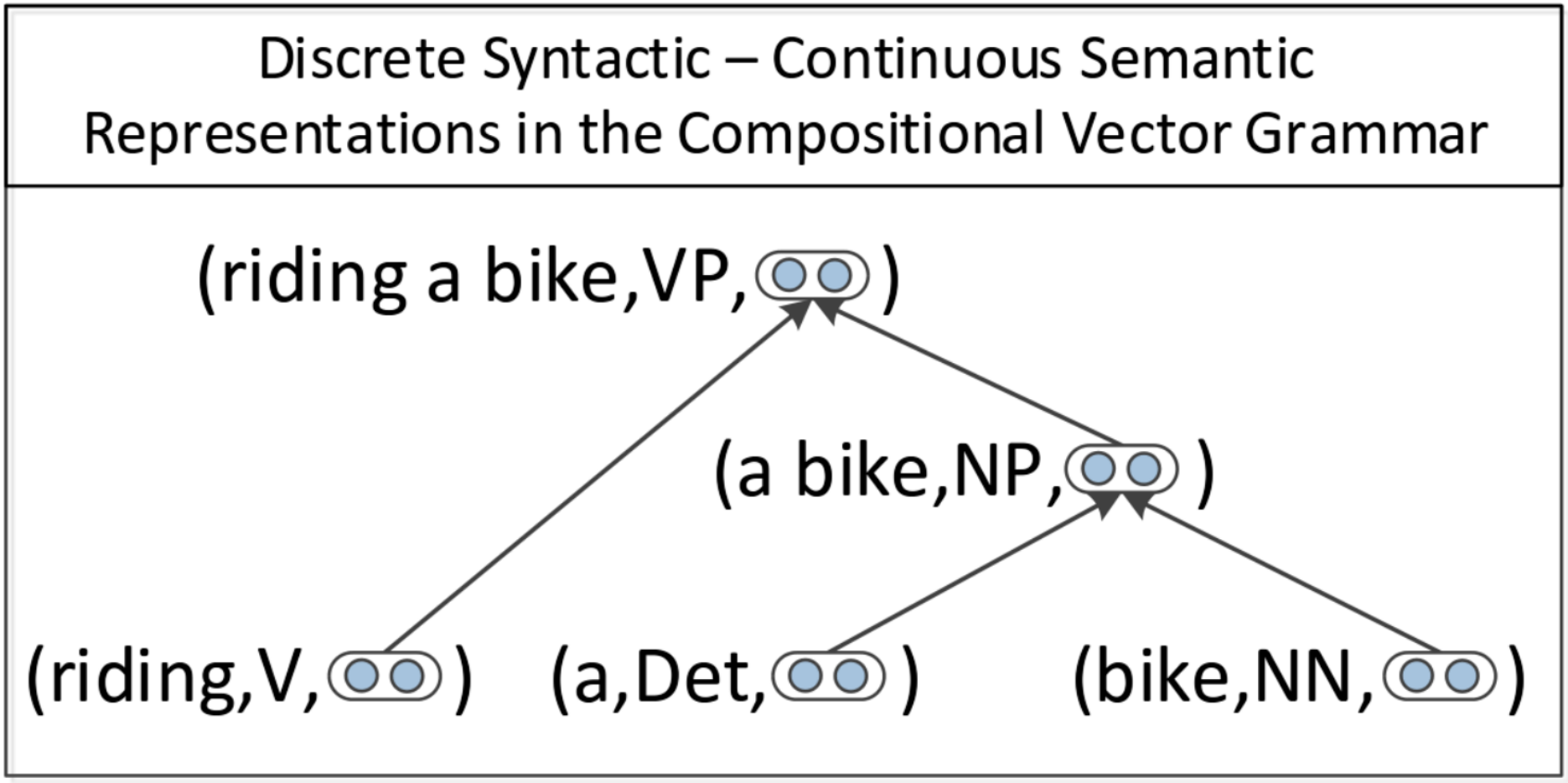
Semi-supervised Compositional Parsing with a Skip-gram Model

