Global Thread-Level Inference for Comment Classification in Community Question Answering

Shafiq Joty, Alberto Barrón-Cedeño, Giovanni Da San Martino, Simone Filice, Lluís Márquez, Alessandro Moschitti and Preslav Nakov
Qatar Computing Research Institute, HBKU

3. Our Solution

- A binary classifier is trained to decide whether a pair of comments in a thread should be in the same class or not.
- The pairwise and the local good-vs-bad classification probabilities are then used in (a) min-cut or (b) ILP models for global inference.

a. Graph Partition

- Find a partition \( P = \{G, B\} \) minimizing
  \[
  C(P) = \lambda \left( \sum_{c_i \in G} s_{iB} + \sum_{c_i \in B} s_{iG} \right) + (1 - \lambda) \sum_{i,j} s_{ij}
  \]
  - \( s_{iG} \) and \( s_{iB} \) are individual scores;
  - \( s_{ij} \) are pairwise scores.

- Max-flow/min-cut: Gives exact solution in polynomial time

b. Linear Programming

- Find an assignment \( \mathbf{A} \) to all variables that minimizes
  \[
  C(\mathbf{A}) = \lambda \sum_{i=1}^{N} (c_{iG} \cdot x_{iG} + c_{iB} \cdot x_{iB}) + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (c_{ijS} \cdot x_{ijS} + c_{ijD} \cdot x_{ijD})
  \]
  - Subject to the constraints:
    - all variables are binary
    - only one label is assigned to each comment
    - the assignments to comments and comment-pairs are consistent

4. Experimental Setup and Results

- Dataset (Qatar Living)

<table>
<thead>
<tr>
<th>Category</th>
<th>Train</th>
<th>Dev</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td>2,600</td>
<td>300</td>
<td>329</td>
</tr>
<tr>
<td>Comments</td>
<td>16,541</td>
<td>1,645</td>
<td>1,976</td>
</tr>
<tr>
<td>Good</td>
<td>8,069</td>
<td>875</td>
<td>997</td>
</tr>
<tr>
<td>Bad</td>
<td>8,472</td>
<td>770</td>
<td>979</td>
</tr>
</tbody>
</table>

- Same vs. Different Classification

<table>
<thead>
<tr>
<th>Classifier</th>
<th>P</th>
<th>R</th>
<th>F1</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline: Same</td>
<td></td>
<td></td>
<td>69.26</td>
<td></td>
</tr>
<tr>
<td>MaxEnt-2C</td>
<td>73.95</td>
<td>90.99</td>
<td>81.59</td>
<td>71.56</td>
</tr>
<tr>
<td>MaxEnt-3C</td>
<td>77.15</td>
<td>80.42</td>
<td>78.75</td>
<td>69.94</td>
</tr>
</tbody>
</table>

- Same-vs.-different better than 3-way classifier
- Small improvement in accuracy over the baseline
- Yet, the classifier is helpful with graph-cut/ILP

Summary

We have shown that using thread-level information in a pairwise classifier + min-cut/ILP improves over the state of the art. Linear-chain CRF model is less helpful.

Future work

(i) Joint models, (ii) exploiting cross-thread information, (iii) use other CQA datasets

References: