Chronicle of Bioresource Management 2023, 7(1):016-020

Chronicle of Bioresource Management
An International E-magazine

March, 2023 Popular Article

Artificial Intelligence Based Prediction Models and Their Applications in Agriculture

Jyostna Bellamkonda[1], Santosha Rathod[2*], Nirmala Bandumula[2], D. Bhanusree[2], M. S. Anantha[2], C. Gireesh[2] and P. Muthuraman[2]

Open Access
Corresponding Author
Santosha Rathod
e-mail: santosha.Rathod@icar.gov.in

Citation: Bellamkonda et al., 2023. Artificial Intelligence Based Prediction Models and Their Applications in Agriculture. Chronicle of Bioresource Management 7(1), 016-020.

Data Availability Statement: Legal restrictions are imposed on the public sharing of raw data. However, authors have full right to transfer or share the data in raw form upon request subject to either meeting the conditions of the original consents and the original research study. Further, access of data needs to meet whether the user complies with the ethical and legal obligations as data controllers to allow for secondary use of the data outside of the original study.

Conflict of interests: The authors have declared that no conflict of interest exists.

Keywords:

Agriculture, AI, ARIMA, ANN, Hybrid models

Article History

Article ID: CBM151
Received on 26th February 2023
Received in revised form on 20th March 2023
Accepted in final form on 28th March 2023

Abstract

Forecasting is used to provide an aid to decision-making and in planning the future effectively and efficiently. It is an important aspect for a developing economy so that adequate planning is undertaken for sustainable development. The Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model is the most widely used classical time series model for forecasting in agricultural and allied sector. The major drawback of classical model is presumption of linearity, hence, no nonlinear patterns can be recognized by these models. Alternatively, intelligence techniques (AI) models are very effective to capture the nonlinear complex and heterogeneous pattern present in the agricultural data. Sometimes, the data under consideration contains both linear and nonlinear patterns, under such condition classical and AI models fails to capture the trend properly. To overcome this problem, a hybrid model or two stage methodology can be employed. Some of the popular classical, AI and hybrid methodologies are described in this article.

1. Introduction

Artificial intelligence is a field of science that creates machines or devices that imitates the intelligent behaviors. Models or group of algorithms comes under AI techniques. Machine learning is a type of artificial intelligence that provides computers with the ability to learn without being explicitly programmed. Machine learning techniques are data driven/phenomenon-based approaches. Artificial intelligence helps to transform the agriculture by integrating advanced technologies to forecast the agricultural productivity. Artificial intelligence can assist the farmers in producing higher yields by estimating agricultural production, soil and nutrient management, choosing the crop varieties and pest and disease forecasting, commodity price forecasting and providing real time information regarding agro-products marketing. Agriculture is a very complex phenomenon where most of the phenomenon's are inter related and are chaotic in nature; to understand such

Author's Address

[1]Agricultural College, Acharya N G Ranga Agricultural University, Bapatla, Andhra Pradesh (522 101), India
[2]ICAR- Indian Institute of Rice Research, Hyderabad, Telangana (500 030), India

Journal Home: *https://pphouse.org/cbm.php*

phenomenon's the classical models may not identify the patterns present in it. Machine learning techniques are possible alternatives which are free of distributional assumptions and they are mostly depends on the pattern present in the data under considerations hence called as data driven techniques. Various AI based classification, regression and pattern recognition techniques were employed in processing agricultural data in the past by various researchers (Zhang, 2003, Saha et al., 2020, Rathod et al., 2022a).

