Abdul Sittar, Hafiz Rizwan Iqbal and Rao Muhammad Adeel Nawab

{ abdulsittar72@yahoo.com, rizwan.iqbal@ciitlahore.edu.pk, adeelnawab@ciitlahore.edu.pk }
Agenda:

- Problem
- Proposed Solution
- Overall System Performance
- Analysis of System Performance
- Conclusion
Problem:
Author Diarization

- Task A - Diarization with two authors
- Task B – Diarization with given number of authors
- Diarization with unknown number of authors
Agenda:

- Problem
- Proposed Solution
- Overall System Performance
- Analysis of System Performance
- Conclusion
Proposed Solution:

- Sentence Segmentation
- Lexical Features
- ClustDist Function
Sentence Segmentation:

- Different combination of sentences ranges from 2 to 15.
Lexical Features:
1. Characters Count
2. Digits Count
3. Uppercase Letters Count
4. Spaces Count
5. Letters Count
6. Tabs Count
Lexical Features:
7. Words Count
8. Ratio of Interrogative Sentence
9. Average Word Length
10. Average Sentence Length
11. Ratio of Digits
12. Ratio of Uppercase Letters
Lexical Features:
13. Ratio of Spaces
14. Ratio of Letters
15. Ratio of Tabs
ClustDist:

- ClustDist Formula
- Document with n sentences
- \( d = \text{distance} \)
- \( V = \text{Feature Vector} \)

\[
ClustDist(x, V) = \frac{\sum_{k} d(x, \vec{v})}{n}
\]
Agenda:

• Problem
• Proposed Solution
• **Overall System Performance**
• Analysis of System Performance
• Conclusion
Overall System Performance:

Task A (combination of 7 sentences)

- Micro-recall 0.1493
- Micro-Precision 0.2655
- Micro-F 0.1911
- Macro-Recall 0.1648
- Macro-Precision 0.2664
- Macro-F 0.2036
Overall System Performance:

Task B (combination of 5 sentences)

- Bcubed-recall 0.4823
- Bcubed-Precision 0.2861
- Bcubed-F 0.3591
Overall System Performance:
Task C (combination of 5 sentences)

- Bcubed-recall 0.5464
- Bcubed-Precision 0.2822
- Bcubed-F 0.3722
Agenda:

• Problem
• Proposed Solution
• Overall System Performance
• Analysis of System Performance
• Conclusion
Analysis of System Performance:

• Results of Task A

<table>
<thead>
<tr>
<th>Sentences</th>
<th>Micro-recall</th>
<th>Micro-precision</th>
<th>Micro-f</th>
<th>Macro-recall</th>
<th>Macro-precision</th>
<th>Macro-f</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.1338</td>
<td>0.2006</td>
<td>0.1605</td>
<td>0.1216</td>
<td>0.2006</td>
<td>0.1514</td>
</tr>
<tr>
<td>3</td>
<td>0.1045</td>
<td>0.1828</td>
<td>0.1330</td>
<td>0.1109</td>
<td>0.1823</td>
<td>0.1379</td>
</tr>
<tr>
<td>4</td>
<td>0.1291</td>
<td>0.2492</td>
<td>0.1701</td>
<td>0.1178</td>
<td>0.2492</td>
<td>0.1600</td>
</tr>
<tr>
<td>5</td>
<td>0.1392</td>
<td>0.2599</td>
<td>0.1813</td>
<td>0.1337</td>
<td>0.2599</td>
<td>0.1766</td>
</tr>
<tr>
<td>6</td>
<td>0.1461</td>
<td>0.2572</td>
<td>0.1864</td>
<td>0.1421</td>
<td>0.2572</td>
<td>0.1830</td>
</tr>
<tr>
<td>7</td>
<td>0.1493</td>
<td>0.2655</td>
<td>0.1911</td>
<td>0.1648</td>
<td>0.2664</td>
<td>0.2036</td>
</tr>
<tr>
<td>8</td>
<td>0.1130</td>
<td>0.1998</td>
<td>0.1444</td>
<td>0.1129</td>
<td>0.1995</td>
<td>0.1442</td>
</tr>
<tr>
<td>9</td>
<td>0.1280</td>
<td>0.2323</td>
<td>0.1651</td>
<td>0.1304</td>
<td>0.2322</td>
<td>0.1670</td>
</tr>
<tr>
<td>10</td>
<td>0.1045</td>
<td>0.1828</td>
<td>0.1330</td>
<td>0.1109</td>
<td>0.1823</td>
<td>0.1379</td>
</tr>
<tr>
<td>11</td>
<td>0.1379</td>
<td>0.2547</td>
<td>0.1875</td>
<td>0.1481</td>
<td>0.2523</td>
<td>0.1866</td>
</tr>
<tr>
<td>12</td>
<td>0.1165</td>
<td>0.2315</td>
<td>0.1550</td>
<td>0.1103</td>
<td>0.2307</td>
<td>0.1492</td>
</tr>
<tr>
<td>15</td>
<td>0.1301</td>
<td>0.2573</td>
<td>0.1728</td>
<td>0.1242</td>
<td>0.2565</td>
<td>0.1674</td>
</tr>
</tbody>
</table>
Analysis of System Performance:

• Results of Task B

<table>
<thead>
<tr>
<th>Sentence Length</th>
<th>bcubed-recall</th>
<th>bcubed-precision</th>
<th>bcubed-f</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.4823</td>
<td>0.2861</td>
<td>0.3591</td>
</tr>
<tr>
<td>10</td>
<td>0.5951</td>
<td>0.1315</td>
<td>0.2154</td>
</tr>
<tr>
<td>12</td>
<td>0.6143</td>
<td>0.1080</td>
<td>0.1838</td>
</tr>
<tr>
<td>13</td>
<td>0.6260</td>
<td>0.1051</td>
<td>0.1800</td>
</tr>
<tr>
<td>14</td>
<td>0.6376</td>
<td>0.0887</td>
<td>0.1558</td>
</tr>
</tbody>
</table>
Analysis of System Performance:

• Results of Task C

<table>
<thead>
<tr>
<th>Sentence Length</th>
<th>bcubed-recall</th>
<th>bcubed-precision</th>
<th>bcubed-f</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.5464</td>
<td>0.2822</td>
<td>0.3722</td>
</tr>
<tr>
<td>10</td>
<td>0.6253</td>
<td>0.1339</td>
<td>0.2206</td>
</tr>
<tr>
<td>12</td>
<td>0.6386</td>
<td>0.1076</td>
<td>0.1842</td>
</tr>
</tbody>
</table>
Agenda:

• Problem
• Proposed Solution
• Overall System Performance
• Analysis of System Performance
• Conclusion
Conclusion:

• Different combination of sentences ranges from 2 to 15.
• Calculation of 17 lexical features.
• Calculation of ClustDist.
THANKS

Abdul Sittar, Hafiz Rizwan Iqbal and Rao Muhammad Adeel Nawab

{ abdulsittar72@yahoo.com, rizwan.iqbal@ciitlahore.edu.pk, adeelnawab@ciitlahore.edu.pk }

Note book for PAN at CLEF 2016