Hybrid model of Bi-LSTM and CNN for Multivariate Time Series Classification for ecommerce sales Forecasting

Sanjay Saini

Final Thesis Presentation for
Masters in AI/ML
Liverpool John Moores University & upGrad
March 2024

Ecommerce Sales Forecasting

E-commerce sales are notoriously unpredictable. One day you could be swamped with orders, the next week facing a lull. This volatility makes it extremely difficult to:

- Manage Inventory
- Set Optimal Prices
- Run Effective Marketing

BENEFITS

Multivariate Time Series Classification

The process of predicting multiple, related variables simultaneously, where each variable is dependent on its own past values and the past values of other variables within the system.

A multivariate approach identifies complex interdependencies, revealing patterns invisible to single-variable models.

Better forecasts inform not just inventory but also pricing, marketing strategies, and overall business planning.

Aim & Objectives

- Highlight how multivariate time series forecasting with a hybrid deep learning model can drive superior business outcomes in e-commerce.
- Promote the use of these innovative techniques to improve e-commerce decision-making and gain a competitive advantage.
- Outline the pattern recognition strengths of CNNs and the temporal dependency modeling of Bi-LSTMs.
- Present performance comparisons of the hybrid model against traditional forecasting methods.

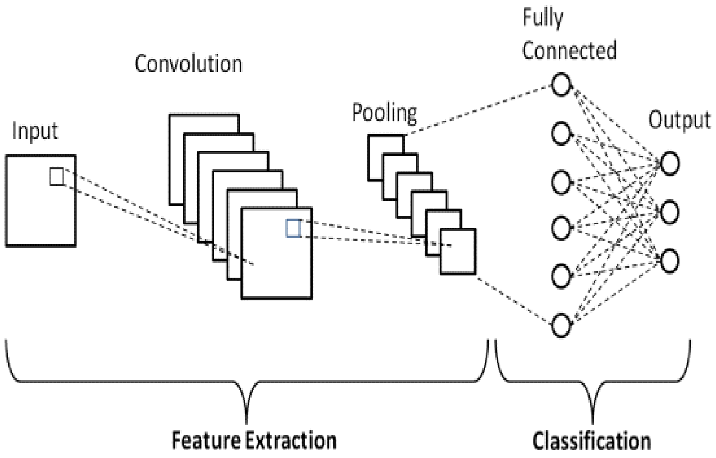
Convolutional Neural Network (CNN)

CNN, or Convolutional Neural Network, is a deep learning architecture which employs convolutional layers to extract spatial patterns and features from input data.

In time series forecasting, CNNs can be adapted to analyze temporal patterns by treating the time dimension as spatial.

This allows CNNs to efficiently capture local dependencies and patterns within the time series data

They can capture both short-term and long-term dependencies in the time series data, enabling accurate forecasting across various temporal scales.

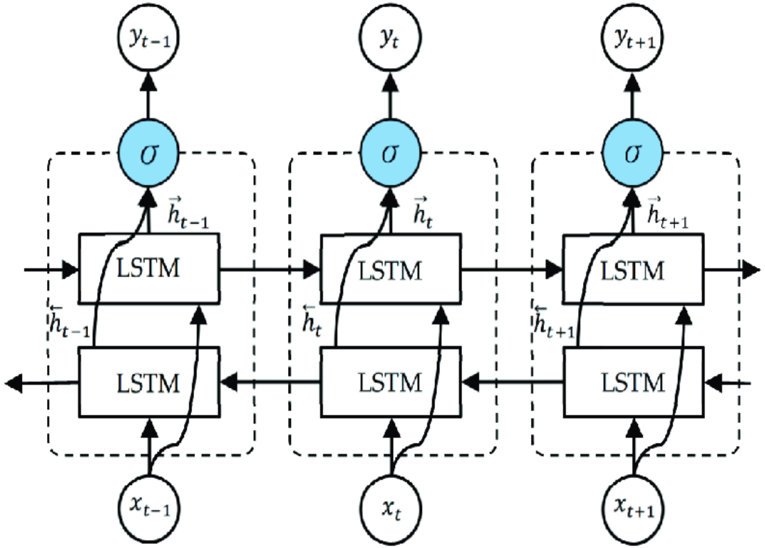


Bidirectional Long Short Term Memory (Bi-LSTM)

Bi-LSTM is a variant of the Long Short-Term Memory (LSTM) architecture, a type of recurrent neural network (RNN), specifically designed to capture long-term dependencies in sequential data.

