Effective Slot Filling Based on Shallow Distant Supervision Methods

Benjamin Roth, Tassilo Barth, Michael Wiegand, Mittul Singh, Dietrich Klakow

Spoken Language Systems (LSV), Saarland University

November 18, 2013
1. Task and System Overview
2. Candidate Generation
3. Candidate Validation
   - Distant Supervision SVM’s
   - Distant Supervision Patterns
4. Per-Component Analysis
5. Conclusion
TAC KBP English Slot Filling

Queries

...<query id="ID_002">
    <name>Marc Bolland</name>
    <enttype>PER</enttype>
</query>
...

...<query id="ID_100">
    <name>Galleon Group</name>
    <enttype>ORG</enttype>
</query>
...

Corpus

... Marc Bolland, 50, former CEO of Morrison Supermarkets PLC, is joining M&S ...

System Response

...ID_002 per:age 50
ID_002 per:title CEO
ID_002 per:employee_or_member_of Morrison Supermarkets PLC
ID_002 per:employee_or_member_of M&S
...

Modular and easily extensible distant supervision relation extractor
Using shallow textual representations and features
Based on LSV 2012 system [Roth et al., 2012]
- same training data
- same architecture
- improved algorithms & context modeling
Data Flow

The data flow process starts with a query. This query initiates a process that involves expansion and document retrieval. The expanded query results in a set of candidate documents. These candidates are then validated and used to generate further information such as alternative names, manual patterns, and distsup patterns. The final output is a post-processed response.
Outline

1. Task and System Overview
2. Candidate Generation
3. Candidate Validation
   - Distant Supervision SVM’s
   - Distant Supervision Patterns
4. Per-Component Analysis
5. Conclusion
Candidate Generation

- Entity expansion based on Wikipedia anchor text language models
  - Query: “Badr Organization”
  - Also used for removing redundant answers (postprocessing)
Candidate Generation

- Entity expansion based on Wikipedia anchor text language models
  - Query: “Badr Organization”
  - Also used for removing redundant answers (postprocessing)

- Document retrieval
  - Lucene index
  - Selection of expansion terms based on point-wise mutual information
Candidate Generation

- Entity expansion based on Wikipedia anchor text language models
  - Query: “Badr Organization”
  - Also used for removing redundant answers (postprocessing)

- Document retrieval
  - Lucene index
  - Selection of expansion terms based on point-wise mutual information

- Candidate matching
  - NE Tagger [Chrupala and Klakow, 2010]
  - NE types from Freebase: CAUSE-OF-DEATH, JOB-TITLE, CRIMINAL-CHARGES, RELIGION
Outline

1. Task and System Overview
2. Candidate Generation
3. Candidate Validation
   - Distant Supervision SVM’s
   - Distant Supervision Patterns
4. Per-Component Analysis
5. Conclusion
Candidate Validation Modules

- Distant Supervision SVM Classifiers
- Distant Supervision Patterns
- Manual Patterns
- Alternate Names from Query Expansion
Candidate Validation Modules

- Distant Supervision SVM Classifiers
- Distant Supervision Patterns
- Manual Patterns
- Alternate Names from Query Expansion
Distant Supervision

Knowledge Base

<table>
<thead>
<tr>
<th>per:city_of_birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>(B. Obama, Honolulu)</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>(M. Jackson, Gary)</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>
Distant Supervision

![Diagram showing the process of distant supervision.]

- **Knowledge Base**
  - `per:city_of_birth`
  - (B. Obama, Honolulu)
  - ...
  - (M. Jackson, Gary)
  - ...

- **Corpus**
  - B. Obama was born in Honolulu
  - B. Obama moved from Honolulu
  - Gary, M. Jackson's birthplace

- **Training Data**
  - B. Obama was born in **Honolulu**
  - B. Obama moved from **Honolulu**
  - Gary, **M. Jackson**'s birthplace
Distant Supervision

Knowledge Base

<table>
<thead>
<tr>
<th>per:city_of_birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>(B. Obama, Honolulu)</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>(M. Jackson, Gary)</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>

Corpus

B. Obama was born in Honolulu
B. Obama moved from Honolulu
Gary, M. Jackson's birthplace

Training Data

B. Obama was born in Honolulu
B. Obama moved from Honolulu
Gary, M. Jackson's birthplace

Classifier
Distant Supervision

Knowledge Base

per:city_of_birth

(B. Obama, Honolulu)

(M. Jackson, Gary)

(N. Chomsky, Philadelphia)