Further the forecasting techniques comes under the domain of time series analysis, where the observations are recorded over of period of time. Again based on the nature of time series data under consideration different time series model has to be chosen; If the series is linear and has short memory then classical linear time series models like autoregressive moving average (ARMA) model is most popularly used and If the series is nonlinear and volatile then GARCH (Generalized Autoregressive Conditional Heteroscedastic) models can be employed to model such series. If the series has long memory and exponential decaying of autoregressive lags, then one should consider long memory models like ARFIMA (Autoregressive Fractionally Integrated Moving Average) model to forecast such time series. For the discrete count time series data, generalized linear model (GLM) based Integer GARCH (INGRACH) can be employed to predict such series (Rathod et al., 2022b). Sometimes the data under consideration are recorded over both space and time are called as spatiotemporal time series data. Spatiotemporal modelling is commonly used in various disciplines viz., Geo-statistics, sociology, economics, environmental, ecological and agricultural science. The ARIMA model slacked in spatial lags are termed as space time autoregressive (STARMA) model and which is the most popular classical model to study the pattern in spatiotemporal time series data (Rathod et al., 2018 and Saha et al., 2020).

The AI based prediction models explained in this article can employed for forecasting agricultural commodity prices, weather parameter prediction under changing climate scenarios, prediction of agricultural production, prediction of genomic estimated breeding values using genome sequence data etc. The pest and disease incidence prediction models can also be developed using suitable activation functions like exponential and sigmoidal functions. The choice of suitable AI models depends on the nature of data under consideration, models like deep learning LSTM can be employed in the situations where data is highly chaotic and complex in nature viz., rainfall prediction, onion and tomato price prediction etc. The popular models used in predicting agricultural phenomenon are described in subsequent sections.

2. Time Series Models

2.1. *ARIMA Model*

The most used classical time series model in forecasting the time series data is the Box-Jenkins ARIMA. ARMA model is useful only when time series data is stationary. To build the ARIMA model the series must be differenced to make it stationary. Later the differenced series is subjected to ARIMA model fitting. The model is represented as ARIMA (p, d, q), where *p*, *d* and *q* represent AR, order of differencing and MA terms respectively. ARIMA model is expressed as follows

$$\Phi(B)(1-B)^d Y_t=\theta(B)e_t \qquad ...(1)$$

$$Y_t=\Phi_1 Y_{t-1}+\Phi_2 Y_{t-2}+...+\Phi_p Y_{t-p}+ \varepsilon t-\theta_1 e_{t-1}-...-\theta_q e_{t-q} \qquad ...(2)$$

Where, is the time series, *d* is the order of differencing, *B* = backshift operator, is the random error, is the random error. *p* is a number of autoregressive terms and *q* is the number of lagged forecast errors. The ARIMA model building consists of 4 stages viz., identification, estimation, diagnostic checking and forecasting (Box and Jenkins, 1970).

2.2. *Spatiotemporal Models*

Spatiotemporal time series are the observations which are recorded over both space and time by considering systematic dependencies across space and time. Spatiotemporal modelling manages the single variables recorded over a timeframe at various locations. Spatiotemporal modelling is commonly used in various disciplines viz., Geo-statistics, sociology, economics, environmental, ecological and agricultural science. The spatiotemporal time series models are same as ARIMA model building, the only difference is in spatiotemporal time series models the neighbourhood effects are incorporated by means spatial weight matrices (Rathod et al., 2018 and Saha et al., 2020).

3. Why AI?

The major limitation of the classical time series models is that they do not have the ability to identify nonlinear patterns, irregular trend and chaotic patterns present in the time series data. To overcome such limitation, machine learning techniques are found to be more promising techniques for forecasting time series data. Machine learning is a type of AI that provides computers with ability to learn without being explicitly programmed. Machine learning techniques are data driven / phenomenon-based approaches.

3.1. Popular AI Models

In machine learning techniques the most extensively used models in forecasting of time series data are ANN, SVM, RF, ELM and LSTM.

