AIDA-Social: Entity Linking on the Social Stream

Yusra Ibrahim, Mohamed Amir Yosef, Gerhard Weikum
{yibrahim|mamir|weikum}@mpi-inf.mpg.de

Motivation

- The tremendous increase in social media popularity drove an abundance of user-generated content.
- Adding semantics to this content assists in subsequent Information Retrieval tasks such as relation extraction and semantic search.
- Named Entity Linking (NEL) disambiguates names to their corresponding canonical entities in Knowledge Bases such as YAGO.
- NEL in social media, more specifically in microblogs, is a challenging task due to the brevity, lack of contextual information, and time-varying importance of entities.
- Twitter is the most popular microblog with 500 million Tweets per day.

Challenges

- Short, no context
- Cryptic Mentions
- Time-Varying Entity Importance
- Such an unbelievable loss. RIP [Amy]. xxx we love you. Xx

Proposed Techniques

- Mention Normalization
- Temporal Importance
- Contextual Enrichment

Experimental Results

<table>
<thead>
<tr>
<th>Microposts2014 annotated corpus</th>
<th>Mention Normalization</th>
<th>Temporal Importance</th>
<th>Contextual Enrichment</th>
<th>P@1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweets 2809</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>58.41%</td>
</tr>
<tr>
<td>Mentions 3392</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>61.07%</td>
</tr>
<tr>
<td>Unique Mentions 2019</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>67.35%</td>
</tr>
<tr>
<td>@ 506</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>62.08%</td>
</tr>
<tr>
<td># 1085</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>69.05%</td>
</tr>
<tr>
<td>Entities 1683</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>66.00%</td>
</tr>
<tr>
<td>Baseline: AIDA - Prior + Contextual Similarity</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>69.07%</td>
</tr>
</tbody>
</table>

Conclusion and Future Work

- Temporal Entity Importance, Contextual Enrichment, and Mention Normalization yields +13% gain over P@1.
- Adding extra context assists in mention-entity similarity measures.
- Temporal Importance improves the accuracy of NEL, especially on plain tweets.
- Adding context from URLs yields better gain than adding context from #Tags and @userMentions.
- Clustering assists in coherence measures.
- TODO: Other Datasets.
- TODO: Other ways to estimate Temporal Importance.
- TODO: More on Clustering.