Mateusz PĘCZEK[1], **Grzegorz LITAWA**[2]

[1] State University of Applied Sciences in Nowy Sącz, Institute of Technology, 33-300 Nowy Sącz, Zamenhofa 1a street, e-mail: mbspeczek@gmail.com
[2] State University of Applied Sciences in Nowy Sącz, Institute of Technology, 33-300 Nowy Sącz, Zamenhofa 1a street, e-mail: glitawa@pwsz-ns.edu.pl

Time series forecasting using the LSTM network

Abstract

Predicting time series is currently one of the problems that can be solved through the use of machine learning methods. A time series is a set of data points in which the sequence is measured at equal time intervals. Predicting the value of the time series can influence your decisions or help you achieve better results. Stock quotes are an example of a time series – the purpose of the created model is attempt to predict their value. One solution to the problem of predicting the results of the time series is the LSTM network. The network contains layered LSTM cells that have the ability to use previously observed relationships in the data set. The number of LSTM layers and cells in each layer is dependent on the designer and is selected based on expert knowledge. The results obtained from the model may seem correct and close to the real ones. Regardless of what values we get and how high the accuracy of the model will be, it should be remembered that stock prices are influenced by parameters and events that cannot be predicted. The predicted values obtained from the model should be treated as a guide or reference information. Stock quotes may change under the influence of geopolitical situations, company involvement, armed conflict or other random and unpredictable phenomenon, therefore, when making decisions, the results of the model should not be taken for granted.

Key words: machine learning, LSTM, value prediction, time series, Keras.

Prognozowanie szeregów czasowych z użyciem sieci LSTM

Streszczenie

Przewidywanie szeregów czasowych jest obecnie jednym z problemów, które mogą zostać rozwiązane poprzez zastosowanie metod uczenia maszynowego. Szeregiem czasowym nazwiemy zbiór danych, w których pomiar odbywał się w jednakowych odstępach czasu. Przewidywanie wartości szeregu czasowego może wpłynąć na podejmowane decyzje lub pomóc w osiąganiu lepszych wyników. Przykładem szeregu czasowego są notowania giełdowe – celem utworzonego modelu jest próba przewidywania ich wartości. Jednym z rozwiązań problemu przewidywania wyników szeregów czasowych jest sieć LSTM. Sieć zawiera warstwowo ułożone komórki LSTM, które mają zdolność do wykorzystywania wcześniej zaobserwowanych zależności występujących w zbiorze danych. Liczba warstw i komórek LSTM w każdej warstwie jest zależna od projektanta i dobiera się ją w oparciu o wiedzę ekspercką. Wyniki otrzymane z modelu mogą wydawać się poprawne i zbliżone do rzeczywistych. Niezależnie od tego, jakie wartości otrzymamy i jak duża będzie dokładność modelu, należy pamiętać, że na notowania giełdowe wpływ mają parametry i zdarzenia, których nie da się przewidzieć. Wartości przewidywane, otrzymane z modelu, należy traktować jako pomoc lub informacje poglądowe. Notowania giełdowe mogą zmieniać się pod wpływem sytuacji geopolitycznej, upadku firmy, konfliktu zbrojnego lub innego losowego i niemożliwego do przewidzenia zjawiska, dlatego przy podejmowaniu decyzji nie należy traktować wyników modelu jako pewne.

Słowa kluczowe: uczenie maszynowe, LSTM, przewidywanie wartości, szeregi czasowe, notowania giełdowe, Keras.

