Abstract

Micha de Rijk

January 6, 2020

Word association is an important part of human language. Many techniques for capturing semantic relations between words exist, but their ability to model word associations is rarely tested. We introduce the game of Codenames with one human player as a word association task to evaluate how well a language model captures this information. We establish the baseline f-score of 0.362 and explore the performance of several collocations and word embedding models on this task. Our best model uses fastText word embeddings and achieves an f-score of 0.789 for Czech and 0.751 for English.