Walmart baseline sales forecasting using LSTM

Anjali Sundareswaran Varier
Eric Johnsson School of Engineering and Computer Science
University of Texas at Dallas
Richardson, USA
axv180027@utdallas.edu

Aparna Cleetus
Eric Johnsson School of Engineering and Computer Science
University of Texas at Dallas
Richardson, USA
axc190011@utdallas.edu

Nachiappa Palaniyappa
Eric Johnsson School of Engineering and Computer Science
University of Texas at Dallas
Richardson, USA
nxp190002@utdallas.edu

Satya Sarvani Baru
Eric Johnsson School of Engineering and Computer Science
University of Texas at Dallas
Richardson, USA
sxb180153@utdallas.edu

***Abstract*—The most crucial part in any E-commerce business is the process of generating accurate and reliable sales forecast. In this project an attempt has been made in predicting the department wide sales for Walmart stores based on the historical sales data for 45 stores located in different regions by training a LSTM model. We use LSTM (Long Short Term Memory) because of it's ability to retain memory. In traditional models which forecasts sales we found that the models were univariant which does not use historical data which is found to be very inefficient .LSTM learns to predict the sales based on not only the features like Store, Department, Temperature, Fuel price etc, but also on how the sales fared in the previous days and weeks. Hence it uses the previous predictions in the current predictions. From our experiments we see that this model achieves good results on departmental and category level sets of data.**

Keywords—LSTM, univariant, sales, historical data, forecast

I. INTRODUCTION

A crucial problem in any E-commerce business is the process of generating accurate and reliable sales forecast. Once an accurate estimate of the sales prediction is available it helps in better demand planning, inventory planning, time phased replenishment plans and helps with promotions. A business that makes use of these strategies with the help of accurate sales prediction is bound to make huge savings and reductions in cost. On the other hand poor predictions can lead to costly losses.

Our aim here is to train LSTM models using Walmart's Recruiting: Store Sales Forecasting dataset in order to predict the weekly sales based on the historical sales data for 45 stores located in different regions. Studies have shown that the sales in E- commerce are highly dynamic and volatile in nature, which is an effect of competition within market, holiday sales and product sales conversion rates. However, it is seen that the times series of these factors are related to each other and they do impact the demand and sales prediction .Hence it is important to consider the correlation for these variables which brings Recurrent Neural Networks(RNN) and Long Short-Term Memory (LSTM) into picture. These are special groups of neural networks that are suited for time series forecasting. These have proven to give high accuracy results for time series prediction which is best suited for our data.

In our experiment we are splitting the Walmart's Recruiting: Store Sales Forecasting dataset into training and test based on the Date then training the model with various epochs to obtain the highest accuracy and least mean square error.

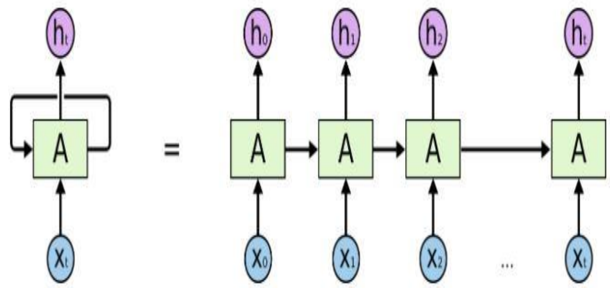
II. BACKGROUND WORK

The Walmart's Recruiting: Store Sales Forecasting dataset is dependent on numerous features like the Store Number, Department Number, Size of the store, whether there is special holiday in that week, the average temperature in the particular region, the cost of fuel in that region, the unemployment rate, the consumer price index(CPI) and the markdown prices. Since there are these many features affecting the sales prediction, we see a lot of challenges in the data pattern such as the data being highly non historical, highly intermediate sales, missing data and many more. As mentioned in the introduction the time series of these features are correlated and this must be taken into consideration while choosing models and making predictions.

Traditional models like neural networks are influenced by univariate statistical forecasting. However, forecasting in E-commerce needs to address the challenges like irregular sales trend, highly busty data and sparse sales data. In addition to this we need to address the multi variant time series correlation. Hence, we go for RNN and LSTM models that take all these into consideration. The details of RNN and LSTM have been explained in the coming sections.

