DOI : https://doi.org/10.56452/7-4-237

ISSN: 0974-5823 Vol. 7 No. 4 April, 2022

International Journal of Mechanical Engineering

LSTM and GRU gathered stock DATA ANALYSIS USING DEEP LEARNING paradigms

Deepak Upadhyaya

Asst. Professor, Department of Computer Science, Graphic Era Hill University, Dehradun, Uttarakhand India 248002

Abstract

Machine learning and deep learning both are evolved from artificial intelligence. Deep learning is more enhanced approach and can solve the issues which couldn't be solved by basic machine learning approaches. Neural network structure designing is the best example of the deep learning which shows the strength and importance of deep learning. Now Deep learning is applied on mostly all of the areas whether it is medical, technology, weather forecasting so on. Deep learning has multiple structures; out of them two structures i.e. LSTM method and GRU method are used to foresee the prices and for modeling the stock exchange. In this paper we take two chronological datasets for analysis which have 5044 of financial data each. To test the data we use 20% data of the datasets and 80% used to train the data. Researcher had evaluated the accuracy and performance of the models by measuring the errors and time consumption.

Keywords: *LSTM, GRU, Deep learning, Datasets.*

1. Introduction

In this modern era technologies is much sought after, which ease human tasks and used in various applications which is highly influential on how the world thinks and interprets activities. Machine Learning and Deep learning have set a major break at how technologies are implemented and are made to look like how human perceives things. Machine Learning methodologies had its own drawbacks and by deploying Deep Learning, most of the companies have witnessed a desirable growth and improvement in their product and services which is one of the major factors that is attributed to the successful implementation of technologies.

Deep learning was inspired by Artificial Intelligence - a subject viewed as a technical dimension that tries to achieve the capability to perceive, think, act in a similar way humans do and the ability to rationalize, take actions for achieving the desired goals [1]. The foundations of their functionality lies in the mathematical and statistical theories [2] designed as neurons which are simple and connected processors, applied as algorithms and implemented in manmade architectures that were inspired from the human cerebral cortex and other brain region Built upon an idea based on manmade neurons, Neural Networks are known to be the backboneof Deep Learning and approach the problem by deciphering them in a different way. The authors compared, analyzed the system and concluded that the best results have been achieved. The system couldbe applicable in other applications pertaining to speech recognition and could provide a platform for devising novel deep learning architectures. Nevertheless its application varies from Natural Language Processing, Forecasting [7] and Logistical optimization to Robotic applications. A paper by S. A. Hasan and Oladimeji Farri [8] gives a brief account of how deep learning is applied to clinical language processing and has also elaborated about the different applications pertaining to clinical data. A brief analysis on the detection of grasping pointsby robotic systems using deep learning was submitted by Shehan Caldera et al [9]. The authors have discussed the many methods where successful implementations have been achieved and the overall benefits, limitations and a promising development in the future for this application while applying deep leaning approaches in the field of robotics.

Neural Networks and its components have undergone various modifications that were measured according to their performances, up gradations and variations in deriving the desired outputs [10]. Miikkulainen et al [11] have put forth the notion that establishing architectures and modeling according to the application is a challenging task. They have suggested a model thatoptimizes deep learning architectures and fitted to cater in the field of object recognition and language modeling, applying in a magazine website by capturing images and proving that the approach can be implemented to get better results in various other applications. Major breakthroughs in Deep Learning are still yet to be achieved and the study of different networkarchitectures [12] allow us to assess the strengths and weaknesses and further facilitate Deep Learning for achieving goals ultimately.

This paper attempts to evaluate two of the neural network architectures and analyze its predictive and computational performance by applying it in two historical datasets from Indianstock exchange. Support and Resistance level in a stock determine the maximum range the pricelevel goes down and up respectively before reverting over a certain period of time. The levels are studied to determine the price points where the investor can choose to buy or sell the sharesat maximum advantage.

