Probabilistic Frame-Semantic Parsing

Dipanjan Das     Nathan Schneider     Desai Chen     Noah A. Smith

School of Computer Science
Carnegie Mellon University

NAACL-HLT
June 4, 2010
In a Nutshell

• Most models for semantics are very local
  (cascades of classifiers)

• This work: towards more global modeling for rich semantic processing
  (feature sharing among all semantic classes)
  (just two probabilistic models)

• Our model outperforms the state of the art

• Our framework lends itself to extensions and improvements
Outline

• Introduction
• Background and Datasets
• Models and Results
• Conclusion
Outline

• Introduction

• Background and Datasets

• Models and Results

• Conclusion
Overview

• Annotate English sentences with semantic representations

• Combination of:
  • semantic frame (word sense) disambiguation
  • semantic role labeling

• Frame and role repository: FrameNet (Fillmore et al., 2003)
Frame Semantics

• Theory developed by Fillmore (1982)

• a word evokes a frame of semantic knowledge
Frame Semantics

- Theory developed by Fillmore (1982)
- a word evokes a *frame* of semantic knowledge

the 1995 book by John Grisham
Frame Semantics

• Theory developed by Fillmore (1982)

• a word evokes a frame of semantic knowledge

the 1995 book by John Grisham
Frame Semantics

• Theory developed by Fillmore (1982)
  
  • a word evokes a *frame* of semantic knowledge

  \[
  \text{the 1995 *book* by John Grisham}
  \]

• a frame encodes a gestalt event or scenario
Frame Semantics

- Theory developed by Fillmore (1982)
  - a word evokes a frame of semantic knowledge

  the 1995 *book* by John Grisham

- a frame encodes a gestalt event or scenario
- it has conceptual dependents filling *roles*
  elaborating the frame instance
FrameNet

MAKE_NOISE

- Sound
- Place
- Time
- Noisy_event
- Sound_source

\textit{cough.v, gobble.v, hiss.v, ring.v, yodel.v, ...}

(Fillmore et al., 2003)
FrameNet

(frame)

MAKE_NOISE

Sound
Place
Time
Noisy_event
Sound_source
cough.v, gobble.v, hiss.v, ring.v, yodel.v, ...

roles

lexical units

(Fillmore et al., 2003)
FrameNet

relationships between frames and between roles
(Fillmore et al., 2003)
FrameNet

- Statistics:
  - 795 semantic frames
  - 7124 roles
  - 8379 lexical units (predicates)
  - 139,000 exemplar sentences containing one frame annotation per sentence
Marco Polo wrote an account of Asian society during the 13th century.
Marco Polo wrote an account of Asian society during the 13th century.

Here, the ambiguous word evokes the $T_{EXT}$ frame.
Marco Polo wrote an account of Asian society during the 13th century.

participants in the event or scenario
Marco Polo wrote an account of Asian society during the 13th century.

participants in the event or scenario

frame-specific
Why Frame-Semantic Parsing?

- Combines lexical and predicate-argument semantics
- Exploits meaningful primitives developed by experts
  - the FrameNet lexicon
- Richer representation than PropBank style SRL
- No inconsistent symbolic tags (ARG2-ARG5)
  (Yi et al. 2007, Matsubayashi et al. 2009)
- Patterns generalizing across frames and roles can be learned
  (Matsubayashi et al. 2009)
Outline

• Introduction
• Background and Datasets
• Models and Results
• Conclusion
Early Work

- Gildea and Jurafsky (2002)
  - Much smaller version of FrameNet
  - Exemplar sentences
SemEval 2007

- Baker et al. (2007) organized the SemEval task on frame structure extraction
  - first set of *full* text annotations available
  - released a corpus of ~2000 sentences with full frame-semantic parses
- Johansson and Nugues (2007) submitted the best performing system
  - *our baseline for comparison* (J&N’07)
SemEval 2007

• SemEval 2007 dataset:
  • training set: 1941 sentences
  • test set: 120 sentences

• Three domains
  • American National Corpus (travel)
  • Nuclear Threat Initiative (bureaucratic)
  • PropBank (news)
SemEval 2007