Corpus

Training Data

B. Obama was born in Honolulu
B. Obama moved from Honolulu
Gary, M. Jackson’s birthplace

Corpus

Instance Candidates

F. Hollande visited Berlin
Born in Philadelphia, N. Chomsky ...
Distant Supervision (DS) SVM Classifiers

- “Workhorse” for candidate validation.
Distant Supervision (DS) SVM Classifiers

- "Workhorse" for candidate validation.
- Argument pairs for training data
  - Freebase
  - Pattern matches

Minimalistic feature set
- n-grams between relation arguments
- n-grams outside relation arguments
- Sparse (or skip) n-grams
- Marking of argument order for every feature

Training scheme:
- Aggregate training
- Global parameter tuning
Distant Supervision (DS) SVM Classifiers

- “Workhorse” for candidate validation.
- Argument pairs for training data
  - Freebase
  - Pattern matches
- Minimalistic feature set
  - n-grams between relation arguments
  - n-grams outside relation arguments
  - *sparse (or skip)* n-grams
  - marking of argument order for every feature

Training scheme:
- aggregate training
- global parameter tuning
Distant Supervision (DS) SVM Classifiers

- “Workhorse” for candidate validation.
- Argument pairs for training data
  - Freebase
  - Pattern matches
- Minimalistic feature set
  - n-grams between relation arguments
  - n-grams outside relation arguments
  - sparse (or skip) n-grams
  - marking of argument order for every feature
- Training scheme:
  - aggregate training
  - global parameter tuning
DS SVMs: Training

- One binary SVM per relation
DS SVMs: Training

- One binary SVM per relation
- *Aggregate training*
  - Training sentences are aggregated per argument pair
  - Feature weights averaged
  - Better generalization than single-sentence training
DS SVMs: Training

- One binary SVM per relation
- **Aggregate training**
  - Training sentences are aggregated per argument pair
  - Feature weights averaged
  - Better generalization than single-sentence training
- **Parameter tuning**
  - Misclassification cost tuning is essential
  - Optimizing per-relation cost parameter does not lead to global optimum
  ⇒ Greedy parameter tuning algorithm for global $F_1$ optimization
Distant Supervision Patterns

- Surface patterns from DS data
  - with “goodness” scores
  - org:alternate_names
    0.9784 [ARG1], abbreviated [ARG2]
    0.4023 [ARG2] is the core division of [ARG1]
Distant Supervision Patterns

- Surface patterns from DS data
  - with “goodness” scores
  - org:alternate_names
    0.9784 [ARG1], abbreviated [ARG2]
    0.4023 [ARG2] is the core division of [ARG1]
- Combination of DS noise reduction models [Roth and Klakow, 2013]
  - discriminative \( \text{at-least-one} \) perceptron model:
    \[
P(relation|pattern, \theta)
    \]
  - generative hierarchical topic model:
    \[
n(pattern, topic(relation))
    \]
  - relative frequency of pattern:
    \[
    \frac{n(pattern, relation)}{n(pattern)}
    \]
Outline

1. Task and System Overview

2. Candidate Generation

3. Candidate Validation
   - Distant Supervision SVM’s
   - Distant Supervision Patterns

4. Per-Component Analysis

5. Conclusion
# Effect of Removing Single Components (one at a time)

<table>
<thead>
<tr>
<th>Component</th>
<th>P</th>
<th>R</th>
<th>F1</th>
<th>F1 gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSV main run</td>
<td>42.5</td>
<td>33.2</td>
<td>37.3</td>
<td></td>
</tr>
<tr>
<td>Query expansion</td>
<td>41.1</td>
<td>17.5</td>
<td>24.5</td>
<td>+12.8</td>
</tr>
<tr>
<td>Distsup SVM classifier</td>
<td>53.3</td>
<td>21.8</td>
<td>30.9</td>
<td>+6.4</td>
</tr>
<tr>
<td>Distsup patterns</td>
<td>39.6</td>
<td>28.6</td>
<td>33.2</td>
<td>+4.1</td>
</tr>
<tr>
<td>Manual patterns</td>
<td>38.2</td>
<td>29.5</td>
<td>33.2</td>
<td>+4.1</td>
</tr>
<tr>
<td>Alternate names</td>
<td>41.1</td>
<td>31.0</td>
<td>35.4</td>
<td>+1.9</td>
</tr>
</tbody>
</table>
## Effect of Removing Single Components (one at a time)

