Crosslingual Induction of Semantic Roles

Ivan Titov and Alexandre Klementiev

Saarland University
From Syntax to Semantics


- However, syntactic analyses are a long way from representing the meaning of sentences.

Specifically, they do not define **Who did What to Whom** (and How, Where, When, Why, …)

- In other words, they do not specify the underlying predicate argument structure.
Semantic Role Labeling (SRL)

Identification of arguments and their semantic roles:

Jack opened the lock with a paper clip

AGENT
an initiator/doer in the event [Who?]

PATIENT
an affected entity [to Whom / to What?]

INSTRUMENT
the entity manipulated to accomplish the goal
 Though syntactic and lexical representations are often predictive of the predicate argument structure, this relation is far from trivial, consider *alternations*:

- John *broke* the window
- The window *broke*
- The window was *broken* by John

Semantic Roles:
- **AGENT** – an initiator/doer in the event [Who?]
- **PATIENT** – an affected entity [to Whom / to What?]
Approaches to SRL

- **Supervised learning approaches** (e.g., [Gildea and Jurafsky, 2002; Johansson, 2008])
  - Large datasets are scarce and provide very low coverage

- **Semi-supervised methods** – combine labeled and unlabeled data success (e.g., [Furstenau and Lapata, 2009; Deschacht and Moens, 2009])
  - Relatively limited success so far

- **Crosslingual annotation projection techniques** (e.g. [Pado and Lapata 2009; van der Plas et. al. 2011])
  - Uses labeled data

- **Unsupervised methods** (e.g. [Titov and Klementiev, 2011, 2012; Lang and Lapata, 2010, 2011; Grenager and Manning, 2006])
Why Crosslingual Semantics?

- **Improvements for individual languages**
  - Crosslingual learning has been successful in syntax [Kuhn, 2004; Snyder et. al., 2009] and morphology [Snyder and Barzilay, 2008]
  - Should be even more beneficial for inducing semantics, as semantics is generally better preserved in translation

- **Induced semantic relationships across multiple languages**
  - Immediately useful for multilingual problems such as machine translation and multilingual web search

Crosslingual (unknown) regularities provide a signal for learning

Can encode directly to drive learning: e.g. one-to-one correspondences between semantic representations
Why Should Crosslingual Work for Semantics?

Helps resolve ambiguity and provide additional evidence

Peter **blamed** Mary for planning a theft
Peter **blamed** planning a theft on Mary

Linkings may be difficult to learn with monolingual data alone

Peter **beschuldigte** Mary einen **Diebstahl zu planen**

Foreign language translations would resolve these ambiguities
Our Approach to Crosslingual SRL

We induce semantic roles across languages using unsupervised monolingual data and parallel texts

- First to consider the crosslingual unsupervised setting for SRL

- Begin with our state-of-the-art nonparametric Bayesian monolingual SRL model [Titov and Klementiev, EACL 2012]

- Propose an agreement penalty for joint learning across languages

- Efficient approximate inference in the multilingual setting
Outline

- Introduction and Motivation
- Monolingual unsupervised semantic role labeling
  - Task definition
  - Overview of the nonparametric Bayesian model
- Multilingual extension
  - Role alignment penalty for joint learning across languages
  - Model inference
- Empirical evaluation
  - Data and metrics
  - Results
Monolingual Unsupervised SRL

- Semantic role labeling involves 2 sub-tasks:
  - Identification: identification of predicate arguments
  - Labeling: assignment of their semantic roles

Goal: induce semantic roles automatically from unannotated texts

- Assume that sentences are (auto-) annotated with syntactic trees
- Equivalent to clustering of argument occurrences (or “coloring” them)

Can be handled with heuristics (e.g. [Lang and Lapata, 2010])

Focus of this work
Role Labeling as Clustering of Argument Keys

- Identify arg occurrences with syntactic signatures or argument keys [Lang and Lapata, 2011]
  - E.g., some simple alternations like locative preposition drop

```
Role 1          Role 2
Mary          up  Mont Ventoux
```

- Argument keys are designed so that to map mostly to a single role

```
Role 1          Role 2
Mary  climbed up  Mont Ventoux
```

- We treat labeling of semantic roles as clustering of argument keys

```
ACTIVE:RIGHT:OBJ
ACTIVE:RIGHT:PMOD_up
```

- Here, we would cluster \texttt{ACTIVE:RIGHT:OBJ} and \texttt{ACTIVE:RIGHT:PMOD\_up} together
  - More complex alternations require multiples pairs of arg keys clustered
Selection preferences:

- Two argument keys are likely to correspond to the same role if the corresponding sets of arguments are similar.

Duplicate roles are unlikely to occur. E.g. this coloring is a bad idea:

```
John taught students math
```

Predicates admit similar alternation patterns (“reuse” them).

How to encode this in a statistical model?
for each predicate $p = 1, 2, \ldots$:
  for each occurrence $l$ of $p$:
    for every role $r \in B_p$:
      if $[n \sim \text{Unif}(0, 1)] = 1$:
        \[ \text{GenArgument}(p, r) \]
      \[ k_{p,r} \sim \text{Unif}(1, \ldots, |r|) \]
      \[ x_{p,r} \sim \theta_{p,r} \]

for each predicate $p = 1, 2, \ldots$:
  for each role $r \in B_p$:
    \[ \theta_{p,r} \sim \text{DP}(\beta, \text{H}(A)) \]
    \[ \psi_{p,r} \sim \text{Beta}(\eta_0, \eta_1) \]

Model factorizes over predicates, can consider a coupled model [Titov and Klementiev, EACL 2012]
Outline

- Introduction and Motivation
- Monolingual unsupervised semantic role labeling
  - Task definition
  - Overview of the model nonparametric Bayesian model
- Multilingual extension
  - Role alignment penalty for joint learning across languages
  - Model inference
- Empirical evaluation
  - Data and metrics
  - Results
Crosslingual Induction of Semantic Roles

- We have additional multilingual resources: texts translated in multiple languages (parallel data)
  - Parliament proceedings, books, etc.
  - Can use standard machine translation techniques to induce word alignments

  Peter beschuldigte Mary einen Diebstahl zu planen

  Peter blamed planning a theft on Mary

- We use aligned data and induce semantics jointly in multiple languages
  - Alignments are only used during learning
Consider an example *blame* alternation
Consider an example *blame* alternation

Learning the corresponding linking is not trivial

- Selectional preferences for these roles are not very restrictive
- Selectional restrictions for Cognizer and Evaluee are overlapping
Consider an example *blame* alternation

However, the alternation does not transfer to German.

Both forms are likely to have the same translation.
Crosslingual Induction of Semantic Roles

- We want induced roles for aligned sentences to be consistent
  - Favoring one-to-one mapping between aligned roles in both languages
Crosslingual Induction of Semantic Roles

- We want induced roles for aligned sentences to be consistent
- Favoring one-to-one mapping between aligned roles in both languages

Consistent roles:
A to 1
Crosslingual Induction of Semantic Roles

- We want induced roles for aligned sentences to be consistent
- Favoring one-to-one mapping between aligned roles in both languages
Crosslingual Induction of Semantic Roles

- We want induced roles for aligned sentences to be *consistent*
  - Favoring one-to-one mapping between aligned roles in both languages

Consistent roles:
- A to 1
- B to 2
- C to 3
Crosslingual Induction of Semantic Roles

- We want induced roles for aligned sentences to be consistent
  - Favoring one-to-one mapping between aligned roles in both languages

Not as good:
A to 1
B to 2 or 3
C to 3 or 2

Should be penalized
Crosslingual Induction of Semantic Roles

- We want induced roles for aligned sentences to be consistent
  - Favoring one-to-one mapping between aligned roles in both languages

- In our example: roles induced for German will be transferred to English resulting in perfect accuracy on both languages
Crosslingual Penalty

- We want roles for aligned sentences to be consistent in languages \((1)\) and \((2)\).
- Favor one-to-one mapping between aligned roles in both languages.
  - Penalize for the lack of isomorphism between the sets of roles in aligned predicates.
  - Penalty is dependent on the degree of violation.
- We augment the joint probability with a penalty term computed on parallel data:

\[
\sum_{r^{(1)}} f_{r^{(1)}} \max_{r^{(2)}} \log \hat{P}(r^{(2)}|r^{(1)})
\]

Similar to the KL expectation criteria [McCallum et al, 08]
Inference

- We use approximate maximum a-posteriori (MAP) decoding to induce semantic representations
  - Efficient: can make use of much more data

- In **monolingual** setup (for each predicate):
  - Greedy procedure for clustering of argument keys
Inference

- We use approximate maximum a-posteriori (MAP) decoding to induce semantic representations
  - Efficient: can make use of much more data

- In **crosslingual** setup (for each aligned predicate pair):
  - Induce roles for the first language (monolingual setup), then take them into account (through the penalty term) when inducing roles in the second language
  - Repeat in reverse direction
  - Choose the solution yielding a higher objective value

i.e. begin with the side which is easier to cluster and provides more clues
Outline

- Introduction and Motivation
- Monolingual unsupervised semantic role labeling
  - Task definition
  - Overview of the model nonparametric Bayesian model
- Multilingual extension
  - Role alignment penalty for joint learning across languages
  - Model inference
- Empirical evaluation
  - Data and metrics
  - Results
Benchmark Dataset: PropBank (CoNLL 08/09)

- Semantic role induction on English
- Purity measures the degree to which each induced role contains arguments sharing the same gold ("true") role

\[
PU = \frac{1}{N} \sum_i \max_j |G_j \cap C_i|
\]

- Collocation evaluates the degree to which arguments with the same gold roles are assigned to a single induced role

\[
CO = \frac{1}{N} \sum_j \max_i |G_j \cap C_i|
\]
Crosslingual Induction of Semantic Roles

- **Experimental setup:**
  - Semantic Role Labeling: identify and cluster predicate arguments
  - Induce jointly in two languages for predicates aligned in parallel data

**Crosslingual (English/German)**

- 2% Improvement for German, little for English

<table>
<thead>
<tr>
<th>Model</th>
<th>English</th>
<th>German</th>
</tr>
</thead>
<tbody>
<tr>
<td>MonoBayes</td>
<td>90</td>
<td>88</td>
</tr>
<tr>
<td>MultiBayes</td>
<td>85</td>
<td>83</td>
</tr>
<tr>
<td>SyntF</td>
<td>80</td>
<td>78</td>
</tr>
</tbody>
</table>

**Legend:**
- **Monolingual modeling**
- **Crosslingual modeling**
- **Syntactic baseline**
Conclusions and Future Work

- First to demonstrate benefits of crosslingual setup for unsupervised semantic induction

- Proposed a technique applicable to any probabilistic semantic model

- Efficient inference procedure

Future work
- Demonstrate method’s viability for other languages
- May need to induce argument keys instead of designing them for each new language

The work is partially supported by a Google Research Award and the MMCI Cluster of Excellence. Thanks to Mikhail Kozhevnikov, Alexis Palmer, Manfred Pinkal and Caroline Sporleder for helpful comments.