David Gaddy
6.864

Compositional Vector Parsing



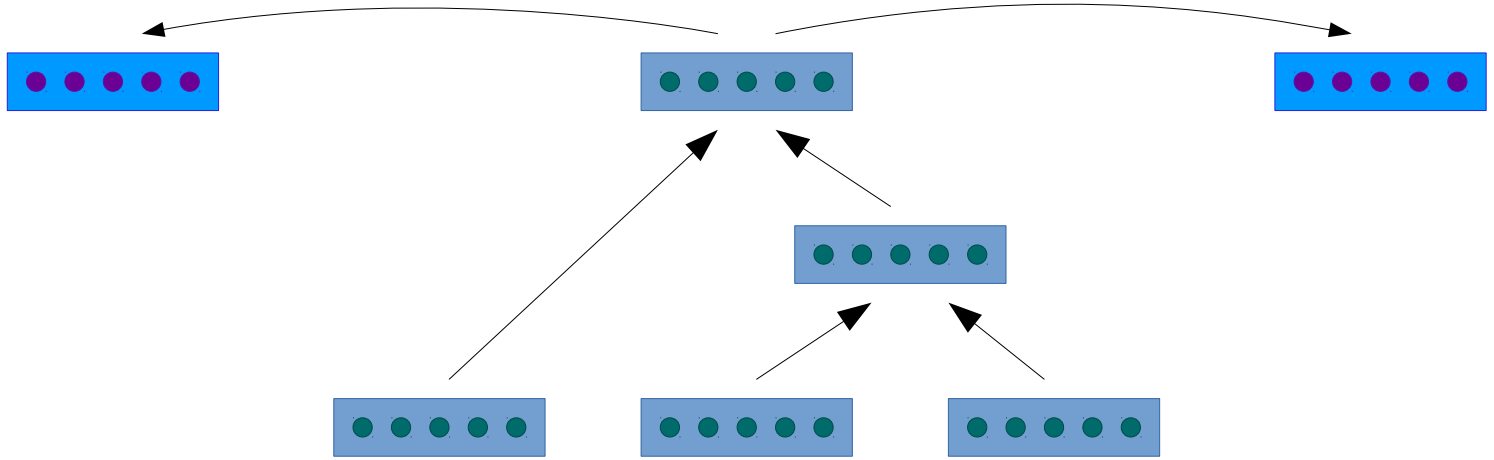
Socher et al. “Parsing with Compositional Vector Grammars”

Motivation

- Goal:
  - Take advantage of unlabeled data to improve parsing performance
- Our approach:
  - Help the parser learn what information to propagate up the compositional tree using a Skip-gram model
  - We hypothesize that the objective of a Skip-gram, determining a vector representation based on context, could also be useful for phrases
  - Context gives clues as to what information should be propagated

Methods

Skip-gram for phrases



Methods continued

- For each training batch of the parser, we add an additional objective term for the Skip-gram on a corpus with no gold tree annotations.
- In order to run the compositional Skip-gram, we need tree structures for the other corpus. We get this by sampling from the top parses of a PCFG parser on the corpus.

Evaluation

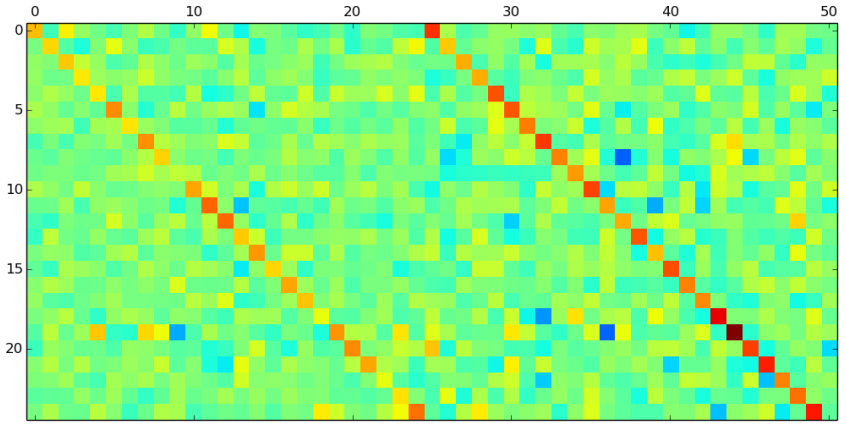
- Evaluated parser on standard constituency parsing setup with the Penn Treebank
- Train on sections 2-21
- Test on section 23 with labeled bracket scoring
- Used section 22 for development
- For Skip-gram model, used the first 500,000 sentences of the English Gigaword corpus

