Application of Long Short-Term Memory Neural Network to Sales Forecasting in Retail—A Case Study

Quan Yu, Kesheng Wang, Jan Ola Strandhagen and Yi Wang

Abstract Sales forecasting is an important task for managers to make replenishment according to historical sales. A flexible and easy to use forecasting solution will benefit retailers from loss of sale, over supply and merchandise waste. Deep learning is a popular topic in many fields in recent years. This paper tests a long short-term memory (LSTM) recurrent neural networks (RNN) on 45 weeks point of sale (POS) data of 66 products without considering the impact of seasonality and promotions. One fourth of products have a relatively low forecasting error, which validates the feasibility of the LSTM network to some degree.

Keywords Deep learning · LSTM · RNN · Sales forecasting

1 Introduction

Sales forecasting is a process to estimate future sales according to historical sales and other relevant information, which enables a company to predict short-term and long term business and make corresponding decisions. In retail industry, sales forecasting is important for inventory control, supply chain management, replenishment, etc. A good forecast that fits the consumption demands can help retailers to extend profits, promote products with respect to consumption patterns, and control the safety stock without being excessive. However, sales forecast methodologies are typically more scientific than an intuitive chart, although different forecasting models are available in various software, forecast executives in companies would not risk tuning the parameters for a potential better forecast [1]. On the other side, machine learning develops dramatically in recent years and contributes in both

Q. Yu (✉) · K. Wang · J. O. Strandhagen
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway
e-mail: Quan.yu@ntnu.no

Y. Wang
School of Business, Plymouth University, Plymouth, UK

© Springer Nature Singapore Pte Ltd. 2018
K. Wang et al. (eds.), *Advanced Manufacturing and Automation VII*,
Lecture Notes in Electrical Engineering 451,
https://doi.org/10.1007/978-981-10-5768-7_2

industry fields and people's daily life, with relatively simpler parameter settings as an alternative to traditional model-based methods.

A big branch of machine learning is neural networks having decades of development and many variations. The Long Short-Term memory (LSTM) neural network is a sort of recurrent neural network (RNN) proposed in 1997 to address the problem of insufficient, decaying error backflow in RNN training [2]. However, research interests on LSTM have been slowly increasing until 2014 as shown in Fig. 1. Publications are dramatically increased along with the application of LSTM to natural language processing by well-known companies such as Google [3]. Since then, LSTM has been applied in fields including robot control [4], speech recognition [5], time series forecasting [6], etc.

Because of the special structure, an LSTM is able to learn from historical information to classify, process and predict time series and reflects important events. This paper tests an LSTM network on real sales data of 66 products collected from a retailer during 45 weeks. As a general experiment, seasonality and promotions in practical are not under consideration. The results reflect the potential of implementing LSTM network in retail sales forecasting.

2 Long Short-Term Memory Neural Network

Similar to neurons in standard neural networks, the central part in a LSTM architecture is a memory cell which can maintain its state over time, and non-linear gating units which regulate the information flow into and out of the cell [7]. Comparing with traditional RNN, an LSTM neural network uses memory units with gates rather than neurons to establish connections between inputs and outputs [8]. Figure 2 shows a standard LSTM memory cell, where

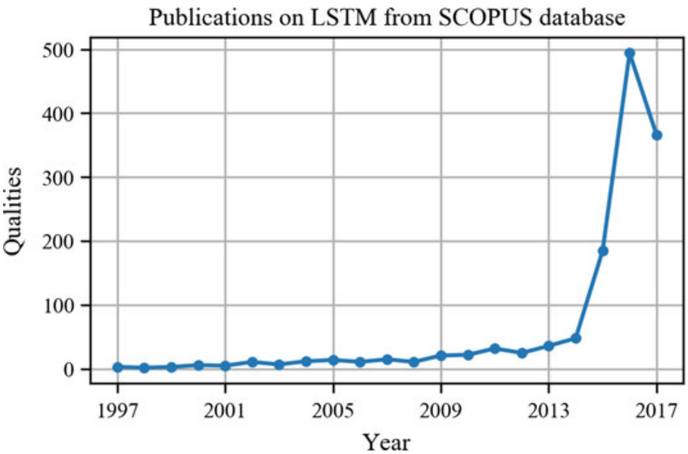

Fig. 1 Trend of research on LSTM

x_t is the input x and time t,
h_t is the output with respect to x_t,
h_{t-1} is the output of previous memory cell, also called hidden state,
C_t is the cell state while C_{t-1} is the cell state of the previous cell.

