

Milestone 2

Team 9: Napat Luevisadpaibul, Dishan Gupta,
Callie Vaughn, Wenyi Wang

Baseline from Milestone 1

Accuracy on input files with improvements to baseline						
		XML 1	XML 2	XML 3	XML 4	Mean
Baseline	% correct c at 1	2/10 = 20% 0.22	1/10 = 10% 0.11	1/10 = 10% 0.11	3/10 = 30% 0.33	7/40 = 17.5% 0.1925
PMI Scoring	% correct c at 1	4/10 = 40% 0.44	2/10 = 20% 0.22	4/10 = 40% 0.44	1/10 = 10% 0.11	11/40 = 27.5% 0.3025
PMI + Removed Answers	% correct c at 1	5/10 = 50% 0.55	2/10 = 20% 0.22	4/10 = 40% 0.44	1/10 = 10% 0.11	12/40 = 30% 0.33

Error Analysis

1. In the baseline code we keep the top 10 candidate sentences according to their similarity score

Problem: Some queries have sentence and score ranked like this:

Sentence 1: 0.9.....

Sentence 2: 0.87....

Sentence 3: 0.79....

Sentence 4: 0.77....

Sentence 5: 0.11....

Sentence 6: 0.08....

Do we have to keep all the candidate sentences?

Error Analysis

2. We planned to use synonyms of named entities and nouns to find candidate sentences and use synonyms in candidate sentences to score answer choice, but we were not able to implement such a system before milestone 1. This week we have implemented a synonym system to address the problem of wording in documents. Our synonym system can support synonyms of phrases too.

ex. Red blood cells -> erythrocytes

Error Analysis

3. PMI scoring works very well in some instances but not all of them. For example, the fourth XML document contains data that works for the baseline similarity measurement but not PMI.

Cosine similarity is another good, standard similarity measurement that has different strengths and weaknesses from PMI

Attempted Improvements

- Adaptive K Sentence
- Synonyms
- Alternative Similarity Scorer

Adaptive K Sentence

First try: get rid of all sentences that have lower score than $\frac{1}{5}$ of previous rank score sentence.

Accuracy on input files with improvements to baseline						
		XML 1	XML 2	XML 3	XML 4	Mean
Baseline	% correct c at 1	2/10 = 20% 0.22	1/10 = 10% 0.11	1/10 = 10% 0.11	3/10 = 30% 0.33	7/40 = 17.5% 0.1925
BaseLine + Adaptive K Sentence	% correct c at 1	2/10 = 20% 0.22	2/10 = 20% 0.22	1/10 = 10% 0.11	3/10 = 30% 0.33	8/40 = 20% 0.22
PMI + Removed Answers	% correct c at 1	5/10 = 50% 0.55	2/10 = 20% 0.22	4/10 = 40% 0.44	1/10 = 10% 0.11	12/40 = 30% 0.33
Above + Adaptive	% correct c at 1	5/10 = 50% 0.55	2/10 = 20% 0.22	4/10 = 40% 0.44	1/10 = 10% 0.11	12/40 = 30% 0.33

Adaptive K Sentence

Second try: get rid of all sentences that have lower score than $1/3$ of previous rank score sentence.

Accuracy on input files with improvements to baseline						
		XML 1	XML 2	XML 3	XML 4	Mean
PMI + Removed Answers	% correct c at 1	5/10 = 50% 0.55	2/10 = 20% 0.22	4/10 = 40% 0.44	1/10 = 10% 0.11	12/40 = 30% 0.33
Above + Adaptive	% correct c at 1	4/10 = 40% 0.44	0/10 = 0% 0	4/10 = 40% 0.44	1/10 = 10% 0.11	9/40 = 22.5%

Third try: get rid of all sentences that have lower score than $1/5$ of max score.
same as second try.

Synonyms

We tried extracting phrase-level synonyms for named entities/noun phrases but the system is trained on newswire corpus and unable to generate good results for AD related queries.

Alternative Similarity Scorer

- Based on the implementation of the default similarity scorer.
- Added cosine similarity, Dice coefficient, and Jaccard coefficient. -- **choice 1**
 - Computed based on **tokens**
 - Computed based on **noun phrases and NERs**
- Slightly better than the original baseline
- Normalization may not be necessary

Alternative Similarity Scorer (cont.)

- Also added normalization to the default similarity scorer, which is merely based on counting. -- **choice 2**
- Slightly better than the original baseline
- Similar to choice 1

Results

Accuracy on input files with improvements to baseline						
		XML 1	XML 2	XML 3	XML 4	Mean
Milestone 1 (PMI + Removed Ans.)	% correct c at 1	5/10 = 50% 0.55	2/10 = 20% 0.22	4/10 = 40% 0.44	1/10 = 10% 0.11	12/40 = 30% 0.33
Cosine + Removed Ans.	% correct c at 1	2/10 = 20% 0.22	2/10 = 20% 0.22	1/10 = 10% 0.11	3/10 = 30% 0.3	8/40 = 20% 0.2125
Milestone 1 + Adaptive	% correct c at 1	5/10 = 50% 0.55	2/10 = 20% 0.22	4/10 = 40% 0.44	1/10 = 10% 0.11	12/40 = 30% 0.33

Next Steps

Sometimes we don't select any choices!

Question: How many persons worldwide are estimated to have a medical condition related to neurodegeneration?

Aggregated counts;

Correct Choice: more than 10 million

Best Choice: null

-We really have high chance to answer this correctly because we've already discarded some answers.

Question: What is the major protease produced by microglia responsible for degrading A?

Aggregated counts;

Correct Choice: extracellular

Best Choice: null