1. Introduction

Machine learning is a certain field of science and computer programming, within the system created by us is able to learn on the basis of the datasets we have provided (Géron, 2020; Brownlee, 2020a; Brownlee, 2020b; Livieris, 2020; Essien and Giannettic, 2020). This is one of the more open definitions of machine learning. However, for the needs of a simple definition of this phenomenon it is fully sufficient. The most common categorization of machine learning is based on three criteria. The first defining characteristic that can be considered is the ability to learn in real time. Incremental learning and batch learning are two examples. The batch learning system accepts a set of data prepared once and starts learning on its basis. A system that uses batch learning cannot learn on the fly. The second method is incremental learning. It allows you to reduce costs and learn continuously. A system taught by means of incremental learning has the ability to learn when data appears at its input. This learning model works well in systems where data flows in and does not need to be archived. The second criterion defining machine learning is the need to supervise the training process. Some topics include supervised learning and unsupervised learning. Supervised learning involves the preparation of a learning data set with solutions to problems, the so-called labels. The classic task of supervised learning is classification. Similarly, the unsupervised learning method is based on the use of a training set without labels. The system will try to create them itself. The third criterion is the workflow. This is done by comparing the data you have with the points you already know. The issues of this criterion are training from examples and training from a model. The training method from examples consists in passing to the system all the examples contained in the data block. Then, with the help of a similarity measurement, the affiliation of the object is determined. Learning from the model consists in generalizing the results contained in the training data and building an appropriate model on their basis. In this method, the key is to choose the right model and its parameters. To find out how good a model is, you need to find a utility function, also known as a fitting function. The key task is to minimize the distance or the cost function between the actual and predicted results (Géron, 2020).

2. General description of the time series

A time series is a set of values that have been observed at regular intervals. So it will be any manifestation of a sequence of information that has been measured at regular intervals. The time series in which the time intervals between measurements are not equal is called a fuzzy time series. Depending on the frequency of observations, the series can be: seconds, minutes, days, etc. Time series can be presented as a function of the value against time.

3. Description of the LSTM cell with its presentation

Long Short-Term Memory networks are example of recursive artificial neural networks used in the field of deep learning. Unlike classical neural networks, they have feedback connections that affect their learning ability. Introduced in 1997, multitasking and the ability to solve difficult problems influenced the development of interest in LSTM networks and spread their application. Currently, LSTM networks are used for handwriting recognition, speech recognition, etc. LSTM networks can be used to predict time series values.

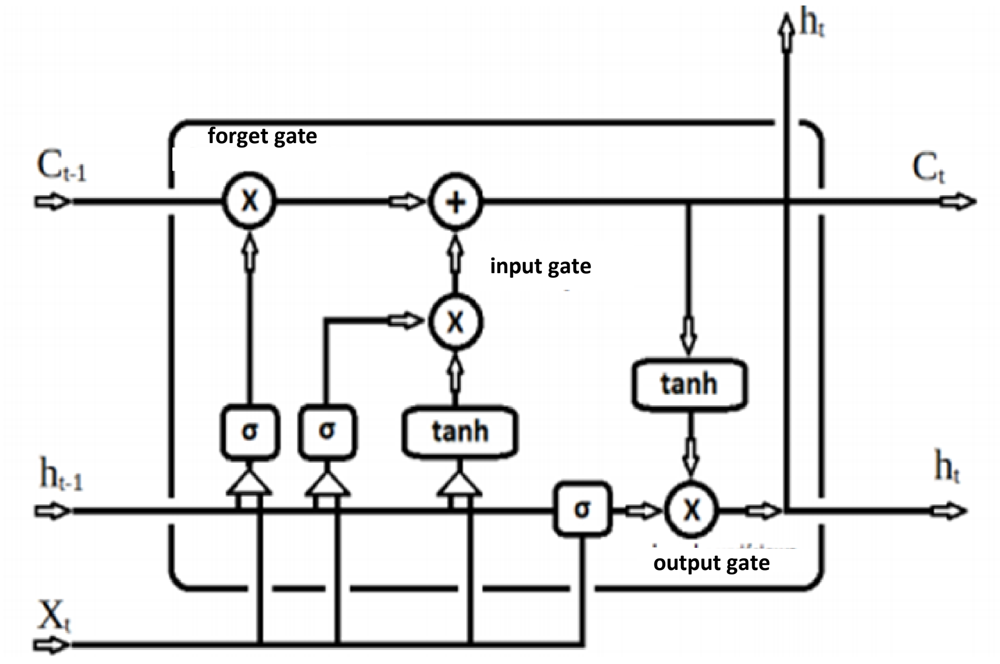

Figure 1. Long Short-Term Memory Cell Model

An LSTM cell can be considered in two ways. The first is to treat it as a black box. In this case, the state of the cell is divided into two vectors: $h_{(t)}$ – corresponds to the short-term state of the cell and $C_{(t)}$ – corresponds to the long-term state of the cell. The second way is to define the long-term state of $C_{(t-1)}$ starting on the left side of the cell and ending on the right side of it. Long-term state, it passes the forget gate and changes its value, then sums up with the entry gate and is passed on without any modification. The long-term state $C_{(t)}$ is copied and passed through the *tanh* function (hyperbolic tangent), the result of this operation is filtered by the output gate, the state $h_{(t)}$ is created – short-term. This state is equal to the output of the cell for this time step (Géron, 2020; Ai et al., 2019; Demertzis et al., 2019).

