Peer Review of Group 54

DD2424 Deep Learning in Data Science
KTH Royal Institute of Technology

By Group 53:
David Alm, 020207-4795, davialm@kth.se
Theofanis Georgakopoulos, 000814-0576 thegeo@kth.se
Antony Zhang, 011231-8159, antonyz@kth.se

June 26, 2025

1 Questions for Group 54 during presentation

- We noticed that the validation loss remains pretty much unchanged in your loss plots. Why do you think this is?
- What was your training time like? Was that an issue for you, i.e. did some of your models take too long to train? If so, what did you do to get around that?
- Related to previous question, what learning scheme did you use (e.g. Adam, cyclical, standard mini-batch)? Maybe we missed it in the report, but can't remember you mentioning anything on this.
- If you had more time and resources, what would be your next bet to improve your models?
- Clarifying question. For the n-gram metrics, what was the actual metric? Was it overlap or diversity?
- Do you think that it comparing the one with the two layer LSTM architectures in your report was a fair comparison, given the increased complexity ?

2 Peer Review Q&As

1. **Question** *In one sentence summarize the main point of this project.*

   Answer Explore and compare the effectiveness of RNNs and LSTMs in NLP and Text Generation.

2. **Question** *Which part of the report was most clearly written?*

   Answer The tables that showed how performance varied according to different metrics with varying parameters were very clear. The clearly showed the isolated impact of the respective parameter. In particular, we are referring to Tables 2, 5, and 6. If one would like to expand on them, maybe heat-maps would interesting with more extensive grid-search.

3. **Question** *Which part/section of the report was least clear? Was it possible to understand the material presented in this section? Please expand with a sentence or two.*

   Answer We do feel that there was a particular section or part that was unclear. We have noticed a couple things that could have been clear up at some point though. For instance:

   - A clearer description of the n-gram measurements. We could not find whether these were diversity or overlap, word- or character-level.

   - Something about the learning scheme used, e.g. Adam optimizer, cyclical learning rate, etc.

   - Something on how long training was conducted (epochs, batch size) and what time it took (time and hardware).

4. **Question** *Did the report give you a better understanding of the problem/deep learning concept investigated? If yes please highlight the subsection that was especially useful and briefly explain why (i.e. Related work, Method section Experiments etc..) If no please explain why.*

   Answer Yes and no. As our group did the same project, we had a lot of similarities and not very big news to us. For instance, our related work sections shared many of the same references and we also implemented a baseline RNN to be improved on by LSTMs.

   With that being said, we did learn something important. We noticed that our n-gram numbers were very different. This lead us to the discovery that we had misunderstood the n-gram definition. We used a character-level n-gram which might be misleading as the word-level arguably tells us more. As an example, if we have the text "I look nice". Our 2-grams look for "I ", " l", "lo", ... whereas we should arguably be looking for ("I", "look"), ("look", "nice"). We are not convinced of either way but this makes our numbers quite high of course and possibly misleading.

5. **Question** *What was the most impressive experimental result presented in the project and why?*

   Answer We were most impressed by your numbers on nucleus sampling (presented in table 5 and 6). They were very high, good work!

6. **Question** *What was the most interesting or surprising experimental result presented in the project and why?*

   Answer The most surprising result in our opinion is the spelling accuracy for most of the models. Rather interesting that such small models are able to spell so well! Our intuition would be that they would need to be a lot bigger, more data, etc. to achieve the type of numbers (all over 55%, some over 90%) that we see.

7. **Question** *Which experiment(s) would you like the project group to complete if they were to continue with this project.*

   Answer The obvious choice here is to continue with some of the proposed extensions for B/A-level. As we did in our project, continuing with word-embeddings (word2vec or GloVe), subword tokenization (BPE), transformers, and data augmentation would be interesting.

   Morover, continuing to expand the LSTM to arbitrary number of layers could be interesting and really compare the trade-off between depth and width would be interesting. Similarly, it would also be interesting to experiment with improving performance even further with the current LSTM implementation - maybe adding some regularization techniques.

8. **Question** *Mention two things you liked about the project report and/or the video presentation.*

Answer Everything was very clear and orderly. Compared with our own report, yours was better structured and more clear. It was easy to follow and had a nice flow.

Your presentation was also good. The slides completed the presentation well and recording it all in one take made for a very natural sounding video.

9. **Question** *(Optional) Is there any reference or paper that you recommend the project group check out after having read the paper*

   Answer Nothing in particular, maybe some of the references regarding the extensions mentioned above in question 7. For instance, [1, 2, 3, 4].

10. **Question** *Other comments*

    Answer We noticed that you went beyond the 6-page limit for the NeurIPS template. Be wary of this.

References

[1] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013. [Online]. Available: https://arxiv.org/abs/1301.3781

[2] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *arXiv preprint arXiv:1310.4546*, 2013. [Online]. Available: https://arxiv.org/abs/1310.4546

[3] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," 2016. [Online]. Available: https://arxiv.org/abs/1508.07909

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023. [Online]. Available: https://arxiv.org/abs/1706.03762