

INFO2049 RNN Assignment

Language Models

Project LM1

Implement an RNN-based language modeling system. Your system your model the languages in newspaper texts (either English or French), which can be crawled. Alternatively, you can model the language from the English (or French) Wikipedia dumps or the European Commission EurLex Corpus.

Experiments:

- Investigate the performance using LSTM vs. GRU
- Investigate the performance using 1-hot encoding and Word2Vec as input.
- Bonus: (you might end up having an extremely large number of parameters), include attention in your model. As inspiration, see Salton et al. (<https://www.aclweb.org/anthology/I17-1045/>). Investigate
- Interpret the relevant scores, in particular the perplexity.