3.1.1. Artificial Neural Network Model

Artificial neural network is most widely used model for forecasting the time series data. ANN model is based on structure of human brain especially central nervous system. A neural network consists of a set of connected cells called neurons. In ANN all the neurons are connected through weights. The simple ANN model consists of input layer, hidden layer and output layer. Input layer receives the information, hidden layer comprised of neuron with non-linear activation functions and output layer aggregates the output from the hidden layer neurons.

$$Y= f\left(\sum_{i=1}^{m} \omega_i x_i\right) \quad ...(3)$$

Y= Activation function (logistic or tangent hyperbolic function)

$$y_{in}=x_1\cdot \omega_1+x_2\omega_2+x_3\cdot \omega_3+...+x_m\cdot \omega_m \quad ... (4)$$

i.e., (net input) $Yin = f((\sum_{i=1}^{m}\omega_i x_i)$. The output can be calculated by applying the activation function over the net input.

$$Y=F(Y_{in}) \quad ...(5)$$

Output= Function (net input calculated)

3.1.2. Support Vector Machine (SVM) Model

SVM is a statistical machine learning proposed by (Cortes et al., 1995) used in structural risk minimization. Support vector machines are applicable for the tasks like pattern classification, non-linear regression. The basic idea of support vector machine is to transform the original variable space into high dimensional variable space to carryout linear classification or linear regression. The general equation of SVR can be written as follows:

$$f(x)=W^T\Phi(x)+b \quad ... (6)$$

Where is function of x, W is weight vector, b is bias term and subscript T denotes the transpose. The coefficients W and b are obtained by minimizing the following regularized risk function.

3.1.3. Random Forest Model

Random forest is an ensemble learning method for both classification and regression problems proposed by (Breiman, 2011). RF model adopts a bagging approach in identifying features where each node is separated randomly by choosing the most dominant possible predictors. This improves the model accuracy without causing overfitting.

1. Using the input predictor variables, n_{trees} of bootstrapping ensembles are generated where n is the number of trees.

2. An unpruned regression tree is built by choosing maximum predictors splitting creating a randomized input predictors sample denoted $m_{try.}$

3. The predictions from n_{trees} are aggregated to forecast future rainfall.

3.1.4. Extreme Learning Machines Model

Extreme Learning Machines (ELM) are suggested as alternative learning algorithms instead of feed forward networks (FNN). The feed forward neural networks created critical bottlenecks among its applications. The former is characterised by single-hidden layer feedforward neural networks. The ELM selects hidden nodes randomly and analytically determines their output weight. Unlike traditional FNN learning methods, ELM is significantly more efficient and it has a greater tendency to achieve a global optimum. ELM's output function for generalised SLFNs is represented by the following equation.

$$f_L(x)=\sum_{i=1}^{L}\beta_i h_i(x)=h(x)\beta \quad ... (7)$$

Where $\beta=[\beta_1,...\beta L]^T$ represents the output weight vector between L nodes' hidden layer to the $m \geq 1$ output nodes, and h(x)=[h1(x),…hL(x)] represents the nonlinear feature. In ELM, the main idea involves the hidden layer weights. Furthermore, the biases are randomly generated and the calculation of the output weights is done using the least-squares estimation.

3.1.5. Long Short-Term Memory Model

LSTM is a Recurrent neural network with memory structures for learning long-term information. Traditional ANNs cannot connect previous information to the current time step when dealing with long term dependencies. LSTM (Hochreiter and Schmidhuber, 1997) was proposed to solve drawbacks in recurrent neural networks (RNNs). LSTM incorporates temporal memory effect by using several recurrently connected memory blocks. A memory block includes three major units: input gate, output gate and forget gate. To control the flow of error these gates can be activated or closed. As the length of input sequence increases, it gets difficult to capture the influence of the earliest stages. The gradients to the first several input points vanish and become equal to zero. Due to its recurrent nature, the activation function of the LSTM is considered as the identity function with a derivative of 1.0. Therefore, the gradient that being back propagated neither vanishes nor explodes but remains same.

4. Hybrid or Two Stage Methodology

The major drawback of any linear time series or long memory time series or Spatiotemporal time series models are presumption of linearity; hence no non-linear patterns can be recognized by these models. Sometimes, time series contain both linear and non-linear components. so, under such circumstances neither ARIMA nor AI models are adequate in modelling and forecasting of time series data. Hybrid modelling method consists of two phases. In the first phase time series is analyzed using the linear time series models. In the second phase the residuals obtained in the previous phase are analyzed by machine learning models. The fitted models obtained in these two phases are summed separately.

