Multilingual Relation Extraction using Compositional Universal Schema

Haw-Shiuan Chang, Abdurrahman Munir, Ao Liu, Johnny Tian-Zheng Wei, Aaron Traylor, Ajay Nagesh, Nicholas Monath, Patrick Verga, Emma Strubell, Andrew McCallum

UMassAmherst

(Team ID: UMass_IESL)

These slides are mostly prepared by Patrick Verga
Wei Li studies at Xinghua U. Her 2008 publications include W. Li. “Scalable NLP” ACL, 2008.
Knowledge Base (KB) Construction

Wei Li studies at Xinghua U. Her 2008 publications include W. Li. “Scalable NLP” ACL, 2008.
Wei Li studies at Xinghua U. Her 2008 publications include W. Li. “Scalable NLP” ACL, 2008.
Wei Li studies at Xinghua U. Her 2008 publications include W. Li. "Scalable NLP" ACL, 2008.
Compositional Universal Schema

January 15, 2000

Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife Melinda French Gates. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.
January 15, 2000

Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife Melinda. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.
January 15, 2000

Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife Melinda French Gates. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.
January 15, 2000

Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife Melinda French Gates. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.
January 15, 2000

Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife Melinda French Gates. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.
January 15, 2000

Tech pioneer Bill Gates stepped down today as chief executive officer of Microsoft, the Seattle-headquartered software giant. His long-time friend, Steve Balmer, will take over as CEO of Microsoft. Gates will now focus on the charitable foundation he runs with his wife Melinda French Gates. Bill and Melinda were married in a ceremony in Hawaii, rather than her hometown of Dallas. Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.
Gates moved his family into their 55,000-square-foot, $54 million house on the shore of Lake Washington, just outside Seattle.
Joe NewEntity worked with his wife Jane NewEntity on relation extraction.
Joe NewEntity worked with his wife Jane NewEntity on relation extraction.

query: **spouse**
Joe NewEntity worked with his wife Jane NewEntity on relation extraction.
Spanish
Relation Extraction
Multilingual Relation Extraction

Steve, Microsoft
Bill, Steve
Bill, Microsoft
Melinda, Bill
Microsoft, Seattle
Melinda, Dallas
Obama, US
Obama, Michelle

Bill, Microsoft
Melinda, Dallas
Obama, Michelle
Bill, Seattle

English

Spanish
Multilingual Universal Schema

Relation Types

structured
English textual
Spanish textual

Steve, Microsoft
Bill, Steve
Bill, Microsoft
Melinda, Bill
Microsoft, Seattle
Melinda, Dallas
Obama, US
Obama, Michelle
Bill, Seattle
null
Entity pair

Relation

spouse

Melinda, Bill

married to

Michelle, Obama

esposa
Search Engine Supervision
Wei Li studies at Xinghua U. Her 2008 publications include W. Li. "Scalable NLP" ACL, 2008.

Search Engine as a Resource

Text docs

Entity Extraction

Entity Mentions

Wei Li (PER)
W. Li (PER)
Xinghua U. (ORG)

Resolution (Coref)

Entitites, Relations

Wei Li
W. Li
Xinghua U.

Relation Extraction

Relation Mentions

Member(Wei Li, Xinghua U.)

KB

query

answer

Google

Freebase

RelationFactory
Distant Supervision

Key phrases for per:spouse
- wife
- and
- president
- lives with
- understands
- married to
Limitations of Distant Supervision

Barack Obama \text{Entity 1} \text{Entity 2} Michelle Obama

- wife \rightarrow \text{High frequency signal}
- and \rightarrow \text{Global background noise} \rightarrow \text{Not expressing a relation}
- president \rightarrow \text{Entity background noise} \rightarrow \text{Freebase selection bias}
- lives with \rightarrow \text{High correlation noise} \rightarrow \text{Only valid in some entity pairs}
- understands \rightarrow \text{Other relation noise} \rightarrow \text{They might have other relations}
- married to \rightarrow \text{Low frequency signal}
Asking in another direction?

