be of great help to grow up in the coming time.The multidisciplinary approach of integrating computer science with agriculture will help in forecasting/managing agricultural crops effectively.

REFERENCES

1. Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner.Estimation of neural network parameters for wheat yieldprediction. In Max Bramer, editor, Artificial Intelligence inTheory and Practice II, volume 276 of IFIP InternationalFederation for Information Processing, 109–118. Springer, July2008.
2. Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner.Optimizing wheat yield prediction using different topologies ofneural 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.
3. A. Mucherino, P. Papajorgji, P.M. Pardalos, A Survey of DataMining Techniques Applied to Agriculture, OperationalResearch: An International Journal 9(2), 121–140, 2009.
4. Mucherino, A.; Papajorgji, P.J.; Pardalos, P. (2009). Data Mining in Agriculture, Springer.
5. Hill, M. G.; Connolly, P. G.; Reutemann, P.; Fletcher, D. (2014-10-01). *"The use of data mining to assist crop protection decisions on kiwifruit in New Zealand"*. Computers and Electronics in Agriculture. 108: 250–257. *doi*:*10.1016/j.compag.2014.08.011*.
6. Mucherino, A.; Urtubia, A. (2010). "Consistent Biclustering and Applications to Agriculture". IbaI Conference Proceedings, Proceedings of the Industrial Conference on Data Mining (ICDM10), Workshop Data Mining in Agriculture (DMA10), Springer: 105–113.

7. Chedad, A.; Moshou, D.; Aerts, J.M.; Van Hirtum, A.; Ramon, H.; Berckmans, D. (2001). "Recognition System for Pig Cough based on Probabilistic Neural Networks". Journal of Agricultural Engineering Research 79(4): 449–457.
8. Abdullah, Ahsan; Hussain, Amir (2006). *"Data Mining a New Pilot Agriculture Extension Data Warehouse"* (PDF). Journal of Research and Practice in Information Technology, Vol. 38, No. 3, August 2006: 229–249.
9. Jain Rajni, Minz, S., V. Rama Subramaniam. 2009. "Machine learning for forewarning cropdiseases". J. Ind. Soc. Agri. Stat. 63(1): pp. 97-107.
10. Jones JW, Tsuji GY, Hoogenboom G, HuntLA, Thornton PK, Wilkens PW, Imamura DT,Bowen WT, Singh U., (1998), "Decision supportsystem for agrotechnology transfer: DSSAT v3".In: Tsuji GY, Hoogenboom G, Thornton PK (eds) ,"Understanding options for agriculturalproduction". Kluwer Academic Publishers,Dordrecht, pp 157–177
11. JianlinJi Dan, Qiu Chen, Jianping Chen, LiHePeng , 2010. "An improved decision treealgorithm and its application in maize seedbreeding". Sixth Internation Conference on NaturalComputation, held at Yantai, Shandon 10-12thJanuary. pp. 117-121.
12. *S.S.Baskar, L.Arockiam, V.Arul Kumar, L.Jeyasimman, "Brief Survey of Application of Data Mining Techniques to Agriculture", Agricultural Journal 5(2): 116:118,2010 ISSN:1816- 9155*
13. ElodieVintrou, Dino Ienco, AgnèsBégué, andMaguelonneTeisseire, "Data Mining, A PromisingTool for Large-Area Cropland Mapping", IEEEjournal of selected topics in appliedearth observations and remotesensing, vol. 6, no. 5, october 2013
14. Latika Sharma, Nitu Mehta, "Data MiningTechniques: A Tool For Knowledge ManagementSystem In Agriculture", International Journal OfScientific& Technology Research Volume 1, Issue 5,June 2012
15. D.Rajesh, "Application of Spatial DataMining for Agriculture", International Journal ofComputer Applications (0975 – 8887) Volume 15–No.2, February 2011.
16. Raorane A.A and Kulkarni R.V, "Review- Roleof Data Mining in Agriculture", InternationalJournal of Computer Science and InformationTechnologies, Vol. 4 (2) , 2013, 270 – 272
17. Han-Wen Hsiao,Meng-Shu Tsai, And Shao-Chiangwang, "Spatial Data Mining of ColocationPatterns for Decision Support in Agriculture",Asian Journal of Health and Information Sciences,Vol. 1, No. 1, pp. 61-72, 2006