Title: A Method for Relation Extraction Using Named Entity Vectors

Abstract:

Relation extraction is a problem in natural language processing that is integral to discovering structured knowledge in unstructured data such as news corpora. We tackle the task of discovering lists of semantically related named entity pairs from the text of news articles.

We first present a naive solution to serve as a benchmark. We demonstrate that applying simple linear transformations in the original vector space to all combinatorial pairs of entities from an article produces poor results. We identify the following two main challenges: a) properly filtering out pairs of unrelated named entities and b) finding vector representations for a single entity pair that we can use for comparison with other entity pairs.

To reduce the number of noisy entity pairs, we use dependency parsing to extract named entity pairs that are syntactically linked as a heuristic for semantically linked entity pairs. To discover a better vector representation for a named entity pair, we employ an autoencoder neural network model. We present the results of these combined methods against the benchmark.