Frame-Semantic Role Labeling with Heterogeneous Annotations

Meghana Kshirsagar  Sam Thomson  Nathan Schneider
Jaime Carbonell  Noah A. Smith  Chris Dyer
Semantic role labeling (SRL)

Input: a sentence
Output: representation of meaning

John stole a big car
Semantic role labeling (SRL)

Input: a sentence
Output: representation of meaning (using “roles”)

Roles

Perpetrator

John stole a big car

Goods
In the case of organic pollutions, the analysis itself took no more than five days.

He sat up and took a piece of mud-coloured rag ...

Predicate: “take”
In the case of organic pollutions, the analysis itself took no more than five days.

He sat up and took a piece of mud-coloured rag...

Predicate: “take”
In the case of organic pollutions, the analysis itself took no more than five days.

He sat up and took a piece of mud-coloured rag...
In the case of organic pollutants, the analysis itself took no more than five days.

He sat up and took a piece of mud-colored rag...

Biggest challenge: **limited annotations** (≈5000 full-text sentences in FrameNet)

Predicate: “take”
Many other resources for SRL

John couldn't take the heat, so he got out of the kitchen.
Many other resources for SRL

**take.02**

John couldn't **take** the heat, so he got out of the kitchen.

PropBank has many predicates, that are not in FrameNet
Ex: attest, involve, nominate ...
Many other resources for SRL

Goal:
• Improve semantic role labeling on FrameNet using other resources

PropBank has many predicates, that are not in FrameNet
Ex: attest, involve, nominate ...

John couldn't take the heat, so he got out of the kitchen.
Many other resources for SRL

• FrameNet full-text (FT) \(\approx 5,000\)
  – document annotations: newswire, emails, transcripts of phone conversations etc.

• FrameNet Exemplars \(\approx 140,000\)
  – single sentences, primarily British National Corpus
  – distribution of roles is “artificial”

• PropBank \(\approx 110,000\)
  – WSJ data, generally coarser sense distinctions
  – different annotation scheme
This work incorporates these resources..

• **FrameNet Exemplars** [\(\approx 140,000\)]
  - single sentences, primarily British National Corpus
  - distribution of roles is “artificial”

• **PropBank** [\(\approx 110,000\)]
  - WSJ data, generally coarser sense distinctions
  - different annotation scheme

• **FrameNet Hierarchy** [Ruppenhofer et al., 2010]
  - relationships such as 
    *inheritance* between roles
A model for Frame SRL

Given: a sentence, context features, POS tags, dependency parse

Output: a set of frame, <argument spans, role label>
A model for Frame SRL

Given: a sentence, context features, POS tags, dependency parse

Output: a set of frame, <argument spans, role label>

SEMAFOR (Das et al., 2010)
A model for Frame SRL

Given: a sentence, context features, POS tags, dependency parse

Output: a set of frame, <argument spans, role label>

SEMAFOR (Das et al., 2010)

Predicate Detection → Frame Disambiguation → Argument detection + Labeling

Focus of this talk!

Given the frame, find the arguments
Objective function

- Goal: Match text-spans with role labels

```
John
John stole
stole a big
stole a big car
a big car
big car
NULL
```

```
Perpetrator
Goods
Victim
Source
Frequency
Manner
Place
```

THEFT

Perpetrator

John stole a big car

Goods
Objective function

- Goal: Match text-spans with role labels
- Score of a span ‘a’

\[
score_w(a \mid x, p, f, r) = w^\top \phi(a, x, p, f, r)
\]
Objective function

- Goal: Match text-spans with role labels
- Score of a span ‘a’
  \[
  \text{score}_w(a \mid x, p, f, r) = w^\top \phi(a, x, p, f, r)
  \]
- Squared hinge loss for \(i^{\text{th}}\) example
  \[
  \text{SqHinge}_w(i) = (\max_{a'} \{ w^\top \phi(a', x, p, f, r) + \text{cost}(a', a) \} - w^\top \phi(a, x, p, f, r))^2
  \]

Adadelta for optimization
Objective function

- Match text-spans with role labels
- Score of a span $a'$
- Squared hinge loss for $i$th example

$$\text{score } w(a',x,p,f,r) = w^T \phi(a',x,p,f,r)$$

$$\text{SqHinge}_w(i) = (\max_{a'} \{w^T \phi(a',x,p,f,r) + \text{cost}(a',a)\} - w^T \phi(a,x,p,f,r))^2$$

Significant benefits in run-time over prior work (1 week -> 9 hours)

Adadelta for optimization
Approaches to incorporate other resources

• Use as additional training data

• Via additional features (feature augmentation)
  – Frustratingly easy domain adaptation [Daumé, 09]
  – Defining “guide features” [Johansson, ‘13]

• Parameter sharing
Feature augmentation using “guides”

[21] [Johansson, ‘13]

Baseline (SEMAFOR)

FrameNet (FT)

Perpetrator

Goods

John stole a big car

feature extraction

“John”

“a big car”

training

Model
Feature augmentation using “guides”

[Johansson, ‘13]

FrameNet (FT)

Perpetrator

John stole a big car

Goods

→ \vec{h}_i

→ \vec{h}_j

→ Model

PropBank

Model

auxiliary model
Feature augmentation using “guides”