By processing input sequences bidirectionally, Bi-LSTMs have the capability to learn from the entire history of the time series, providing a comprehensive understanding of the underlying patterns.

Bi-LSTMs offer improved performance over traditional LSTM models by incorporating information from both past and future contexts.



Literature Review on Multivariate Time Series Forecasting

As referenced by Ensafi et al. (2022), and Khan et al. (2021), sales data contains temporal and spatial dependencies, To manage these we tried to optimize the blend of CNN and Bi-LSTM as CNNs are excellent at extracting hierarchical spatial features, while Bi-LSTMs are adept at capturing long-term dependencies and relationships.

From the studies of Akhtar & Shah (2021), Karim et al. (2017), Khan et al. (2021), Ko et al. (2021), Sirisha et al. (2022), and Wei et al. (2022), comparing the effectiveness of hybrid model with that of traditional and singular-architecture models, we evaluated how well the hybrid model can forecast e-commerce sales by considering multiple variables.

Traditional Methods:

- Autoregression (AR)
- Moving Average (MA)
- ARIMA
- SARIMA

Methodology: Data Collection and Exploratory Data Analysis

The data used in the study was sourced from Corporación Favorita, a large Ecuadorian-based grocery retailer, which is an extensive dataset obtained from Kaggle.

This dataset contains a wealth of relevant features and predictors that can be used in sales forecasting.

The data is about store sales forecasting containing 54 stores having 33 products in 16 states.

Methodology: Model Implementation

Layer (type)	Output Shape	Param #
conv1d_29 (Conv1D)	(None, 1, 32)	992
bidirectional_26 (Bidirectional)	(None, 1, 64)	16640
bidirectional_27 (Bidirectional)	(None, 64)	24832
dropout_11 (Dropout)	(None, 64)	0
flatten_14 (Flatten)	(None, 64)	0
dense_14 (Dense)	(None, 1)	65

Total params: 42,529
Trainable params: 42,529
Non-trainable params: 0

Architecture of Proposed Model

Model Architecture

- **Convolution Layer**: Performs convolution on the input.
- **Pooling Layer**: Reduces dimensionality
- **bi-LSTM Layer**: process the sequence in both forward and backward to capture long-term dependencies and context.
- **Dropout Layer**: Regularization technique to prevent overfitting.
- **Dense Layer**: For mapping the input to single output.

Hyperparameter Tuning

Hyperparameter tuning helps identify the optimal combination of hyperparameters that yield the best performance on the validation set. This can lead to higher accuracy and better generalization of the model.

Performed cross validation using Grid Search to search for the best hyperparameters for model such as different number of neurons, dropout rate, optimizer and number of epochs.

Models were ranked on the basis of MSE (Mean Squared Error) values.

Selected top two best performing models after the search and trained them.

Results

Model	Optimizer	Neurons	Dropout Rate	Epochs	RMSE
Model_1	adam	32	0.1	100	121.539159826848
Model_2	adam	64	0.2	100	123.952789968161

Results of the CNN/Bi-LSTM Model

Authors	Methodology	RMSE
(Saar-Tsechansky and Provost., 2007)	Random Forest	282.07174328758134
(Saar-Tsechansky and Provost., 2007)	Gradient Boost	206.54946863732793
(Kaunchi et al., 2021)	LSTM/CNN	810.39743331262634
(Chu and Zhang., 2003)	Linear Regression	1002.44253781232982

Results of Traditional Forecasting Algorithms

Conclusion

- This research demonstrated the effectiveness of a hybrid CNN/Bi-LSTM model for addressing the challenges of Multivariate Time Series Classification.
- The CNN component successfully extracted meaningful features from complex multivariate input data, while the Bi-LSTM effectively captured long-term dependencies within the time series.
- Experimental results on sales data indicated that the proposed model surpasses the performance of other state-of-the-art MTSC methods in terms of MSE.

Future Scope

- Incorporating attention mechanisms to further enhance the model's ability to focus on crucial features.
- Investigating alternative hyperparameter optimization techniques such as Bayesian optimization or random search.
- Investigating the explainability of the model for better understanding of its decision-making processes.
- Using better grading metrics like RMSLE (Root Mean Squared Logarithmic Errors).
- Investigating other hyperparameters like learning rate, adding more dropout layers etc.

Thank you