Cross-lingual Transfer of Semantic Role Labeling Models

Mikhail Kozhevenkov and Ivan Titov
Semantic Role Labeling

Dependency-based, like in CoNLL 2009 ST
The Low-resource Setting

- Training requires large amounts of annotated data
- Even large corpora face coverage problems
- Very little or no data for many new languages
Unsupervised SRL

Aileen quickly showed Rob the book

The film was shown last night

Show me the map
Cross-lingual Approaches: Projection

- Run source-language model on the source side
- Propagate annotations through word-alignment links
- Train a target-language model on the output

Source-language model

Target-language model

The film was shown last night

La película se mostró anoche

[van der Plas et al., 2011]

[Annesi and Basili, 2010]

[Pado and Lapata, 2009]
Cross-lingual Approaches: Model Transfer

The film was shown last night.

La película fue muy aburrida.

Successfully applied to dependency parsing, NER, etc.

[Zeman and Resnik, 2008] [Durrett et al., 2012]
[Søgaard, 2011] [McDonald et al., 2011]
Overview

Purpose

- Create a simple model
- Compare against the alternatives in low-resource setting
- Figure out which features are useful

Model

- Independent linear classifiers for each argument
- No feature selection, no second-order features
Outline

- Motivation
- Cross-lingual Approaches
- Shared Feature Representation
- Evaluation and Baselines
- Results
- Conclusions and Outlook
Features of an argument instance include attributes of:
- Argument word
- Predicate word
- Parent
- Children
- Siblings
- Preceding and following words

Dog

Gloss: c123

Cross-lingual cluster ID: [Täckström et al., 2012]

Universal POS tag: NOUN [Petrov et al., 2012]

Distributed word representation: 0.84, -0.02, -0.11, ...

[Klementiev et al., 2012]
Dependency features

★ (Unlabeled) dependency structure

★ Gold-standard dependencies stripped of dependency relations

★ Direct transfer [McDonald et al., 2011]

Highlight: Dependency relations

★ Currently only PCEDT

★ Need more homogeneous treebanks [Zeman et al., 2012] [McDonald et al., 2013]

<table>
<thead>
<tr>
<th>Pair</th>
<th>UAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>En-Zh</td>
<td>35%</td>
</tr>
<tr>
<td>Zh-En</td>
<td>42%</td>
</tr>
<tr>
<td>En-Cz</td>
<td>36%</td>
</tr>
<tr>
<td>Cz-En</td>
<td>39%</td>
</tr>
<tr>
<td>En-Fr*</td>
<td>67%</td>
</tr>
</tbody>
</table>

En-Fr*: evaluation against predicted, not gold syntax
Evaluation Data

Need: different languages, similar annotation

- English-Czech
  - Prague Czech-English Dependency Treebank 2.0 \cite{Hajic2012}
  - Similar dependency relations on the two sides

- English-French
  - Annotation projection from English (PropBank)
  - 1000 sentences manually corrected \cite{vanderPlas2011}

- English-Chinese
  - Chinese Treebank, guidelines similar to PropBank
  - Core roles only (no modifiers)
Baselines

- **Annotation projection baseline**
  - Apply annotation projection to parallel data (except for French)
  - Train a lexicalized model on the output
  - Compare in terms of F1 score

- **Unsupervised baseline** [Titov and Klementiev, 2012]
  - Compare to an unsupervised SRL system using cluster measures
  - Classification only
Argument Identification

- Projection (t)
- Transfer (t)
- Projection (o)
- Transfer (o)

(t): transferred syntax
(o): original syntax
Argument Classification: Top Ten PCEDT 2.0 Labels
Proper projection model is somewhat better

Comparable performance
Unsupervised model performs better with original dependencies.
Feature Group Contribution: En-Fr, transferred syntax

- POS, Synt, Cls, Gloss
- POS, Cls, Gloss
- POS, Synt, Gloss
- POS, Synt, Cls
- POS, Gloss
- POS, Cls
- POS, Synt
- POS

$F_1$
Performance Drop Due to Transfer

$F_1$

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN-ZH</td>
<td></td>
</tr>
<tr>
<td>ZH-EN</td>
<td></td>
</tr>
<tr>
<td>EN-CZ</td>
<td></td>
</tr>
<tr>
<td>CZ-EN</td>
<td></td>
</tr>
<tr>
<td>EN-FR</td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

- Transfer performs comparably to annotation projection
- Easy to implement and trivial to apply
- Does not require high-quality parallel data
- May work better than unsupervised SRL where no accurate dependency parser is available for the target language
Current and Future Work

- Better shared representation for dependency features
- Model inter-argument dependencies
- Multiple source languages
- Domain adaptation techniques
  - Refine using parallel data
Acknowledgments

- Alexandre Klementiev
- Ryan McDonald
- Oscar Täckström
- Slav Petrov