RESEARCH PAPER | Engineering | Volume : 5 | Issue : 4 | April 2015 | ISSN - 2249-555X

Agricultural Recommender Using Data Mining Techniques

KEYWORDS: Data Mining, Association Rule, Genetic Algorithm, Agriculture recommender.

Mr.Omkar B. Bhalerao
Computer Science and Engineering Department
Walchand Institute of Technology, Solapur India.

Prof. L. M. R. J. Lobo
Computer Science and Engineering Department,
Walchand Institute of Technology, Solapur. India.

ABSTRACT *The Agriculture sector is a backbone of Indian economy and its lifeline for farmers. But, today's Agriculture is not the kind of farming as our forefathers did. The strong climatic changes due to factors like global warming have caused a difficulty to predict the climatic condition properly. Novices or beginner in the agriculture sector, need to take a decision on growing a new crop in his field. The main problem faced by this farmer is to decide, "Which crop to select, So that the production will improve?" By understanding the soil condition it is possible to provide a proper crop recommendation to a farmer; Farmers will able to maximize the use of their land by planting the right crop at a right place which will generate an improved yield. The digital world has brought a dramatic change due to influence in the IT field at each and every sector. Recent technologies are able to provide a lot of information on agriculture related activity, with which we are able to analyze agricultural data in order to find out important information and knowledge. This is extracted with the major goal of increasing profitability in agriculture by applying various data mining techniques.*

Introduction

I n this high technical era, nowadays all fields are computerized. Agriculture also having a large amount of data becomes a candidate for data mining. If we apply the proper technique on it this data no longer remains only pieces of data but, it gives lot of pattern from it. In this we need to make inferences from immense data so that we can make decisions driven by knowledge. Various factors which affect the production of crops like soil type, crop price and other factors are taken into consideration.

Data mining is the process of knowledge discovery in database. It is art and science of intelligent analysis of large data sets for meaning and previously unknown insights and is nowadays actively applied in a wide range of disciplines related to agriculture. With the help of Knowledge Discovery in Database (KDD) and data mining we extract the meaningful data sets from the large amount of data. So, on the large data sets when we applied data mining techniques then it gives results into improved quality of mined data. Data mining is popularly known as "Knowledge Discovery in Database (KDD)".

Data mining in agriculture is an important research field. Data mining tools are powerful generating rules from vast & diversified data such as in agriculture datasets which are in a large amount. Generally, data mining is the process of analyzing data from different prospective and summarizing it into useful information.

In the data mining there are various methods which are applied over the huge amoung of data and we got some pattern or knowledge from it.For optimization of solution or result we uses Genetic algorithm.In the data mining technique we can use various techniques such as Neural Network,Association Rules,k-means,k-nearest neighbour,ID3 algorithm. Genetic Algorithm is a randomized algorithm that could be run for a very long time to obtain an optimal solution.

Association Rule Mining

Association Rule Mining is the process of finding new interesting Correlations , frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories[1].In data mining, association rule mining is an important and easy method to find frequent item sets from large dataset.

It is intended to identify strong rules discovered in databases using two different measures of interestingness. The first one is support which generates frequent item set from the provided database and the other one is confidence which is focuses on rule generation.

Frequent item sets- A set of attributes is termed as frequent item set if the occurrence of the set within the database is more than a user given threshold.

Support- Support determines how often a given rule is applicable to a given data set.

Confidence- Confidence determines how frequently items in Y appear in transactions that contain X.

$$\text{Support, } s(X \longrightarrow Y) = \frac{\sigma(X \cup Y)}{N};$$
$$\text{Confidence, } c(X \longrightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}.$$

Where, X and Y disjoint item set.

Genetic Algorithm

Genetic Algorithms (GA) are direct, parallel method for global search and optimization. GA is one of the most commonly used Evolutionary Algorithms (EA). The Genetic Algorithms are direct and efficient method for optimization. As they use populations with allowed number of solutions (individuals), they are added in the group of parallel algorithms. Main ingredients of GA are Chromosomes, Selection, Recombination and Mutation.

Selection

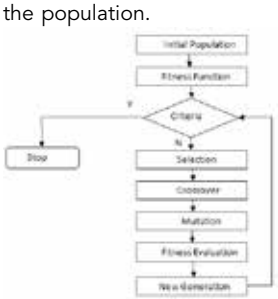
- During each successive generation, a proportion of the existing population is selected to breed a new generation. Fitness-based process is used to select individual solutions where fitter solutions (as measured by a fitness function) are typically more likely to be selected. At this stage elitism could be used – the best n individuals are directly transferred to the next generation. The elitism ensures,

Volume : 5 | Issue : 4 | April 2015 | ISSN - 2249-555X

that the value of the optimization function cannot get worst (once the extremum is reached it would be kept).

