A NOVEL APPROACH IN CROP YIELD PREDICTION BASED ON SEASONAL CLIMATE FORECASTS USING FFO_RF

[1]S. Gokila, [2]Mr. S. Omprakash,
[1]M. Phil., Scholar, [2]Assistant Professor,
[1]Department of Computer Science,
[1]Kovai Kalaimagal College of Arts & Science, Coimbatore, India.

Abstract:

The perfect knowledge of yield before harvest has been a wish puzzling human being since the beginning of agriculture because seasonal forecast of crop yield plays a critical role in decision making for different stakeholders from farmers to policy makers to governments for food security, to commodities traders. Different methods have been used to forecast yield with different levels of granularity, accuracy and timing. Yield forecast are mainly based on field surveys, statistical regressions between historical yield and in-season variables, crop simulation models, or on integration between statistical modeling with dynamic process-based crop simulation models. The proposed approach use the firefly optimization (FFO) for feature selection algorithm and the Random Forest (RF) technique in predicting the crop yield prediction based on Seasonal climate forecasts. The performance of the proposed approach is evaluated using the implementation result.

Keywords: Crop yield prediction, Accuracy, Climate forecast, Firefly algorithm, Random forest.

Introduction

Crop yield prediction is an important component of an early warning system for food security related planning. It is also useful for trade, development policies and humanitarian assistance related to food security. Similarly, it helps farmers and/or decisions makers to prepare for the upcoming growing season. Moreover, it serves as an important indicator of the national income where agriculture contribution to gross domestic product (GDP) is high [1]. The situation is further aggravated when the food habit is less diversified and people depend on a single crop to fulfill most of their dietary requirements.

The promise of reliable seasonal climate forecasts is that they would enhance a farmer's ability to flexibly adjust investments in farm enterprises by taking full advantage of the few good seasons and by avoiding the risks associated with seasons which ultimately turn out poorly. Climate forecasts in August for the remainder of the cropping season have even higher skill due to the impact of the season to date on yield potential [2]. An August forecast can be used to make the final nitrogen application decisions and, when conditions have been below average, to help decide whether to cut the crop for hay while it is green or to keep the crop in the hope of a more profitable grain harvest.

There are challenges involved in using a climate model over a local historical weather record or statistical approaches. Outputs from climate models are generated on a coarser grid than what is required for farm-scale crop forecasts and must be locally downscaled and calibrated to reduce inherent biases and overcome the 'connectivity problem'[3]. Nevertheless, the scientific understanding and skill of dynamic seasonal climate models continues to improve.

Dynamic model approaches are expected to be superior to statistical methods as they allow for forecasts that incorporate the effects of a number of climate modes, rather than just the Southern Oscillation Index which is not the only driver of climate variability [4]. It is therefore timely to explore whether these models can be used directly as inputs to force crop models and whether these forecasts provide any advantage over using climatology to inform crop models.

In this study feature selection and classification algorithm is applied for predicting hospitalizations due to the two leading chronic diseases: heart disease and diabetes. The following section discusses the literature of chronic disease prediction by several researchers. Section III gives the detailed explanation of proposed methodology for feature selection and prediction process. The experimental result and the performance evaluation are discussed in Section IV. Section V describes the summary of the study and future enhancement in this research.

Related works

Nagahamulla *et al.* [5] used an ANN to predict seasonal monsoon precipitation in Srilanka. The developed models targeted estimating rainfall for four months: May, June, July and August where climate indices were used as possible predictors. Correlation analyses were used to determine the predictors of each month. The generated networks were compared against each other and the best accuracy was obtained in June.

Meknik *et al.* [6] utilized weather attributes to forecast spring rainfall in Victoria, Australia. A MLR and an ANN were utilized as the prediction models. Data were collected from nine weather stations. Better accuracy was recorded with the neural network in eight out of the nine stations compared to MLR.

Thi-Thu-Hong Phan [7] aims first to build a framework for forecasting meteorological univariate time series and then to carry out a performance comparison of different univariate models for forecasting task. Six algorithms are discussed: Single exponential smoothing (SES), Seasonal-naive, Seasonal-ARIMA (SARIMA), Feed-Forward Neural Network (FFNN), Dynamic Time Warping-based Imputation (DTWBI), and Bayesian Structural Time Series (BSTS). Four performance measures and various meteorological time series are used to determine a more customized method for forecasting. Through experiments results, FFNN method is well adapted to forecast meteorological univariate time series with seasonality and no trend in consideration of accuracy indices and DTWBI is more suitable as considering the shape and dynamics of forecast values.

Ali Haidar and Brijesh Verma [8] propose a new forecasting method that uses a deep convolutional neural network (CNN) to predict monthly rainfall for a selected location in eastern Australia. To our knowledge, this is the first time applying a deep CNN in predicting monthly rainfall. The proposed approach was compared against the Australian Community Climate and Earth-System Simulator-Seasonal Prediction System (ACCESS), which is a forecasting model released by the Bureau of Meteorology. In addition, the CNN was compared against a conventional multi-layered perceptron (MLP). The better mean absolute error, root mean square error (RMSE), Pearson correlation (r), and Nash Suttcliff coefficient of efficiency values were obtained with the proposed CNN. A difference of 37.006 mm was obtained in terms of RMSE compared with ACCESS and 15.941 compared with conventional MLP. Further investigation revealed that the CNN was generally performing better in months with higher annual averages, while ACCESS was performing better in months with low annual averages. The generated output is promising and can be widely extended in this type of applications.

