International Conference on Paradigms of Computing, Communication and Data Sciences (2020)

Hybrid ARIMA-Deep Belief Network model using PSO for Stock Price Prediction

Presented by Shaikh Sahil Ahmed

Authors

Shaikh Sahil Ahmed

Mahesh Kankar

Nagaraj Naik

Biju R. Mohan

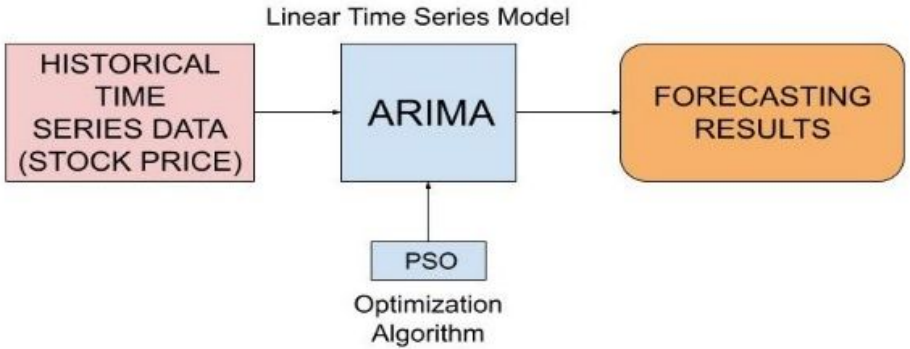
Contents

Introduction

- Accuracy is now a major factor in stock price prediction.
- Models used:
  - ARIMA (AutoRegressive Integrated Moving Average)
  - LSTM (Long Short Term Memory)
  - ARIMA-LSTM
  - ARIMA-DBN (ARIMA-DEEP BELIEF NETWORK)
- Performance Metrics:
  - Mean Squared Error (MSE)
  - Mean Absolute Error (MAE)
  - Root Mean Squared Error (RMSE)
  - Root Mean Squared Logarithmic Error (RMSLE)
  - Mean Absolute Percentage Error (SAME)
  - Symmetric Mean Absolute Percentage Error (SMAPE)

Introduction (Contd...)

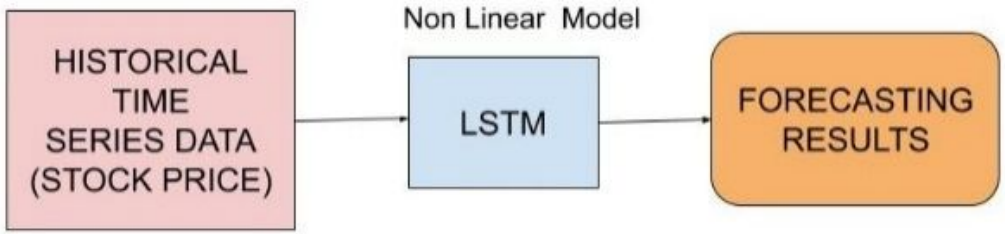
- ARIMA (AutoRegressive Integrated Moving Average)



- ARIMA (p,d,q)
  - p: AutoRegressive Term
  - d: order of differencing
  - q: Moving average term

Introduction (Contd...)

❖ LSTM (Long Short Term Memory)



- Category of Recurrent Neural Networks (RNN).
- LSTMs are efficient in capturing long term data and are well designed to keep away the problem called long-term dependency.

