Bootstrapping Semantic Role Labelers from Parallel Data

Mikhail Kozhevnikov and Ivan Titov
Jack easily opened the lock with a paper clip

[Jack] [easily] opened [the lock] [with a paper clip]

Agent  Manner  Patient  Instrument
Challenges with supervised learning for SRL

- Rely on large annotated datasets
  - Limited coverage (esp. across domains)
- Little or no data available for many languages

- How can we reduce reliance of SRL methods on labeled data?
Cross-lingual Approaches

- Leverage resources available for a related language
- Annotation projection
  - transfer annotations through alignment links

  \[ \text{Peter and Mary} \quad \text{left} \quad \text{Peter und auch Maria gingen} \]

- Model transfer
  - create a shared feature representation for the two languages
  - train a model on source language and apply to the target one

  \[ \text{[Kozhevnikov and Titov, ACL 2013]} \]

\[ \text{Kozhevnikov and Titov, ACL 2013} \]
Our Setting

- What if we have a little data for the target language, but with different annotation guidelines?

Our setting:
- Different annotation schemes
- Parallel data available
- Some annotated data for each language

Optimize a joint objective:

\[
\max \left( L_1(\theta_1, D_1) + L_2(\theta_2, D_2) + R(\theta_1, \theta_2, \Sigma, D_p) \right)
\]

We need a role correspondence model (RCM)!

How to define the agreement score?
Outline

- Motivation
- Model and Joint inference
- Evaluation
Role correspondence model (RCM)

- *RCM translates* a role representation in one language to a role representation in another language
  - for a given src / tgt predicate pair
  - preserving uncertainty

- What kind of correspondences can be captured by the RCM model?
  - Correspondence between roles in different annotation schemes
  - Translation shifts
Divergence in annotation schemes

- How many roles to define?
- A common set of roles across predicates?
- Roles specific to verbs, their senses or groups of verbs?
- Granularity of these groups / senses?
- Argument / modifier distinction

Different answers to these (and other) questions lead to very different annotation schemes
Annotation Schemes

- **PropBank:**

  [Jack] [easily] opened [the lock] [with a paper clip]

  \[
  \begin{array}{c|c|c}
  A0 & AM-MNR & open.01 \\
  \end{array}
  \]

  **open.01:**
  - A0 – opener
  - A1 – thing opening
  - A2 – instrument
  - A3 – benefactive

  - Roles are specific to framesets (open.01)
  - A0, A1 - Proto-Agent and Proto-Patient (Dowty 91)
  - AM* - adjuncts
Annotation Schemes

FrameNet:

[Jack] [easily] opened [the lock] [with a paper clip]

Agent Manner Closure Container_Portal Instrument

Closure:
• Agent
• Container_Portal
• Containing_Object
• Fastener
• ....

(Core) Roles are specific to a frame
'Syntactically equivalent' roles (Container_Portal and Containing_Object) are distinguished
## Example: granularity of adjuncts

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFHL</td>
<td>for how long</td>
</tr>
<tr>
<td>TFRWH</td>
<td>from when</td>
</tr>
<tr>
<td>THL</td>
<td>[after] how long</td>
</tr>
<tr>
<td>THO</td>
<td>how often</td>
</tr>
<tr>
<td>TOWH</td>
<td>to when</td>
</tr>
<tr>
<td>TPAR</td>
<td>during what time</td>
</tr>
<tr>
<td>TSIN</td>
<td>since when</td>
</tr>
<tr>
<td>TTILL</td>
<td>until when</td>
</tr>
<tr>
<td>TWHEN</td>
<td>when</td>
</tr>
</tbody>
</table>

\[=\] \textbf{AM-TMP}

English, PropBank/PTB

Czech, PCEDT 2.0
Cross-lingual Issues

- Parallelism (in projection)
  - How much of the structure is preserved in translation?
  - Translation shifts
    - “Robin sold a car to Abby” vs. “Abby bought a car from Robin”
    - “the commission set the price at 95$” vs. “the price rose to 95$”
    - “the contest ended” vs. “the winner was named”

- Roles are specific to predicates
  - Map roles between predicates
We would like to know their names and their faces.

Nos gustaría conocer sus nombres y sus rostros.

I do not have these concerns.

Yo no tengo tales preocupaciones.

