Rudolf Rosa, Zdeněk Žabokrtský
{rosa,zabokrtsky}@ufal.mff.cuni.cz

MSTParser Model Interpolation for Multi-source Delexicalized Transfer

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

IWPT, Bilbao, 22 July 2015
Outline

- Intro and motivation
- MSTParser and its delexicalization
- Single-source delexicalized parser transfer
  - $KL_{cpos3}$ language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Results and discussion
Semi-supervised parsing

- fully supervised dependency parsing
  - requires training data (treebank) or a grammar
  - there are ~100 treebanks (manually annotated)
  - there are ~7000 languages
  - + various domains, language evolution...

- semi-supervised parsing
  - utilize existing resources, avoid new annotations
    - treebanks for other langs (HamleDT: 30 langs)
    - unannotated data (here: POS tagged)
(Lexicalized) MSTParser
(Lexicalized) MSTParser

Rudolf Rosa – MSTParser Model Interpolation for Multi-source Delexicalized Transfer
Delexicalized MSTParser

#root → NOUN

NOUN → VERB

VERB → NOUN

#root → NOUN

NOUN → VERB

VERB → NOUN

16.1

7.4

-4.3

10.8

-1.4

9.7

2.2

-2.4

9.7

10.8

5.2
Single-source delex parser transfer

- (Zeman and Resnik, 2008)
- train a delexicalized parser on a source language treebank (e.g. Czech)
- apply it to a target language, without a treebank but with a POS tagger (e.g. Slovak)
Utilizing multiple treebanks

- recall: we have 30 treebanks available
- How do we choose the source treebank?
- Can we use more/all source treebanks?
Choosing the source treebank

- src should be as similar to tgt as possible
  - WALS (Naseem et al., 2012)
  - POS n-gram model (Søgaard and Wulff, 2012)
  - $KL_{cpos^3}(tgt, src)$: KL divergence of POS trigram distributions (Rosa and Žabokrtský, 2015)
Multi-source delex parser transfer

- treebank concatenation (McDonald et al., 2011)
  - weighted by lang. sim. (Søgaard and Wulff, 2012)
- parse tree combination (Sagae and Lavie, 2006)
  - crosslingual transfer (Rosa and Žabokrtský, 2015)
- parser model interpolation (this work)
Parse tree combination

src 1:  
#root  
VERB  
PREP  
NOUN  

+ src 2:  
#root  
VERB  
PREP  
NOUN  

+ src 3:  
#root  
VERB  
PREP  
NOUN  

= tgt:  
#root  
VERB  
PREP  
NOUN
Parse tree combination

calculation of the parse tree combination:

\[
\text{src 1: } \quad \text{#root} \quad \text{VERB} \quad \text{PREP} \quad \text{NOUN} \\
+ \text{src 2: } \quad \text{#root} \quad \text{VERB} \quad \text{PREP} \quad \text{NOUN} \\
+ \text{src 3: } \quad \text{#root} \quad \text{VERB} \quad \text{PREP} \quad \text{NOUN} \\
= \text{tgt: } \quad \text{#root} \quad \text{VERB} \quad \text{PREP} \quad \text{NOUN}
\]
Weighted parse tree combination

KL_{cpos}^{-4}:

 src 1: 
  #root
  VERB
  PREP
  NOUN

 src 2: 
  #root
  VERB
  PREP
  NOUN

 src 3: 
  #root
  VERB
  PREP
  NOUN

 tgt: 
  #root
  VERB
  PREP
  NOUN

KL_{cpos}^{-4}:

src 1: x 1.9
src 2: x 1.7
src 3: x 0.5

= tgt:
Weighted parse tree combination

\[ KL_{cpos^3} : \]

\[ x 1.9 \quad x 1.7 \quad x 0.5 \]

src 1:
- #root
- VERB
- PREP
- NOUN

src 2:
- #root
- VERB
- PREP
- NOUN

src 3:
- #root
- VERB
- PREP
- NOUN

tgt:
- #root
- VERB
- PREP
- NOUN

Weights:
- 4.1
- 2.4
- 2.2

Product of weights:
- 1.9
- 1.7
- 0.5
Parser model interpolation

- motivation: maybe the parser is more sure with some edges than other?
- the score assigned to the edge might show that
  - MSTParser before running the MST algorithm:
**Parser model interpolation**

- **motivation**: maybe the parser is more sure with some edges than other?
- **the score assigned to the edge** might show that
  - **MSTParser** before running the MST algorithm:
Parser model interpolation

- score normalization!