2. Deep Learning Architectures

Neural networks have a plethora of architectures that are built with varying rules pursued by the distinct characteristics of the inputs and the specific output as demanded by the applications.Initially developed from perceptrons, the simplest networks are the FFNN, CNN, Multilayer Perceptron, RBFNN, RNN [13]. Eventually the networks evolved to much more advanced architectures [14] like Alexnet, VGG net, Goolge Net, ResNet, Region based CNN, developed for applications that need more flexibility or adaption according to the desired outcome. For applications related to sequential data the architecture based on Recurrent Neural Networks (RNN) [15], have been preferred over than the previous architectures in Deep Learning. But its application has been limited by the “vanishing” gradient problem and to overcome this issue, architectures based on retaining memory had been introduced. LSTM and GRU are the two architectures which retained the memoryof the previous layers thereby providing a solution to the problem of vanishing gradient. This paper uses the above mentioned two models and predicts the price level in two stock market historical datasets and analyzes its strengths and limitations.

DOI : https://doi.org/10.56452/7-4-237

1. Long Short-Term Memory(LSTM) Model

The architecture of LSTM is a modification of RNN to preserve the data for a longer duration. Basically a LSTM unit comprises of a memory cell that supports the layers in the network to retain the information which eventually does not permit loss of information. And the result is more accurate in sequential data or any other application that needs to be more precise. It contains of one memory cell, three gates and one hidden state. Discarding unwanted data is the responsibility of the forget gate and the data that has to be displayed is determined by the output gate. The input gate as depicted in Fig 1 isa circle filled with yellow color, determines what data should be allowed into the cell along with the memory cell which is shown as blue filled circles that contains the activation functions.The data that should enter the next sequence is taken care of by the hidden state and the orange filled circles are the output gates as depicted in Fig 1.

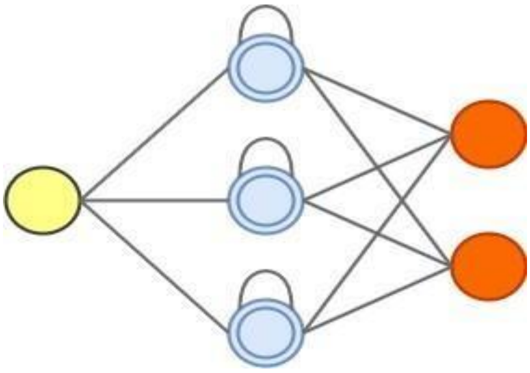

Fig. 1. Structure of LSTM network

The following transition equations represent the basic architecture of LSTM, where it is the input gate, ft is the forget gate, ot is the output gate, gt is the candidate hidden state, ht is the output hidden state, ct is the internal memory state and U & W are the weights used for training

the gates and t denotes the time.

$$it = \sigma(xtU^{i} + ht - 1W^{i}) \quad (1)$$

$$ft = \sigma(xtU^{f} + ht - 1W^{f}) \quad (2)$$

$$ot = \sigma(xtU^{o} + ht - 1W^{o}) \quad (3)$$

$$gt = \tanh(xtU^{g} + ht - 1W^{g}) \quad (4)$$

$$ht = \tanh(ct).o \quad (5)$$

$$ct = ct - 1.f + g.i \quad (6)$$

There had been many variants [16], [17] of LSTM since its inception and is used in applications like time series analytics, classification problems, natural language processing, communication,forecasting and prediction.

a. Gated Recurrent Unit Model

Gated Recurrent Unit (GRU) also a modification of e-current neural network is almost similar to LSTM in architecture except that it has two gates instead of three and has fewer parametersas depicted in Fig 2. There is no output gate but has an update gate and a reset gate along witha current memory state. The amount of information that should flow into memory is controlledby the update gate. The amount of information that should flow out of the memory is controlledand the effect of the previous data that has on the present data is suppressed by the reset gate.

