ReQuery

Lisa Gress Kathryn Nichols Shannon Watanabe

May 16, 2013

Overview

Query classification/reformulation

Information retrieval

Answer extraction/Document retrieval

Results

Next steps

Query Classification

- Li & Roth coarse-grained types
- MaxEnt classifier
- Training
  - Li & Roth 5500 labeled queries
- Devtest
  - TREC-2004
  - TREC-2005
- Vectors
  - Binary
  - Lower-case unigrams, targets included
- Accuracy:
  - TREC-2004: 86.5%
  - TREC-2005: 85.6%

Query Reformulation

- Exact queries
  - Attempted to follow Lin 2007 and use POS tags to create rewrite rules
  - In testing, lowered results and POS tagging took time
  - This approach not optimized for IR
- Inexact
  - Head chunk boosting
  - NE boosting
  - Target boosting
  - Inflectional expansion
  - Wh-word expansion
  -

Information retrieval

- Document retrieval
  - Lucene index of AQUAINT corpus
  - 10 documents per query
  - 160-character fragment(s) containing query terms

- Web search
  - xgoogle module
  - baseline queries (no reformulation)
  - 100 snippets per question
  - cached with pickle

Answer extraction

- Redundancy approach from Lin (2007)
- Web snippets and AQUAINT fragments pooled together
  - no weighting - simplification
- Document retrieval
  - 20 highest scoring n-grams selected
  - for each n-gram, document with highest Lucene score is returned

Results

TREC 2006

	Strict	Lenient	Q answered
D2 w/ Bing	0.00017	0.019	250
D2 w/ Google	0.04052	0.14531	383
D3	0.05492	0.14013	402

Next steps

- Different queries for web versus AQUAINT corpus
- Implement (passage) retrieval for AQUAINT documents
- Answer extraction
  - Use question type information
- Weighting system for AQUAINT answer candidates and web answer candidates