PAN@CLEF 2019
Style Change Detection Task

Eva Zangerle, Michael Tschuggnall, Martin Potthast, Benno Stein
Task Description

Author Identification

• Authorship Attribution
• Style Change Detection

<table>
<thead>
<tr>
<th>Author 1</th>
<th>Author 2</th>
<th>Author 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonumy eirmod tempor incididunt ut labore et dolore magna aliqua erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Sed ut diam nonumy eirmod tempor incididunt ut labore et dolore magna aliqua erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum.</td>
<td>Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonumy eirmod tempor incididunt ut labore et dolore magna aliqua erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Sed ut diam nonumy eirmod tempor incididunt ut labore et dolore magna aliqua erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum.</td>
<td>Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonumy eirmod tempor incididunt ut labore et dolore magna aliqua erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Sed ut diam nonumy eirmod tempor incididunt ut labore et dolore magna aliqua erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum.</td>
</tr>
</tbody>
</table>

expected answer: no, yes, yes
Task Description

Given a document, participants should apply intrinsic style analyses to hierarchically answer the following questions:

(a) Is the document written by one or more authors, i.e., do style changes exist or not?
(b) If it is multi-authored, how many authors have collaborated?
Dataset

• Realistic, non-artificial and comprehensive dataset

• Requirements
  • Find multiple authors that write about the same topic
  • Find texts that are freely available and of sufficient length
  • Multi-authored texts need to contain the same topic / subtopic

• Q&A platform StackExchange fulfills these requirements
Dataset

StackExchange consists of several sites (170+ sites), data freely available

Each question/answer associated with
• site
• set of tags (subtopics)

Example: photography site – subtopics: lens, canon, nikon, lightroom, photoshop, ...
Dataset

• Cleaning
  • Remove links
  • Remove images
  • Remove code snippets
  • Remove bullet lists
  • Remove block quotes
  • Remove very short questions/answers
  • Remove edited questions/answers
• Using the raw texts, a **training** (50%), **validation** (25%) and **test** (25%) dataset has been created
• Each dataset contains 50% single-author documents and 50% multi-authored documents
## Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Configuration Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of style changes</td>
<td>0-10</td>
</tr>
<tr>
<td>Number of collaborating authors</td>
<td>1-5</td>
</tr>
<tr>
<td>Document length</td>
<td>300-1500 tokens</td>
</tr>
<tr>
<td>Change positions</td>
<td>End of paragraph, within paragraph, mixed</td>
</tr>
<tr>
<td>Segment length distribution</td>
<td>100-1500</td>
</tr>
</tbody>
</table>
## Dataset

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Docs</th>
<th>Authors</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Training</td>
<td>2,546</td>
<td>1,273 (50%)</td>
<td>325 (13%)</td>
</tr>
<tr>
<td>Validation</td>
<td>1,272</td>
<td>636 (50%)</td>
<td>179 (14%)</td>
</tr>
<tr>
<td>Test</td>
<td>1,210</td>
<td>605 (50%)</td>
<td>147 (12%)</td>
</tr>
</tbody>
</table>
Evaluation

• Two subtasks, scored individually
  • Task a (binary classification): accuracy
  • Task b (classification on number of authors): Ordinal Classification Index (OCI)
• Overall score = $\frac{\text{accuracy} + (1 - \text{OCI})}{2}$
Approaches

5 registrations, 2 submissions to TIRA:

**Treshold-Based and Window-Merge Clustering Methods** (Sukanya Nath)
- both tasks tackled at same time
- two clustering approaches for windows
- clustering on pair-wise distance of windows - windows in same cluster are assumed to be written by same author

**Feed-forward Neural Networks** (Chaoyuan Zuo, Yu Zhao, Ritwik Banerjee)
- subtasks are dealt with individually
- binary classification utilizing multi-layer perceptron (single layer) on tf/idf word vector
- second task: features based on winning solution of 2018 (lexical features (POS, etc.), contracted word form, readability scores, ...), added tf/idf, three different clustering methods applied (k-means on tf-idf, hierarchical clustering on all features, MLP)
Baselines

Baseline-RND
  • „advanced“ guessing using text length statistics

Baseline-C99
  • Utilize C99 text segmentation algorithm (Choi, 2000)
  • Let the algorithm determine the number of segments
  • If #segments = 1: predict no style changes, otherwise predict changes
Results

<table>
<thead>
<tr>
<th>Participant</th>
<th>Accuracy</th>
<th>OCI</th>
<th>Rank</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nath</td>
<td>0.848</td>
<td>0.865</td>
<td>0.491</td>
<td>00:02:23</td>
</tr>
<tr>
<td>Zuo et al.</td>
<td>0.604</td>
<td>0.809</td>
<td>0.398</td>
<td>00:25:50</td>
</tr>
<tr>
<td>Baseline-RND</td>
<td>0.600</td>
<td>0.856</td>
<td>0.372</td>
<td>-</td>
</tr>
<tr>
<td>Baseline-C99</td>
<td>0.582</td>
<td>0.882</td>
<td>0.350</td>
<td>00:00:30</td>
</tr>
</tbody>
</table>
Results (#tokens)

Subtask 1

Subtask 2
Results (#authors)

![Graphs showing accuracy and number of correct authors for different methods (Baseline-RND, Nath, Zuo et al.) with varying number of authors (1 to 5).]
Conclusion

• Style change detection task
• Two subtasks were tackled
• Unfortunately only two submissions
• Many exciting plans for next year, looking forward to your submissions next year 😊