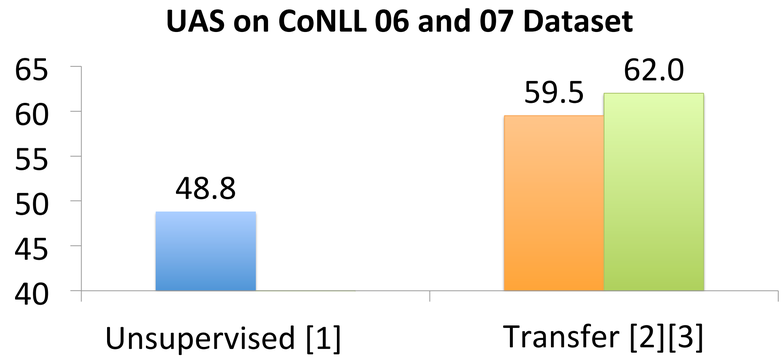
Hierarchical Low-Rank Tensors for Multilingual Transfer Parsing

Yuan Zhang, Regina Barzilay
CSAIL, MIT

Motivation

- Multilingual transfer improves unsupervised parsing performance even in the absence of parallel data



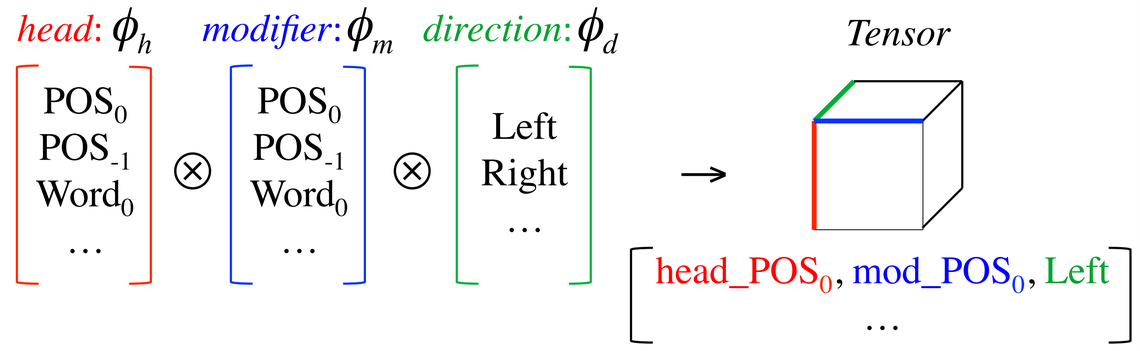
- Automatize feature engineering with tensor scoring

Manual features: Heavy engineering → *Automatic feature representation* → Tensor scoring w/o prior knowledge: invalid features

Our goal: tensor scoring with prior knowledge

Background

- Tensor scoring captures all feature combinations



- Tensor scoring alleviates the parameter explosion problem via a low-rank assumption
  - Projecting sparse feature vectors into dense r-dimensional embeddings

$$S_{tensor}(h \to m) = \sum_{i=1}^{r} [U\phi_h]_i [V\phi_m]_i [W\phi_d]_i$$

Hierarchical Low-rank Tensor Scoring for Transfer Parsing

Traditional Multi-way Tensors:

- Directly capture all feature combinations

typology SV, Prep, Adj-N, … ⊗ *head* POS0, POS-1, Word0, … ⊗ *modifier* POS0, POS-1, Word0, … ⊗ …

