MIM QA System D3

Antariksh Bothale

Julian Chan

Yi-Shu Wei

D3

- Question classification

- Answer reranking based on question and answer type

- Query expansion using redundancy-based web-boosting

- Other improvements

lenient
0.364

strict
0.241

Question Classification

- SVM Question Classification
  - Trained on UIUC question
  - 5500 questions; 50 fine-grained classes
  - Features: unigrams, wh-words, lemmatized headwords
  - Tested on TREC 2004 & 2005 data (only labeled with 6 coarse-grained classes)
  - Accuracy: 87%

Reranking based on question type

- Classified answers based on regex matches
  - Time
  - People

- Answers that match the question type are moved higher on the list.

Query Expansion with Web-boosting

- We implemented query expansion using a redundancy-based web-boosting strategy.

- The original question is passed to the Ask.com search engine and the snippets of the first 40 results are extracted and cached. (Thanks, CJ!)

- From these snippets we obtain bigram counts and keep only the top 10 most frequent ones that do not contain any words from the original question.

- Often times, we found these top bigrams to contain the correct answer to the questions.

- We then harvest unique terms from these top bigrams and add them to the query to the document retrieval system.

- Boosted our accuracy by 10%.

Other improvement

- Added target words in queries
  - Criterion: All target words in the article, or some target word in the headline

- Use sentence as the unit for segmenting documents

- Parallel processing to speed up devtest run