International Journal of Engineering and Techniques - Volume 6 Issue 2, March2020

ARMA BASED CROP YIELD PREDICTION USING TEMPERATURE AND RAINFALL PARAMETERS WITH GROUND WATER LEVEL CLASSIFICATION

Mr.S.Jagadeesan, M.E.,[1]Mr. T.Gopinath, MCA[2]

Assistant Professor[1], Final MCA[2]

Department of ComputerScience and Engineering[1]

Department of Computer Applications[2]

Nandha Engineering College (Autonomous),

Erode, Tamil Nadu, India

jagadeesan12398@gmail.com[1], tgobinathsvg@gmail.com[2]

Abstract: The main aim of this project is to provide a methodology for crop yield production based on the historical climatic and production data. Crop yield prediction based on the previous years of temperature and rainfall can help farmers take necessary steps to improve crop yield in the coming season. Understanding crop yield can help ensure food security and reduce impacts of climate change. Crops are sensitive to various weather phenomena such as temperature and rainfall. Therefore, it becomes crucial to include these features when predicting the yield of a crop. Weather the forecasting are complicated process. In this work, ARMA (Auto Regressive Moving Average) method is used to forecast crop yield. Past ten years of data set is taken for temperature, rainfall and ground water level for our country. Yield prediction is then carried out using a Fuzzy logic algorithm to better judge the crop yield. In addition, this project classifies the ground water level data set records using KNN to predict the model for future test record data sets. It will be helpful in analyzing the ground water levels in the past and so as to predict the future levels.

Keywords: ARMA, yield predicition, KNN Algorithm..

I. INTRODUCTION

Although information mining remains in its infancy, businesses in a wide range of industries - along with retail, finance, fitness care, production transportation, and aerospace - arealreadyusinginformationminingtoolsandtechniquestotakebenefitofhistoricalrecords.Data mining derives its call from the similarities between attempting to find valuable information in a huge database and mining a mountain for a vein of precious ore. Both processes require either sifting through an immense quantity of material, or intelligently probing it to find wherein the price resides.

Data mining or information discovery is the computer-assisted procedure of digging through and analyzing sizable sets of records and then extracting the that means of the facts. Tools of information mining expect behaviors and destiny trends, allowing groups to make proactive, expertise-driven decisions. They can answer business questions that traditionally were too time eating to resolve. They scour databases for hidden patterns, locating predictive information that experts may additionally miss because it lies out of doors their expectations.

Several models had been used within the beyond to are expecting temperature and rainfall. ARIMA and Multiple Linear Regression may be used to expect the version for rainfall records. In MLR equation, parameters are taken from the dataset and variables are extracted from the

dataset by correlation. ARIMA is used for modelling time series and rainfall prediction.

II. RELATED WORK

The purpose of this paper is to explore the dynamics of neural networks in forecasting crop (tomato) yield using environmental variables; here they aimed at giving accurate yield amount. They used the Adaptive Neuro-Fuzzy Inference System (ANFIS). The input to ANFIS is several parameters derived from the crop growth model (temperature, Co2, vapor pressure deficit (VPD), yield, and radiation).

ANFIS has only one output node, the yield. One of the difficult issues in predicting yield is that remote sensing data do not go long back in time. Therefore any predicting effort is forced to use a very restricted number of past years in order to construct a model to forecast future values. The system is trained by leaving one year out and using all the other data. They then evaluate the deviation of their estimate compared to the yield of the year that is left out. The procedure is applied to all the years and the average forecasting accuracy is given.

ANFIS is considered as a class of adaptive networks that perform as a framework for adaptive fuzzy inference [6] systems. Generally, it is a multilayer feed forward adaptive network where each node realizes a particular node function of its corresponding inputs and the nodes in ANFIS include adaptive and fixed ones, and ANFIS is characterized with the parameter set that is the union of the parameter sets associated with all adaptive nodes. The use of a neuro-fuzzy system for crop yield [7] estimate has some interesting characteristics. All the variables that are input into the system are associated with varying degrees of accuracy. Some ambiguity comes from measurement error and generality.

