International Journal of Computer Science and Engineering Communications,
Vol.4, Issue.2, Page.1311-1313, (2016)
www.scientistlink.com

Analysis of Data Mining Techniques for Agriculture Data

E.Manjula*, S.Djodiltachoumy
[1]Pachaiyappas College, India
**manjulababu123@gmail.com*

Abstract: *Data mining is the process of extracting important and useful information from large sets of data. The goal of the data mining process is to extract knowledge from an existing data set and transform it into a unique human understandable format for some advance use. Data mining in agriculture is in relation to novel research field. Some efficient techniques can be developed and tailored for solving complex agricultural problems using data mining. Agriculture and allied activities constitute the single largest component of India's gross domestic product, contributing nearly 25% of the total and nearly 60% of Indian population depends on this life's precise profession. Due to vagaries of climate factors the agricultural productivities in India are continuously decreasing over a decade. The reasons for this were studied mostly using regression analysis. In this paper an attempt has been made to compile the research findings of different researchers who used agriculture data. This paper summarizes the application of data mining techniques such as k-means, bi clustering, k nearest neighbor, Neural Networks, Support Vector Machine and Naïve Bayes Classifier in the agriculture field.*

Keywords: *Agriculture, Data mining, k-means, bi clustering, k nearest neighbor, Artificial Neural Network, Support Vector Machine, Naïve Bayesian Classifier.*

Introduction: Data mining is the process of extracting important and useful information from a large sets of data [20]. The goal of the data mining process is to extract knowledge from an existing data set and transform it into human understandable format for advance use. The intention of this paper is to give details about different data mining techniques and their applications to agricultural related areas. Different techniques have been proposed for mining data over the years. The 10 most used data mining techniques are discussed in a recent paper [22]. Among the algorithms we can concentrate on the most used data mining techniques for solving agricultural problems.

Data Mining Techniques

Data mining techniques can be divided in two groups: Classification and Clustering techniques. Classification techniques are designed for classifying unknown samples using information provided by a set of classified samples. This set is usually referred to as a training set, because in general it is used to train the classification technique how to perform its classification. For instance, Neural Networks and Support Vector Machines exploit training sets for tuning their parameters in order to solve a particular classification problem. In other words, these two classification techniques learn from a training set how to classify unknown samples. Another classification technique the k nearest neighbor does not have any learning phase, because it uses the training set every time a classification must be performed. In the event a training set is not available, there is no previous knowledge about the data to classify. In this case, clustering technique can be used to split a set of unknown samples into cluster. One of the most used clustering techniques is the k – means method. This paper mainly focuses on the most used techniques in agriculture related fields.

Applications of Data Mining Techniques in Agriculture

There are number of studies which have been carried out on the application of data mining techniques for agricultural data sets. Naïve Bayes data mining technique is used to classify soils that analyze large soil profile experimental data sets [3]. Decision tree algorithm in data mining is used for predicting soil fertility [4]. By using clustering techniques based on Partitioning Algorithms and Hierarchical Algorithm the land utilization for agriculture and non-agriculture areas for the past ten years have been determined [5].

ISBN: 978-93-5254-256-7

Methodology	Applications
K-means	Forecasts of pollution in atmosphere Classifying soil in combination with GPS
k-nearest Neighbor	Simulating daily precipitations and other weather variable
Support Vector Machine	Analysis of different possible change of the weather scenario
Decision Tree Analysis	Prediction soil dept
Unsupervised Clustering	Generate cluster and determine any existence of pattern
WEKA Tool	Classification system for sorting and grading mushrooms.

Table: 1 Data mining methodologies and its use in Agriculture domain

Crop yield has been analyzed using k-means approach [6]. The k-means algorithm is also used for soil classification using GPS-based technology, classification of plant, soil and residue regions of interest by color images, grading apples before marketing, monitoring water quality changes, detecting weeds in agriculture, [8,9,10,11,12]. The prediction of wine fermentation problems can also be performed by using a k-means approach. Knowing in advance that the wine fermentation process could get stuck or be slow can help the enologist to correct it and ensure a good fermentation process [13]. The k – nearest algorithm is used in simulating daily precipitations and other weather variables and also for estimating soil water parameter and climate forecasting [14,2]. The neural network is used in prediction of flowering and maturity dates of soybean and is also used in forecasting of water resources variables [15,16]. The application of support vector machine is the crop classification [17] .The support vector machine technique is also used in the analysis of the climate change scenarios [18]. Some data mining methodology which are used in agricultural domain is shown in table 1. Soil classification has been done using Naïve Bayes classifier [7].