• Evaluation is done using the official SemEval script
  • Measures precision, recall and F₁ score for frames and arguments
  • Features a partial matching criterion for frame identification
    • assigns score between 0 and 1 to closely related frames in the FrameNet hierarchy
Outline

- Introduction
- Background and Datasets
- Models and Results
- Conclusion
Challenges

- Several times more labels than traditional shallow semantic parsing
- Annotated data does not have gold syntactic annotation
- Very little labeled data
  - Identifying semantic frames for unknown lexical units
- Very sparse features
Desired Structure

Everyone in Dublin seems intent on changing places with everyone else.
Everyone *in* Dublin *seems intent* on *changing places* with everyone else.
Three Subtasks:

- **Target identification**
  - Identifying frame-evoking predicates (nontrivial!)

- **Frame identification**
  - Labeling each target with a frame type (795 possibilities; \(\sim\)WSD)

- **Argument identification**
  - Finding each frame's arguments (\(\sim\)SRL; roleset is frame-specific)
Three Subtasks:

- **Target identification**
  - Identifying frame-evoking predicates (nontrivial!)

- **Frame identification**
  - Labeling each target with a frame type (795 possibilities; ~WSD)

- **Argument identification**
  - Finding each frame's arguments (~SRL; roleset is frame-specific)
Three Subtasks:

• **Target identification**
  - Identifying frame-evoking predicates (nontrivial!)

• **Frame identification**
  - Labeling each target with a frame type (795 possibilities; \(\sim\)WSD)

• **Argument identification**
  - Finding each frame's arguments (\(\sim\)SRL; roleset is frame-specific)
Outline

• Introduction

• Background and Datasets

• Models and Results
  • Target Identification
  • Frame Identification
  • Argument Identification
  • Final Results

• Conclusion
Target Identification

Everyone *in* Dublin *seems intent* on *changing places* with everyone else.
Target Identification

Everyone *in* Dublin *seems intent* on *changing places* with everyone else.

- Rule-based identification
- list of all morphological variants of predicates in the lexicon
- all prepositions filtered
- support verbs were not identified
- J&N’07 filtered these
Outline

- Introduction
- Background and Datasets
- Models and Results
  - Target Identification
  - Frame Identification
  - Argument Identification
  - Final Results
- Conclusion
Everyone *in* Dublin *seems intent* on *changing places* with everyone else.
Frame Identification

Everyone in Dublin seems intent on changing places with everyone else.

J&N’07 used several classifiers for this subtask.
Frame Identification
(Johansson and Nugues, 2007)

Seen LUs

\[ place \]
\[ locale \quad placing \]
\[ book \]
\[ in \]
\[ \ldots \]

1 classifier
Frame Identification
(Johansson and Nugues, 2007)

Seen LUs

Unseen LUs from WordNet-extended set

\[ \text{intent} \in \text{PURPOSE?} \]

\[ \begin{array}{c}
\text{Y} \\
\text{N}
\end{array} \]

\[ \text{intent} \in \text{AIMING?} \]

\[ \text{intent} \in \text{INGESTION?} \]

\[ \ldots \]

8 classifiers

795 classifiers
Frame Identification

Everyone *in* Dublin *seems intent* on *changing places* with everyone else.

**LOCATIVE_RELATION**  **APPEARANCE**  **PURPOSE**  **EXCHANGE**  **LOCATE**

Our approach:

One single model for frame identification
Frame Identification

Everyone in Dublin seems intent on changing places with everyone else.

Assume POS tags and dependency trees to be given.
Assume that target $t$ is connected to the frame $f$ through a prototype unit $l$.
 Everyone in Dublin seems intent on changing places with everyone else.

Assume that target $t$ is connected to the frame $f$ through a prototype unit $l$. 
Frame Identification

• Consider the **PURPOSE** frame

![PURPOSE Diagram]

- **PURPOSE**
  - Agent
  - Goal
  - Means
  - Attribute
  - Value

aim.n, goal.n, object.n, objective.n, purpose.n, target.n
Frame Identification

- Consider the PURPOSE frame

\[ \ell \in \{ \text{aim.n, goal.n, object.n, objective.n, purpose.n, target.n} \} \]

- **PURPOSE**
  - Agent
  - Goal
  - Means
  - Attribute
  - Value
Frame Identification