Within the memory cell, the input x_t will be processed together with the previous output h_{t-1} by three steps respectively, i.e. forget step, input step and output step.

- The forget step contains a sigmoid function, which returns a value between 0 and 1, while 0 means no information pass and 1 means all information pass.
- The input step has a sigmoid function and a tanh function, with the purpose of adding new information according to x_t and h_{t-1}.
- The output step is another sigmoid function to decide how much information is included in the output h_t.

Additionally, the cell state C_t is updated according to the previous cell state C_{t-1}, the output of the forget step for h_{t-1} and x_t, and the output of input step for h_{t-1} and x_t, thereafter feed to the output h_t and the next cell as a reference.

By stacking memory cells, information of original input $\boldsymbol{x}$ is kept in the final output $\boldsymbol{y}$ to some degree, carried by cell state $\boldsymbol{C}$ and cell by cell output $\boldsymbol{h}$. To be noticed, the stack of the memory cells can mimic a time series. Based on Fig. 2, correspondence between a time series and an LSTM network is shown in Fig. 3. Suppose we use an LSTM network with three memory cells for time series forecasting and $\boldsymbol{x}$ is a time series, cells with the same internal structure are stacked and each of them holds an input x_i sequentially. When an x_i is input to a memory cell, it is processed and output to be the cell state C_i and hidden state h_i, as inputs for the next cell together with next x_{i+1}. The last hidden state h_3 is taken as the final output y_4 corresponding to x_4. By sliding this three cells LSTM network along the time

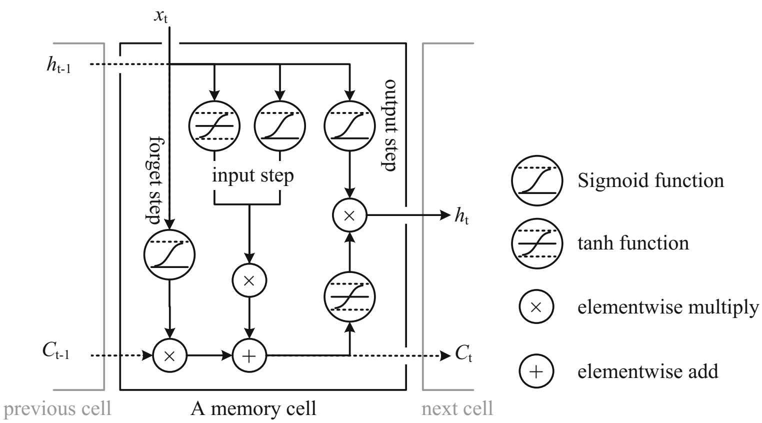

Fig. 2 An LSTM memory cell

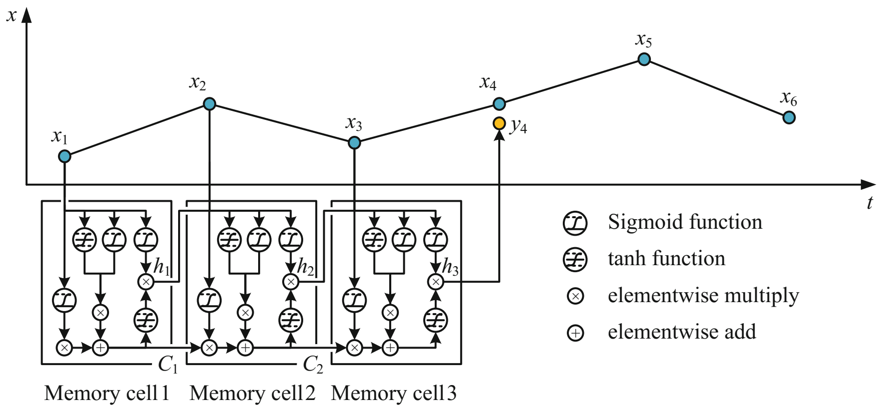

Fig. 3 A three cells LSTM network for time series forecasting

axis and aligning to the time stamp, every three inputs x will have the fourth x as the output, which can be used as training data to train the network and fix the parameters, thereafter to be ready for the forecasting later on.

