

Scope of Data Mining Concepts in Agricultural Sector

Sharon James, Sini Thomas , Sanal Jacob

Saintgits College Of Applied Sciences Kottayam ,Kerala

Abstract - Agriculture plays an integral role in economy of many countries like India .It is not only a source of food but also a sector that employs billions of people. This Scenario rise the need to apply new methods and techniques to improve the yield from agriculture .This paper gives an insight to the usage of data mining concepts that would make a sound approach to agricultural sector by addressing the issues faced and generating reliable predictions to increase productivity.

Keywords - Data Mining ,Knowledge Discovery Database(KDD),K-means algorithm ,K-nearest neighbor algorithm , Artificial Neural Networks, Support Vector Machines

1. Introduction

Agriculture is the practice of cultivating crops and rearing the livestock. Agriculture plays a key role in maintaining food security. But still factors like growing population, adverse changes in climate, over exploitation of the natural resources are the pressing problems globally that require a systematic overhaul to ensure food security. Experts have been in an effort to increase agricultural yield by analysing the vast and diverse raw data from the agricultural sector to find useful information and the hidden relations they hold. Data warehouses stores the huge data from agricultural sector that is required to perform the data analysis

Data Mining is solely for the exploration of useful information from large amount of data. It is also known as Knowledge Discovery Databases or KDD. Two categories of Data mining tasks are predictive and descriptive. Predictive helps to make predictions based on data whereas general properties of data in database is characterized in descriptive mining tasks. The primary objective of Data mining is to make large sets of data into readily understandable format.

Data Mining in Agricultural sector is an emerging concept with immense potential. In countries like India, agriculture contributes upto 50% of the employment and about 17-18% of the country's GDP. But the

productivity of the sector remains low when compared with other sectors. So the Governmental Organizations have focused on improving the productivity of the agricultural sector.

Agricultural production depends on various factors such as weather conditions ,diseases,pest ,planning of harvest etc. So this paper attempts to compile different data mining techniques and implement it on the various factors that affect agriculture so as to improve the yield.

2. Techniques used in Data mining

The Book ‘Data Mining in Agriculture’ by Mucherino A.,Papajorgi and Pardalos P (2009) classifies data mining techniques as classification techniques, clustering technique and statistical approaches such as association rules and regression.

This paper discusses about the various classification techniques such as K-nearest neighbor approach, Support Vector machines, Artificial Neural Networks and clustering techniques like K-means clustering and bi-clustering techniques.

2.1 Classification Techniques

Classification in Data mining is used to generate models that predict the trends in future from the data. Classification is the process by which a model learns from the training set of data and then discrete class labels of new samples are predicted. It is a supervised learning task in which machines are trained using labeled set of data and according to this input fed, they learn and works to yield the data desired.

K-Nearest Neighbour, Artificial Neural Network and Support Vector machines are classification techniques.

2.1.1 K-Nearest Neighbour.

K-Nearest Neighbour method is a classification algorithm in which training set is classified to find patterns in unknown sample by considering only K-data instance that are similar. The nearest neighbor is defined according to Euclidean distance.

2.1.2 Artificial Neural Network

An Artificial Neural Network, popularly known as Neural Network is a model in Mathematics inspired by the neural network in the Human body[13]. Each neuron in Neural System perform simple task and they are combined to perform complex activities .In ANN, nodes represents the neuron and links between

neurons shows the interactions among the two. It works typically like a black box as a hidden layer finding the input output relations in data to draw useful information.

2.1.3 Support Vector Machines

It was in early 1960's the Support Vector Machines (SVM) were suggested for the first time[2]. SVMs classify samples of information. It could be applied on linear and non linear data. In SVM method a hyper plane separates the two classes of samples, in which linear separability of classes are considered. Hyper planes in SVM are found using support vectors and margins. It can be used for both classification and prediction of data.

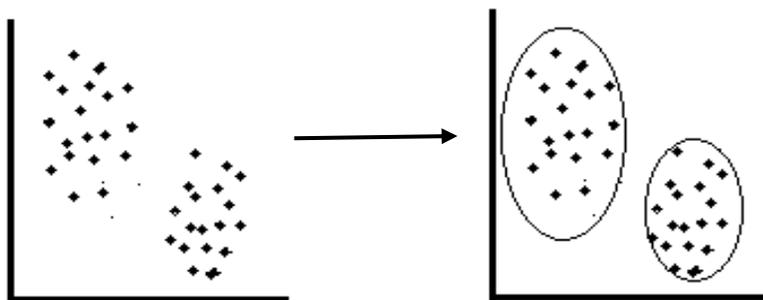
2.2 Clustering Technique

Clustering is a technique in which data records with similar data are arranged into clusters and those differ are grouped into different clusters. This helps in analysing the trends in information.

K-Means algorithm and bi-clustering are some clustering techniques.

2.2.1 K-Means Algorithm

It is one of the extensively used clustering techniques in data mining. This algorithm finds the mean value of the cluster by centroid of the cluster. Initially it takes K objects in datasets that contain N objects, which represents the mean of the cluster. All the other objects are clustered according to the similarity and the algorithm iterates with the difference in the cluster and finds the mean of each cluster.



