1. Answer Selection

**Question**

GOOD* 

POTENTIAL* 

BAD* 

SemEval-2015 Task 3: Answer Selection in cQA

2. The Issues

**Good Scissor**

Dears,

Anyone have an idea where to find a good scissor for hair and beard trimming please??

**Visit Family food center**

Visit Family food center

**Al rawnaq airport road...U'll**

find all types of scissors there

**Thank you all**

Thanks for all... I will try that.

3. The Baseline

**Similarity Features**

- greedy string tiling
- longest common subseq.
- Jaccard coefficient
- word containment
- idf-weighted intersection
- cosine similarity
- PTK similarity between syntactic trees

**Heuristic Features**

The comment (c) includes...
- URL or email
- "yes", "sure", "no", @, etc.
- 3 or more repeated chars.
- longer than expected words
- thank* ack*
- c is posted by the same user who asked the question (uq)
- the category of the question
- length of the comment

4. The Alternatives

**Structured Prediction Models**

<table>
<thead>
<tr>
<th>SVM</th>
<th>SVM\textsuperscript{hmm}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>CRF</td>
</tr>
</tbody>
</table>

**Thread-Level Global Features**

- Identification of dialogues
- |comments by uq|
- position of the comment
- pseudo-ranking

5. The Outcome

**Comment- and Thread-Level Evaluation: Test | Cross-validation**

<table>
<thead>
<tr>
<th></th>
<th>F\textsubscript{1}</th>
<th>A</th>
<th>F\textsubscript{1,avg}</th>
<th>A\textsubscript{avg}</th>
<th>F\textsubscript{1}</th>
<th>A</th>
<th>F\textsubscript{1,avg}</th>
<th>A\textsubscript{avg}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Features</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>76.89</td>
<td>74.39</td>
<td>66.52</td>
<td>76.13</td>
<td>74.98 ±0.7</td>
<td>72.90 ±1.0</td>
<td>64.56 ±1.0</td>
<td>75.32 ±0.4</td>
</tr>
<tr>
<td>SVM\textsuperscript{hmm}</td>
<td>78.49</td>
<td>76.37</td>
<td>68.55</td>
<td>77.58</td>
<td>75.28 ±1.0</td>
<td>73.75 ±1.6</td>
<td>65.25 ±1.2</td>
<td>74.68 ±1.0</td>
</tr>
<tr>
<td>LogReg</td>
<td>75.96</td>
<td>70.85</td>
<td>68.84</td>
<td>74.79</td>
<td>74.42 ±0.8</td>
<td>69.99 ±0.9</td>
<td>66.00 ±1.3</td>
<td>73.04 ±1.0</td>
</tr>
<tr>
<td>CRF\textsubscript{aap}</td>
<td>78.94</td>
<td>76.67</td>
<td>67.17</td>
<td>76.55</td>
<td>74.42 ±1.3</td>
<td>72.66 ±1.7</td>
<td>63.90 ±1.7</td>
<td>73.51 ±0.7</td>
</tr>
<tr>
<td>CRF\textsubscript{mpm}</td>
<td>77.38</td>
<td>75.15</td>
<td>66.54</td>
<td>75.42</td>
<td>74.93 ±1.2</td>
<td>73.20 ±1.8</td>
<td>64.53 ±1.4</td>
<td>74.32 ±0.9</td>
</tr>
</tbody>
</table>

**Baseline+Thread-based Features**

| SVM | 79.96 | 78.44 | 67.65 | 76.02 | 77.59 ±0.9 | 76.23 ±1.4 | 66.41 ±1.3 | 76.23 ±0.4 |
| SVM\textsuperscript{hmm} | 78.82 | 77.43 | 66.61 | 77.06 | 77.21 ±1.2 | 76.20 ±1.8 | 65.33 ±1.1 | 76.43 ±0.9 |
| LogReg | 79.39 | 77.33 | 68.10 | 75.57 | 77.22 ±1.1 | 75.43 ±1.5 | 66.57 ±1.5 | 75.05 ±0.7 |
| CRF\textsubscript{aap} | 79.31 | 77.43 | 66.37 | 76.08 | 76.75 ±1.3 | 75.14 ±1.7 | 65.36 ±1.4 | 75.61 ±0.6 |
| CRF\textsubscript{mpm} | 79.51 | 77.78 | 67.36 | 76.63 | 77.00 ±1.3 | 75.43 ±1.8 | 65.57 ±1.5 | 75.71 ±0.7 |

6. Remarks

- Answer dependencies are helpful to improve in this task.
- The dependencies can be exploited with thread-level features, causing the context of a comment to affect quality estimation.
- If opting for an ad-hoc learning algorithm (e.g., CRF or SVM\textsuperscript{hmm}), the class of a comment depends on the class assigned to its neighbours as well.
- Robust context-aware features are the best way to push up the performance of the classifier.

Acknowledgements

This research is developed by the Arabic Language Technologies (ALT) group at the Qatar Computing Research Institute (QARI). It is part of the Interactive Systems for Answer Search (Iyas) project.