Semantic Frame Identification with Distributed Word Representations

Karl Moritz Hermann†¹  Dipanjan Das†
Jason Weston†  Kuzman Ganchev†

†Google Inc., New York
‡ Department of Computer Science, University of Oxford

Baltimore, 25.06.2014

¹The majority of this research is result of an internship at Google.
We investigate features and classifiers for frame-semantic parsing

Input

John  runs  the show
We investigate features and classifiers for frame-semantic parsing
We investigate features and classifiers for frame-semantic parsing.
We investigate features and classifiers for frame-semantic parsing.
We investigate features and classifiers for frame-semantic parsing

Input

John  runs  the show

Features

nsubj  dobj  ...

Model

Log-linear Classifier
We investigate features and classifiers for frame-semantic parsing
Frame-Semantic Parsing

The task of extracting semantic predicate-argument structures from text.

John sold Mary a car.
Frame-Semantic Parsing

The task of extracting semantic predicate-argument structures from text.

\[ \text{COMMERCE\_BUY} \]

\[ \text{sell.V} \]

\[ \text{John} \quad \text{sold} \quad \text{Mary} \quad \text{a} \quad \text{car} \quad . \]
Frame-Semantic Parsing

The task of extracting semantic predicate-argument structures from text.

John sold Mary a car.
Frame-Semantic Parsing

The task of extracting semantic predicate-argument structures from text.
Frame parsing as a two-stage task

Frame-Semantic Parsing is a combination of two tasks:

- Frame Identification
- Argument Identification

This paper focuses on Frame Identification

However, we also present results on the full pipeline task.
Representing frame instances

Default Approach

- Frame instances represented by candidate and context
- Context: sparse features based on parse tree

Distributed Approach

- Replaces binary features with word embeddings
- Context: same features, but represented distributionally
Step 1: Parsing

- Dependency parse of input sentence
- Paths are considered relative to candidate word

Jane recently bought flowers from Luigi's shop
Instances to Vectors

Step 1: Parsing

- Dependency parse of input sentence
- Paths are considered relative to candidate word
Step 2: Context word selection strategy

- Direct dependents based on parse
- Argument dependency paths learned from gold data
Step 3: Embeddings

- Replace words with embeddings
Instances to Vectors

Step 4: Single vector creation

- Merge embeddings into a unified vector representation
- Effectively concatenation; zeros for empty slots
Joint-space Model (Wsabie) — Learning

Joint-space Model

- Instances represented in $\mathbb{R}^d$ based on pre-trained embeddings.
Joint-space Model (Wsabie) — Learning

Joint-space Model

- Labels represented as discrete values given lexicon (size $F$).
Joint-space Model (Wsabie) — Learning

Joint-space Model

- Learn a linear mapping $M : \mathbb{R}^d \rightarrow \mathbb{R}^m$. 

Frame Instance Map $M : \mathbb{R}^d \rightarrow \mathbb{R}^m$

Joint Space $\mathbb{R}^m$

Frame Instance Space $\mathbb{N} \times \mathbb{R}^d$

Label Space (discrete)

- Commemorative
- Commerce_buy
- Commerce_collect
Joint-space Model (Wsabie) — Learning

- Learn a matrix $Y \in \mathbb{R}^{F \times m}$ to represent labels.
Joint-space Model (Wsabie) — Learning

Joint-space Model

- Objective function:

\[ \sum_x \sum_{\tilde{y}} L(\text{rank}_y(x)) \max(0, \gamma + s(x, y) - s(x, \tilde{y})). \]
Joint-space Model (Wsabie) — Classification

- Project candidate into joint-space

Joint-Space $\mathbb{R}^m$
Joint-space Model (Wsabie) — Classification

- Project candidate into joint-space
- Only consider label projections
Joint-space Model (Wsabie) — Classification

- Project candidate into joint-space
- Only consider label projections
- Restrict labels using lexicon
Joint-space Model (Wsabie) — Classification

- Project candidate into joint-space
- Only consider label projections
- Restrict labels using lexicon
- Classify candidate using suitable distance metric
Experiments

Learning Setup

• Neural language model (~ Bengio et al., 2003) trained on over 100 billion tokens to learn 128-dimensional word embeddings
• FrameNet 1.5 and Ontonotes 4.0 (PropBank, WSJ-only) used for training the actual Wsabie models
• Hyperparameters optimised on development data
Baselines: Where is the power coming from?

Input

John  runs  the show

Features

nsubj:John  dobj:show  ...

Model

Log-linear Classifier

Models Evaluated
Baselines: Where is the power coming from?

Models Evaluated

- Log-Linear Words
Baselines: Where is the power coming from?