Model without relying on linguistic resources (e.g., SemLink) or manually-annotated parallel data.
Outline

- Motivation
- Model and Joint inference
- Evaluation
Algorithmic view

1. **Raw Parallel Data (Source)**
2. **Labeled Parallel Data (Source)**
3. **RCM**
4. **Labeled Parallel Data (Target)**
5. **Raw Parallel Data (Target)**

- **Source Model**
- **Initial Source Data**
- **Refined Parallel Data (Source)**
- **Refined Parallel Data (Target)**
- **Target Model**
- **Initial Target Data**

**Repeat the process?**
Role Correspondence

Source role, source and target predicates

Target role

Source role

Target role, source and target predicates

RCM

Source-to-Target Mapping

Target-to-Source Mapping

(A1, have, tener): arg2-atr

Yo no tengo tales preocupaciones

A0

A1

do not have these

concerns

no tengo

preocupaciones

arg1-tem

arg2-atr

A1

A1

(arg2-atr, have, tener): A1

16
Joint inference:

- **Projection set-up:**
  - A stronger source model

\[
\hat{y}_s = \arg \max_{y_s} f_s(x_s, y_s) \\
\hat{y}_t = \arg \max_{y_t} f_t(x_t, y_t) + f_{st}(y_t, \hat{y}_s)
\]

- **Symmetric set-up:**

\[
\hat{y}_s, \hat{y}_t = \arg \max_{y_t, y_s} f_t(x_t, y_t) + f_{st}(y_s, y_t) + f_{ts}(y_t, y_s) + f_s(x_t, y_t)
\]

How to perform this inference?
Joint inference:

- Assume that RCM model factorizes over role pairs

- Projection set-up:
  - RCM scores bias the target model predictions – easy

- Symmetric set-up:
  - In a similar way, the problem can be 'reduced' to ensuring agreement between 2 models on \( \hat{y}_t \)
    
    \[
    \hat{y}_s, \hat{y}_t = \arg \max_{y_s, y_t} f_s(x_s, y_t) + f_{st}(y_s, y_t) + f_{ts}(y_t, y_s)
    \]
    
    \[
    \hat{y}_t = \arg \max_{y_t} f_t(x_t, y_t)
    \]
    
    and tackled using dual decomposition

\[x_s, x_t \] - sentences
\[y_s, y_t \] - roles structures

In fact, we do it slightly differently; details are in the paper
Joint Inference: Intuition

[On] byl jmenován [prezidentem] …

[He] was named [president] …

In practice, useful for predicting the entire predicate argument structure
Outline

- Goals
- Model and Joint inference
- Evaluation
Evaluation

- Four language pairs
  - English-German
  - English-Chinese
  - English-Spanish
  - English-Czech
- In-domain and out-of-domain evaluation

- Symmetric set-up
- Asymmetric (projection) set-up
- Better informed, “oracle” RCM

For RCM estimation, the parallel data was annotated using a model trained on a large dataset.
Evaluation
Results: English-Spanish

Source data: 20,000 instances
Parallel data: 50,000 instances

F_1

Initial training set size (argument instances)
Results: Projection

Source data: 20,000 instances
Target data: 600 instances (~100 sentences)
Parallel data: 50,000 instances

F₁

Self-training baseline
Bootstrapping

With vs. without RCM

Out of domain test set

Larger gains on out-of-domain data

Source data: 20,000 instances
Target data: 600 instances (~100 sentences)
Parallel data: 50,000 instances

Self
Joint
Combining joint model output with the initial dataset yields no significant improvement.
Results: Oracle RCM

RCM quality significantly affects the performance

Self-training

Initial

Oracle

Joint

en-cz*

en-cz

en-de*

en-de

en-es

en-zh
Results: Symmetric Setup

Source and target data: 1400 instances

$F_1$

en-cz*  en-cz  en-de*  en-de  en-es  en-zh

Self  Joint
Conclusions

- Consistent improvements over the self-training baseline
  - also in the symmetric setup
- Role correspondence model is important
  - prior knowledge could help a lot
- Depends strongly on
  - parallel data quality
  - annotation schemes involved
- Can be used in a monolingual setting
  - 20% absolute error reduction on a simpler task from agreement-based training (Titov and Kozhevnikov, ACL 2010)

Thanks to Alex Klementiev, Ryan McDonald, Alexis Palmer, Manfred Pinkal
Funded by MMCI Cluster of Excellence, DFG and a Google Research award.