K-Means

2.2.2 Bi-clustering

Bi-clustering is a clustering technique for classifying information from the training set in a data set. Under bi-clustering approach, a data set is analyzed to find properties which are similar in samples and it is continuously grouped into partitions called bi-clusters. This partitioning performed on the data set is called bi-clustering.

3. Application of Data Mining Techniques

Data mining techniques are studied widely to find the application in the field of agriculture.

Yield from crops are analyzed using K- Means approach. This technique is also used to classify soil with the help of GPS technology, checks the quality of water, forecasting of pollution in atmosphere and even used in processing of wine to ensure adequate fermentation.

K-Nearest Neighbor or KNN algorithm is applied to predict weather conditions and rainfall. This field has been further enhanced by the development of remote satellite technology so that the scientist can use these data mining techniques to forecast the results to the farmers via Kissan portals.

The Neural Network technique is used to understand the flowering and maturity time period of soya beans. ANNs are combined with the Internet of Things(IoT) to analyse the chance of floods in river basins. This advancement can benefit countries that experience flood every year by reducing the damages caused by floods in agriculture. Even the diseases in animals can be detected by the sound that animals produce in farms. This helps to cure those diseases at an early stage itself.

Digital Images of plants are analysed by support Vector Machines[11] to differentiate weeds from crops and found to have 97% accuracy in results obtained . This technology could help farmers to control weeds with minimal efforts. Yield from crops can be predicted through Support Vector Machines or SVMs, the spatial autocorrelation methods have further led to the advancement of this field. Cultivation of crops primarily depends on weather especially rainfall. So prediction regarding rainfall by SVMs aid the farmers in decision making.

4. Literature Survey about the Research conducted on Coriander using K-Means Algorithm

This paper gives brief review about usage of clustering technique in Data Mining such as K-Means Approach [7]. Coriander is traditionally used to enrich the taste and aroma of food. There are about 9 varieties of coriander grown in Indian subcontinent. Coriander plants require loamy well drained soil with pH value of 6-8 and temperature range of 20-25°C. Coriander cultivation is risky and non-profitable due to improper modelling of farming methods. This is tackled by K-means algorithm. It is applied in large data set to maintain appropriate conditions such as fertility of soil and suitability of climate which are the crucial factors in coriander cultivation. It is also used to calculate the yield from the coriander leaves.

5. Challenges of applying Data mining in Agricultural Sector

There are several problems that hinder the usage of data mining in agricultural sector. There is vast and heterogeneous raw data that is obtained from various sources which are prone to missing values, invalid and inconsistent data. Further it may have errors due to inaccurate assessments. Data Storage construction in Data mining requires a lot of time and money. These complexities in agricultural data lays hurdles for applying data mining techniques.

Secondly, Data mining techniques are not devised for a specific field. It could be applied on any kind of data except transactional databases and data warehouses, which use specific techniques. Hence the data mining in agriculture generates more of generalized results. And those results derived from data mining could be understood and utilized by those with expert knowledge in that field.

Finally in a country like India where most farmers cultivate on small patches of land and rural areas have no access to internet and the application of these techniques remains as a challenge.

6. Conclusion

Agricultural sector could be revived in sustainable way by the ad hoc intervention of information technology. Data mining techniques play a crucial role in making decisions regarding the issues in the field of agricultural and to derive effective solutions to declining productivity. As we have seen there are several factors that hinder the application of data mining techniques, yet these approaches in agricultural sector holds an immense scope.

7. References

- [1] Han, J, Kamber, M., & Pei, J. (2006). Data mining: concepts and techniques. Morgan kaufmann
- [2] Mucherino, A., Papajorgji, P., & Pardalos, P. (2009), Data mining in agriculture (Vol. 34), Springer.
- [3] Raorane A.A and Kulkarni R.V, Review- Role of Data Mining in Agriculture, International Journal of Computer Science and Information Technologies,
- [4] S.S.Baskar, L.Arockiam, V.Arul Kumar,L.Jeyasimman, Brief Survey of Application of Data Mining Techniques to Agriculture
- [5] D Ramesh, B Vishnu Vardhan, Data Mining Techniques and Applications to Agricultural Yield Data, International Journal of Advanced Research in Computer and Communication Engineering
- [6] Hetal Patel, Dharmendra Patel A brief survey of data mining technique applied to agricultural data.
- [7] Research of Production and Growth of Coriander in Various Seasons using K-Means Algorithm by N.Santhosh, R.Shankar and R.Narendranath, K.Srinivasan
- [8] Data mining of agricultural yield Data: A comparison of regression models” George RuB express.
- [9] Lei Shi¹, Qiguo Duan², Xinming Ma¹, and Mei Weng¹, The Research of Support Vector Machine in Agricultural Data Classification
- [10] Classification algorithm in Data Mining : An Overview by S.Neelamegam, Dr. E. Ramaraj.
- [11] F Ahmed, HA Al-Mamun, ASMH Bari, E Hossain, Classification of crops and weeds from digital images: A support vector machine approach.
- [12] B. Milović¹ and V. Radojević², Application of Data Mining in Agriculture
- [13] Surabhi Sindhu, Divya Sindhu, Role of Data mining Techniques in Agriculture Improvements
- [14] https://en.wikipedia.org/wiki/data_mining_in_agriculture
- [15] <http://Javatpoint.org>