Conclusion

Use of information technology in agriculture can change the scenario of decision making and farmers can yield in a better way. Several data mining techniques related to agriculture domain has been discussed in this paper.

References

[1] Hetal Patel & Dharmendra Patel. (2014). A Brief survey of Data Techniques Applied to Agricultural Data, International Journal of Computer Applications, 95(3).

[2]Mucherino, A., Papajorgji, P., & Pardalos, P.(2009). Data mining in agriculture (Vol. 34). Springer.

[3] Bhargavi, P, & Jyothi, S. (2009). Applying Naive Bayes data mining technique for classification of agricultural land soils. International journal of computer science and network security, 9(8), 117-122.

[4]Jay Gholap. (2012). Performance tuning of j48 algorithm for prediction of soil fertility. Asian Journal of Computer Science And Information Technology 2: 8 (2012) 251– 252.

[5] Megala, S., & Hemalatha, M. (2011). A Novel Datamining Approach to Determine the Vanished Agricultural Land in Tamilnadu. International Journal of Computer Applications, 23.

[6] D Ramesh, B Vishnu Vardhan, (2013). Data Mining Techniques and Applications to Agricultural Yield Data. International Journal of Advanced Research in Computer and Communication Engineering 2(9).

[7] V. Ramesh and K. Ramar, 2011. Classification of Agricultural Land Soils: A Data Mining Approach. Agricultural Journal, 6: 82-86.

[8] Verheyen, K., Adriaens, D., Hermy, M., & Deckers, S. (2001). High-resolution continuous soil classification using morphological soil profile descriptions. Geoderma, 101(3), 31-48.

[9] Meyer, G. E., Camargo Neto, J., Jones, D. D., & Hindman, T. W. (2004). Intensified fuzzy clusters for classifying plant, soil, and residue regions of interest from color images. Computers and electronics in agriculture, 42(3), 161-180.

[10] Leemans, V., & Destain, M. F. (2004). A real-time grading method of apples based on features extracted from defects. Journal of Food Engineering, 61(1), 83-89.
[11] K.A. Klise and S.A. McKenna.(2006). Water Quality Change Detection: Multivariate Algorithms. Proceedings of SPIE 6203, Optics and Photonics in Global Homeland Security II, T.T. Saito,D. Lehrfeld (Eds.)
[12] Tellaeche, A., BurgosArtizzu, X. P., Pajares, G., & Ribeiro, A. (2007). A vision-based hybrid classifier for weeds detection in precision agriculture through the Bayesian and Fuzzy k-Means paradigms. In Innovations in Hybrid Intelligent Systems (pp. 72-79). Springer Berlin Heidelberg.
[13] Urtubia, A., Pérez-Correa, J. R., Soto, A., & Pszczolkowski, P. (2007). Using data mining techniques to predict industrial wine problem fermentations. Food Control,18(12), 1512-1517.
[14] Rajagopalan, B., & Lall, U. (1999). A k–nearest-neighbor simulator for daily precipitation and other weather variables. WATER RESOURCES RESEARCH,35(10), 3089-3101.
[15] Elizondo, D. A., McClendon, R. W., & Hoogenboom, G. (1994). Neural network models for predicting flowering and physiological maturity of soybean. Transactions of the ASAE (USA).
[16] Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental modeling & software, 15(1), 101-124.
[17] Camps-Valls, G., Gómez-Chova, L., Calpe-Maravilla, J., Soria-Olivas, E., Martín-Guerrero, J. D., & Moreno, J. (2003). Support vector machines for crop classification using hyper spectral data. In Pattern recognition and image analysis(pp. 134-141). Springer Berlin Heidelberg.
[18] Tripathi, S., Srinivas, V. V., & Nanjundiah, R. S. (2006). Downscaling of precipitation for climate change scenarios: a support vector machine approach. Journal of Hydrology, 330(3), 621-640.
[19] Mucherino, A., Papajorgji, P., & Pardalos, M.P.(2009). A Survey of Data Mining techniques applied to agriculture. Spinger.
[20] Abello J, Pardalos PM, Resende M (2002) Handbook of massive data sets. Kluwer, New York.
[21] Mucherino A, Papajorgji P, Pardalos PM (2009) Data mining in agriculture. Springer, New York (in press).
[22] Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou Z-H, Steinbach M, Hand DJ, Steinberg D (2008) Top 10 algorithms in data mining. Knowl Inf Syst 14:1–37.
[23] Yethiraj N G (2012) Applying Data Mining Techniques In the field of Agriculture and Allied Sciences, International Journal of Business Intelligent.