<table>
<thead>
<tr>
<th>Component</th>
<th>P</th>
<th>R</th>
<th>F1</th>
<th>F1 gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSV main run</td>
<td>42.5</td>
<td>33.2</td>
<td>37.3</td>
<td></td>
</tr>
<tr>
<td>Query expansion</td>
<td>41.1</td>
<td>17.5</td>
<td>24.5</td>
<td>+12.8</td>
</tr>
<tr>
<td>Distsup SVM classifier</td>
<td>53.3</td>
<td>21.8</td>
<td>30.9</td>
<td>+6.4</td>
</tr>
<tr>
<td>Distsup patterns</td>
<td>39.6</td>
<td>28.6</td>
<td>33.2</td>
<td>+4.1</td>
</tr>
<tr>
<td>Manual patterns</td>
<td>38.2</td>
<td>29.5</td>
<td>33.2</td>
<td>+4.1</td>
</tr>
<tr>
<td>Alternate names</td>
<td>41.1</td>
<td>31.0</td>
<td>35.4</td>
<td>+1.9</td>
</tr>
</tbody>
</table>
Effect of Removing Single Components (one at a time)

<table>
<thead>
<tr>
<th>Component</th>
<th>P</th>
<th>R</th>
<th>F1</th>
<th>F1 gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSV main run</td>
<td>42.5</td>
<td>33.2</td>
<td>37.3</td>
<td></td>
</tr>
<tr>
<td>Query expansion</td>
<td>41.1</td>
<td>17.5</td>
<td>24.5</td>
<td>+12.8</td>
</tr>
<tr>
<td>Distsup SVM classifier</td>
<td>53.3</td>
<td>21.8</td>
<td>30.9</td>
<td>+6.4</td>
</tr>
<tr>
<td>Distsup patterns</td>
<td>39.6</td>
<td>28.6</td>
<td>33.2</td>
<td>+4.1</td>
</tr>
<tr>
<td>Manual patterns</td>
<td>38.2</td>
<td>29.5</td>
<td>33.2</td>
<td>+4.1</td>
</tr>
<tr>
<td>Alternate names</td>
<td>41.1</td>
<td>31.0</td>
<td>35.4</td>
<td>+1.9</td>
</tr>
</tbody>
</table>
### Per-Component Analysis

#### Effect of Removing Single Components (one at a time)

<table>
<thead>
<tr>
<th>Component</th>
<th>P</th>
<th>R</th>
<th>F1</th>
<th>F1 gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSV main run</td>
<td>42.5</td>
<td>33.2</td>
<td>37.3</td>
<td></td>
</tr>
<tr>
<td>- Query expansion</td>
<td>41.1</td>
<td>17.5</td>
<td>24.5</td>
<td>+12.8</td>
</tr>
<tr>
<td>- Distsup SVM classifier</td>
<td>53.3</td>
<td>21.8</td>
<td>30.9</td>
<td>+6.4</td>
</tr>
<tr>
<td>- Distsup patterns</td>
<td>39.6</td>
<td>28.6</td>
<td>33.2</td>
<td>+4.1</td>
</tr>
<tr>
<td>- Manual patterns</td>
<td>38.2</td>
<td>29.5</td>
<td>33.2</td>
<td>+4.1</td>
</tr>
<tr>
<td>- Alternate names</td>
<td>41.1</td>
<td>31.0</td>
<td>35.4</td>
<td>+1.9</td>
</tr>
</tbody>
</table>
## Effect of Removing Single Components (one at a time)

<table>
<thead>
<tr>
<th>Component</th>
<th>P</th>
<th>R</th>
<th>F1</th>
<th>F1 gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSV main run</td>
<td>42.5</td>
<td>33.2</td>
<td>37.3</td>
<td></td>
</tr>
<tr>
<td>– Query expansion</td>
<td>41.1</td>
<td>17.5</td>
<td>24.5</td>
<td>+12.8</td>
</tr>
<tr>
<td>– Distsup SVM classifier</td>
<td>53.3</td>
<td>21.8</td>
<td>30.9</td>
<td>+6.4</td>
</tr>
<tr>
<td>– Distsup patterns</td>
<td>39.6</td>
<td>28.6</td>
<td>33.2</td>
<td>+4.1</td>
</tr>
<tr>
<td>– Manual patterns</td>
<td>38.2</td>
<td>29.5</td>
<td>33.2</td>
<td>+4.1</td>
</tr>
<tr>
<td>– Alternate names</td>
<td>41.1</td>
<td>31.0</td>
<td>35.4</td>
<td>+1.9</td>
</tr>
</tbody>
</table>
### Effect of Removing Single Components (one at a time)

