UNED Online Reputation Monitoring Team at RepLab 2013

Damiano Spina, Jorge Carrillo-de-Albornoz, Tamara Martín, Enrique Amigó, Julio Gonzalez and Fernando Giner
(damiano, jcalbornoz, tmartin, enrique, julio)@lsi.uned.es, fginer3@alumno.uned.es

RepLab at CLEF 2013
Valencia, Spain. September 29-26, 2013

Monitoring Task

- Input: entity of interest + set of tweets + representative URLs.
- Example: Apple Inc. -> tweets containing "Apple" and www.apple.com
- Filtering: Binary classification of tweets (related/unrelated)
  - Polarity for Reputation: classified each tweet according to its polarity for reputation (positive/negative/neutral)
  - Topic Detection: Group tweets by topics
  - Topic Priority: Rank topics, reputation alerts go first
- Output: Monitoring summary (ranking of topics) for the reputation manager

Filtering Subtask

- Two-step classification algorithm
  - Step 1: Automatic Keyword Discovery
    - Each term is classified as positive keyword / negative keyword / other
  - Step 2: Automatic Tweet Classification
    - Tweets containing keywords are used to feed a binary Bkt classifier that classifies the remaining tweets as related/unrelated

Instance-based learning + Heterogeneity-Based Ranking (HBR)

- Similar to the RepLab 2013 official baseline
- Each tweet in the test set is labeled as the most similar tweet in the training set
- Combination of rankings given by multiple text similarity measures
- Applicable to all the subtasks (Topic Detection, Polarity, Priority,...)

Semantic Graphs for Domain-specific Affective Lexicon Adaptation

- SentiSense
  - Affective Lexicon of 5,496 words and 2,190 onsets from WordNet labeled with emotional categories
- Domain-specific Lexicon Adaptation
  - For each domain, WordNet concepts are extracted from the training data. The graph is generated upon semantic relations between concepts. Emotional categories are propagated using SentiSense as seed.
- Polarity Classification
  - Tweets represented as a Vector of Emotional Intensities (VEI) feed a Machine learning classifier.

Polarity for Reputation Subtask

- Approach
  - Accuracy
  - Reliability
  - Sensitivity
  - F(R,S)
  - Run (out of 50 runs)

<table>
<thead>
<tr>
<th>Approach</th>
<th>Accuracy</th>
<th>Reliability</th>
<th>Sensitivity</th>
<th>F(R,S)</th>
<th>Run (out of 50 runs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RepLab 2013 Best System</td>
<td>0.91</td>
<td>0.73</td>
<td>0.45</td>
<td>0.49</td>
<td>1</td>
</tr>
<tr>
<td>Filter Keywords (Tweet Classification Step)</td>
<td>0.06</td>
<td>0.43</td>
<td>0.38</td>
<td>0.34</td>
<td>19</td>
</tr>
<tr>
<td>RepLab 2013 Official Baseline</td>
<td>0.87</td>
<td>0.49</td>
<td>0.32</td>
<td>0.33</td>
<td>21</td>
</tr>
<tr>
<td>Instance-based Learning + HBR</td>
<td>0.07</td>
<td>0.47</td>
<td>0.33</td>
<td>0.30</td>
<td>27</td>
</tr>
<tr>
<td>Filter Keywords (training: same entity)</td>
<td>0.04</td>
<td>0.67</td>
<td>0.26</td>
<td>0.25</td>
<td>42</td>
</tr>
<tr>
<td>Filter Keywords (training: other entities)</td>
<td>0.50</td>
<td>0.17</td>
<td>0.29</td>
<td>0.14</td>
<td>61</td>
</tr>
</tbody>
</table>

LDA-based Clustering

- Based on Twitter LDA and Topics over time models
- Transfer learning: target tweets + background tweets to establish the right number of clusters

Term Clustering

- Step 1: Term Clustering
  - Learned similarity function (content-based, meta-data, time-aware features)
  - Hierarchical Agglomerative Clustering
- Step 2: Tweet clustering
  - Assign tweets according to maximal term overlap (highest accord similarity)

Wikified Tweet Clustering

- Representation: Tweets are linked to Wikipedia pages/entities
- Clustering: second centroid over Wikipedia entities

Topic Detection Subtask

- Approach
  - Reliability
  - Sensitivity
  - F(R,S)
  - Accuracy
  - Run (out of 64 runs)

<table>
<thead>
<tr>
<th>Approach</th>
<th>Reliability</th>
<th>Sensitivity</th>
<th>F(R,S)</th>
<th>Accuracy</th>
<th>Run (out of 64 runs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RepLab 2013 Best System</td>
<td>0.46</td>
<td>0.32</td>
<td>0.30</td>
<td>0.30</td>
<td>1</td>
</tr>
<tr>
<td>SentiSense (training: same entity)</td>
<td>0.36</td>
<td>0.10</td>
<td>0.15</td>
<td>0.62</td>
<td>21</td>
</tr>
<tr>
<td>Domain-specific Adaptation (training: same entity)</td>
<td>0.33</td>
<td>0.11</td>
<td>0.14</td>
<td>0.62</td>
<td>22</td>
</tr>
<tr>
<td>Instance-based Learning + HBR</td>
<td>0.32</td>
<td>0.29</td>
<td>0.30</td>
<td>0.59</td>
<td>26</td>
</tr>
<tr>
<td>RepLab 2013 Official Baseline</td>
<td>0.32</td>
<td>0.29</td>
<td>0.30</td>
<td>0.58</td>
<td>28</td>
</tr>
<tr>
<td>SentiSense + Domain-specific Adaptation (training: same entity, balanced)</td>
<td>0.34</td>
<td>0.12</td>
<td>0.16</td>
<td>0.58</td>
<td>31</td>
</tr>
</tbody>
</table>

Conclusions

- Full Task. Large room for improvement. Filtering is crucial for the overall performance of a monitoring system.
- Filtering. Use entity-specific training data when available: +78% F(R,S), +68% accuracy for Filter Keywords.
- Polarity for Reputation. Different from traditional sentiment analysis. Domain-adaptive affective lexicons less competitive than other RepLab submissions.
- Topic Detection. Three approaches perform competitively w.r.t. other RepLab submissions.
- Topic Priority (future work). Challenging due to the difficulty of combining dissimilar and imperfect signals (computed automatically): polarity, novelty, centrality, etc.