4. Application of LSTM

The LSTM network can be used to forecast time series values. The way in which information is modified by this network plays a key role in this process. In an LSTM network there is a relationship between cells. As in any recursive network, it is possible to use previously obtained values for the currently performed task. Stock quotes are a special case of the time series. Quotation values can be analysed and predicted. The aim of the research is attempt to predict the value of stock exchange quotes based on the previous results. This means that we will try to get new results based on the LSTM network and data from earlier days. The key role in this task is played by the appropriate selection of parameters for the network. The key parameters will be: the optimizer, the number of generations's, the number of samples for a given generation, and the number of neural network layers along with the number of cells in each layer and the dropout rate.

5. Description of the prediction method based on the LSTM network

The LSTM network used for prediction will use pre-prepared datasets. The scaled and prepared data set was divided into groups: the training set and the validation set. In order to be able to compare the results predicted by the network with the actual results, part of the general set has been reduced by the number of predicted days. Before training the neural network, it is necessary to define the loss function and the optimizer. The Keras library provides many prepared and ready to use loss functions: "mean squared error", "mean absolute error", "mean absolute percentage error", "mean squared logarithmic error", "cosine similarity" etc. The following optimizers are available: SGD, RMSprop, Adam, Adadelta, Adagrad, Adamax, Nadam, Ftrl (Chollet, 2019; Albon, 2018). The optimizer used in the model and the most commonly used optimization algorithm is ADAM. Apart from the optimizer parameters and the loss function, there are other parameters that significantly affect the behavior of the model and the correctness of the results obtained. Before starting the prediction, it is necessary to build an appropriate model based on the layers and number of LSTM cells. An important parameter is also the number of generations and the size of the samples. The sample size defines the number of iterations that will be performed during one generation (Loukas, 2020; TensorFlow org., 2021).

The first of several models that will be presented includes 3 LSTM network layers with 40 cells each. It is common to fix the number of cells for each network tier for the number of days to be predicted.

