

ANLP Seminar Week 3: Neural Language Modelling

Julie Weeds

February 2, 2026

Kim et al. (2016) provide a comparison of character-aware neural language models with word-level models. Note that this is a difficult paper if you are not familiar with neural network architectures - but you can still read it a high level and not worry (too much) about the details of the different architectures and the hyper-parameters! Once you have read the paper, consider the following questions.

1. What problem with using n-gram models is addressed by the use of neural language models? Why is it not completely addressed? How might character-aware models help?
2. Why are LSTMs generally preferred over vanilla RNNs in language modelling?
3. In a character-level CNN, what is the purpose of a filter or kernel? How many filters do they state are typically used in NLP applications? How many do the authors use in each of their **small** and **large** architectures?
4. How is the character-level CNN incorporated into the overall architecture of the RNN-LM?
5. How are OOV words handled in these experiments? What potential improvement could the authors have made and why didn't they do it?
6. Which model(s) performs best in the optimization experiments on the Penn Treebank?
7. Why do the authors expect the performance gains to be more in other languages such as Arabic than in English? Are their expectations met in the experimental results?
8. What observations can you make of the nearest neighbours of 'richard' using each of the word representations?
9. What are the main conclusions of the paper? Are you convinced?

References

Yoon Kim, Yacine Jernite, David Sontag, and Alexander M. Rush. 2016. Character-aware neural language models. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*.