

Grounding Computational Models of Language in the Social Environment: A New Direction for Models of Language Processing

Brendan T. Johns (McGill University)

Linguistic experience varies across individuals and is impacted by both demography and personal preferences, leading to differences in word meanings across people. An active area of study in the cognitive sciences that examines the impact of varied knowledge across individuals is the wisdom of the crowd effect, where it is found that the aggregate judgement of a group of individuals is often better than the judgement of the best individual. The goal of this talk is to demonstrate a wisdom of the crowd effect in lexical semantic memory, such that the aggregated word similarity values from many individual language users exceeds the fit of the best fitting individual. This was accomplished by training 500 different distributional models from 500 high-level commenters on the internet forum Reddit. By deriving aggregated word similarity values across individuals, a strong wisdom of the crowd effect was found where the aggregated similarity values exceeded the performance of the best fitting individual for each dataset tested. Additionally, it was found that even aggregating only a small number of users provided a large increase in fit relative to the individual corpora, but with the best fitting measure including word similarity values from all possible users.