3 Experiment and Results

In this paper, the LSTM network is established in python environment with Keras deep learning library running on top of TensorFlow library developed by Google. The sales data is of 66 products in 45 weeks in 2015, covering milk products, fruits and vegetables, meat products, etc. The forecasting is performed in week level, while sales of four consecutive weeks are used to forecast the sale of the fifth week. Accordingly, the LSTM network is set to have four memory cells. Raw data of each product is sequentially divided in the ratio 2:1 to be training data and test data. In other words, the data of first 30 weeks is taken as training data, while the latter 15 weeks data is used as test data. Furthermore, for better evaluating the performance of LSTM network, the sales values are scaled down between [0, 1] in the test so that they are normalized and 10 experiments are performed for each product and the average mean square error (MSE) is calculated as the final performance measure. For each product, the work flow is shown in Fig. 4. An example of the experiment results is shown in Fig. 5. There is no significant pattern in raw data and no large bias between raw data and forecasted values at the beginning, the forecasting accuracy is decreased along with a fluctuation of historical data at time

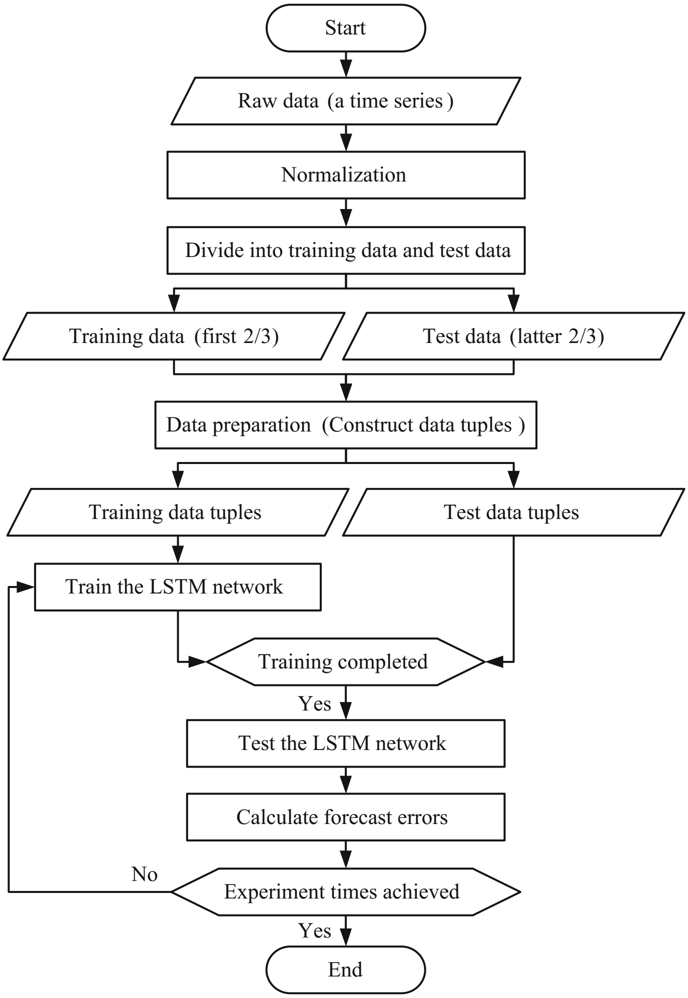

Fig. 4 Work flow of the experiment of one product

stamp 6. Regarding all products, 10 forecasting experiments are performed for each product and the final results are shown in Fig. 6, measured by average MSE values. There are about one fourth products having an MSE lower than 0.3, which means the forecasting accuracy for other products is not good enough.

