Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications

Manpreet Kaur

Heena Gulati

Harish Kundra

ABSTRACT

In agriculture crop price analysis, Data mining is emerging as an important research field. In this paper, we will discuss about the applications and techniques of Data mining in agriculture. There are various data mining techniques such as K-Means, K-Nearest Neighbor (KNN), Artificial Neural Networks (ANN) and Support Vector Machines (SVM) which are used for very recent applications of Data Mining techniques. This paper will consider the problem of price prediction of crops. Price Prediction, nowadays, has become very important agricultural problem which is to be solved only based on the available data. Data Mining techniques can be used to solve this problem. This work is based on finding suitable data models that helps in achieving high accuracy and generality for price prediction. For solving this problem, different Data Mining techniques were evaluated on different data sets.

Keywords

Data Mining, K-Means, K-Nearest Neighbour, Artificial Neural Networks, Support Vector Machines, Price Prediction

1. INTRODUCTION

The process of extracting important and useful information from large sets of data is called Data Mining. In agricultural field, Data Mining is an important research field. In this paper, Description and overview of data mining techniques which are applied to agriculture and their applications to agriculture related areas is described. Price prediction is a very important problem for any farmer as he is the one who should know how much cost he would expect for his crops. In past years, price prediction was done by judging farmer's experience on particular crop and field. Suppose we have the previous data available in which various corresponding price predictions are recorded and these recorded price predictions are used to classify future price predictions..

2. DATA MINING TECHNIQUES

From long time, different techniques were used for mining data. Researchers have discussed a detailed and elaborated 10 Data Mining Techniques [1]. This paper present the most used Data Mining Techniques in agricultural field.

Classification and Clustering Techniques are two types of Data Mining Techniques [3]. For classifying unknown samples in which information is provided by a set of classified samples, Classification Techniques are designed. Neural Networks [2] and Support Vector Machines [4] are two classification techniques that are used generally to classify unknown samples. The technique that does not have any learning set is the K-Nearest Neighbor (KNN)[5] but it has the training set that is used for classification and in this technique similar samples should have similar classification.

The parameter K in K-Nearest Neighbor is used to show the number of similar known samples. The K-Nearest Neighbor

uses the training set, in case, if any training set is not available, clustering techniques can be used to split a set of unknown samples into clusters. K-Means algorithm [6] is one of the most used clustering techniques. In a set of data with unknown classification, we will find a partition of the set in which we have same data which is grouped in same cluster. The parameter K present in K-Means algorithm specifies the number of clusters in which data is to be partitioned. The main reason behind K-Means algorithm is the centers of clusters that can be computed as means of all samples belonging to a cluster. The representative of the cluster can be considered as the center of cluster because the center is quite close to all samples. But one of the disadvantages in using the K-Means algorithm can be the choice of parameter K. Another important issue is the computational cost of the algorithm. Other data mining techniques such as Principle Component Analysis (PCA), Regression Model [7] and Bi clustering techniques have some applications in agriculture or agricultural-related fields.

3. APPLICATIONS

Several applications of Data Mining Techniques are used in the field of agriculture like techniques related to weather situations and forecasts. For example, to forecast the atmospheric pollution, we can use K-Means algorithm. For simulating daily precipitations and other weather variables, we can use K-Nearest Neighbor and SVMs can be used to analyze the possible changes of the weather.

Sound recognition problems can also be solved by Data Mining Techniques. The sound of birds and other sounds can be classified using SVMs and this technique was used by Fagerlund S [8]. To evaluate forest inventories and to estimate forest variables for observing satellite imagery, Holmgren et al. [9] used K-Nearest Neighbor. For classifying eggs as fertility, Das KC et al. [10] used Artificial Neural Networks and for recognition of cracks in eggs, Patel VC et al. [11] used Computer Vision. For classifying Pizza sauce spread, Du CJ et al. [12] used SVMs and for detection of weed and nitrogen stress in corn, Karimi Y et al. [13] used SVMs. K-Means approach is used for classification of soils with GPS based technologies [14]. To classify soils and plants, Meyer GE et al. [15] used K-Means approach. SVMs were used to classify crops by Camps Valls et al. [16].

A Neural Network is used for making difference between good and bad apples and Shahin MA et al. [17] used X-rays images of apples to observe the presence of water cores. A supervised Bi-clustering technique was applied by Mucherino et al. [18] to a set of wine fermentations in order to select and discover the features to predict the quality of new fermentations. ANNs [19] use taste sensors to obtain data from fermentation process. Also, SVMs [20] use sensors to smell milk.

4. LITERATURE SURVEY

This paper attempts to forecast the prices of vegetable from changes in the price of crude oil. The analysis would help. Price prediction helps the farmers and also Government to make effective decision. Data mining classification techniques can be used to develop an innovative model to predict the market price of respective commodity.

Price prediction is highly useful in agriculture for forecasting, and other involved players manage their inventories effectively and influence their purchasing decisions.

