Relation Extraction with Neural Nets and Graph Propagation

Golnar Sheikhshab

Relation Extraction

To develop a broad understanding of the causes and patterns of occurrence of wheezing associated respiratory infections, we analyzed data from an 11-year study of acute lower respiratory illness in a pediatric practice. Although half of the WARI occurred in children less than 2 years of age, wheezing continued to be observed in 19% of children greater than 9 years of age who had lower respiratory illness. Males experienced LRI 1.25 times more often than did females; the relative risk of males for WARI was 1.35. A nonbacterial pathogen was recovered from 21% of patients with WARI; respiratory syncytial virus, parainfluenza virus types 1 and 3, adenoviruses, and Mycoplasma pneumoniae accounted for 81% of the isolates. Patient age influenced the pattern of recovery of these agents. The most common cause of WARI in children under 5 years of age was RSV whereas Mycoplasma pneumoniae was the most frequent isolate from school age children with wheezing illness. The data expand our understanding of the causes of WARI and are useful to diagnosticians and to researchers interested in the control of lower respiratory disease.

Approach

- Consider one candidate pair at a time
- Input:
    - Embedding of the two entities
    - Embedding of the sentence concatenated to embeddings of two entities
    - Same as above only the two entities are removed from the sentence and replaced by Arg1 and Arg2
- Output
    - 1 if there is a relationship, 0 if there is not.
- Classifier
    - A logistic regression of W.E (dot product of weight matrix and input embedding)
    - Propagate the output on a graph where embeddings are vertices and edge weights are cosine similarity

Embeddings

- Entities
    - Average of word entities
    - Word entities come from word2vec
- Sentence embedding
    - Skip-thought vectors
    - GRU is used in an encoder-decoder where the task is to predict the surrounding sentences