Predicting cloze task results with language models

Lopukhina A.*, Lopukhin K.**, Laurinavichyute A.*

* Neurolinguistics Laboratory, National Research University Higher School of Economics, Moscow, Russia; ** Scrapinghub, Moscow, Russia

nastya.lopukhina@gmail.com

When reading or listening, people can anticipate upcoming words based on the previous context — thus, each word has a degree of predictability that is usually measured by the cloze task (Taylor, 1953). Being convenient, this task is criticized for its biases and the lack of reliable information about very unprobable continuations, which may affect the processing difficulty, according to the surprisal theory (Hale, 2001; Levy, 2008). To overcome the limitations of the cloze task several studies turn to corpus data and language models to quantify predictability (Smith and Levy, 2013). Although corpus probabilities are sometimes used as a substitute for cloze, we still don’t know how well they match both cloze data and readers’ online predictions. A predictability model that better matches cloze probabilities may provide better results in predicting readers’ behaviour and may be used as an alternative to the cloze task. This study aims at investigating the performance of two language predictability models (an n-gram model and an LSTM recurrent neural network) and the correlation between their predictions and cloze probabilities. We found that models’ perplexity influences the correlation between cloze and models’ probabilities and that low-predictability words show better correlation, which means that answers that no human managed to guess correctly also have low corpus probability. Overall, the neural network model has lower perplexity than the 5-gram model and produces probabilities that are more similar to cloze task results (Pearson correlation is 0.59; Spearman correlation is 0.95 — if we include zero probabilities).

Keywords: predictability, cloze probability, neural networks.