Rudolf Rosa – MSTParser Model Interpolation for Multi-source Delexicalized Transfer
Parser model interpolation

= tgt (Σ):

#root → VERB

VERB → PREP

PREP → NOUN

29.9 → 17.7 → 25.4

12.9 → 3.5 → 7.8

2.7 → -1.1
Parser model interpolation

= tgt:

#root → VERB → PREP → NOUN

29.9 → 17.7 → 25.4

2.7 12.9 3.5

-1.1 16.8 7.8
Weighted parser model interpol.

- multiply each edge score with $KL_{cpos3}^{-4}(tgt, src)$

$$KL_{cpos3}^{-4}(tgt, src1) = 0.5$$
Weighted parser model interpolation.

- multiply each edge score with $KL_{cpos3}^{-4}(tgt, src)$

$$KL_{cpos3}^{-4}(tgt, src1) = 0.5$$
Why “model interpolation”? 

- MSTParser edge score = $w \cdot f$
Why “model interpolation”?  

- MSTParser edge score = \( w \cdot f \)  
- unweighted model interpolation  
  - edge score = \( \sum_{src} (w_{src} \cdot f) \)
Why “model interpolation”?

- MSTParser edge score = \( w \cdot f \)
- unweighted model interpolation
  - edge score = \( \sum_{src} (w_{src} \cdot f) = (\sum_{src} w_{src}) \cdot f \)
Why “model interpolation”?

- MSTParser edge score = $w \cdot f$
- unweighted model interpolation
  - edge score = $\sum_{src} (w_{src} \cdot f) = (\sum_{src} w_{src}) \cdot f$
  - interpolated model $w_{int} = (\sum_{src} w_{src})$
  - edge score = $w_{int} \cdot f$
Why “model interpolation”?

- MSTParser edge score = $w \cdot f$
- unweighted model interpolation
  - edge score = $\sum_{src} (w_{src} \cdot f) = (\sum_{src} w_{src}) \cdot f$
  - interpolated model $w_{int} = (\sum_{src} w_{src})$
  - edge score = $w_{int} \cdot f$
- weighted model interpolation: $KL_{cpos3}^{-4} (tgt, src)$
  - edge score = $\sum_{src} (KL_{src} \cdot w_{src} \cdot f) = (\sum_{src} KL_{src} \cdot w_{src}) \cdot f$
  - interpolated model $w_{int} = (\sum_{src} KL_{src} \cdot w_{src})$
Evaluation on HamleDT (30 langs)

- **Treebank concatenation**
  - Unweighted: 46.1%
  - Weighted: 52.9%

- **Tree combination**
  - Unweighted: 48.9%
  - Weighted: 52.9%

- **Model interpolation**
  - Unweighted: 46.5%
  - Weighted: 52.9%
Conclusion

- Multi-source delexicalized parser transfer
  - parse tree combination
  - MSTParser model interpolation
  - $KL_{cpos3}$: language similarity for src selection/weighting
- Weighted model interpolation
  - similar accuracy to tree combination
  - faster inference
  - edge score not a good indicator of parser confidence; better methods exist (Mejer and Crammer, 2012)
Thank you for your attention

Rudolf Rosa, Zdeněk Žabokrtský
{rosa,zabokrtsky}@ufal.mff.cuni.cz

MSTParser Model Interpolation for Multi-source Delexicalized Transfer

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

http://ufal.mff.cuni.cz/rudolf-rosa/