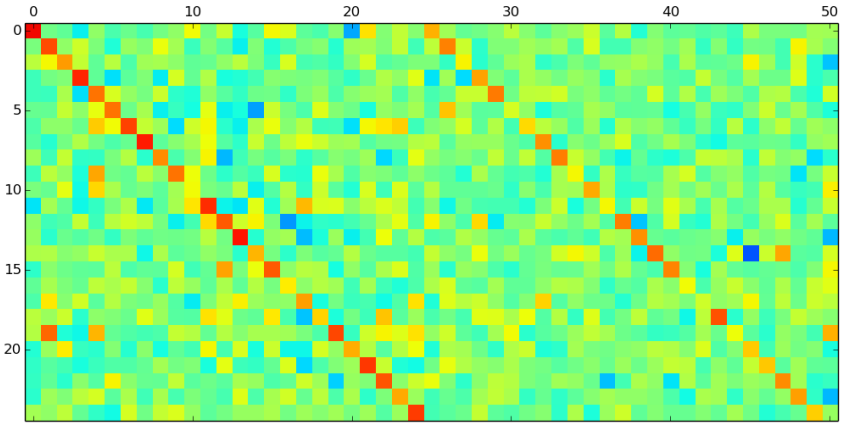
Results

	Development	Test
Original	90.05	89.19
With Skip-gram objective weight=1	88.80	88.15
With Skip-gram weight=0.1	90.23	89.09

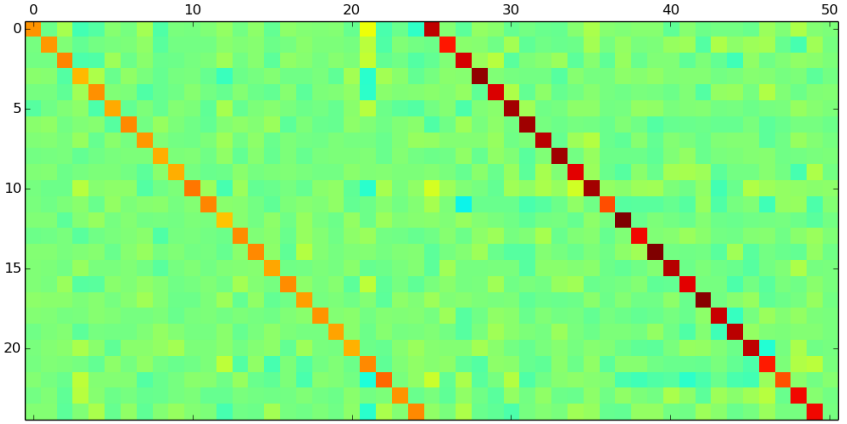
Learned Composition Matrices

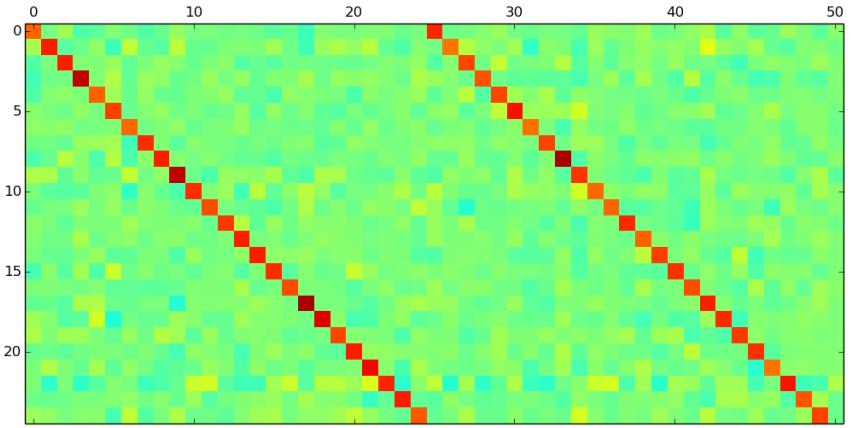
Original parsing objective





Skip-gram objective





Future Work

- Further analyze reasons for differences in learned matrices
- More data
- Larger vectors
- Syntactically untied context vectors