Frame-Semantic Parsing

Das, Chen, Martins, Schneider & Smith
2012
Frame semantics

- A theory of linguistic meaning; relates linguistic semantics to encyclopaedic knowledge - cannot understand the meaning of a single word without all the essential knowledge that relates to that word.

- A word activates, or evokes, a frame of semantic knowledge relating to the specific concept it refers to.
Frame semantics (example)
FrameNet

- A linguistic resource storing considerable information about lexical and predicate-argument semantics in English.

- FrameNet 1.5 release:
  - 877 frames
  - 1068 role types
  - 154,607 exemplar sentences
  - 5676 fully annotated sentences
Frame-Semantic Parsing

- Aims to extract shallow semantic structure from text.
- Identify target (frame evoking element)
- Identify frame (disambiguate word sense)
- Identify role-filling phrases
- Annotate phrases with roles from the frame
In that time more than 1.2 million jobs have been *created* and the official jobless rate has been *pushed* below 17% from 21%.
Semantic Analyzer of Frame Representations

- Task 1: target identification
- Task 2: frame identification
- Task 3: argument identification

Results evaluated using the standard SemEval 2007 evaluation script
- Calculates precision, recall and F1 for frames and arguments
- Partial credit for a related frame

Johansson & Nugues results from the SemEval 2007 shared task used as the baseline system.
Target Identification

- Create master list of all morphological variants of targets that appear in the exemplar sentences and in the training set.
  - Do not capture discontinuous frame targets
- Candidate targets in new string are only phrases that appear in master list
- Prune candidate target by applying rules similar to J&N07, except:
  - Prune all prepositions
  - Keep candidates marked as support verbs for other targets
- Cannot identify targets for completely new LUs
Target Identification - results

- 85 distinct LUs which the baseline fails, but SEMAFOR succeeds
  - Most contain more than one token, which the baseline doesn’t model
  - Variants of *there be.v*
  - Month names
  - LUs of the ORIGIN frame
  - Directions

Table 3: Target identification results for our system and the baseline on the SemEval’07 dataset. Scores in bold denote significant improvements over the baseline ($p < 0.05$).
Frame identification

- LUs are lemmatized, POS-tagged words and phrases that evoke relevant frames.
- Base model on targets instead of LUs to avoid possible lemmatization errors.
- \( L \) - set of auto-POSIed, non-lemmatized targets found in training data
- \( L_f \) - subset for frame \( f \)
- \( L_f^l, L^l \) - lemmatized versions
Frame identification

- For each target $t_i$ of a given sentence $x$
  - $t'_i$ - lemma of $t_i$
  - $F = \{f' \mid t'_i \text{ is in } L_{f'}\}$
  - If $t'_i$ is not in any $L_{f'}$, $F = \{\text{all known frames}\}$

- Use latent variable $l$ to allow identification of targets with unseen lemmas
  - Can be thought of as prototypes for the expression of the frame
  - $l$ are elements in $L_f$ for a given frame
  - $l = (w, \pi)$ where $w$ is the phrase and $\pi$ is the POS tagging
Frame identification

- Return $f$ in $F$ which maximizes:

$$f_i \leftarrow \arg \max_{f \in F_i} \sum_{\ell \in L_f} p_\theta(f, \ell | t_i, x)$$

where

$$p_\theta(f, \ell | t_i, x) = \frac{\exp \theta^T g(f, \ell, t_i, x)}{\sum_{f' \in F} \sum_{\ell' \in L_{f'}} \exp \theta^T g(f', \ell', t_i, x)}$$

- $\theta$ is a vector of weights
  - Set as the value which maximizes the model on the training data
- $g$ is a feature vector
- both are consistent across all targets, frames and prototypes
Frame identification

- the POS of the parent of the head word of $t_i$
- the set of syntactic dependencies of the head word of $t_i$
- if the head word of $t_i$ is a verb, then the set of dependency labels of its children
- the dependency label on the edge connecting the head of $t_i$ and its parent
- the sequence of words in the prototype, $w_\ell$
- the lemmatized sequence of words in the prototype
- the lemmatized sequence of words in the prototype and their part-of-speech tags $\pi_\ell$
- WordNet relation $^21$ holds between $\ell$ and $t_i$
- WordNet relation $^21$ holds between $\ell$ and $t_i$, and the prototype is $\ell$
- WordNet relation $^21$ holds between $\ell$ and $t_i$, the POS tag sequence of $\ell$ is $\pi_\ell$, and the POS tag sequence of $t_i$ is $\pi_\ell$
## Frame identification - results