Models Evaluated
- Log-Linear Words
- Log-Linear Embeddings
Baselines: Where is the power coming from?

Models Evaluated
- Log-Linear Words
- Log-Linear Embeddings
- Wsabie
Evaluation

Evaluation Settings

- We evaluate on FrameNet (here) and PropBank (see paper)
- FrameNet setup follows Das et al. (2014), with a restricted lexicon during training (Semafor)
- Multiple evaluations
  - **All** evaluates all frames
  - **Rare** predicates with frequency $\leq 11$ in the training data
  - **Unseen** predicates not observed in the training data
Frame Identification Results (FrameNet - All Predicates)

Figure: Frame Identification results on FrameNet dataset. We restrict the training data to the Semafor lexicon for comparability with Das et al., 2014.
Figure: Frame Identification results on FrameNet dataset. We restrict the training data to the Semafor lexicon for comparability with Das et al., 2014.
Figure: Frame Identification results on FrameNet dataset. We restrict the training data to the Semafor lexicon for comparability with Das et al., 2014.
Full pipeline

Full frame-semantic analysis

In addition to the frame identification experiments, we also use our models as part of a full frame-semantic parsing setup together with a standard argument identification method.

Argument Identification System

- Standard set of discrete features
- Local log-linear model
- Global inference via hard constraints and ILP
Full pipeline results (FrameNet)

Figure: Full frame-structure prediction results for FrameNet dataset (Semafor).
Conclusion

• Novel approach to frame identification
• Model outperforms prior state of the art on frame identification
• In a pipeline setting with a standard argument identification system, the model also sets a new state of the art on various semantic parsing tasks
• General approach, could easily be extended for alternative frame-semantic parsing frameworks
The End

Thank you!

Questions?
### Frame Identification Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Semafor Lexicon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
</tr>
<tr>
<td>Das et al., 2014 supervised</td>
<td>82.97</td>
</tr>
<tr>
<td>Das et al., 2014 best</td>
<td>83.60</td>
</tr>
<tr>
<td>Log-Linear Words</td>
<td>84.53</td>
</tr>
<tr>
<td>Log-Linear Embed.</td>
<td>83.94</td>
</tr>
<tr>
<td>Wsabie Embedding</td>
<td><strong>86.49</strong></td>
</tr>
</tbody>
</table>

**Table**: Frame identification results on the FrameNet test data.
# Frame Identification Results

<table>
<thead>
<tr>
<th>Model</th>
<th>All</th>
<th>Ambiguous</th>
<th>Rare</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-Linear Words</td>
<td>87.33</td>
<td>70.55</td>
<td>87.19</td>
</tr>
<tr>
<td>Log-Linear Embed.</td>
<td>86.94</td>
<td>70.26</td>
<td>86.56</td>
</tr>
<tr>
<td>Wsabie Embedding</td>
<td><strong>88.41</strong></td>
<td><strong>73.10</strong></td>
<td><strong>88.93</strong></td>
</tr>
</tbody>
</table>

**Table**: Frame identification results on the FrameNet test data
# Full Structure Prediction Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Semafor Lexicon</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
<td>$F_1$</td>
<td></td>
</tr>
<tr>
<td>Das et al. supervised</td>
<td>67.81</td>
<td>60.68</td>
<td>64.05</td>
<td></td>
</tr>
<tr>
<td>Das et al. best</td>
<td>68.33</td>
<td>61.14</td>
<td>64.54</td>
<td></td>
</tr>
<tr>
<td>Log-Linear Words</td>
<td>71.21</td>
<td>63.37</td>
<td>67.06</td>
<td></td>
</tr>
<tr>
<td>Wsabie Embedding</td>
<td><strong>73.00</strong></td>
<td><strong>64.87</strong></td>
<td><strong>68.69</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Full structure prediction results for FrameNet test data. We compare to the prior state of the art (Das et al., 2014).
Full Structure Prediction Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-Linear Words</td>
<td>73.31</td>
<td>65.20</td>
<td>69.01</td>
</tr>
<tr>
<td>Wsabie Embedding</td>
<td><strong>74.29</strong></td>
<td><strong>66.02</strong></td>
<td><strong>69.91</strong></td>
</tr>
</tbody>
</table>

Table: Full structure prediction results for FrameNet test data. We compare to the prior state of the art (Das et al., 2014).
Figure: Argument only evaluation (argument identification metrics) using the CoNLL 2005 shared task evaluation script (Carreras and Màrquez, 2005). Results from Punyakanok et al. (2008) are taken from Table 11 of that paper.
Frame Identification Results (PropBank)

Figure: Frame Identification results on PropBank datasets.
Full pipeline results (PropBank)

Figure: Full frame-structure prediction results for PropBank dataset.