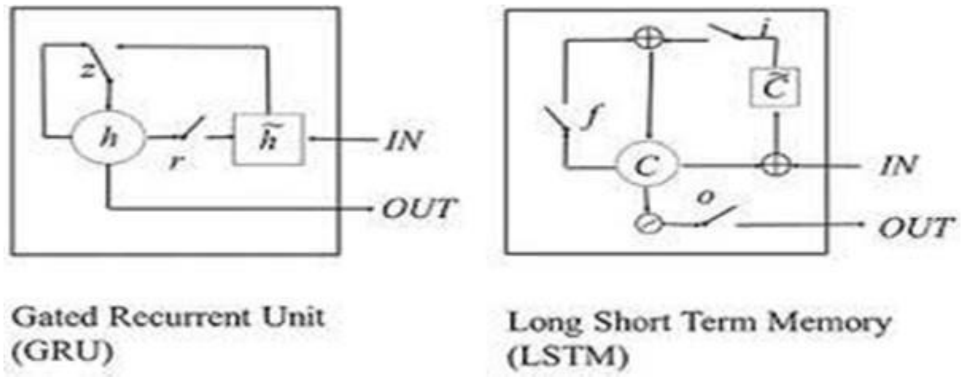

Fig. 2. Architecture of GRU and LSTM

The architecture of GRU`s implementation is denoted by the following equations where rt is the reset gate, zt is the update gate, k is the output state, ht is the hidden state U &W are the weights assigned to train the gates and t denotes time.

$$rt = \sigma(xtU^{r} + ht - 1W^{r}) \quad (7)$$

$$zt = \sigma(xtU^{z} + ht - 1W^{z}) \quad (8)$$

$$k = \tanh(xtU^{k} + (ht - 1.r)W^{k}) \quad (9)$$

$$ht = (1 - z).k + z.ht - 1 \quad (10)$$

This architectural modification enables GRU to execute faster, use less memory, applicable for short time sequences and works better for small and sparse datasets. So for specific applications like polyphonic music modeling it shows better results than LSTM [18] and variants of GRU

[19] creates a possibility to extend its applications related to bioinformatics, network intrusion,health monitoring [20] and various other fields.

b. Description of Dataset

Two datasets that were taken for analysis were obtained from Kaggle which is an online repository for datasets. These datasets are of historical type that contains financial equity stockdetails dated from 3rd Jan 2000 till 28th Feb 2020. Each of the dataset contains a total of

DOI : https://doi.org/10.56452/7-4-237

5044 values from which 80% of the total data was taken for training data and the rest 20% as testingdata. The variables Date and Close indicating the date and the price at which the stock was closed were taken for predicting the prices and the rest of the variables that define the datasets are Previous closing Price, Opening price, the highest and the lowest prices of the stock for the day, volume weighted average price, total volume of sales and turnover. The condensed set of samples from both the datasets are shown as Table 1 and Table 2 respectively and the format of these datasets is Excel comma separated value (csv).

Table 1: Sample Data from dataset1 – AU BANK

Date	**Open**	**High**	**Low**	**Last**	**Close**
1/3/2000	26.7	26.7	26.7	26.7	26.7
1/4/2000	27	28.7	26.5	27	26.85
1/5/2000	26	27.75	25.5	26.4	26.3
1/6/2000	25.8	27	25.8	25.9	25.95
1/7/2000	25	26	24.25	25	24.8

Table 2: Sample Data from dataset2 – YES BANK

Date	**Open**	**High**	**Low**	**Last**	**Close**
1/3/2000	166	170	166	170	170
1/4/2000	182	183.45	171	174	173.8
1/5/2000	170	173.9	165	168	166.95
1/6/2000	168	170	165.3	168.95	168.3
1/7/2000	162.15	171	162.15	170.75	168.35

2. Experimental Results and Performance Analysis

The effectiveness of any neural network architecture depends on how successfully it was modeled and trained. Here LSTM presented in [21] and GRU were modeled to predict the prices that indicate the support and

resistance levels through Fibonacci Retracement for two companies that were taken from Indianstock exchange and the level of accuracy attained is taken up for analysis to establish how theyhave predicted. Major trend lines were considered for assessment of the support and resistanceprice levels retraced with three of Fibonacci percentages i.e., 23.6, 38.2, 61.8 along with two other percentages 0 and 100 indicating the lowest and highest price levels.

The accuracy is measured by the error metrics and the hyper parameters assess the behavior of a model. Graphs were drawn for illustrative purposes to compare the models accuracy level. Computational time is taken up to determine the level of efficiency achieved by the models so as to judge how well the models have adapted the datasets. For the purpose of training 4035 samples were taken out of 5044 data and the rest 1009 samples for validation purpose. For training the model Adam is set as the optimizer. Google cloud engine was used as a training platform [Machine type: n1-standard-2 (2 vCPUs, 7.5 GB memory), CPU platform: Intel Corei5] and used Windows 7, Keras (Frontend) and Tensorflow (Backend) as the learning environment.

a. Error Metrics

For determining the accuracy level of the models the column date was taken as independent variables and the closing price as dependent variables from the dataset. The range of the values for dataset1 for the column closing price lies between 21 to 2050 and 163 to 2566 for dataset2. The error metrics [22] applied were Root Mean Square Error (RSME), Bais, Mean Absolute Error (MAE), Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). The residual obtained is the difference between the actual values and the predicted values. When RSME residuals are taken for analysis it is widely accepted that lower the difference when compared to the lowest range of the dependent variable, higher the level of accuracy.