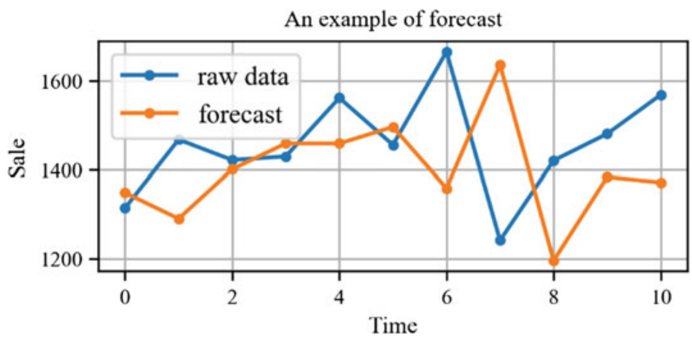

Fig. 5 A forecasting example from the results

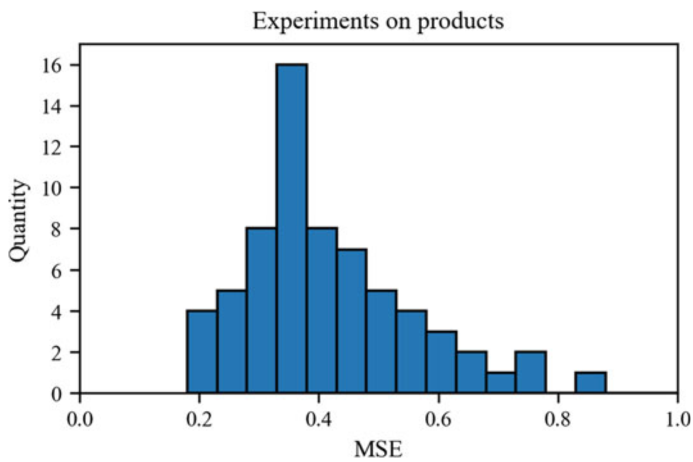

Fig. 6 Experiment results on products

4 Discussion and Conclusions

Sales data in retail industry can be seen as a sort of time series, which have been researched for decades. Various models are proposed and implemented, however, executives tend to not tune parameters in practical use. As a machine learning techniques, LSTM networks are relatively understandable and easy to mimic the time series forecasting process. Regarding the experiment on 66 products, there are only one fourth products having a relatively low forecasting error, while results of other products are not positive enough, this is because

- The length of raw data is maximum 45 weeks, even less for some products in datasets, which is further divided into training data and test data. It is not sufficient to obtain a robust neural network in this case.

- Seasonality is not considered in the test, moreover, 45 weeks data only contains some short-term seasonality not the long-term seasonality.
- In practical, promotions impacts much on the sales in a way of large fluctuation, which is neither included in the experiment.
- The experiment uses one LSTM network for all products, which means it is not optimized with respect to characteristics of the product.

Without considering the impacts of promotion and seasonality, LSTM network still shows potential for one fourth products in sales forecasting with limited datasets. Further works will be developed in consideration of promotions and seasonality to improve the stability of the network.

Acknowledgements This work was supported by the Research Council of Norway through the Retail Supply Chain 2020.

References

1. Economist Intelligence Unit (2012) Data versus instinct—perfecting global sales performance. http://www.oracle.com/us/products/applications/eiu-oracle-sales-1719684.pdf?sc=EMEAOU13045591MPP008C147
2. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
3. Gonzalez-Dominguez J, Lopez-Moreno I, Sak H, Gonzalez-Rodriguez J, Moreno PJ (2014) Automatic language identification using long short-term memory recurrent neural networks. In: Proceedings of the annual conference of the international speech communication association (INTERSPEECH), pp 2155–2159. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84910072596&partnerID=40&md5=a73a851422179feea8df33e70f2b6af9
4. Bakker B, Linåker, F, Schmidhuber J (2002) Reinforcement learning in partially observable mobile robot domains using unsupervised event extraction. In: IEEE international conference on intelligent robots and systems, pp 938–943. https://doi.org/10.1109/IRDS.2002.1041511
5. Gers FA, Eck D, Schmidhuber J (2001) Applying LSTM to time series predictable through time-window approaches. In: Dorffner G, Bischof H, Hornik K (eds) Artificial neural networks—ICANN 2001: international conference Vienna, Austria, August 21–25, 2001 proceedings, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 669–676. https://link.springer.com/chapter/10.1007/978-1-4471-0219-9_20
6. Wöllmer M, Marchi E, Squartini S, Schuller B (2011) Multi-stream LSTM-HMM decoding and histogram equalization for noise robust keyword spotting. Cogn Neurodyn 5:253. https://dx.doi.org/10.1007%2Fs11571-011-9166-9
7. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J (2015) LSTM: a search space odyssey. ArXiv e-prints. https://arxiv.org/ct?url=http%3A%2F%2Fdx.doi.org%2F10%252E1109%2FTNNLS%252E2016%252E2582924&v=57f036b2
8. http://colah.github.io/posts/2015-08-Understanding-LSTMs/