NLP using RNNs, LSTMs, and Transformers
Self Assessment

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 Knowledge and Skills Acquired

1.1 David

- I got a much greater understanding of, in particular, LSTMs and their inner workings as we implemented them on our own for an arbitrary number of layers.
- I have familiarized myself with more advanced concepts within NLP such as word embeddings, subword tokenization, and transformer networks. In my opinion, this is very important with the rise to prominence of LLMs that rely on variations of these techniques.
- I improved my knowledge of `Tensorflow` and `Keras`. I especially gained experience in their data handling capabilities and how to implement ”off-the-shelf-models” with them.
- I improved my report writing skills and familiarized myself with journal templates, in particular that of NeurIPS.

1.2 Theofanis

- I got the chance to fulfill my learning expectations regarding RNNs, LSTMs and backpropaggation through time.
- I got the chance to dive deeper into different tokenization schemes on word-level (GloVe) and subword level (BPE). I had the chance to experiment with different hyperparameters such as the vocabulary size on BPE and observe its effect on the performance of the trained model.
- Lastly, I got the chance to learn more about the Architecture of Transformers and how they work. If given enough computational resources and time, indeed "Attention is all you need".
- Regarding technicalities, I acquired a lot in using different Deep Learning libraries. I also managed to better

reconsider how to structure a paper, think of Future work and meaningful questions, as well as, meaningful results.

1.3 Antony

- I fulfilled my aim of diving deeper into the theory behind RNN and LSTMs, and learning more about the intricacies of implementing them .

- I also improved my skills in using TensorFlow and Keras instead of PyTorch, especially in building and training models.

- I gained more insight into how different word embeddings and subword tokenizations affect NLP models such as the transformers we implemented, as well as the importance of hyperparameter tuning when implementing these models.

- Overall, this project gave me both practical experience with deep learning tools and a clearer picture of the theory behind NLP models.

2 Grading

We believe we are deserving of the grade A because ...

- We successfully implemented our project proposal and the level of difficulty of our project exceeded that of the B/A level. That is, we fulfilled the E-grade requirements by implementing the baseline RNN and LSTM. We then went beyond this by implementing the LSTM for an arbitrary number of layers, word-embeddings (GloVe), subword tokenization (BPE), and transformers networks. The project proposal only requires the implementation of two of these three (word-embeddings, subword tokenization, transformers networks) for reaching the B/A-level.

- We conducted a thorough set of experiments throughout the project. We explored the issues of regularization (L2, dropout), sampling methods (temperature and nucleus sampling), width and depth exploration (varying architecture params) for the LSTM, data augmentation, among other things.

- The project in its entirety is the culmination and combination of all the skills and knowledge we have acquired throughout the course. From the fundamental theory and first principles of deep learning used in implementing the LSTMs from scratch and the heuristics of optimizing their performance, to the practicalities of leveraging deep learning libraries and exploring more advanced, current state-of-the-art, concepts within the subject area such as transformers and subword tokenization. We feel that the project has been all-encompassing in this sense.

- Our report is clearly written with a thorough related work section where we summarise the seminal papers relevant to NLP and the project, showing the big achievements and important steps taken over the years. This shows that we have thoroughly investigated the subject area and reviewed the relevant literature. The report is also well structured with the work we have done and the results we have achieved presented in a clear manner with accompanying analysis and conclusions. Moreover, the code used for experiments has been cleaned up to be presentable as part of a GitHub repo.

- Our video presentation was clear and to the point. All major findings and work we did was accounted for in a concise manner.