Proposed Methodology

The proposed system

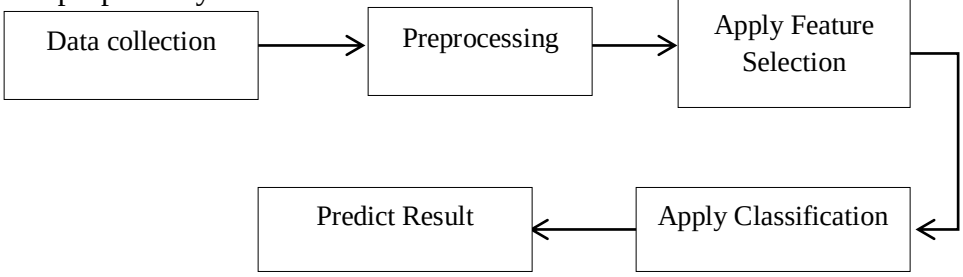

Figure 1. Proposed Framework

Data collection

The data collection is initial process in crop yield prediction. The dataset could contain missing values. The data was sampled every minute, computing and uploading it smoothed with 15 minute means. The header of the data file is a commentary indicating which data is stored at which column. The data is time information is in UTC.

Preprocessing

Data preprocessing step is one significant step in knowledge discovery process. Most health care data contain missing value, noisy and inconsistency data. Pima Indian data set has missing value and inconsistency data. Missing value are identified then replaced or handled by attribute mean value. Inconsistence data identified and removed manually. The dataset doesn't consist of noisy data.

Firefly Algorithm

Firefly Algorithm (FA) was first developed by Xin-She Yang in late 2007 and 2008 at Cambridge University [9], which was based on the flashing patterns and behavior of fireflies. In essence, FA uses the following three idealized rules:

• Fireflies are unisex so that one firefly will be attracted to other fireflies regardless of their sex.

• The attractiveness is proportional to the brightness, and they both decrease as their distance increases. Thus for any two flashing fireflies, the less bright one will move towards the brighter one. If there is no brighter one than a particular firefly it will move randomly.

• The brightness of a firefly is determined by the landscape of the objective function.

The pseudo-code of FA is given in Algorithm 1. All fireflies are generated randomly in the search space (Step 2). And the light intensity of fireflies is calculated by the objective function (Step 3). Each firefly will be compared with the rest of fireflies in the population, and the least bright one of two fireflies will update its position, and then the light intensity of the new position is calculated. Finally, all fireflies are ranked according to the light intensity to find the global optimal solution.

Initialization: the population number is N_p, the maximum number of iterations is

ItMax, β_0=1.0.

t = 1 and γ= 1.0.

x_i (i = 1, 2, 3..., N_p) is generated

Calculate the light intensity of each x_i (i = 1, 2, 3, ..., N_p)

While t <= *ItMax* do

For i = 1 to N_p − 1 do

For j = i to N_p do

If f(x_j) < f(x_i)

x_i is updated Else

x_j is updated End

Evaluate the light intensity of the new solution

End

End

Rank all fireflies and find the current best solution

t = t + 1

End while

Random forest Classifier

Random Forest is believed to be one of the best ensemble classifiers for high-dimensional data. Random forests are a mixture of tree predictors such that each tree depends on the values of a random vector sampled autonomously and with the same distribution for all trees in the forest. The generalization error for forests converges to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the association between them. A different subset of the training data is selected, with replacement, to train each tree. Remaining training data are used to estimate error and variable importance. Class assignment is made by the number of votes from all of the trees and for regression, the average of the results is used. [10] [11]

Algorithm:

Each tree is constructed using the following algorithm:

1. Let the number of training cases be N, and the number of variables in the classifier is M.

2. We are told the number m of input variables to be used to determine the decision at a node of the tree; m should be much less than M.

3. Choose a training set for this tree by choosing n times with replacement from all N available training cases (i.e. take a bootstrap sample). Use the rest of the cases to estimate the error of the tree, by predicting their classes.

4. For each node of the tree, randomly choose m variables on which to base the decision at that node. Calculate the best split based on these m variables in the training set.

5. Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier). For prediction, a new sample is pushed down the tree. It is assigned the label of the training sample in the terminal node it ends up in. This procedure is iterated over all trees in the ensemble, and the average vote of all trees is reported as random forest prediction. [10] [11].

Experimental Result

The proposed system is developed in java language. The netbeans IDE is utilized for front end design. NetBeans is an integrated development environment (IDE) for Java. NetBeans allows applications to be developed from a set of modular software components called modules. MYSQL is used for database access. MYSQL is an open source relational database management system (RDBMS). The experiment evaluation of the research is shown in following results.