<table>
<thead>
<tr>
<th>Component</th>
<th>P</th>
<th>R</th>
<th>F1</th>
<th>F1 gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSV main run</td>
<td>42.5</td>
<td>33.2</td>
<td>37.3</td>
<td></td>
</tr>
<tr>
<td>Query expansion</td>
<td>41.1</td>
<td>17.5</td>
<td>24.5</td>
<td>+12.8</td>
</tr>
<tr>
<td>Distsup SVM classifier</td>
<td>53.3</td>
<td>21.8</td>
<td>30.9</td>
<td>+6.4</td>
</tr>
<tr>
<td>Distsup patterns</td>
<td>39.6</td>
<td>28.6</td>
<td>33.2</td>
<td>+4.1</td>
</tr>
<tr>
<td>Manual patterns</td>
<td>38.2</td>
<td>29.5</td>
<td>33.2</td>
<td>+4.1</td>
</tr>
<tr>
<td>Alternate names</td>
<td>41.1</td>
<td>31.0</td>
<td>35.4</td>
<td>+1.9</td>
</tr>
</tbody>
</table>
## Single Component Performance

<table>
<thead>
<tr>
<th>Component</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distsup SVM classifier</td>
<td>34.7</td>
<td>23.6</td>
<td>28.1</td>
</tr>
<tr>
<td>Distsup Patterns</td>
<td>42.7</td>
<td>15.6</td>
<td>22.9</td>
</tr>
<tr>
<td>Manual patterns</td>
<td>50.2</td>
<td>10.3</td>
<td>17.1</td>
</tr>
<tr>
<td>Alternate names</td>
<td>54.2</td>
<td>1.8</td>
<td>3.4</td>
</tr>
</tbody>
</table>
Bottleneck: Candidate Generation

- Lost recall on candidate level cannot be undone by validation modules.
- Query and argument matching is of crucial importance.
- Recall analysis (on 2012 queries):
  - good recall on document level
  - big potential on candidate sentence extraction

<table>
<thead>
<tr>
<th>Query expansion</th>
<th>document recall</th>
<th>candidate recall</th>
<th>end-to-end F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>90.2</td>
<td>58.8</td>
<td>32.1</td>
</tr>
<tr>
<td>no</td>
<td>87.7</td>
<td>34.4</td>
<td>23.0</td>
</tr>
</tbody>
</table>
Outline

1. Task and System Overview
2. Candidate Generation
3. Candidate Validation
   - Distant Supervision SVM’s
   - Distant Supervision Patterns
4. Per-Component Analysis
5. Conclusion
Conclusion

- TAC KBP English Slot-Filling
Conclusion

- TAC KBP English Slot-Filling
  - Query-driven relation extraction

More details and analysis in our workshop paper!
Conclusion

- TAC KBP English Slot-Filling
  - Query-driven relation extraction
  - Subtasks
Conclusion

- TAC KBP English Slot-Filling
  - Query-driven relation extraction
  - Subtasks
    - Candidate retrieval and matching

More details and analysis in our workshop paper!
Conclusion

- TAC KBP English Slot-Filling
  - Query-driven relation extraction
  - Subtasks
    - Candidate retrieval and matching
    - Relation modeling
Conclusion

- TAC KBP English Slot-Filling
  - Query-driven relation extraction
  - Subtasks
    - Candidate retrieval and matching
    - Relation modeling
- LSV system

More details and analysis in our workshop paper!
Conclusion

- **TAC KBP English Slot-Filling**
  - Query-driven relation extraction
  - Subtasks
    - Candidate retrieval and matching
    - Relation modeling

- **LSV system**
  - Modular, shallow approach

More details and analysis in our workshop paper!
Conclusion

- TAC KBP English Slot-Filling
  - Query-driven relation extraction
  - Subtasks
    - Candidate retrieval and matching
    - Relation modeling

- LSV system
  - Modular, shallow approach
  - Query expansion: Wikipedia anchor text

More details and analysis in our workshop paper!
Conclusion

- TAC KBP English Slot-Filling
  - Query-driven relation extraction
  - Subtasks
    - Candidate retrieval and matching
    - Relation modeling

- LSV system
  - Modular, shallow approach
  - Query expansion: Wikipedia anchor text
  - N-gram-based distant supervision SVMs

More details and analysis in our workshop paper!
Conclusion

- **TAC KBP English Slot-Filling**
  - Query-driven relation extraction
  - Subtasks
    - Candidate retrieval and matching
    - Relation modeling

- **LSV system**
  - Modular, shallow approach
  - Query expansion: Wikipedia anchor text
  - N-gram-based distant supervision SVMs
  - Scored distant supervision surface patterns

More details and analysis in our workshop paper!
Conclusion

- **TAC KBP English Slot-Filling**
  - Query-driven relation extraction
  - Subtasks
    - Candidate retrieval and matching
    - Relation modeling

- **LSV system**
  - Modular, shallow approach
  - Query expansion: Wikipedia anchor text
  - N-gram-based distant supervision SVMs
  - Scored distant supervision surface patterns

- More details and analysis in our workshop paper!