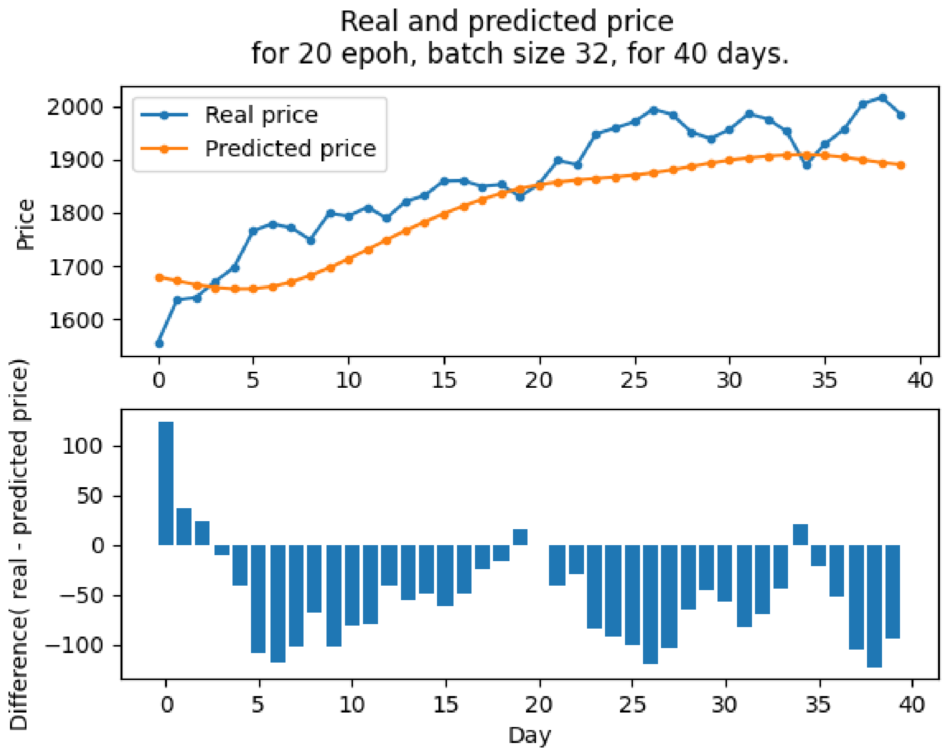

Figure 2. Predicted and actual value for 5 generations 40 days ahead

Increasing the size of the data batch resulted in little change in the fitting. The above examples show that the matching of the graph of the predicted value and the actual value depends on the appropriate selection of the appropriate parameters of the model.

6. Construction of the LSTM network model with hyperparameters

Enabling the model as Sequential, the network model will be built on its basis.

model = Sequential()

Adding LSTM layers with the number of cells determined by the number of steps, determining the shape of the input based on the shape of the input data.

model.add(LSTM(units=number_steps,return_sequences=True, input_shape=(training_set_x.shape[1],1)))

Adding two more layers along with information on the number of cells that will be deleted.

model.add(LSTM(units = number_steps, return_sequences = True))
model.add(Dropout(0.2))
model.add(LSTM(units = number_steps))
model.add(Dropout(0.2))

Output layer with one output. Setting the loss function with the optimizer.

model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='ADAM')

The above model is an example of how a model can be made to make predictions based on the LSTM network. However, this is not the only model that can be created. The accuracy of approximation of the output data to the actual value depends on the construction of the model and the selection of its parameters.

7. Conclusions

Stock quotations depend on many factors, generally changing in time, they may be: geopolitical situation, random events, fluctuations in global markets, conflicts, etc. The model presented above does not take into account any of these parameters. It is worth remembering that the model while learning is based on previous values. The results obtained from the model show that when the model has enough data, it is possible to make a prediction based on this data. As the previous model configurations show, it is possible to predict stock market values, but this should be treated more as information about the relationships between earlier quotation values than as an accurate way of predicting reality. It can be noticed that the model was able to process the data and return the results, the graph of which is able to reflect the behavior of the quotations in the future. Nevertheless, such a solution should be treated more as a form of support than a real market forecasting tool. The data used to train the model were based on the market quotations, the behavior of which did not record sudden changes, and the quotation values increased. When making predictions, keep in mind that the stock market can change at any time.

References

Ai, Y., Li, Z., Gan, M., Zhang, Y., Yu, D., Chen, W., Ju, Y. (2019). A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. *Neural Comput Appl, 31*(5), 1665-1667.

Albon, C. (2018). *Uczenie maszynowe w Pythonie*. Receptury. Gliwice: Helion.

Brownlee, J. (2020a). *How to Develop LSTM Models for Time Series Forecasting*. Download from: https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/.

Brownlee, J. (2020b). *Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras*. Download from: https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/.

Chollet, F. (2019). *Deep Learning. Praca z językiem Python i biblioteką Keras*. Gliwice: Helion.

Demertzis, K. Iliadis, L. Bougoudis, I. (2019). *Gryphon: a semi-supervised anomaly detection system based on one-class evolving spiking neural network*. Download from: https://doi.org/10.1007/s00521-019-04363-x.

Essien, A. Giannettic, C. (2020). A Deep Learning Model for Smart Manufacturing Using Convolutional LSTM Neural Network Autoencoders. *IEEE*, 6069-6078.

Géron, A. (2020). *Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow*. Gliwice: Helion.

Livieris, I.E. (2020). A CNN–LSTM model for gold price time-series forecasting. *Neural Computing and Applications, 32*, 17351-17360.

Loukas, S. (2020). *Time-Series Forecasting: Predicting Stock Prices Using An LSTM Model*. Download from: https://towardsdatascience.com/lstm-time-series-forecasting-predicting-stock-prices-using-an-lstm-model-6223e9644a2f.

TensorFlow org. (2021). *Prognozowanie szeregów czasowych*. Download from: https://www.tensorflow.org/tutorials/structured_data/time_series.