Table 3: Error Metrics for dataset1 - AU BANK

Architectur etype	**Error Metrics**				
	RMSE	**Bias**	**MAE**	**MSE**	**MAPE**
LSTM	14.01	6.08	10.71	196.46	10.71
GRU	19.70	-6.77	15.25	388.10	15.25

Table 4: Error Metrics for dataset2 - YES BANK

Architectur e type	**Error Metrics**				
	RMSE	**Bias**	**MAE**	**MSE**	**MAPE**
LSTM	53.23	-25.59	28.61	2834.40	28.61
GRU	77.02	35.39	46.74	5932.75	46.74

From Table 3 representing dataset1 it can be seen that the residual values for LSTM is less than the values of GRU though the values for both the models is near the lowest closing price value which implies that the accuracy level is acceptable. To judge the level of accuracy for dataset2, Table 4 displays the residual values obtained from different error metrics. The residualvalues of LSTM are again less than the values of GRU and the difference between the lowest value of the dependent variable and the residual values is less for both the models. It can be seen that the level of accuracy is high for both LSTM and GRU for dataset1 and dataset2 but in comparison LSTM is more accurate than GRU.

The graphical representation for both the datasets is drawn to analyze the models LSTM and GRU. Through the graphs it can validate the error residuals attained as represented by Table 3 for dataset1 and Table 2 for dataset2. X-AU is plotted with independent variable, in this case the variable date is chosen and the variable close representing the closing price is taken as dependent variable which is plotted along the Y-AU.

