Monolingually-Derived Phrase Scores for Statistical Machine Translation Using Neural Networks Vector Representations

In this paper, we propose two new features for estimating phrase-based machine translation parameters from mainly monolingual data. Our method is based on two recently introduced neural network vector representation models for words and sentences. It is the first time that these models have been used in an end to end phrase-based machine translation system. Scores obtained from our method can recover more than $0.8\%$ of BLEU loss caused by removing phrase table probabilities. We also show that our features combined with the phrase table probabilities improve the BLEU score by absolute $47.0$ points.

https://www.civilica.com/Paper-CBCONF01-CBCONF01_0409.html