The data for the study, namely oil and vegetable were collected from syndicated agencies and extend over a 3 year period, from 2009-2011. The methodology included data cleansing and conformity wherein missing values were accounted for by linear extrapolation. The forecasting methods used included Naïve, and regression (based on multiple predictors like oil and time function) and were compared on accuracy measured by minimal errors in the forecasted values.

The data for oil was observed to have an increasing pattern over time while the vegetable had a periodic high every April. The key findings validated this higher price (April) which necessitates inventory replenishment during December when vegetable prices are comparatively lower. The caveats which need to be highlighted are the limitations in data collection, wherein larger samples could have helped us in validating the model better. Ancillary variables like the seasonal rainfall or economic indicators could also have had an impact in forecasting vegetable prices.

5. FORECASTING

Forecasting is the process of making statements about events whose actual outcome have not yet been observed. A common place example might be estimation of some variable of interest at some specified future date. Prediction is a similar but more general term. Both might refer to formal statistical methods employing time series, cross-sectional or longitudinal data, or alternatively to less formal judgmental methods. Usage can differ between areas of application: for example, in hydrology, the terms "forecast" and "forecasting" are sometimes reserved for estimation of values at certain specific future times, while the term "prediction" is used for more general estimates such as the number of times floods will occur over a long period. Risk and uncertainty are central to forecasting and prediction; it is generally considered good practice to indicate the degree of uncertainty attaching to forecasts. In any case, the data must be up to date in order for the forecast to be as accurate as possible ...

6. CONCLUSION

In this paper certain Data Mining techniques were adopted in order to estimate crop price analysis with existing data. The applications that use the K-Means approach, utilize only the basic algorithm, while many other improvements are available. Some Data Mining techniques have not yet been applied to agricultural problems. As an example, Regression techniques may be employed for discovering important information from agricultural-related sets of data.

In future the Genetic algorithm based neural network will be constructed for price prediction to increase the accuracy percentage. The BP neural network prediction model of vegetable market price is established. We have taken Coimbatore market price of tomato as an example and simulated the result using MATLAB.

7. 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] 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.
- [3] 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.
- [4] 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.
- [5] Cover TM, Hart PE, Nearest Neighbor pattern classification. IEEE Trans Info Theory 13(1): 21-27, 1967.
- [6] J. Hartigan, Clustering Algorithms, John Wiles & Sons, New York, 1975.
- [7] A. Mucherino, A. Urtubia, Consistent Bi clustering and Applications to Agriculture, Ibal Conference Proceedings, Proceedings of the Industrial Conference on Data Mining (ICDM10), Workshop "Data Mining in Agriculture" (DMA10), Berlin, Germany, 105-113, 2010.
- [8] Fagerlund S Bird species recognition using Support Vector Machines. EURASIP J Adv Signal Processing, Article ID 38637, p 8, 2007.
- [9] Holmgren P, Thuresson T Satellite remote sensing for forestry planning: a review. Scand J For Res 13(1):90– 110, 1998.
- [10] Das KC, Evans MD Detecting fertility of hatching eggs using machine vision II: Neural Network classifiers. Trans ASAE 35(6):2035–2041, 1992.
- [11] Patel VC, McClendon RW, Goodrum JW Crack detection in eggs using computer vision and neural networks. Artif Intell Appl 8(2):21–31, 1994.
- [12] Du C-J, Sun D-W Pizza sauce spread classification using colour vision and support vector machines. J Food Eng 66:137–145,2005.
- [13] Karimi Y, Prasher SO, Patel RM, Kim SH Application of support vector machine technology for Weed and nitrogen stress detection in corn. Computer Electronics Agriculture 51:99–109, 2006.
- [14] Verheyen K, Adriaens D, Hermy M, Deckers S Highresolution continuous soil classification using morphological soil profile descriptions. Geoderma 101:31–48, 2001.
- [15] 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.
- [16] Camps-Valls G, Gomez-Chova L, Calpe-Maravilla J, Soria- Olivas E, Martin-Guerrero JD, Moreno J Support

International Journal of Computer Applications (0975 – 8887) Volume 99– No.12, August 2014

Vector Machines for crop classification using hyperspectral data. Lect Notes Comp Sci 2652:134–141 , 2003.

- [17] Shahin MA, Tollner EW, McClendon RW Artificial intelligence classifiers for sorting apples based on watercore. J Agric Eng Res 79(3):265–274, 2001.
- [18] A. Mucherino, A. Urtubia, Feature Selection for Datasets of Wine Fermentations, I3M Conference Proceedings, 10th International Conference on Modeling and Applied Simulation (MAS11), Rome, Italy, September 2011.
- [19] Riul A Jr, de Sousa HC, Malmegrim RR, dos Santos DS Jr, Carvalho ACPLF, Fonseca FJ, Oliveira Jr ON, Mattoso LHC Wine classification by taste sensors made from ultra-thin films and using Neural Networks. Sens Actuators B98:77–82, 2004.
- [20] Brudzewski K, Osowski S, Markiewicz T Classification of milk by means of an electronic nose and SVM neural network. Sens Actuators B98:291–298, 2004.