[FrameNet (FT) \( \downarrow \)]

Perpetrator

\( \text{John} \) \( \text{stole} \) \( \text{a big car} \)

\( A_0 \) \( A_1 \)

\( \vec{h}_i, g('A0') \)

\( \vec{h}_j, g('A1') \)

Output from auxiliary model is used as features in the target task

[PropBank \( \downarrow \)]

Model

auxiliary model

Model

[auxiliary model \( \downarrow \)]
Feature augmentation using “guides”

\[ g(x) = [1, x, x \Lambda FNrole, x \Lambda frame(x), x \Lambda FNframe] \]

[Johansson, ‘13]
Feature augmentation using “guides”

\[ g(x) = \left[ 1, x, x \wedge \text{FNrole}, x \wedge \text{frame}(x), x \wedge \text{FNframe} \right] \]

\[ g(\text{‘A0’}) = \left[ 1, \text{A0}, \text{A0} \wedge \text{Perpetrator}, \text{A0} \wedge \text{stole.01}, \text{A0} \wedge \text{Theft} \right] \]
Parameter sharing using the FrameNet hierarchy

Please refer to paper for details!
Parameter sharing using the FrameNet hierarchy

Parameters are shared between all siblings  
Sharing involving higher levels did not work as well

Please refer to paper for details!
Evaluation

• FrameNet 1.5
  – test set from Das et al. 2010
  – 2420 sentences, 7210 overt arguments
• For frame:
  – assume gold frame is known
  – use frames from SOTA frame-identification
    [Hermann et al 2014]
$F_1$ on test set (given gold frame)
$F_1$ on test set (given gold frame)

Baseline (FT)  |  PropBank | FrameNet Hierarchy | Exemplars | PropBank + Exemplars | Hierarchy + Exemplars
--- | --- | --- | --- | --- | ---
59.12 | 59.47 | 60.36 | 61.9 | 62.8 | 63.07

3.95% improvement over SEMAFOR

Additional resource used
**F₁ on test set (given gold frame)**

<table>
<thead>
<tr>
<th>Method</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (FT)</td>
<td>59.12</td>
</tr>
<tr>
<td>PropBank</td>
<td>59.47</td>
</tr>
<tr>
<td>FrameNet Hierarchy</td>
<td>60.36</td>
</tr>
<tr>
<td>Exemplars</td>
<td>61.9</td>
</tr>
<tr>
<td>PropBank + Exemplars</td>
<td>62.8</td>
</tr>
<tr>
<td>Hierarchy + Exemplars</td>
<td>63.07</td>
</tr>
</tbody>
</table>

**Full-system improvement:**  66.8 → 67.9

Additional resource used
Role-wise $F_1$

- Baseline (FT)
- FT + Exemplars
- FT + Exemplars + PB
- FT + Exemplars + Siblings

Frequency

Roles (ordered by frequency)

Common roles

Rare roles
Role-wise $F_1$

![Graph showing $F_1$ values for different role sets: Baseline (FT), FT + Exemplars, FT + Exemplars + PB, FT + Exemplars + Siblings. The graph plots roles ordered by test set frequency against $F_1$ scores. There are two clusters: common roles and rare roles.](image-url)

- **Common roles**
- **Rare roles**
Role-wise $F_1$

- **Baseline (FT)**
- **FT + Exemplars**
- **FT + Exemplars + PB**
- **FT + Exemplars + Siblings**

Roles (ordered by frequency):
- STATEMENT.Speaker
- ARRIVING.Goal
- CALENDRIC_UNIT.Unit
- CHATTING.Topic
- HEAR.Message
- ACCOMPLISHMENT.Goal

Frequency:
- Common roles
- Rare roles
Test sentences with gains

BOARD_VEHICLE

Vehicle
Can _ get on a plane and fly to Paris?
Traveller

BODY_MOVEMENT

Agent Body_part Purpose
Passengers crane their necks for dizzying glimpses of the harbor

Arguments in blue colour are missed by the baseline, but found by our model
Conclusion

• Contributions:
  – we exploit multiple diverse resources for better coverage
  – side-effect: faster training using hinge loss

• Future work:
  – incorporate additional resources
  – combine with other models as the baseline
    [Tackstorm et al., 2015]
Prior work

• Using FrameNet hierarchy
  – Matsubayashi et al., ‘09
  – Johansson ‘12

• Other directions
  – Pavlick, ‘15
  – Fezabadi & Pado, ‘15
Thank you!
# Sizes of the resources

<table>
<thead>
<tr>
<th>Resource</th>
<th>Number of sentences</th>
<th>Number of overt arguments</th>
</tr>
</thead>
<tbody>
<tr>
<td>FrameNet (FT)</td>
<td>2,780</td>
<td>25,918</td>
</tr>
<tr>
<td>FrameNet Exemplars</td>
<td>137,515</td>
<td>278,985</td>
</tr>
<tr>
<td>PropBank</td>
<td>112,831</td>
<td>541,759</td>
</tr>
</tbody>
</table>
Frustratingly easy domain adaptation

[Daume et al., 2009]

Maintain task-specific and general copies of features