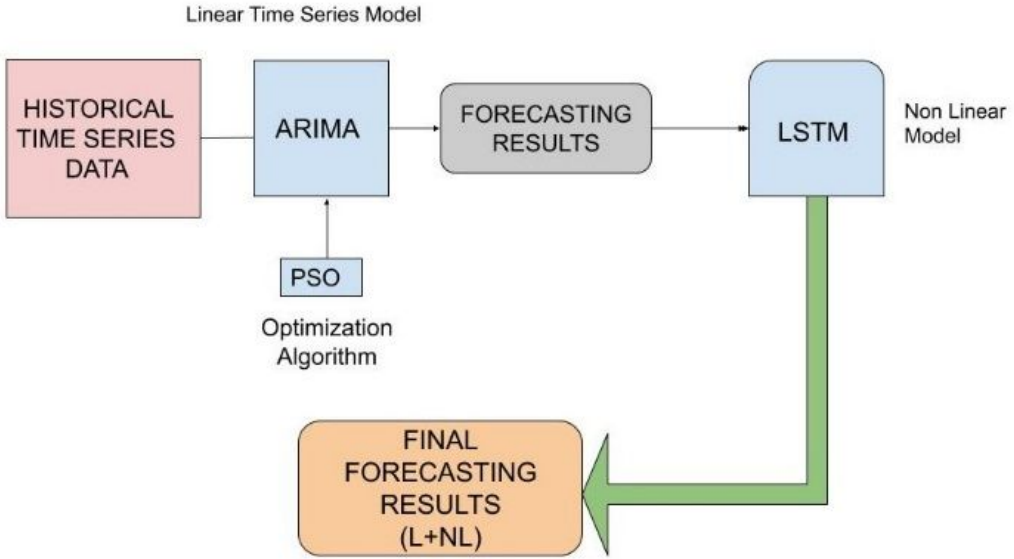
Introduction (Contd...)

- DBN (Deep Belief Network)
  - DBN is a sort of deep neural network which is built up of multiple RBMs (Restricted Boltzmann Machines).
  - Each RBM consists of 2 layers, one is the visible layer and the other one is the hidden layer.
  - The visible layer of first RBM takes input from the time series data and the hidden layer is computed based on weights, biases, and input of the visible layer.
  - In the next RBM, the output of the first hidden layer is given as input to the visible layer and repeatedly this process is executed for all the RBMs present in the DBN network.

Introduction (Contd...)

- ARIMA-LSTM

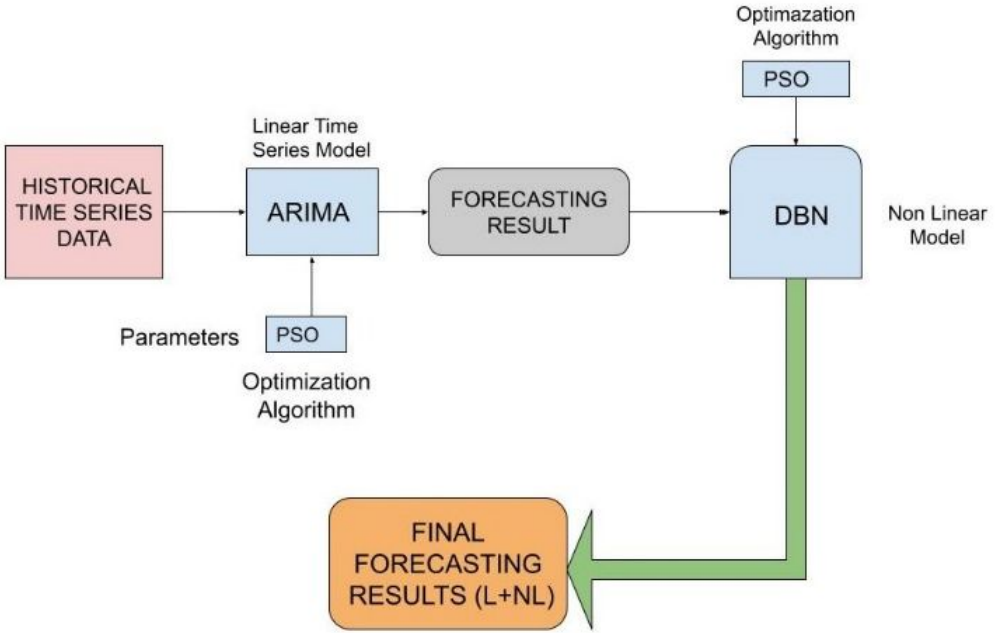
ARIMA model handles the linear component of the time series and gives the non-linear residual output which is given further input to LSTM. LSTM handles the non-linear pattern and the output prediction is merged with the estimations of the ARIMA model to give the final output of the series hybrid ARIMA-LSTM model.