<table>
<thead>
<tr>
<th>Frame Identification (§5.2)</th>
<th>exact matching</th>
<th>partial matching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P$</td>
<td>$R$</td>
</tr>
<tr>
<td><strong>SemEval 2007 Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gold targets</td>
<td>60.21</td>
<td>60.21</td>
</tr>
<tr>
<td>automatic targets (§4)</td>
<td>69.75</td>
<td>54.91</td>
</tr>
<tr>
<td>J&amp;N’07 targets</td>
<td>65.34</td>
<td>49.91</td>
</tr>
<tr>
<td>Baseline: J&amp;N’07</td>
<td>66.22</td>
<td>50.57</td>
</tr>
<tr>
<td><strong>FrameNet 1.5 Release</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gold targets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– unsupported features</td>
<td>82.97</td>
<td>82.97</td>
</tr>
<tr>
<td>&amp; – latent variable</td>
<td>80.30</td>
<td>80.30</td>
</tr>
<tr>
<td>75.54</td>
<td>75.54</td>
<td>75.54</td>
</tr>
</tbody>
</table>

Table 5: Frame identification results on both the SemEval 2007 dataset and the FrameNet 1.5 release. Precision, recall, and $F_1$ were evaluated under exact and partial frame matching; see §3.3. Bold indicates best results on the SemEval 2007 data, which are also statistically significant with respect to the baseline ($p < 0.05$).
Semi-Supervised Lexicon Expansion

- Construct graph from both labeled and unlabeled data
  - Vertices are LUs - each vertex corresponds to a lemmatized word or phrase appended with a coarse POS tag
    - Both from FrameNet (labeled) and from (Lin 1998) (unlabeled)
- Weighted edges between LUs
  - For each LU calculate distribution over all frames in FrameNet 1.5
  - Euclidean distance between distribution of 2 LUs
  - Interpolate with similarity measure from (Lin 1998)
- Link each vertex to $K$ nearest neighbours
- Final graph contains 64,480 vertices:
  - 9,263 labeled
  - 55,217 unlabeled
Figure 4: Excerpt from our constructed graph over LUs. Green LUs are observed in the FrameNet 1.5 data. Above/below them are shown the most frequently observed frame that these LUs associate with. The black LUs are unobserved and graph propagation produces a distribution over most likely frames that they could evoke as target instances.
Semi-Supervised Lexicon Expansion

- Propagate distributions from labeled to unlabeled vertices using one of 2 methods:
  - Uses normalized probability distributions at each vertex and is a Gaussian field; it also employs a uniform $l_2$ penalty (third term)
  
  \[
  \text{NGF-}l_2 : \arg \min_{q \in \mathcal{Q}, \text{s.t. } \sum q_v = 1} \sum_{v \in V} \| \hat{q}_v - q_v \|_2^2 + \mu \sum_{v \in V, u \in N(v)} \| q_v - q_u \|_2^2 + \nu \sum_{v \in V} \| q_v - \hat{q}_v \|_2^2
  \]

  - Uses unnormalized probability measures at each vertex and is a Jensen-Shannon field, employing pairwise Jensen-Shannon divergences and a sparse $l_{1,2}$ penalty (third term)
  
  \[
  \text{UJSF-}l_{1,2} : \arg \min_{q \geq 0} \sum_{v \in \hat{V}} D_{JS}(\hat{q}_v \| q_v) + \mu \sum_{v \in V, u \in N(v)} w_{uv} D_{JS}(q_v \| q_u) + \nu \sum_{v \in V} \| q_v \|_1^2
  \]
Semi-Supervised Lexicon Expansion

- Once the objective function is minimized, we arrive at $q^*$, the optimal set of frame distributions.