- Consider the PURPOSE frame

\[ \ell \in \{ \text{aim.n, goal.n, object.n, objective.n, purpose.n, target.n} \} \]

note that the target \textit{intent} is unseen
• Consider the **PURPOSE** frame

\[ l \in \{ \text{aim.n, goal.n, object.n, objective.n, purpose.n, target.n} \} \]

Note that the target *intent* is unseen

*but lexical semantic relationships between some* \( l \) *and the target exist*

**purpose \( \approx \) intent**
Thus, we define a probabilistic model:

\[ p_\theta(f, \ell | t, x) \propto \exp \theta^\top g(f, \ell, t, x) \]
Thus, we define a probabilistic model:

$$ p_\theta(f, \ell | t, x) \propto \exp \theta^\top g(f, \ell, t, x) $$

Some features looking at the lexical and semantic relationships between $\ell$ and $f$. 

WordNet relationships!
Frame Identification

Everyone in Dublin seems **intent** on changing places with everyone else.

Thus, we define a probabilistic model:

\[
p_{\theta}(f, \ell \mid t, x) \propto \exp \theta^\top g(f, \ell, t, x)
\]

other features looking at the whole sentence structure \(x\).
Thus, we define a probabilistic model:

\[ p_\theta(f, \ell \mid t, x) \propto \exp \theta^\top g(f, \ell, t, x) \]

Note that \( \ell \) is unknown.
Thus, we define a probabilistic model:

\[
p_{\theta}(f, \ell | t, x) \propto \exp \theta^\top g(f, \ell, t, x)
\]

Marginalization of latent variable:

\[
p_{\theta}(f | t, x) \propto \sum_{\ell} \exp \theta^\top g(f, \ell, t, x)
\]
Frame Identification

Everyone in Dublin seems \textit{intent} on changing places with everyone else.

\begin{equation}
\hat{f} \leftarrow \arg\max_f \sum_\ell \exp \theta^\top g(f, \ell, t, x)
\end{equation}
Frame Identification

Everyone in Dublin seems **intent** on changing places with everyone else.

Inference:

\[ \hat{f} \leftarrow \text{argmax}_f \sum_{\ell} \exp \theta^\top g(f, \ell, t, x) \]

Training:

maximum conditional likelihood
Frame Identification

Results

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>J&amp;N’07</strong></td>
<td>73.9</td>
<td>64.0</td>
<td>68.3</td>
</tr>
<tr>
<td><strong>This Work</strong></td>
<td>77.5</td>
<td>61.0</td>
<td>74.2</td>
</tr>
<tr>
<td><strong>This Work</strong></td>
<td>74.2</td>
<td>68.3</td>
<td>74.2</td>
</tr>
<tr>
<td><strong>(Model Targets)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(Oracle Targets)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Frame Identification

Results

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>J&amp;N’07 (Model Targets)</td>
<td>73.9</td>
<td>56.4</td>
<td>64.0</td>
</tr>
<tr>
<td>This Work (Model Targets)</td>
<td>77.5</td>
<td>61.0</td>
<td>68.3</td>
</tr>
<tr>
<td>This Work (Oracle Targets)</td>
<td>74.2</td>
<td>74.2</td>
<td>74.2</td>
</tr>
</tbody>
</table>

Significant improvement
Frame Identification

- For gold standard targets, 210 out of 1058 lemmas were unseen
- 190 of these get some positive score for partial frame matching
- 4 of these exactly match
- 44 get 0.5 or more, indicating close match
Outline

- Introduction
- Background and Datasets
- Models and Results
  - Target Identification
  - Frame Identification
  - Argument Identification
  - Final Results
- Conclusion
Argument Identification

Everyone in Dublin seems intent on changing places with everyone else.
**Argument Identification:** The traditional approach

*Candidate spans*

- Everyone in Dublin
- in Dublin
- on changing places
- changing places
- with everyone else
- places
- everyone

**Two steps:**
**Argument Identification:** The traditional approach

*Candidate spans*

每个人在都柏林

每个人 else

改变地方

改变地方

所有地方

所有地方

……

Two steps:

- Argument Identification:

- The traditional approach

- Candidate spans

- Everyone in Dublin

- in Dublin

- on changing places

- changing places

- with everyone else

- places

- everyone

- ✔

- ✗

- ✔

- ✗

- ✗

- ✔

- ✗

- ✗

- ✗

- ✔

- binary filtering

- potential

- arguments

Das, Schneider, Chen and Smith, NAACL-HLT 2010
Argument Identification: The traditional approach

Candidate spans

Everyone in Dublin

in Dublin

on changing places

changing places

with everyone else

places

everyone


Two steps:

Exchanger_1

classification of arguments into different roles

Exchanger_2

Themes

......
Argument Identification: The traditional approach

Candidate spans

Everyone in Dublin ✔

in Dublin ✗

on changing places ✗

changing places ✗

with everyone else ✔

places ✔

everyone ✗

Two steps: unnecessary

Exchanger_1

Exchanger_2

Themes
**Argument Identification:** Our approach

**Roleset for** \textsc{exchange}

- Exchanger\_1
- Exchanger\_2
- Themes
- Exchangers
- Theme\_1
- Theme\_2
- Manner
- Means

**Candidate spans**

- Everyone in Dublin
- in Dublin
- on changing places
- changing places
- with everyone else
- places
- everyone

......
Argument Identification: Our approach

Roleset for $\text{EXCHANGE}$

- Exchanger$_1$
- Exchanger$_2$
- Themes
- Exchangers
- Theme$_1$
- Theme$_2$
- Manner
- Means

Candidate spans

- Everyone in Dublin
- on changing places
- changing places
- with everyone else
- places
- everyone

one step!

......
Argument Identification

A probabilistic model:

\[ p_\psi(r \rightarrow s \mid f, t, x) \propto \exp \psi^\top h(r, s, f, t, x) \]
Argument Identification

A probabilistic model:

\[ p_\psi(r \rightarrow s \mid f, t, x) \propto \exp \psi^\top h(r, s, f, t, x) \]

features looking at the span, the frame, the role and the observed sentence structure
Argument Identification

A probabilistic model:

$$p_\psi(r \rightarrow s \mid f, t, x) \propto \exp \psi^\top h(r, s, f, t, x)$$

Decoding:

Best span for each role is selected

For each frame, the best set of non-overlapping arguments is decoded together
Argument Identification

A probabilistic model:

\[ p_\psi(r \rightarrow s \mid f, t, x) \propto \exp \psi^\top h(r, s, f, t, x) \]

Training:

Maximum conditional likelihood
Argument Identification

Results

![Bar chart showing precision, recall, and F1 scores for Model Spans and Oracle Spans.]

Model Spans:
- Precision: 78.7
- Recall: 60.6
- F1: 68.5

Oracle Spans:
- Precision: 88.3
- Recall: 74.8
- F1: 81.0

Argument identification only, with gold targets and frames
Outline

• Introduction

• Background and Datasets

• Models and Results
  • Target Identification
  • Frame Identification
  • Argument Identification

• Final Results

• Conclusion
Full Frame-Semantic Parsing

Results

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>J&amp;N’07</td>
<td>56.0</td>
<td>38.4</td>
<td>45.6</td>
</tr>
<tr>
<td>This work</td>
<td>62.8</td>
<td>41.9</td>
<td>50.2</td>
</tr>
</tbody>
</table>

full frame-semantic parsing
Full Frame-Semantic Parsing

Results

<table>
<thead>
<tr>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>56.0</td>
<td>38.4</td>
<td>45.6</td>
</tr>
<tr>
<td>62.8</td>
<td>41.9</td>
<td>50.2</td>
</tr>
</tbody>
</table>

J&N’07 vs. This work

Significant improvement in F1 score for full frame-semantic parsing.
Conclusion

• Best results to date on frame-semantic parsing

• Only two probabilistic models instead of a cascade of classifiers for the frame-semantic parsing task

• Latent variable model for frame identification

• Better modeling of the argument identification (SRL) stage using only one model instead of two

• Publicly available software: http://www.ark.cs.cmu.edu/SEMAFOR
Thanks!

http://www.ark.cs.cmu.edu/SEMAFOR
Thanks!

JUDGMENT_DIRECT_ADDRESS

http://www.ark.cs.cmu.edu/SEMAFOR