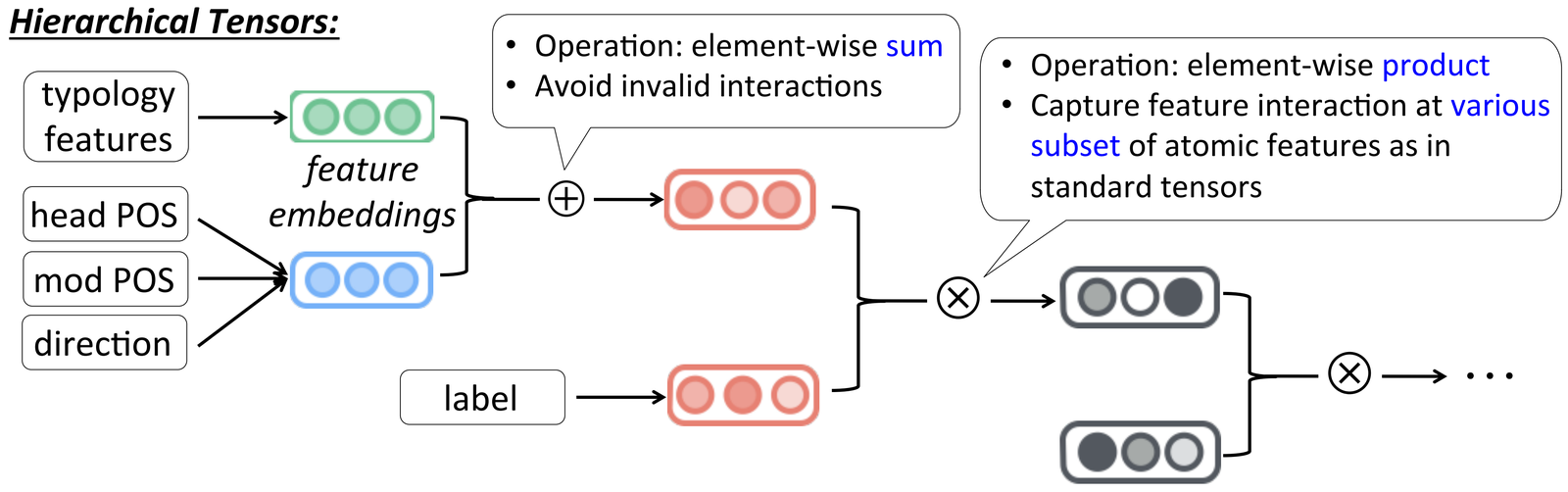
- Invalid feature combinations

typology=Adj-N, head_POS=Verb, mod_POS=Noun

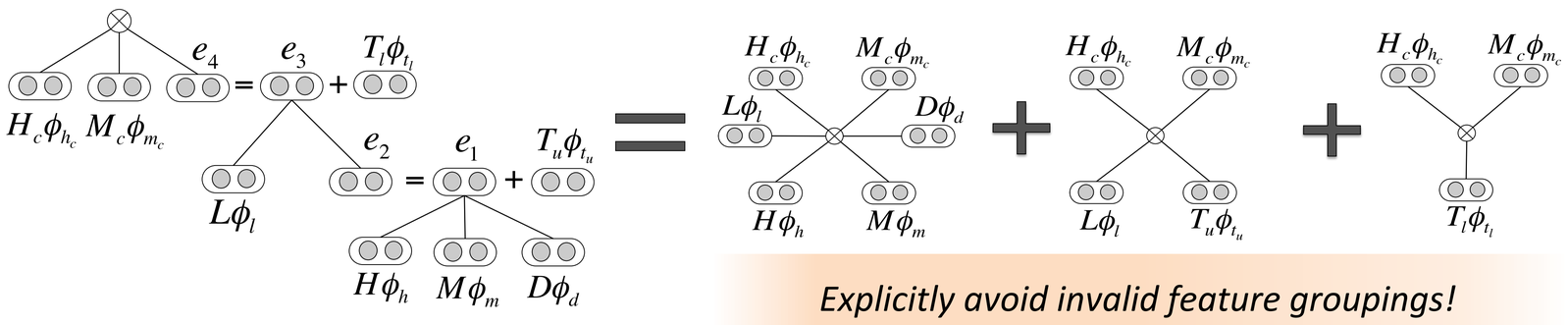
  - Adj-N indicates the ordering preference of an Adj-Noun arc, not a Verb-Noun arc.

- Issue: consider interaction at the same subset of atomic features (e.g. head POS).
- Solution: avoid such construction via a hierarchical tensor scoring structure

Hierarchical Tensors:



Algebraic Equivalence:



Explicitly avoid invalid feature groupings!