Introduction (Contd...)

- ARIMA-DBN (ARIMA-Deep Belief Network)

ARIMA handles the linear component and the residuals are given further input to DBN. PSO optimizes the p,d,q values and hidden units in RBM. DBN handles the non linear component and the output prediction is merged with the estimations of the ARIMA model to give the final output.



Problem Statement

- As there are already existing model that is used for stock price prediction. We want a good accuracy in this field. So, various existing models such as ARIMA, LSTM. For better results we are going for series hybrid model, i.e combining two models, such that one of the model solves the linear and the other model solves the nonlinear pattern present in the time series.
- Building such type of model gives better accuracy as compared to single existing models.

Methodology

- **Particle Swarm Optimization (PSO)** is an optimization algorithm that is used to find an appropriate solution and speed up the training process.
- PSO technique is used to optimize the parameters for ARIMA p, d, q value and number of neurons in Deep Belief Network where p is the AutoRegressive term, d is differencing and q is Moving Average term such that Mean Squared Error is minimized.
- PSO helps in finding the optimum number of hidden units present in RBM (Restricted Boltzmann Machine) layer.
- PSO for finding the best parameters of p, d, q is used in all the models having ARIMA.

Experiment and Results

- Infosys Dataset
  - Using PSO we got (p, d, q) values of ARIMA as (2,1,0) best fit.
  - Also we got 512 hidden units present in the RBM layer.
  - Results over various performance metrics:

Performance Indicator	Models			
	ARIMA	LSTM	ARIMA-LSTM	ARIMA-DBN
MSE	117.191	270.334	139.759	**117.176**
MAE	7.670	13.316	8.678	**7.668**
RMSE	92.644	**16.442**	92.189	92.644
RMSLE	0.015	0.022	0.017	**0.015**
MAPE	0.107	**0.018**	0.107	0.107
SMAPE	10.655	**1.851**	10.572	10.654

Experiment and Results (Contd...)

- Mahindra & Mahindra Dataset
  - Using PSO we got (p, d, q) values of ARIMA as (3,1,2) best fit.
  - Also we got 512 hidden units present in the RBM layer.
  - Results over various performance metrics:

Performance Indicator	Models			
	ARIMA	LSTM	ARIMA-LSTM	ARIMA-DBN
MSE	175.278	265.064	204.015	**175.261**
MAE	9.897	12.620	10.975	**9.896**
RMSE	196.696	**16.281**	196.943	196.696
RMSLE	0.019	0.024	0.021	**0.019**
MAPE	0.232	**0.019**	0.231	0.232
SMAPE	22.383	**1.829**	22.519	22.383

Conclusion

- As PSO is used an optimization technique helps the hybrid models reaching their efficient goals.
- Based on Mean Squared Error, the series hybrid ARIMA-DBN model shows the least among all the other models.
- These results were compared to other existing models that was better than those models.

References

- Biju R Mohan, G Ram Mohana Reddy.: A Hybrid ARIMA-ANN Model for Resource Usage Prediction, International Journal of Pure and Applied Mathematics, Volume 119 No. 12, (2018).
- Mehdi Khashei, Zahra Hajirahimi.: A comparative study of series arima/mlp hybrid models for stock price forecasting, Communication in Statistics – Simulation and Computation, (2018).
- Mengjiao Qin, Zhihang Li, Zhenhong Du.: Red tide time series forecasting by combining ARIMA and deep belief network, Knowledge-Based Systems, (2017).
- Ina Khandelwal, Ratnadip Adhikari, Ghanshyam Verma.: Time Series Forecasting using Hybrid ARIMA and ANN Models based on DWT Decomposition, International Conference on Intelligent Computing, Communication & Convergence (2015).