Review of
"Visualizing and Understanding Recurrent Networks"

Anuar Maratkhan

The paper itself is well written for some reasons: it gives explicit description of concepts used, and provides the good work flow by showing what work has been done specifically. The paper reflects the name of topic, and the topic gives relatively good understanding of what the paper is about. The authors give thorough definitions of recurrent network models, and the theory lying above it by providing readers with mathematical meaning of those network models. The style of writing is well also because of abbreviations to repeatedly long terms such as RNN, LSTM, GRU. However, the citation style is quite misunderstandable due to adding citations in sentence itself and not separating by writing it in brackets. The paper is also relevant because it cites a lot (37) of reliable sources and publications from conferences.

The topic itself is relatively hard to understand at first sight. Although, paper shows that the authors have good understanding and background knowledge in the area. As mentioned above, the concepts used in the paper are explicitly described. For instance, the authors first start by explaining reccurent networks in general, and further continue by giving brief explanation of Recurrent Neural Networks, Long Short-Term Memory, and Gated Recurrent Unit. These explanations that include mathematical meanings clearly shows the advantages and disadvantages of one network model over others.

The optimization done seems good. However, more knowledge needed to understand such concepts like epoch, learning rate, or decay. The initialization of parameters in some range is also needs attention.

The datasets used in the experimentation process were outstanding because in total they included more than 6 million characters one from the big and complex russian literature book, and the other from the source code, which actually showed out of the ordinary approach to the problem. Data was also split quite well into train/val/test fractions.

The comparison of three network models was nicely described. The use of different number of cells and parameters looks reliable. Moreover, the visualization of network models' results as clusters clearly showed that RNN has really distinct predictions, and in consequence gave relatively less efficient performance.

Gate activiation statistics illustrated the benefit of applying LSTM clearly on more intuitive level by giving explanations of observations. However, the sentence "We struggle to explain this finding but note that it is present across all of our models." leads to a little doubts on that because even no citation was provided there.

During error analysis, the researchers observed error patterns and provided readers with several oracles that in consequence may influence the LSTM to improve further. However, one notable thing was that the authors defined the predicted character to be an error if the probablity assigned to it by a model on the previous time step is below 0.5. This generates some concerns and questions like finding the reason of using exactly such probability rate. There was no explanation on that provided.

Lastly, the last similar notable thing is that in the explanation of Vanilla Recurrent Netural Network, the authors omitted the bias vector. Even though, they state that the reason for that was brevity. However, it may seem a bit unclear to readers with no specific knowledge on the topic. It is better to show on the example why non-usage of bias vector does not make much difference in the results.

REFERENCES

[1] Andrej Karpathy, Justin Johnson, Li Fei-Fei *Visualizing and Understanding Recurrent Networks*. Stanford University, 2015