III. THEORETICAL AND CONCEPTUAL STUDY

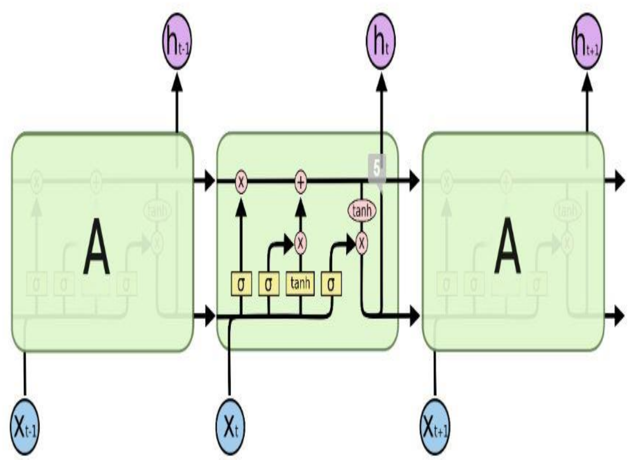
Basically, we started the implementation of RNN (Recurrent Neural Network) for our dataset. So, RNN is an abstraction of Neural Network with a memory in it. It is recurrent in nature, that means the previous data is also passed along with the input to predict the next state and this continuous till we get a required output.



An unrolled recurrent neural network.

But then we observed a problem that it can only be used if the current data is dependent on the recent one and cannot be used if the data rely on the past which is further back. So, we thought of implementing LSTM which is extended from RNN. LSTM refers to Long Term Short Memory which solves the above said issue. It can store the relevant information form the past and use it in the present. IT memorizes the information which can be used for future and forgets which ever are not relevant.

Following is the architecture of LSTM

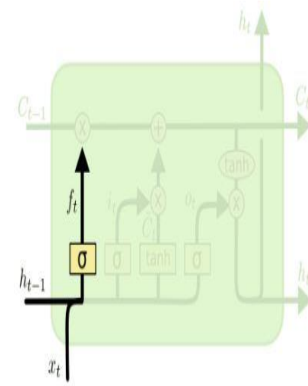


The repeating module in an LSTM contains four interacting layers.

There are 3 gate layers for LST which are "input gate", "forget gate" and "output gate".
The architecture of each of these layers are described below:

A. Forget layer

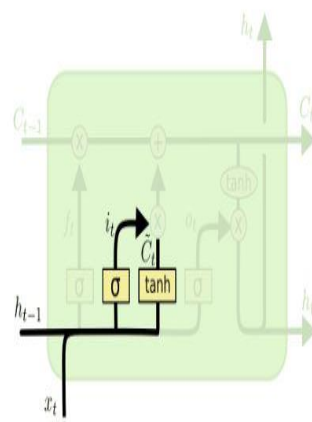
Forget layer focuses on removing all the information which are not relevant and will not be used in future. This is done with the help of sigmoid function which outputs 1 or 0 for each of them.



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right)$$

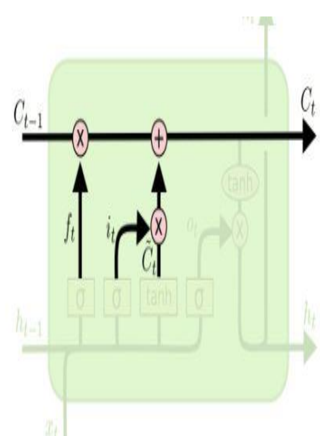
B. Input layer

Input layer focuses on the information that needs to be stored and all the valid suggestions that are applicable for a particular prediction. So, the sigmoid function will allow the relevant information to flow by outputting 0 or 1 and tanh gives the rank for each of them by the level of importance.



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right)$$

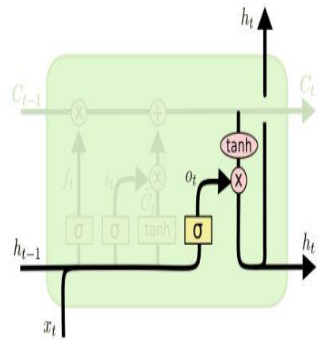
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

C. *Output layer*

Output layer gives the prediction for the current scenario. So, the output is decided by sigmoid function which output 1 or 0 that shows what should be outputted and the tanh function gives the rank based on the relevance of the information that needs to be displayed.



$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

IV. DATA PREPROCESSING

The Walmart's Recruiting: Store Sales Forecasting dataset is dependent on numerous features like the Store Number, Department Number, Size of the store, whether there is special holiday in that week, the average temperature in the particular region, the cost of fuel in that region, the unemployment rate, the consumer price index(CPI) and the markdown prices. The data set is available for the years 2010-02-05 to 2012-11-01. Before using this dataset, we need to do some amount of data preprocessing in order to address the issues of missing data, negative value and dates in different formats. Each of this is explained below:

a) Missing data: Missing data is addressed by filling these cells with zero values in order to avoid data mismatch.

b) Negative values: On data exploration it is seen that there are negative values for mark down and sales prediction on test data. We address this by replacing all negative values by zero inorder to avoid incorrect calculations.

c) Categorical and Boolean data: Categorical and Boolean data are given categorical values by performing the one hot and label encoder using sklearn's LabelEncoder.

d) Normalisation: Finally we are normalizing the dataset using Min-Max normalization and obtain a linear transformation of the dataset.

