

Application of Machine Learning for Sequential Data

Peter UHEREK*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
peteruherek@gmail.com*

Nowadays Artificial Neural Networks (ANN) gain an increasing interest after periods of disappointments and several cycles of hype. They are being used in ordinary real life applications, dealing with various types of problems such as natural language processing, handwriting recognition and loans applications. Every year new forms of ANNs applications are being brought into practice. There is still a potential to discover new applications based on ANNs, especially when combined with interesting data.

In our study we are focusing on using ANNs with data that originates from the paywall of foreign news portal. These data are sequentially storing user history of the web browsing. They also are storing information about user's payments for articles. According to these data, we want to research the possibilities of the various predictions in different approaches. Our major object of interest is in the predictions of the user's payments for articles.

Our approach for this problem is based on supervised machine learning, especially Recurrent Neural Networks (RNN). We designed model of RNNs with using the Long Short Term Memory (LSTM) architecture. This type of architecture consists of several cell memories which can help when there are very long time lags of unknown size between important events [1]. Design of our network can be seen on the figure 1.

Our network has 2 hidden layers. The LSTM block is located on the second hidden layer, enclosed by two hidden linear layers. The final layer is Softmax with 2 outputs, which represent a probability distribution of the user's payments for articles. An error in model is defined by cross entropy. There is also applied the dropout technique for all connections.

We have made several experiments to evaluate our solution. We researched the effect of the entries into our model and the effect of the LSTM block on the outcome. Our train dataset was composed from history of 7250 users. There were 721 payers and

* Supervisor: Michal Barla, Institute of Informatics, Information Systems and Software Engineering

6529 non-payers. For this reason, we had to use oversampling and duplicated paying users. Test dataset consisted from 100 payers and 100 non-payers.

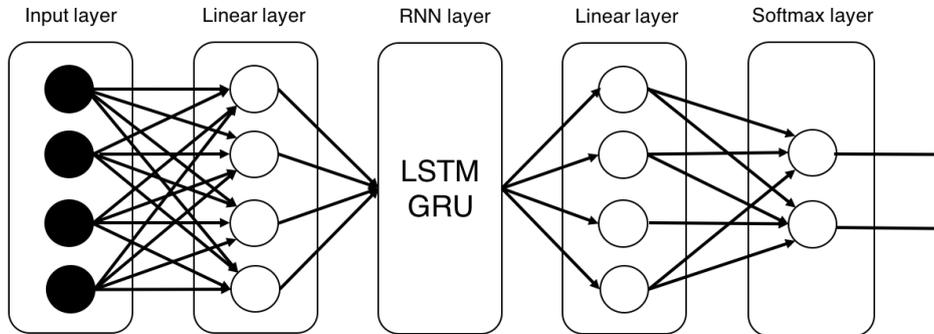


Figure 1. The architecture of our ANN for prediction user's payments for articles.

We identified the most appropriate inputs by several experiments. We also found out that stacked several LSTM block in row can help minimalizing error of the model. The different between using of 1 LSTM block and 3 LSTM block is 2 % in the error in test dataset. We also experimented with simpler type of the LSTM architecture, the Gated Recurrent Unit (GRU) architecture. We replaced the LSTM block with the GRU block and simple RNN block in several experiments. Results from experiments can be seen on the table 1.

Table 1. The comparison of multiple type of architectures.

MODEL	TRAIN	TEST	ACCURACCY	PRECISION	RECALL	FNR	EPOCH
RNN	0,37	0,39	0,79	0,83	0,8	0,2	4
1 GRU	0,35	0,36	0,82	0,86	0,76	0,24	5
1 LSTM	0,32	0,36	0,81	0,86	0,74	0,26	6
2 GRU	0,35	0,35	0,82	0,87	0,75	0,25	5
2 LSTM	0,32	0,35	0,81	0,87	0,73	0,27	8
3 GRU	0,36	0,34	0,83	0,94	0,72	0,28	5
3 LSTM	0,32	0,34	0,82	0,88	0,75	0,25	14

In our study we also research two other problems. First one is the prediction of the popularity of articles. Second one is the predictions of articles that user visited. In the same way we use similar architecture of network with LSTM, GRU or RNN blocks like in the figure 1.

References

- [1] Chung, J., Gulcehre, C., Cho, K. Empirical evaluation of gated recurrent neural networks on sequence modelling, In: Deep Learning and Representation Learning Workshop (2014)
- [2] Sutskever, I., Vinyals, O, Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems (2014), pp. 3104-3112.