$$\hat{Y}_t=\hat{L}_t+\hat{N}_t \quad \text{.... (8)}$$

Where, $\hat{L}$ and $\hat{N}$ represents the predicted linear and nonlinear component respectively. The schematic representation of hybrid methodology is expressed in following figure;

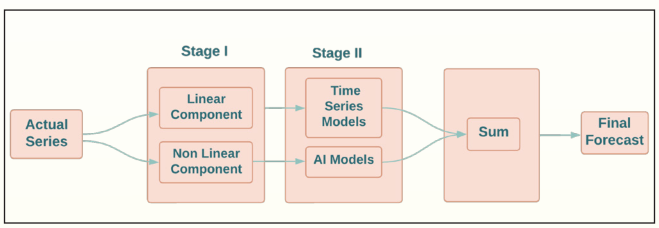

Figure 1: Schematic representation of time series – AI hybrid methodology

5. Statistical Software Packages to Implement AI Models

Statistical software uses the statistical analysis tools and techniques for collecting and analysing data to provide insight into patterns and trends and present the data in an easy-to-understand form. Different statistical software packages to implement AI models are Python, R, SAS, SPSS, MATLAB, Google Colab *etc.* The following R packages are most popularly used to build classical and AI based prediction models in R open source software.

Table 1: List of R packages used to build prediction models in R software

Sl. No.	R package	Used for
1.	ggplot2	To create elegant data visualization using the grammar of graphics.
2.	lmtest	Testing of parameters of the models.
3.	tseries	For basic time series modelling and operations.
4.	fNonlinear	For analysing the nonlinear data
5.	forecast	To build various time series models including neural net-works.
6.	fGarch	Analyse and model heteroskedastic behaviour in financial time series.
7.	FinTS	Analysis of financial time series.
8.	e1071	Functions for latent class analysis, support vector machines.
9.	nnet	To develop the feed forward neural networks and multinomi-al log linear models.
10.	random Forest	Developing classification and regression tree models. Breiman and Cutler's Random Forests for Classification and Regres-sion.

6. Conclusion

Most of the times, the agricultural time series data contains both linear and nonlinear pattern and such data cannot be modelled with either classical time series models or AI based models alone. Under such circumstances, hybrid models or two stage models help to determine both linear and nonlinear pattern present in the data sets for accurate prediction of agricultural phenomenon under consideration.

7. References

Box, G.E.P., Jenkins, G., 1970. Time series analysis, Forecasting and control, Holden-Day, San Francisco, CA.

Breiman, L., 2001. Random forests. Machine Learning 45(1), 5–32.

Cortes, C., Vapnik, V., 1995. Support vector networks, Machine Learning 20(3), 273–297.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation 9(8), 1735–1780.

Rathod, S., Gurung, B., Singh, K.N., Ray, M., 2018. An improved space-time autoregressive moving average (STARMA) model for modelling and forecasting of spatio-temporal time-series data. Journal of the Indian Society of Agricultural Statistics 72, 14–25.

Rathod, S., Chitikela, G., Bandumula, N., Ondrasek, G., Ravichandran, S., Sundaram, R.M., 2022a. Modeling and forecasting of rice prices in India during the COVID-19 Lockdown Using Machine Learning Approaches. Agronomy 12, 2133. https://doi.org/10.3390/agronomy12092133.

Rathod, S.,Yerram, S., Arya, P., Katti, G., Rani, J., Padmakumari, A.P., Somasekhar, N.,Padmavathi, C., Ondrasek, G., Amudan, S., Malathi, S., Rao, N.M., Karthikeyan, K., Mandawi, N., Muthuraman, P., Sundaram, R.M., 2022b. Climate-based modeling and prediction of rice gall midge populations using count time series and machine learning approaches. Agronomy 12, 22. https://doi.org/10.3390/agronomy12010022.

Saha, A., Singh, K.N., Ray, M., Rathod, S., 2020. A hybrid spatio-temporal modelling: an application to space-time rainfall forecasting. Theoretical and Applied Climatology 142, 1271–1282.

Zhang, G.P., 2003. Time-series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159–175.