Using fuzzy sets instead of the actual values as inputs, we aim at shifting to the semantics of the data rather than its measure [6]. It is well known that with neuro–fuzzy modeling there is the alternative to use a fuzzy set as the output. In this case, yield would be expressed for example as low, normal of high with each of those three borders corresponding to a fuzzy set. They did not imply however that seeking a crisp (non-fuzzy) value is a more exact approach than seeking a trend expressed in fuzzy sets (low, medium, etc.). But although the accuracy of prediction is probably the same in both expressions of desired output, people are more used to and feel more confident in looking at number rather than a membership function.

ANFIS is a system that accepts numerical inputs and produces a single out value. ANFIS is susceptible to the “curse of dimensionality”. The training time increases exponentially with respect to the number of fuzzy sets per input variable used. To illustrate this let us consider a system with 8 input variables that are coded into two fuzzy sets (e.g. low, high) and has 256 rules. If we chose now to use three linguistic variables instead of two (say low, medium, high) the number of fuzzy rules becomes 6561. This phenomenon limits in practice the choice of input variables as well as the expression of those variables into meaningful fuzzy sets.

Fuzzy logic was introduced by [6] [8] to represent and manipulate data and information in which there are various forms of uncertainty. Fuzzy rule-based systems use linguistic variables to reason using a series of logical rules that contain IF-THEN rules which connect antecedent(s) and consequent(s), respectively. An antecedent is a fuzzy clause with certain degree of membership (between 0 and 1), see Fig.2.1. Fuzzy rules can have multiple antecedents connected with AND or OR operators, where all parts are calculated simultaneously and resolved into a single number. Consequents can also be comprised of multiple parts, which are then aggregated into a single output of a fuzzy set [9]. Fuzzy inference [10] is a process of mapping from a given input to an output using the fuzzy set methods.

III. EXISTING WORKS

In existing system, the data set is taken from Indian meteorological sites saved in excel files as 'csv' files. It contains temperature, rainfall and ground water level data. These data are taken for text pre-processing first and then converted into time series data set format. Then ARMA model is used to predict the future years' levels and then a fuzzy logic is applied to classify the crop yield in future years.

IV. PROPOSED SYSTEM

In proposed system, all the existing methodology is carried out. Like existing system, here also, the data set is taken from Indian meteorological sites saved in excel files as 'csv' files. It contains temperature, rainfall and ground water level data. These data are taken for text pre-processing first and then converted into time series data set format. Then ARMA model is used to predict the future years' levels and then a fuzzy logic is applied to classify the crop yield in future years.

In addition, KNN and SVM is applied for classification and so it is found to be suitable especially if the data set is having more number of records is contains outlier data. A wide variety of ground water level records can be taken for classification purpose and predicting a new model at the same time increasing the efficiency.

V. DESCRIPTION

ARMA MODEL BASED PREDICTION FOR RAINFALL

In this module, rainfall water data set is taken for Indian data for past ten years. The data is converted into data frame and preprocessed such that zero values in all columns records are eliminated. The data is converted into time series format such that twelve records (for each month) for all years present in the data set. Then using 'arima' function, the model is prepared for the given data set and predicted for upcoming years. Using ts.plot() the upcoming years values are plotted.

ARMA MODEL BASED PREDICTION FOR TEMPERATURE

In this module, temperature data set is taken for Indian data for past ten years. The data is converted into data frame and preprocessed such that zero values in all columns records are eliminated. The data is converted into time series format such that twelve records (for each month) for all years present in the data set. Then using 'arima' function, the model is prepared for the given data set and predicted for upcoming years.

Using ts.plot() the upcoming years values are plotted.

CROP YIELD PREDICTION BASED ON RAINFALL AND TEMPERATURE

In this module, using the previous rainfall and temperature outcomes, fuzzy logic based crop yield prediction is carried out the following algorithm.

