

COMSATS INSTITUTE OF INFORMATION TECHNOLOGY, LAHORE, PAKISTAN

Abdul Sittar, Hafiz Rizwan Iqbal and Rao Muhammad
Adeel Nawab

{ abdulesittar72@yahoo.com, rizwan.iqbal@ciitlahore.edu.pk,
adeelnawab@ciitlahore.edu.pk }

Note book for PAN at CLEF 2016

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 = distance
- V = Feature Vector

$$ClustDist(\vec{x}, V) = \frac{\sum d(\vec{x}, \vec{v})}{n}$$

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 2: Task A Results

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

Analysis of System Performance:

- **Results of Task B**

Table 3: Task B Results

Sentence Length	bcubed-recall	bcubed-precision	bcubed-f
5	0.4823	0.2861	0.3591
10	0.5951	0.1315	0.2154
12	0.6143	0.1080	0.1838
13	0.6260	0.1051	0.1800
14	0.6376	0.0887	0.1558

Analysis of System Performance:

- **Results of Task C**

Table 4: Task C Results

Sentence Length	bcubed-recall	bcubed-precision	bcubed-f
5	0.5464	0.2822	0.3722
10	0.6253	0.1339	0.2206
12	0.6386	0.1076	0.1842

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

{ abdulesittar72@yahoo.com, rizwan.iqbal@ciitlahore.edu.pk,
adeelnawab@ciitlahore.edu.pk }

Note book for PAN at CLEF 2016