Learning to Embed Words in Context for Syntactic Tasks

Lifu Tu; Kevin Gimpel; Karen Livescu
Toyota Technological Institute at Chicago

How To Capture Linguistic Characteristics Of Tokens?

1. Same syntactic category, different senses:
   He robbed 9 banks. vs. It washed up on the banks.
2. Different POS tag and sense:
   I was unable to police the situation. vs. I was unable to contact the police. _

How to solve this? Each word type can have a different vector representation in different contexts!

Models and Loss Function

- Reconstruction of input
  - Hidden layers
  - Decode
  - Enode
  - Word type embedding
  - Copy
  - Follow

Figure 2. seq2seq token embedding model

Loss Function

- Weighted Reconstruction Error:
- Input word sequence s, encoder f, decoder g

Figure 3. Baseline DNN Tagger

Token Embedding Models

- Output probability for each tag
- Binary feature vector
- W or w' baseline DNN TE Seq2seq TE
- Attachment F1 (%) on validation set using different models and window sizes.
- For TE columns, the input does not include any type embeddings at all, only token embeddings.
- Dependency parsing unlabelled attachment F1 (%) on test sets for baseline parser and results when augmented with token embedding features.

Contact

Lifu Tu
Toyota Technological Institute at Chicago
Email: lifu@ttic.edu

Qualitative Analysis

Table 1. Nearest neighbours for token embeddings, where we consider neighbors that may have 1, w'=1.

Datasets

- Part-of-Speech Tagging: from Gimpel et al. (2011) and Owoputi et al. (2013)
  - OCT27TRAIN, OCT27DEV, OCT27TEST
  - DAILY547
- Dependency Parsing: from Kong et al. (2014)
  - 717 training tweets
  - 201 tweets TEST-NEW

Figure 1.

Figure 4. Token Embedding Tagger

References

2. Olutobi Owoputi, Brendan O’Connor, Chris Dyar, Kevin Gimpel, Nathan Schneider, and Noah A. Smith. 2013. Improved part-of-speech tagging for online conversational text with word clusters. In Proc. of NAACL.