Multifacet Relation Extraction
Rohan Paul*, Haw-Shiuan Chang*, Andrew McCallum

Introduction

- Previous work in Distant Supervised Relation Extraction:
  - Training signal - Co-occurrence of entity pairs with sentence patterns
  - Universal Schema - assumes one embedding per sentence pattern
  - Compositional Universal Schema uses LSTM to encode sentence pattern

- Limitations:
  - A sentence pattern could co-occur with different facets (i.e., entity pairs with different relation)
  - All facets compressed into a single embedding for each sentence pattern

Main Idea – Multifacet Embeddings

- Represent each pattern using multiple embeddings
- Each embedding represents a facet
- One facet can be represented by multiple embeddings
- A facet of a pattern is similar to a facet of another pattern if the patterns share same entity pairs

Seq2Seq Model

- Transformer encoder to encode sentence into a single embedding
- Transformer decoder to decode encoded embedding into multiple facet embeddings

Objective function

- Choose closest facet for each entity pair
- Minimize the distance of the co-occurred entity pair and sentence pattern
- Maximize all the other distances
- Similar to kmeans clustering, encouraging pattern embeddings to become the centers of entity pairs that co-occur with the pattern

Scoring functions

- \( \text{Symm}(\ell_{1}, \ell_{2}) \) computes the average of cosine similarities from every center in \( \{\ell_{1}\} \) to its closest center in \( \{\ell_{2}\} \)
- Checking if \( \{\ell_{1}\} \) is more specific than \( \{\ell_{2}\} \): \( \text{Symm}(\ell_{1}, \ell_{2}) > \text{Symm}(\ell_{2}, \ell_{1}) \)?
- Symmetry between \( \{\ell_{1}\} \) and \( \{\ell_{2}\} \): \( \text{Symm}(\ell_{1}, \ell_{2}) + \text{Symm}(\ell_{2}, \ell_{1}) \)/2

Our Method

- Embeddings of SARG1’s partner SARG2 and Entity Pair Embeddings

Experiments

- Distant supervised relation extraction:
  - Setting is the same as CUSchema (Verga et al., 2016)
  - Training data: text with linked entities + Freebase + a few rules (No label data)
  - Validation data: TAC slot filling 2012
  - Testing data: TAC slot filling 2013 2014

- Unsupervised entailment detection:
  - Dataset: Sentence pattern pairs where the words in premise and hypothesis are hypernymy; e.g., president -> leader
  - Collect labels of 1,500 sentence pattern pairs
  - Retrieve entailment out of other relations (including paraphrase)
  - Predict the entailment direction of patterns

Conclusion

- We propose a novel method to improve the compositional universal schema, CUSchema, by modeling different facets of a sentence pattern
- Representing every sentence pattern using the cluster centers, we can achieve better symmetric and asymmetric similarity measurement between sentence patterns for relation extraction and entailment detection

References