V. RESULTS AND ANALYSIS

The following are the results obtained from various experiments. The size of the dataset is 421570.

Based on the split, training data and testing data are obtained from the dataset and the mean squared error is calculated for both training as well as testing datasets.

Experiment Number	***Parameters Used***	***Results***	***Values***
1.	learning rate = 0.00000001	Dataset Size:	421570
	No. of epochs = 1	Train/Test Split:	0.965:0.035
	memory count = 14	Training Total Squared Error:	6.37E+02
		Test Total Squared Error:	1.91E+02
		Training RMSE:	1.56E-03
		Test RMSE:	1.28E-02
2.	learning rate = 0.1	Dataset Size:	421570
	No. of epochs = 1	Train/Test Split:	0.965:0.035
	memory count = 15	Training Total Squared Error:	4.98E+02
		Test Total Squared Error:	3.58E+02
		Training RMSE:	1.22E-03
		Test RMSE:	2.41E-02
3.	learning rate = 0.1	Dataset Size:	421570
	No. of epochs = 1	Train/Test Split:	72.6:27.4
	memory count = 15	Training Total	4.98E+02

Experiment Number	*Parameters Used*	*Results*	*Values*
		Squared Error:	
		Test Total Squared Error:	1.96E+03
		Training RMSE:	1.22E-03
		Test RMSE:	1.32E-01
4.	learning rate = 0.1	Dataset Size:	421570
	No. of epochs = 1	Train/Test Split:	72.6:27.4
	memory count = 15	Training Total Squared Error:	4.98E+02
		Test Total Squared Error:	1.96E+03
		Training RMSE:	1.62E-03
		Test RMSE:	1.70E-02
5.	learning rate = 0.5	Dataset Size:	421570
	No. of epochs = 1	Train/Test Split:	79.0:21.0
	memory count = 25	Training Total Squared Error:	5.40E+02
		Test Total Squared Error:	1.66E+03

Experiment Number	*Parameters Used*	*Results*	*Values*
		Training RMSE:	1.62E-03
		Test RMSE:	1.86E-02
6.	learning rate = 0.5	Dataset Size:	421570
	No. of epochs = 1	Train/Test Split:	79.0:21.0
	memory count = 25	Training Total Squared Error:	5.40E+02
		Test Total Squared Error:	1.66E+03
		Training RMSE:	1.62E-03
		Test RMSE:	1.86E-02

VI. CONCLUSION AND FUTURE WORK

We have used features like Store, Dept, Date, Unemployment, IsHoliday, Type, Size, Temperature, Fuel_Price, MarkDown1, MarkDown2, MarkDown3,MarkDown4,MarkDown5, CPI to predict the weekly sales of a particular store. After implementing the algorithm to the dataset, we can observe that as we give the number of iterations, the loss is decreasing in each iteration during the training, which seems to be good. And when we test the algorithm with the test data, the calculated loss less, so, we can conclude that the predicted values are close to the real values and so there is very high accuracy.

As future work, we could also extend this project to predict the products that will be sold more and less and category of the products in which most of the products are sold along with the weekly sales. Same algorithm can be used for the sale of a product from a particular category.

.

REFERENCES

[1] https://www.analyticsvidhya.com/blog/2019/01/fundamentals-deep-learning-recurrent-neural-networks-scratch-python/)

[2] https://www.kaggle.com/nsawal/walmart-baseline-sales-forecasting-lstm#5.-Data-pre-processing

[3] http://wp.firrm.de/index.php/2018/04/13/building-a-lstm-network-completely-from-scratch-no-libraries/

[4] https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[5] Chawla, Aditya & Singh, Amrita & Lamba, Aditya & Gangwani, Naman & Soni, Umang. (2019). Demand Forecasting Using Artificial Neural Networks—A Case Study of American Retail Corporation. 10.1007/978-981-13-1822-1_8.

[6] Kasun Bandara1, Peibei Shi2, Christoph Bergmeir1, Hansika Hewamalage1, Quoc Tran2, Brian Seaman2 "Sales Demand Forecast in E-commerce using a Long Short-Term Memory Neural Network Methodology