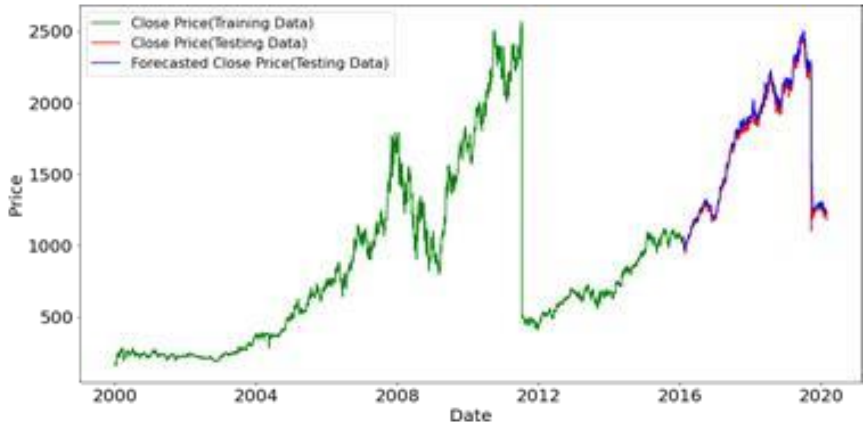

Fig. 3. Result of LSTM model for dataset1-AU Bank

Fig. 4. Result of LSTM model for dataset2 - YES Bank

DOI : https://doi.org/10.56452/7-4-237

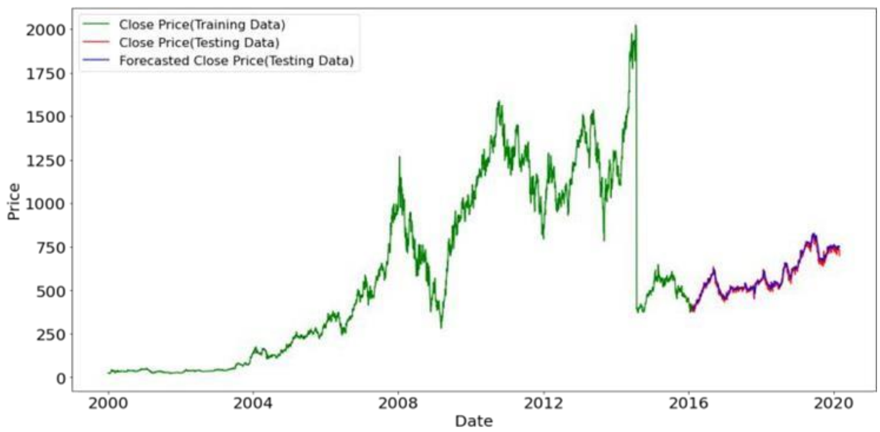

Fig. 5. Result of GRU model for dataset1-AU Bank

Fig. 6. Result of GRU model for dataset2-YES Bank

The green lines represent the 4035 instances which are the training data and the testing data with 1009 instances are represented as red lines. The blue lines indicate the values as forecastedby the models otherwise known as unseen data. The testing data and the unseen data for dataset1 and dataset2 implemented by LSTM are very close and the difference between them is negligible as seen from Fig 3 and Fig 4 which clearly indicates the level of accuracy achieved. The accuracy level of GRU for dataset1 and dataset2 is depicted in Fig 5 and Fig6 respectively indicating the difference between testing data and the unseen data is under the acceptable rangebut in comparison LSTM proves to be more efficient than GRU.

b. Hyper parameters

The process of learning during training the data are set through hyper parameters. The batch size, number of neurons used, number of hidden layers and epoch were set for the models to assess the learning process in this paper. Hyper parameters [23] are deemed to be important due to the fact that the process of training must be properly tuned considering its effect it has on the performance to get the best possible result and recently adaption to the method of training has been taken for further study so that the users can reduce their unnecessary effort and time [19].

Table. 5. Hyper parameters set for dataset1 - AU BANK

Architecture type	**Epoch**	**Batch size**	**Neurons**	**Hidden Layers**
LSTM	2	1	50	2
GRU	2	1	50	4

Table. 6. Hyperparameters set for dataset2 - YES BANK

Architectur e type	**Epoch**	**Batch size**	**Neurons**	**Hidden Layers**
LSTM	3	1	50	2
GRU	3	1	50	4

To attain the desired level of accuracy for dataset1 as can be referred from Table 5 the hyperparameters epoch, batch size and the number of neurons is at minimum and same for both LSTM and GRU differing only inthe number of hidden layers where GRU needed two more layers to predict. Table 6 lists the hyperparameters

set for dataset2 and the value of epoch was raised to one more when compared to the set epochfor dataset1, for both the models. Batch size, number of neurons and number of hidden layers used for dataset2 is the same set values as dataset1 for both the models and differs only in the number of the hidden layers. Considering the values set for the training process it can be inferred that for both the models, not much of an effort was taken for a database with 5044 instances.

c. Computation time

In Artificial neural networks the computational complexity is assessed by the parameter weight, time taken to train the dataset, the size and length of the input. In case of neural networks considering the number ofoperations required for a forward and backward passes, is one of the likely methods to assess the time taken. ForDeep Learning methods the total training time is taken up so that a fair judgment on how the models perform may be considered. On the above mentioned basis, in this paper the running time is represented in seconds to show how the models had trained and predicted from the test data.

DOI : https://doi.org/10.56452/7-4-237

Table 7: THE COMPUTATION TIME FOR THE DATASETS

Computation time	**LSTM (Time in Sec.)**	**GRU (Time in Sec.)**
Dataset1 AU BANK	76	168
Dataset2 YES BANK	80	146

