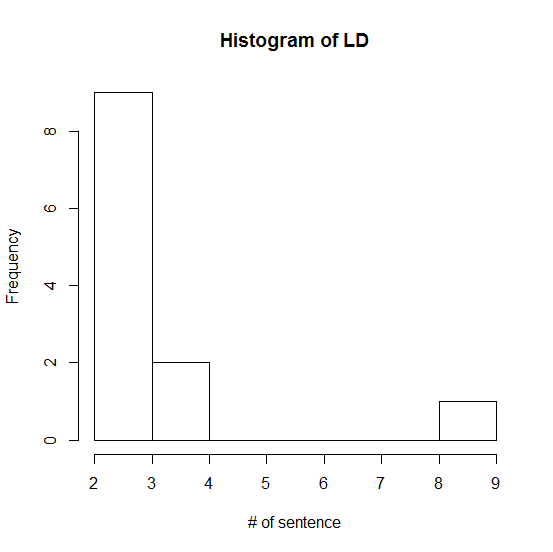
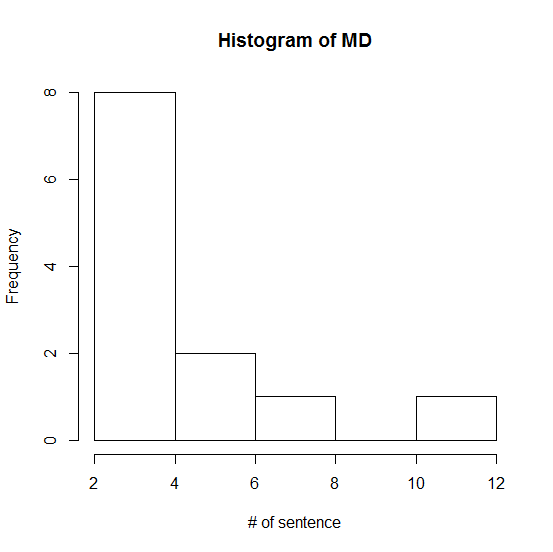
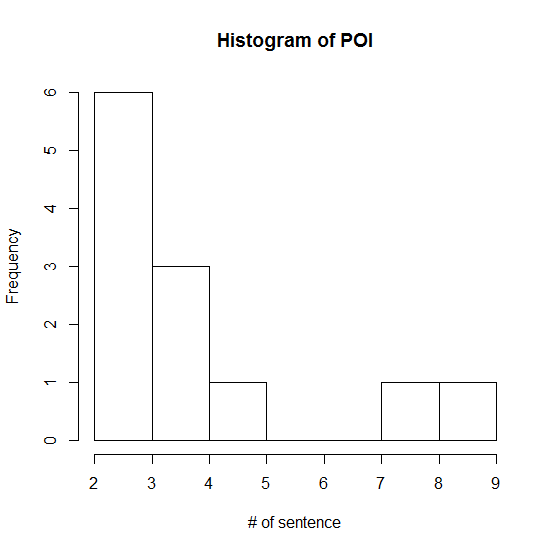
# Summary Report (07/17/2014)

## Distribution of the length of TA’s summary



Most of them are 2 sentences or 3 sentences.

## The length of Mead’s Summary: # of words

The # of words in Mead’s summary for POI (p<0.05) and LP (p <0.001) is significantly smaller than TA’s summary (paired t-test); the p-value for MD is 0.057, which is close to 0.05.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | word count | | | word count | | |
| Week | TA:POI | TA:MP | TA:LP | Mead:POI | Mead:MP | Mead:LP |
| 1 | 19 | 12 | 15 | 76 | 85 | 46 |
| 2 | 15 | 30 | 11 | 47 | 99 | 50 |
| 3 | 30 | 25 | 40 | 48 | 77 | 37 |
| 4 | 17 | 31 | 17 | 55 | 53 | 59 |
| 5 | 19 | 45 | 25 | 38 | 71 | 25 |
| 6 | 17 | 19 | 24 | 49 | 38 | 45 |
| 7 | 27 | 31 | 10 | 43 | 65 | 52 |
| 8 | 27 | 32 | 15 | 30 | 67 | 34 |
| 9 | 59 | 83 | 13 | 60 | 71 | 57 |
| 10 | 35 | 55 | 22 | 37 | 38 | 53 |
| 11 | 95 | 111 | 84 | 57 | 43 | 77 |
| 12 | 65 | 45 | 15 | 48 | 46 | 47 |