Example of Prior Knowledge

- The weight of
  typology=Adj-N, head_POS=Verb, mod_POS=Noun
  should be zero
- The weight of
  typology=Subj-Verb, head_POS=Verb, label=obj
  should be zero

Features

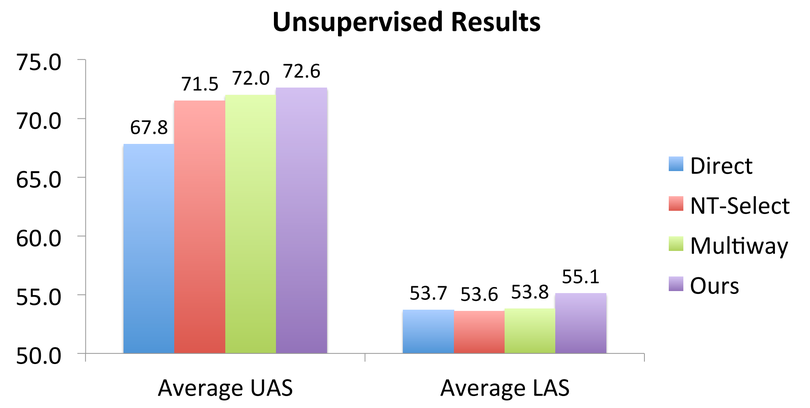
- Typological features [5]
  - Inspired by [2] and [3]
  - Include arc labels

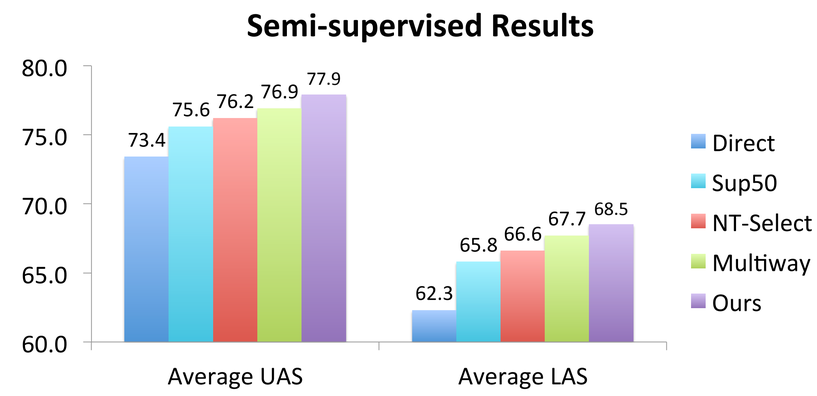
Feature	Description
82A	Order of Subject and Verb
83A	Order of Object and Verb
85A	Order of Adposition and Noun Phrase
86A	Order of Genitive and Noun
87A	Order of Adjective and Noun

- Traditional features
  - Linear scoring features: following MST Parser
  - Tensor scoring features: POS tags, labels, directions and distances etc.

Experiments

- Dataset: multilingual universal dependency treebank v2.0 [6]
  - 10 languages
- Results:
  - Unsupervised: no sentence in target language
  - Semi-supervised: 50 sentences in target language
- Findings:
  - Our model achieves best UAS and LAS on 7 out of 10 languages
  - On average, our model outperforms variants of baselines





References

[1] Valentin Spitkovsky, Hiyan Alshawi, Daniel Jurafsky. "Breaking out of local optima with count transforms and model recombination" EMNLP. 2013.
[2] Tahira Naseem, Regina Barzilay, Amir Globerson. "Selective sharing for multilingual dependency parsing." ACL, 2012.
[3] Oscar Täckström, Ryan McDonald, Joakim Nivre. "Target language adaptation of discriminative transfer parsers." NAACL (2013).
[4] Manaal Faruqui, Chris Dyer. "Improving vector space word representations using multilingual correlation." EACL, 2014.
[5] Matthew Dryer, David Gil, Bernard Comrie, Hagen Jung, Claudia Schmidt. "The world atlas of language structures." 2005.
[6] Ryan McDonald, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das et al. "Universal Dependency Annotation for Multilingual Parsing." ACL, 2013.

Mail: {yuanzh, regina}@csail.mit.edu Code: https://github.com/yuanzh/TensorTransfer