IF (rain=='very good' and temp=='very good') or (rain=='very good' and temp=='good'): THEN

yield = 'very good'

IF (rain=='very good' and temp=='average') or (rain=='good' and temp=='very good') or (rain=='good' and temp=='good') or (rain=='good' and temp=='average'): THEN

yield = 'good'

IF (rain=='very good' and temp=='bad') or (rain=='very good' and temp=='very bad') or (rain=='good' and temp=='bad') or (rain=='good' and temp=='very bad') or (rain=='average' and temp=='very good') or (rain=='average' and temp=='good') or (rain=='average' and temp=='average'): THEN

yield = 'average'

IF (rain=='average' and temp=='bad') or (rain=='average' and temp=='very bad') or (rain=='bad' and temp=='very good') or (rain=='bad' and temp=='good') or (rain=='bad' and temp=='average') or (rain=='bad' and temp=='bad'): THEN

yield = 'bad'

IF (rain=='bad' and temp=='very bad') or (rain=='very bad' and temp=='very good') or (rain=='very bad' and temp=='good') or (rain=='very bad' and temp=='bad') or (rain=='very bad' and temp=='average') or (rain=='very bad' and temp=='bad') or (rain=='very bad' and temp=='very bad'): THEN

yield = 'very bad'

CLASSIFICATION	AVERAGE YEARLY RAIN	AVERAGE SEASON TEMPERATURE
Very Good	62-87 cm	20-25 C
Good	50-62 cm or 87-100cm	15-20 C
Average	25-50 cm or 100-150 cm	12-15 C or 25-33 C
Bad	10-25 cm or 150-200cm	8 -12 C or 33-40 C
Very Bad	All other values	All other values

ARMA MODEL BASED PREDICTION FOR GROUND WATER

In this module, groundwater level data set is taken for Indian data for past ten years. The data is converted into data frame and

preprocessed such that zero values in all columns records are eliminated. The data is converted into time series format such that twelve records (for each month) for all years present in the data set. Then using 'arima' function, the model is prepared for the given data set and predicted for upcoming years. Using ts.plot() the upcoming years values are plotted.

GROUND WATER LEVEL CLASSIFICATION BASED ON KNN MODEL

In this module, groundwater level data set is taken for Indian data for past ten years. The data is converted into data frame and preprocessed such that zero values in all columns records are eliminated. The data from 'MONSOON', 'POMKH', 'POMRB' and 'PREMON' column are taken, average value calculated, middle value is found out and values below middle value are considered as zero and above as one. These two values are applied in class factor column. Then KNN classification is predicted using the class factor column. For test run, K value is given as 6 and the model is predicted. The accuracy is calculated and displayed for the KNN model.

V. CONCLUSION

According to the results, temperature is best predicted by the ARIMA model and the accuracy of predictions made for rainfall by ARMA model is also good. Rainfall, which is an important factor for the prediction of crop yield is difficult to estimate precisely. Climate factors may change due to other remaining variables which may influence the prediction of rainfall.

Also, the proposed work makes use of fuzzy logic to estimate crop yield which works on a set range rather than discrete values, therefore, the error in predicted rainfall data does not cause problems as long as the difference between actual and estimated values is not drastic.

VI. REFRENCES

[1] P. Kumar "Crop yield forecasting by adaptive neuro fuzzy inference system." Mathematical Theory and Modeling 1.3, pp. 1-7, 2011.

[2] I. Kaushik, and S.M. Singh. "Seasonal ARIMA model for forecasting of monthly rainfall and temperature." Journal of Environmental Research and Development 3, no. 2, pp. 506-514, 2008.

[3] S. Veenadhari, B. Misra, and C. D. Singh. "Machine learning approach for forecasting crop yield based on climatic parameters." 2014 International Conference on Computer Communication and Informatics, pp. 1-5. IEEE, 2014.

[4] S. Prabakaran, P.N. Kumar, and P.S.M. Tarun. "Rainfall prediction using modified linear regression", 2006.

[5] S.K. Mohapatra, A. Upadhyay, and C. Gola. "Rainfall prediction based on 100 years of meteorological data." 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), pp. 162-166. IEEE, 2017.

[6] L. A. Zadeh, "Fuzzy sets," Information and Control, pp. 338-353, 1965.