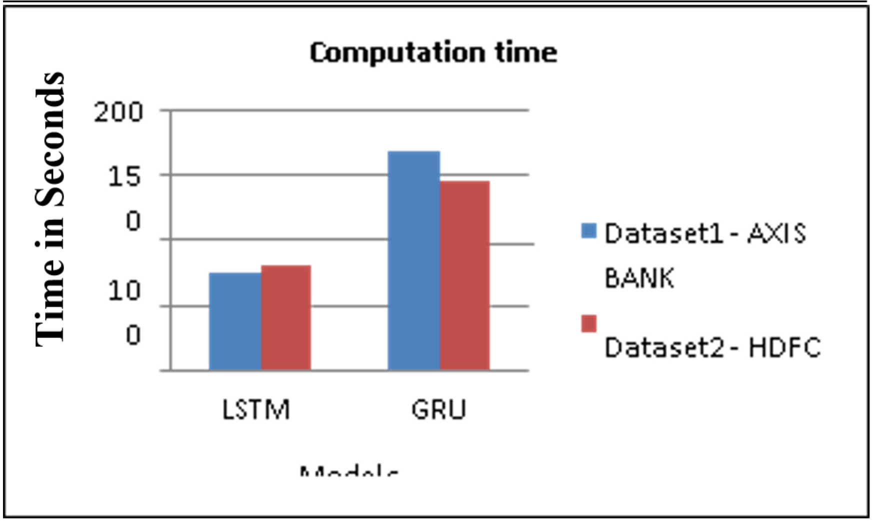

Fig 7: Comparison of the computation time of the models

Table 7 displays the time taken for training for both the models. The values from Table 7 are depicted as bar diagram in Fig 7. The vertical bars represent the two models where the blue barshows the running time for dataset1 and the red bar shows the running time for dataset2. The time in seconds is plotted at Y-AU and the time taken by GRU is higher for both the datasets when compared to the training time taken by LSTM.

3. Conclusion

Deep Learning has a major impact in modern technology and is considered as a cutting-edge solution provider in most of the applications where machine learning meets its limitations. Predicting stock prices is a risky business and the Neural Network paradigms of Deep Learning have aided in reducing the errors while forecasting. The popular architectures of Neural Networks learning, LSTM and GRU, were designed to predict the support and resistance levels for two stocks from Indian Stock Exchange containing 5022 data each, to determine the exit or the entry price point. The structure of the model was set through four

DOI : https://doi.org/10.56452/7-4-237

hyper parameters and is taken up to analyze the training process. Long short-term memory achieved more accuracy than GRU and the computation time taken by LSTM is much lower than the time taken by GRU. The differences in performance can be attributed to one of the sets hyper parameter and the size of the dataset. Though both of the architecture has been modeled well, it can be concluded that LSTM is a better fit than GRU both in terms of accuracy and computation time.This research work predicted support and resistance levels and work can be extended further by implementing various multiple technical factors for predicting the stock prices.

References

[1] T. J. Sejnowski, "The unreasonable effectiveness of deep learning in artificial intelligence", Proceedingsof the National Academy of Sciences, 2020.

[2] R. Vidal, J. Bruna, R. Giryes and S. Soatto, "Mathematics of deep learning", arXiv preprintarXiv:1712.04741, 2017.

[3] M. A. Nielsen, "Neural networks and deep learning", San Francisco, CA: Determination press, Vol. 25,2015.

[4] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning", nature 521, No. 7553, pp. 436-444, 2015.

[5] T. Velmurugan and T. Indhumathy, "Comparative study on Prediction of Support and Resistance Levels with k-Nearest Neighbor and Long Short-Term Memory Methods", Journal of Information and Computational Science, ISSN: 1548-7741, Vol. 10, Issue 09, pp. 243-252, 2020.

[6] Epistemological View: Data Ethics, Privacy Trust on Digital Platform, Harsh, R., Acharya, G., Chaudhary, S., 2018 IEEE International Conference on System, Computation, Automation and Networking, ICSCA 2018, 2018, 8541166

[7] Enhance the Data Security in Cloud Computing by Text Steganography, Sanghi, A., Chaudhary, S., Dave, M., Lecture Notes in Networks and Systemsthis link is disabled, 2018, 18, pp. 241–248

[8] H. M. Fayek, M. Lech and L. Cavedon, "Evaluating deep learning architectures for Speech Emotion Recognition", Neural Networks, Vol. 92, pp. 60-68, 2017.

[9] T. Indhumathy and T. Velmurugan, "Performance of k-Nearest Neighbor and Gated Recurrent Unit in Stock Market", Journal of the Maharaja Sayajirao University of Baroda, Vol. 54, No.2 (V), pp. 122-127, 2020.

[10] S. A. Hasan and Oladimeji Farri, "Clinical natural language processing with deep learning", In Data Science for Healthcare, Springer, Cham, pp. 147-171, 2019.

[11] S. Caldera, A. Rassau, and D. Chai, "Review of deep learning methods in robotic grasp detection", Multimodal Technologies and Interaction, Vol. 2, No. 3, p. 57, 2018.

[12] J. Schmidhuber, "Deep learning in neural networks: An overview", Neural networks, Vol. 61, pp 85-117, 2015.