Recurrent LSTM Neural Networks for Language Modelling and Speech Recognition

Piotr Kłosowski
Faculty of Automatic Control, Electronics and Computer Science
Silesian University of Technology
Akademicka 16, 44-100 Gliwice, Poland
Email: pklosowski@polsl.pl

***Abstract*—This paper examines interesting natural language modelling tasks, such as word-based and subword-based language modelling, where deep learning methods are making some progress. Language modelling helps to predict the sequence of recognised words or subwords and thus can be used to improve the speech recognition process. However, the field of language modelling is currently witnessing a shift from statistical methods to recurrent neural networks and deep learning techniques. This article focusses on an example of using recurrent LSTM neural networks for language modelling and speech recognition. The new research results presented in this paper, following on from previous papers, focus on how to develop word-based and subword-based LSTM language models and how to use them together. The simultaneous use of both LSTM language modelling methods allows for the development of hybrid language models that have even better properties and can further improve the speech recognition process. The results presented in this paper apply to Polish language modelling, but the results obtained and the conclusions formulated on their basis can also be applied to language modelling applications for other languages.**

***Keywords*—artificial intelligence, neural networks, language modelling, speech recognition.**

Statistical language modelling enables the development of probabilistic models that can predict the next word in a sequence given the words that precede it. Language modelling is the task of assigning a probability to sentences in a language. In addition to assigning a probability to each sequence of words, language models also assign a probability for the likelihood that a given word (or a sequence of words) will follow a sequence of words.

A language model learns the probability of word occurrence based on examples of text. Simpler models may look at the context of a short sequence of words, whereas larger models may work at the level of sentences or paragraphs. Most commonly, language models operate at the level of words. The notion of a language model is inherently probabilistic. A language model is a function that puts a probability measure on strings drawn from some vocabulary. A language model can be developed and used stand-alone, such as to generate new sequences of text that appear to have come from the corpus. Language modelling is a root problem for a wide range of natural language processing tasks. More practically, language models are used on the front- or back-end of a more sophisticated model for a task that requires language understanding. Language modelling is a crucial component in real-world applications such as machine translation and automatic speech recognition. For these reasons, language modelling plays a central role in natural language processing, AI, machine learning and speech recognition research.

A good example is speech recognition, where audio data are used as input to the model and the output requires a language model that interprets the input signal and recognises each new word within the context of the words already recognised. Speech recognition is principally concerned with the problem of transcribing the speech signal as a sequence of words. From this point of view, speech is very often represented by a language model that provides estimates of $P(W)$ for all word strings W independently of the observed signal. The main goal of speech recognition is to find the most likely word sequence given the observed acoustic signal.

Developing better language models often results better on natural language processing and speech recognition task. It is therefore a motivation to develop ever better and more accurate language models. The main objective of the research presented in this paper is to develop language models using recurrent neural networks with a special focus on Long-Short-Term Memory (LSTM) networks. The results presented in this article relate to the development of an LSTM language model using word-based and subword-based methods. In addition, it was proposed to use both LSTM language modelling methods simultaneously to develop a hybrid language model, which has even better properties and can significantly improve speech recognition. The accuracy, loss, and perplexity parameters were used as a measure of language model performance. The hybrid model performs quite well in comparison to other models developed. The performance of the hybrid model is clearly better than the word-based or subword-based language model used alone. The calculated parameters were only used to compare the developed language models. The language models were trained on a limited size test corpus to reduce the training time. Computing the real parameters of the language models would require a long learning process using corpora of very large size and sufficient quality to reflect the specifics of the language. The use of relevant language corpora as training data for developed language models provides the opportunity to simultaneously use word-based and subword-based types of models, as a hybrid language model, to predict word more effectively in the speech recognition process.