## The Average word per sentence in Mead’s summary

### Mead prefers longer sentences

* The average word length per sentence in a mead’s summary is significantly bigger than the average word length in the students’ responses. (p < 0.001)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Response Average Word Count | | | Mead's Average Word Count | | |
| Week | POI | MP | LP | POI | MP | LP |
| 1 | 9 | 9 | 9 | 25 | 28 | 15 |
| 2 | 11 | 11 | 10 | 15 | 33 | 16 |
| 3 | 9 | 9 | 12 | 16 | 25 | 12 |
| 4 | 7 | 7 | 8 | 18 | 17 | 19 |
| 5 | 7 | 8 | 7 | 12 | 23 | 8 |
| 6 | 6 | 7 | 7 | 16 | 12 | 15 |
| 7 | 8 | 7 | 11 | 14 | 21 | 17 |
| 8 | 6 | 7 | 6 | 10 | 22 | 11 |
| 9 | 10 | 7 | 8 | 20 | 23 | 19 |
| 10 | 5 | 6 | 11 | 12 | 12 | 17 |
| 11 | 8 | 9 | 7 | 19 | 14 | 25 |
| 12 | 7 | 10 | 8 | 16 | 15 | 15 |
| 13 | 7 | 7 | 6 | 16 | 20 | 20 |
| 14 | 6 | 7 | 6 | 17 | 16 | 12 |
| 15 | 7 | 9 | 6 | 14 | 23 | 16 |
| 16 | 7 | 8 | 4 | 12 | 17 | 13 |
| 17 | 9 | 6 | 8 | 19 | 19 | 25 |
| 18 | 6 | 7 | 6 | 19 | 20 | 19 |
| 19 | 6 | 8 | 6 | 17 | 21 | 20 |
| 20 | 6 | 7 | 6 | 14 | 20 | 18 |
| 21 | 6 | 8 | 7 | 17 | 11 | 30 |
| 22 | 6 | 7 | 6 | 20 | 18 | 21 |
| 23 | 6 | 6 | 6 | 14 | 22 | 19 |
| 24 | 6 | 5 | 6 | 15 | 13 | 28 |
| 25 | 13 | 9 | 12 | 36 | 23 | 34 |

## Multiple-document Summarization

Instead of concatenating all the students’ responses into one single document, each of the responses are one single document. Thus, it is a multiple document summarization task instead of single document summarization. The evaluation results are shown in the following sessions.

## Evaluation metric

ROUGE-2 (N=2): The larger the better.

### 

### ROUGE-SU4

ROUGE-S is a skip-bigram version of ROUGE-2. And ROUGE-SU4 is an extension of ROUGE-S by including unigrams and skip-bigram by up to four words.

## Results

All the evaluations are performed with the official ROUGE software [1] with the same setting used in [2].

### Single-doc summarization vs. multi-doc summarization

Even though most of the scores for multi-doc summarization are lower than scores in single-doc summarization, they are not significantly different. (p > 0.05)

* The scores for R2 and R-SU4 are not that good. (The state-of-the-art of the R2 is around 0.10~0.14, and R-SU4 is around 0.16~0.19 depending on the dataset.)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | single document summarization | | | | | |  |  | multiple document summarization | | | | | |
|  | POI | | MD | | LP | |  |  | POI | | MD | | LP | |
| week | R2 | R-SU4 | R2 | R-SU4 | R2 | R-SU4 |  | week | R2 | R-SU4 | R2 | R-SU4 | R2 | R-SU4 |
| 1 | 0.000 | 0.015 | 0.000 | 0.002 | 0.000 | 0.011 |  | 1 | 0.000 | 0.014 | 0.000 | 0.002 | 0.000 | 0.007 |
| 2 | 0.058 | 0.013 | 0.093 | 0.042 | 0.000 | 0.004 |  | 2 | 0.058 | 0.013 | 0.084 | 0.025 | 0.000 | 0.006 |
| 3 | 0.182 | 0.060 | 0.091 | 0.021 | 0.012 | 0.016 |  | 3 | 0.157 | 0.077 | 0.062 | 0.028 | 0.015 | 0.024 |
| 4 | 0.000 | 0.019 | 0.123 | 0.098 | 0.000 | 0.004 |  | 4 | 0.000 | 0.029 | 0.083 | 0.078 | 0.000 | 0.005 |
| 5 | 0.094 | 0.044 | 0.018 | 0.028 | 0.000 | 0.010 |  | 5 | 0.094 | 0.044 | 0.030 | 0.016 | 0.000 | 0.010 |
| 6 | 0.000 | 0.029 | 0.056 | 0.041 | 0.030 | 0.024 |  | 6 | 0.000 | 0.032 | 0.068 | 0.037 | 0.000 | 0.010 |
| 7 | 0.000 | 0.036 | 0.136 | 0.054 | 0.000 | 0.000 |  | 7 | 0.026 | 0.035 | 0.136 | 0.054 | 0.000 | 0.003 |
| 8 | 0.050 | 0.021 | 0.063 | 0.036 | 0.000 | 0.003 |  | 8 | 0.045 | 0.014 | 0.088 | 0.043 | 0.000 | 0.003 |
| 9 | 0.112 | 0.050 | 0.080 | 0.119 | 0.030 | 0.032 |  | 9 | 0.048 | 0.058 | 0.080 | 0.117 | 0.029 | 0.026 |
| 10 | 0.024 | 0.031 | 0.020 | 0.025 | 0.021 | 0.023 |  | 10 | 0.022 | 0.028 | 0.020 | 0.025 | 0.018 | 0.015 |
| 11 | 0.295 | 0.099 | 0.101 | 0.069 | 0.012 | 0.042 |  | 11 | 0.089 | 0.054 | 0.057 | 0.055 | 0.012 | 0.038 |
| average | 0.074 | 0.038 | 0.071 | 0.049 | 0.010 | 0.015 |  | average | 0.049 | 0.036 | 0.064 | 0.044 | 0.007 | 0.013 |

## TODO

Data Analysis

## Reference

1. Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries." *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*. 2004.
2. Wang, Lu, et al. "A Sentence Compression Based Framework to Query-Focused Multi-Document Summarization." *ACL (1)*. 2013.