The experimental result of the proposed approach is given based on before and after feature selection.

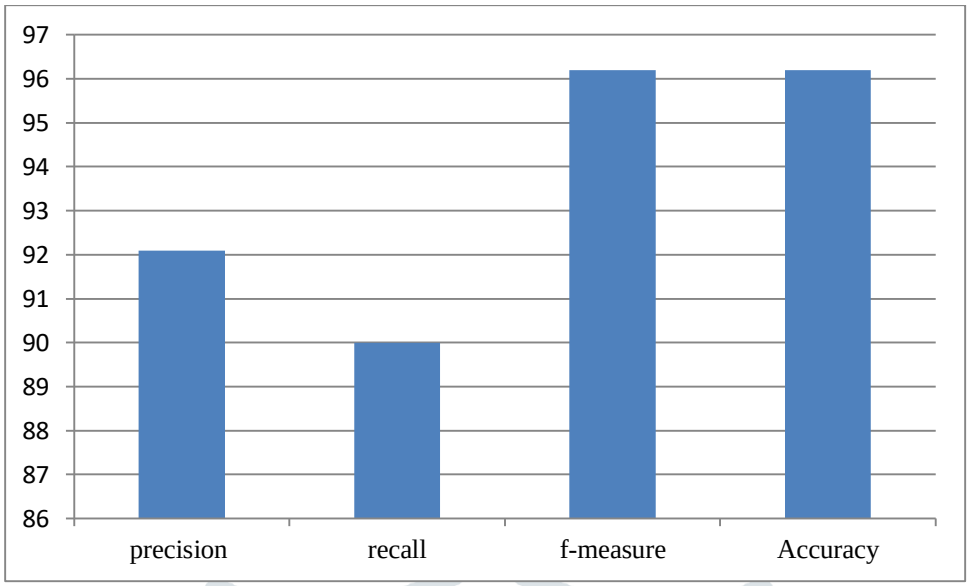

Figure 2. Performance report with feature selection

Conclusion

Agriculture yield forecasts are a very useful tool for farm management and can help stakeholders to perform critical decisions in their agricultural operations. Many available methods provide yield forecast information, Machine-learning is one of the techniques gaining popular- ity for agriculture applications, especially with the increasing number of new data sources being developed in the latest years. We propose a machine learning system that provides pre-season yield forecasting, meaning farmers can make farm management decisions. Our results show that farmers and agriculture stakeholders can benefit from useful information with significantly fewer data requirements and maintain useful accuracy values. The global availability of the input datasets also allows the system to easily scale across different regions if local yield data is present. Although the used input datasets allow for relatively high-resolution forecasts

References

[1]. FAO, "Wfp (2015), the state of food insecurity in the world 2015. meeting the 2015 international hunger targets: taking stock of uneven progress," Food and Agriculture Organization Publications, Rome, 2016.

[2] S. S. Dahikar and S. V. Rode, "Agricultural crop yield prediction using artificial neural network approach," International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engi- neering, vol. 2, no. 1, pp. 683–686, 2014.

[3] B. Ji, Y. Sun, S. Yang, and J. Wan, "Artificial neural networks for rice yield prediction in mountainous regions," The Journal of Agricultural Science, vol. 145, no. 3, pp. 249–261, 2007.

[4] A. X. Wang, C. Tran, N. Desai, D. Lobell, and S. Ermon, "Deep transfer learning for crop yield prediction with remote sensing data," in Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies. ACM, 2018, p. 50.

[5] H. R. K. Nagahamulla, U. R. Ratnayake, and A. Ratnaweera, ``Monsoon rainfall forecasting in Sri Lanka using articial neural networks," in Proc. 6th Int. Conf. Ind. Inf. Syst., 2011, pp. 305309.

[6] F. Mekanik, M. A. Imteaz, S. Gato-Trinidad, and A. Elmahdi, ''Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes,'' J. Hydrol., vol. 503, pp. 11–21, Oct. 2013

[7] Caillault, É.P. and Bigand, A., 2018, September. Comparative Study on Univariate Forecasting Methods for Meteorological Time Series. In *2018 26th European Signal Processing Conference (EUSIPCO)* (pp. 2380-2384). IEEE.

[8] Haidar, A. and Verma, B., 2018. Monthly Rainfall Forecasting Using One-Dimensional Deep Convolutional Neural Network. *IEEE Access*, *6*, pp.69053-69063.

[9]X. S. Yang, Firefly algorithms for multimodal optimisation, Proc. 5th Symposium on Stochastic Algorithms, Foundations and Applications, (Eds. O. Watanabe and T. Zeugmann), Lecture Notes in Computer Science, 5792: 169-178 (2009).

[10] Lakshmi Devasena C, Comparative Analysis of Random Forest, REP Tree and J48 Classifier for Credit Risk Prediction, International Journal of Computer Applications (0975 – 8887).

[11] Jehad Ali, Rehanullah Khan, Nasir Ahmad, Imran Maqsood, Random Forests and Decision Trees, International Journal of Computer Science Issues(IJCSI), Vol. 9, Issue 5, No 3, September 2012, ISSN (Online) 1694-0814.