Crossover – The most common type is single point crossover. In single point crossover, we choose a locus point at which you swap the remaining alleles from one parent to the other. The children take one section of the chromosome from each parent. Chromosome is broken based on the randomly selected crossover point. This particular method is called single point crossover because only one crossover point exists. Sometimes only one child is created, but generally both offspring are created and put into the new population. Crossover does not always occur. Sometimes, based on a set probability, no crossover occurs and the parents are copied directly to the new population.

Mutation – After selection and crossover, we have a new population full of individuals where some are directly copied, and others are produced by crossover. In order to ensure that the individuals are not all exactly the same, we allow a small chance of mutation. We go through all the alleles of all the individuals, and if that allele is selected for mutation, we either change it by a small amount or replace it with a new value. Mutation is fairly simple. Mutation is, however, vital to ensuring genetic diversity within the population.



Basic block diagram of Genetic Algorithm is:
Figure 1: Genetic Algorithm
Genetic Algorithm is a randomized algorithm that could be run for a very long time to obtain an optimal solution.

LITERATURE REVIEW

In the previous research, different techniques were presented for historical agriculture data analysis-

Lida Xu, Ning Liang, and Qiong Gao [3] proposed a systematic approach based on integrated information systems (IISs) for agricultural ecosystem management is proposed. In this paper, they extracts data on terrain, land use, planting, and others, and integrates them for the purpose of agricultural and ecosystem management. In short in this paper they concludes that, for effective management of agriculture and ecosystems, a systematic approach is essential in which Integrated Information Systems play a crucial role.

S.Veenadhari, Dr.Bharat Misra, Dr.C.D.Singh [11] presented Data mining techniques for predicting crop productivity. In that they, attempted to review the research studies on application of data mining techniques in the field of agriculture. This is the review article in which it gives various techniques which predict crop productivity.

Arvind Jaiswal, Gaurav Dubey[12] find out the best association rules and their optimization using genetic algorithm. They proposed genetic algorithm based method for finding frequent item sets. Repeatedly transforms the population by executing the following various steps: Fitness Evaluation: The fitness (i.e., an objective function) is calculated for each individual. Selection: Individuals are chosen from the current population as parents to be involved in recombination. Recombination: New individuals (called offspring) are produced from the parents by applying genetic operators such as crossover and mutation. Replacement: Some of the offspring are replaced with some individuals (usually with their parents).

Sumitha Thankachan, Dr.S.Kirubakaran [14] worked on E-Agriculture Information Management System. They proposed some news or information about agriculture SMS passed on daily basis. They also proposed system for information passed on seasonal basis as well as other details information regarding agriculture. In methodology, they explain as in the database they have information of farmer's details and crops details. Depending upon the climatic condition and other agriculture parameter they inform about fertilizers, price in market etc information sent to that farmer through SMS via SMS Gateway. This paper estimates the benefits of the Indian farmers if the market and weather information is delivered to their mobile phones.

In literature review we have reviewed different ideas of Integrated IT for Agriculture applications & different machine learning technique for Association rule mining and genetic algorithm.

METHODOLGY

Our proposed methodology is a two stage model. In first stage we apply association rule mining on the agriculture historical data and generate rules from frequent item sets by applying the proper support and confidence for each rule. The user then gives a minimum support and confidence and based on this initial best rules that form the initial population for GA are extracted. In the second stage, we apply Genetic algorithm to optimize the initial population rules which we get from association rule mining. So that, we will get best rules that predict output as an optimized agriculture crop.

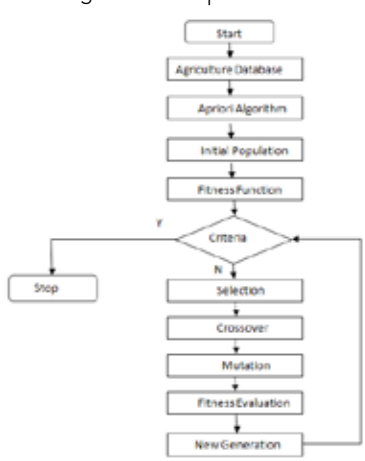


Figure 3: Architecture
The figure 3 shows, the first stage of proposed methodology where we are going to apply apriori algorithm on agriculture database. After applying apriori algorithm we will find the frequent item sets with minimum support, which we will send to Genetic algorithm as an input which initializes its population with frequent item sets. Then GA does the selection, crossover & mutation on population & returns the best population. After that, input the termination condition for genetic algorithm & it will test for the desired output .If desired output is found then it stops Genetic Algorithm otherwise GA continues until termination condition is met.