Barack Obama ... wife ... ______
Barack Obama ... and ... ______

P( ___ = Michelle Obama)

Hard to accurately estimate this when corpus is not large enough
Distant supervision (after applying several noises removal techniques)

After applying search engine supervision

Lower their influence on Universal Schema
Results
### TAC 2016 Component Testing

The recall is low because we don’t include Chinese query and Chinese corpus.

<table>
<thead>
<tr>
<th>KB, XLING</th>
<th>Hop0</th>
<th></th>
<th>Hop1</th>
<th></th>
<th>Total</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec</td>
<td>Recall</td>
<td>F1</td>
<td>Prec</td>
<td>Recall</td>
<td>F1</td>
</tr>
<tr>
<td>USchema</td>
<td>0.475</td>
<td>0.015</td>
<td>0.028</td>
<td>0.071</td>
<td>0.003</td>
<td>0.005</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.421</td>
<td>0.038</td>
<td>0.069</td>
<td>0.106</td>
<td>0.018</td>
<td>0.030</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SF, ENG</th>
<th>Hop0</th>
<th></th>
<th>Hop1</th>
<th></th>
<th>Total</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec</td>
<td>Recall</td>
<td>F1</td>
<td>Prec</td>
<td>Recall</td>
<td>F1</td>
</tr>
<tr>
<td>DS</td>
<td><strong>0.244</strong></td>
<td>0.168</td>
<td>0.199</td>
<td><strong>0.500</strong></td>
<td><strong>0.013</strong></td>
<td><strong>0.025</strong></td>
</tr>
<tr>
<td>SES+DS</td>
<td>0.229</td>
<td><strong>0.203</strong></td>
<td><strong>0.215</strong></td>
<td>0.444</td>
<td>0.013</td>
<td>0.025</td>
</tr>
</tbody>
</table>
# TAC 2016 Results

## LDC Max All Micro Scores

<table>
<thead>
<tr>
<th>SF</th>
<th>Hop0 (ENG: 350, SPA: 150)</th>
<th>Hop1 (ENG: 185, SPA: 53)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec</td>
<td>Recall</td>
<td>F1</td>
</tr>
<tr>
<td>2015 system</td>
<td>0.302</td>
<td>0.154</td>
<td>0.204</td>
</tr>
<tr>
<td>2016 system</td>
<td>0.229</td>
<td>0.203</td>
<td>0.215</td>
</tr>
<tr>
<td>Spanish</td>
<td>0.1364</td>
<td>0.285</td>
<td>0.187</td>
</tr>
</tbody>
</table>

## 2015 KB evaluation

<table>
<thead>
<tr>
<th>2015 system</th>
<th>Prec</th>
<th>Recall</th>
<th>F1</th>
<th>Prec</th>
<th>Recall</th>
<th>F1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015 system</td>
<td>0.227</td>
<td>0.149</td>
<td>0.192</td>
<td>0.038</td>
<td>0.097</td>
<td>0.054</td>
<td>0.120</td>
</tr>
</tbody>
</table>

## 2016 KB evaluation

<table>
<thead>
<tr>
<th>2015 system</th>
<th>Hop0</th>
<th>Hop1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prec</td>
<td>Recall</td>
<td>F1</td>
<td>Prec</td>
</tr>
<tr>
<td>2015 system</td>
<td>0.390</td>
<td>0.128</td>
<td>0.192</td>
</tr>
<tr>
<td>2016 system</td>
<td>0.271</td>
<td>0.154</td>
<td>0.196</td>
</tr>
<tr>
<td>Spanish</td>
<td>0.280</td>
<td>0.121</td>
<td>0.169</td>
</tr>
</tbody>
</table>
Result Summary

• Compositional Universal Schema (LSTM) improves recall significantly.

• Search engine supervision cleans training data well.

• Universal Schema can be effectively applied to the language with low resources.

• Our 2015 system is scored much better in this year.
  • Need further investigation.
Q & A