Deliverable #3: Document and Passage Retrieval

Ling 573
NLP Systems and Applications
May 10, 2011
Main Components

- Document retrieval
  - Evaluated with Mean Average Precision

- Passage retrieval/re-ranking
  - Evaluated with Mean Reciprocal Rank (MRR)
Mean Average Precision (MAP)

- Traverse ranked document list:
  - Compute precision each time relevant doc found
    - Average precision up to some fixed cutoff
    - $R_r$: set of relevant documents at or above $r$
    - $\text{Precision}(d)$: precision at rank when doc $d$ found

$$\frac{1}{|R_r|} \sum_{d \in R_r} \text{Precision}_r(d)$$

- Mean Average Precision: 0.6
  - Compute average of all queries of these averages
  - Precision-oriented measure

- Single crisp measure: common TREC Ad-hoc
Baselines

- Indri:
  - Default settings: #combine
  - 2003:
    - MAP: 0.23
  - 2004:
    - No processing: MAP: 0.13
    - Simple concatenation: MAP: 0.35
    - Conservative pseudo-relevance feedback:
      - 5 top docs, 5 terms, default weighting: MAP: 0.35
      - Per-query variation
MRR

- Classical:
  - Return ranked list of answer candidates
  - Idea: Correct answer higher in list => higher score

- Measure: Mean Reciprocal Rank (MRR)
  - For each question,
    - Get reciprocal of rank of first correct answer
      - E.g. correct answer is 4 => $\frac{1}{4}$
      - None correct => 0
    - Average over all questions

$$MRR = \frac{\sum_{i=1}^{N} \frac{1}{rank_i}}{N}$$
Baselines

- 2004:
  - Indri passage retrieval 100 term passages
    - Strict: MRR: 0.19
    - Lenient: MRR: 0.28
Pattern Matching

- Litkowski pattern files:
  - Derived from NIST relevance judgments on systems
  - Format:
    - Qid answer_pattern doc_list
    - Passage where answer_pattern matches is correct
    - If it appears in one of the documents in the list
Pattern Matching

- Litkowski pattern files:
  - Derived from NIST relevance judgments on systems
  - Format:
    - Qid answer_pattern doc_list
    - Passage where answer_pattern matches is correct
      - If it appears in one of the documents in the list

- MRR scoring
  - Strict: Matching pattern in official document
  - Lenient: Matching pattern
Examples

- Example
  - Patterns
    - 1894 (190|249|416|440)(\s|\-\-\-)million(\s|\-\-\-)miles?
      APW19980705.0043 NYT19990923.0315
      NYT19990923.0365 NYT20000131.0402
      NYT19981212.0029
    - 1894 700-million-kilometer APW19980705.0043
    - 1894 416 - million - mile NYT19981211.0308
  - Ranked list of answer passages
    - 1894 0 APW19980601.0000 the casta way weas
    - 1894 0 APW19980601.0000 440 million miles
    - 1894 0 APW19980705.0043 440 million miles
Evaluation Issues

- Exhaustive vs Pooled scoring
  - Exhaustive
    - Every document/passage evaluated for every query
  - Pooled scoring
    - All participant responses are collected and scored
    - All correct responses form basis for patterns/qrels
      - Scores usually well-correlated with exhaustive

- Exhaustive:
  - More thorough; MUCH!! more expensive

- Pooled:
  - Cheaper, faster; penalizes non-conforming systems
Presentations

- Group 8
  - Mowry, Srinivasan, Wong

- Group 4
  - Alotayq, Pham, Wang

- Group 9
  - Hermsen, Lushtak, Lutes