- Improve frame identification over the supervised method by selecting only the $M$ best candidate frames for each target.
Semi-Supervised Lexicon Expansion - results

Table 6: Exact and partial frame identification accuracy on the FrameNet 1.5 dataset with the size of lexicon (in terms of non-zero frame components in the truncated frame distributions) used for frame identification, given gold targets. The supervised model is compared to alternatives in Table 5. Bold indicates best results. **UJSF-ℓ₁,₂** produces statistically significant results ($p < 0.001$) for all metrics with respect to the supervised baseline for both the unseen LUs as well as the whole test set. Although the **NGF-ℓ₂** and **UJSF-ℓ₁,₂** models are statistically indistinguishable, it is noteworthy that the **UJSF-ℓ₁,₂** objective produces a much smaller lexicon.
Argument Identification

- Identify a set $S$ of spans in a given sentence $x$ which are candidate role-fillers:
- span $s$ is in $S$ if:
  - $s$ is a single word
  - $s$ is a valid subtree in the dep-parse of $x$ (MST parser)
  - $s$ is the empty span (to cover NIRs)
  - Covers ~80% of the dev data
  - In training, a labeled argument is included even if not valid subtree
Argument Identification

- To match spans to roles given a target and a frame:

\[ A_i(r_k) \leftarrow \arg\max_{s \in S} p_{\theta}(s | r_k, f_i, t_i, x) \]

where

\[
p_{\theta}(A_i(r_k) = s | f_i, t_i, x) = \frac{\exp \psi^T h(s, r_k, f_i, t_i, x)}{\sum_{s' \in S} \exp \psi^T h(s', r_k, f_i, t_i, x)}
\]

- Assigns spans independently
  - Allows sharing spans between frames to fulfill different roles

- Selects and annotates roles in the same step
Argument Identification

- Using the model as is to identify spans+roles may result in overlapping spans for different roles (due to independence of identification).

- Add non-overlap constraint by using beam search on the space of (span, role) matches (beam width = 10,000).
## Argument Identification

<table>
<thead>
<tr>
<th>ARGUMENT IDENTIFICATION</th>
<th>targets</th>
<th>frames</th>
<th>decoding</th>
<th>exact matching</th>
<th>partial matching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$P$</td>
<td>$R$</td>
</tr>
<tr>
<td>Argument identification (full)</td>
<td>gold</td>
<td>gold</td>
<td>naive</td>
<td>77.43</td>
<td>60.76</td>
</tr>
<tr>
<td>Parsing (oracle targets)</td>
<td>gold</td>
<td>supervised ($$5.2$)</td>
<td>beam</td>
<td>49.68</td>
<td>42.82</td>
</tr>
<tr>
<td>Parsing (full)</td>
<td>auto</td>
<td>supervised ($$5.2$)</td>
<td>beam</td>
<td>58.98</td>
<td>38.76</td>
</tr>
<tr>
<td>Parsing (I&amp;N’07 targets and frames)</td>
<td>auto</td>
<td>supervised ($$3.4$)</td>
<td>beam</td>
<td>56.26</td>
<td>36.63</td>
</tr>
<tr>
<td>Baseline: I&amp;N’07</td>
<td>auto</td>
<td>supervised ($$3.4$)</td>
<td>N/A</td>
<td>51.59</td>
<td>35.44</td>
</tr>
</tbody>
</table>

### SemEval’07 Data

### FrameNet 1.5 Release

| Argument identification (full) | gold | gold | naive | beam | 82.00 | 76.36 | 79.08 | 83.83 | 76.28 | 79.88 |
| Parsing (oracle targets) | gold | supervised ($\$5.2$) | beam | 67.51 | 60.68 | 64.05 | 72.47 | 64.85 | 68.45 |
| Parsing (full) | gold | SSL (NGF-$l_2$, $\$5.5$) | beam | 68.22 | 61.04 | 64.43 | 72.87 | 65.20 | 68.82 |
| Parsing (I&N’07 targets and frames) | gold | SSL (UJSF-$l_1, l_2$, $\$5.5$) | beam | 68.33 | 61.14 | 64.54 | 72.98 | 65.30 | 68.93 |