Volume : 5 | Issue : 4 | April 2015 | ISSN - 2249-555X

IV. CONCLUSION

Each and every sector in this digital world is undergoing a dramatic change due to the influence of IT field. But, till date not much work has been done in the agricultural sector. The use of various data mining techniques in agricultural sector will be a continuing area of research. The ultimate goal is to increase the yield of the agricultural sector.

We have proposed to get an optimized result from various agricultural databases by applying the data mining and optimizing techniques and understanding the soil condition and give the proper recommendation of crop on the basis of soil condition and other factors. It also recommends use of fertilizers and pesticides for a specific recommended crop. With respect to this recommendation, we also suggest side business for formers which will be helpful to improve his economical life.

In conclusion we can say that if proper and optimized recommendations are given to farmers it will definitely help in building the economical status of agricultural produce dependent countries.

REFERENCE

[1] Han J., Kamber M. Data Mining: Concepts & Techniques, Morgan & | Kaufmann, 2000. | | [2]Sumitha Thankachan, Dr.S.Kirubakaran,"E-Agriculture Information Management System", IJCSMC, Vol. 3, Issue. 5, May 2014, page: 599-607. | | [2] Agrawal, R., Imielinski, T. and Swami, A.N. (1993) "Database mining: a performance Perspective." IEEE Transactions on Knowledge and Data Engineering, Vol. 5,914–925. | | [3] Lida Xu, Ning Liang, and Qiong Gao, "An Integrated Approach for Agricultural Ecosystem Management", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 38, NO. 4, JULY 2008 | | [4] P.Bhargavi, Dr.S.Jyothi, "Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils", IJCSNS, VOL.9 No.8, August 2009 | | [5] Ahsan Abdullah, Amir Hussain, "Data Mining a New Pilot Agriculture Extension Data Warehouse", Journal of Research and Practice in Information Technology, Vol. 38, No. 3, August 2006. | | [6] Sam Y. Sung, Member, IEEE Computer Society, Zhao Li,Chew L. Tan, and Peter A. Ng, "Forecasting Association Rules Using Existing Data Sets", IEEE transactions on knowledge and data engineering, vol 15, no.6, November/December 2003. | | [7] Farah Khan, Dr.Divakar Singh, "Knowledge Discovery on Agricultural Dataset Using Association Rule Mining", IJETAE, Volume 4, Issue 5, May 2014. | | [8] Darcy Miller, Jaki McCarthy, Audra Zakzeski,"A Fresh Approach to Agricultural Statistics: Data Mining and Remote Sensing", Section on Government Statistics â€"JSM 2009. | | [9] Dr.Deshmukh Nilesh Kailasrao, "An Overview on ICT for Indian Agricultural Informatics Developments", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June 2012. | | [10]K. S. Rathnamala, "Analysis of Self Organizing Maps Using Visual DM Techniques in Agro Database for Prediction of Yield", International Journal of Advanced Computer Science, Vol. 3, No. 10, Pp. 508-511, Oct., 2013 | | [11] Nasira,G.M and Hemageetha,N., "Vegetable price prediction using data mining classification technique" Proceedings of the International Conference on pattern Recognition, Informatics and Medical Engineering (PRIME 2012), PP. 99-102 ISBN No:978-1-4673-1038-3. © 2012 IEEE. | | [12] Arvind Jaiswal,Gaurav Dubey.,"Identifying Best Association Rules and Their optimization using Genetic Algorithm", IJESE,ISSN:2319-6378, VOL.1,Issue-7,May 2013 | | [13]D.Rajesh,"Application of Spatial Data Mining for Agriculture", IJCA Volume 15â€"No.2, February 2011. | | [14]S.Veenadhari, Dr.Bharat Misra, Dr.C.D.Singh,"Data mining Techniques for Predicting Crop Productivity â€"A review article", IJCST Vol. 2, Issue 1, March 2011. | | [15]D.Ramesh, 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. | | [16]J. Rajendra Prasad, P. Ravi Prakash, S.Sai Kumar, M. Sundara Babu K. Swarupa Rani,"Identification of Agricultural Production Areas in Andhra Pradesh",IJEIT Volume 2,Issue 2,August 2012. | | [17]Antonio Mucherino, Petrq J. Papajorgji, Panos M.Pardalos,"Data Mining in Agriculture", Springer. |