Table 8: Argument identification results on both the SemEval’07 data as well as the full text annotations of FrameNet 1.5. For decoding, “beam” and “naive” indicate whether the approximate joint decoding algorithm has been used or local independent decisions have been made for argument identification, respectively. On the SemEval 2007 data, for full parsing (automatic target, frame and argument identification), bold scores indicate best results, which are also significant improvements relative to the baseline ($p < 0.05$). On the FrameNet 1.5 dataset, bold scores indicate best results on automatic frame and argument identification—this is achieved by the frame identification model that uses the UJSF-$l_1, l_2$ graph-objective and automatic argument identification using beam search. This result is statistically significant over the supervised results shown in row 9 ($p < 0.001$). In terms of precision and $F_1$ score measured with partial frame matching, the results with the UJSF-$l_{1, 2}$ model is statistically significant over the NGF-$l_2$ model ($p < 0.05$). For recall with partial frame matching, and for all the three metrics with exact frame matching, the results with the two graph objectives are statistically indistinguishable. Note that certain partial match results are missing because in those settings, gold frames have been used for argument identification.
Argument identification with constraints

- Argument identification above does not capture all semantic knowledge included in FrameNet

- Attempt to identify the set of arguments that best satisfies all constraints
  - Uniqueness - each role is filled by at most one span (96.4%)
  - Overlap - different roles cannot be fulfilled by overlapping spans
  - Pairwise exclusion - some roles are declared as mutually exclusive in FrameNet (204 frames)
  - Pairwise requirement - some roles are declared as mutually inclusive in FrameNet (54 frames)
Argument identification with constraints

- Pairwise exclusion/inclusion:

1. A blackberry \textit{resembles} a loganberry.
   \begin{align*}
   \text{Entity}_1 & \text{\underline{\textit{resembles}}} \text{ Entity}_2 \\
   & \text{Entities}
   \end{align*}

2. Most berries \textit{resemble} each other.
   \begin{align*}
   \text{Entities} & \text{\underline{\textit{resemble}} \text{ each other.}} \\
   & \text{Entities}
   \end{align*}
Argument identification with constraints

- Solve argument identification using ILP/LP, where constraints of the problem are the linguistic constraints from FrameNet.

- Solve argument identification using dual decomposition (specifically Alternating Directions Dual Decomposition - AD³ (Martins et al. (2011b))):
  - Solves relaxed problem
  - Often tight, but can find exact solution of original problem using branch-and-bound technique
    - In theory can be exponential, in practice always converges quickly
Argument identification with constraints - results

<table>
<thead>
<tr>
<th>Method</th>
<th>P</th>
<th>R</th>
<th>F₁</th>
<th>Violations</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>naive</td>
<td>82.00</td>
<td>76.36</td>
<td>79.08</td>
<td>441 45 15</td>
<td>1.26 ± 0.01</td>
</tr>
<tr>
<td>beam = 2</td>
<td>83.68</td>
<td>76.22</td>
<td>79.78</td>
<td>0  49 0</td>
<td>2.74 ± 0.10</td>
</tr>
<tr>
<td>beam = 100</td>
<td>83.83</td>
<td>76.28</td>
<td>79.88</td>
<td>0  50 1</td>
<td>29.00 ± 0.25</td>
</tr>
<tr>
<td>beam = 10000</td>
<td>83.83</td>
<td>76.28</td>
<td>79.88</td>
<td>0  50 1</td>
<td>440.67 ± 5.53</td>
</tr>
<tr>
<td>CPLEX, LP</td>
<td>83.80</td>
<td>76.16</td>
<td>79.80</td>
<td>0  1 0</td>
<td>32.67 ± 1.29</td>
</tr>
<tr>
<td>CPLEX, exact</td>
<td>83.78</td>
<td>76.17</td>
<td>79.79</td>
<td>0  0 0</td>
<td>43.12 ± 1.26</td>
</tr>
<tr>
<td>AD³, LP</td>
<td>83.77</td>
<td>76.17</td>
<td>79.79</td>
<td>2  2 0</td>
<td>4.17 ± 0.01</td>
</tr>
<tr>
<td>AD³, exact</td>
<td>83.78</td>
<td>76.17</td>
<td>79.79</td>
<td>0  0 0</td>
<td>4.78 ± 0.04</td>
</tr>
</tbody>
</table>

Table 9: Comparison of decoding strategies in Section 7.3 on the dataset released with the FrameNet 1.5 Release, given gold frames. We evaluate in terms of precision, recall and F₁ score on our test set containing 4,458 targets. We also compute the number of constraint violations each model makes: the three values are the numbers of overlapping arguments and violations of the “requires” and “excludes” constraints of Section 7.1. Finally, decoding time (without feature computation steps) on the whole test set is shown in the last column averaged over 5 runs.