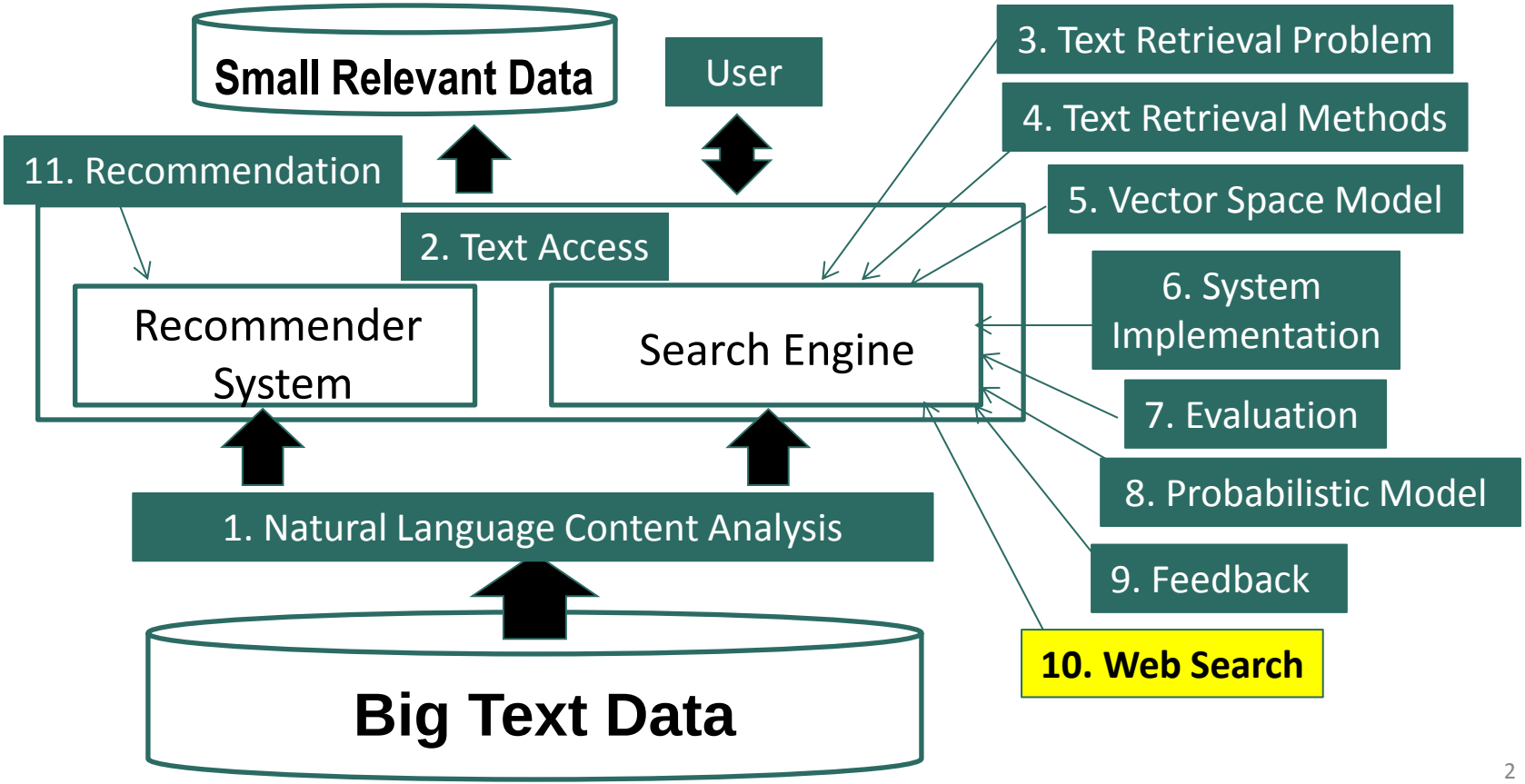
Text Retrieval and Search Engines

Web Search: Learning to Rank - Part 1 - 3

ChengXiang “Cheng” Zhai
Department of Computer Science
University of Illinois at Urbana-Champaign

Web Search: Learning to Rank



How Can We Combine Many Features? (Learning to Rank)

- General idea:
  - Given a query-doc pair (Q,D), define various kinds of features Xi(Q,D)
  - Examples of feature: the number of overlapping terms, BM25 score of Q and D, p(Q|D), PageRank of D, p(Q|Di), where Di may be anchor text or big font text, "does the URL contain '~'?"….
  - Hypothesize p(R=1|Q,D)=s(X1(Q,D),…,Xn(Q,D), λ) where λ is a set of parameters
  - Learn λ by fitting function s with training data, i.e., 3-tuples like (D, Q, 1) (D is relevant to Q) or (D,Q,0) (D is non-relevant to Q)

训练语料的格式 → (Doc, Query, judge)

Regression-Based Approaches

Logistic Regression: Xi(Q,D) is feature; β's are parameters

$$\log \frac{P(R=1|Q,D)}{1-P(R=1|Q,D)} = \beta_0 + \sum_{i=1}^{n} \beta_i X_i$$

$$P(R=1|Q,D) = \frac{1}{1+\exp(-\beta_0 - \sum_{i=1}^{n} \beta_i X_i)}$$

Estimate β's by maximizing the likelihood of training data

	X1(Q,D) BM25	X2 (Q,D) PageRank	X3(Q,D) BM25Anchor
D1 (R=1)	0.7	0.11	0.65
D2 (R=0)	0.3	0.05	0.4

$$p(\{(Q,D_1,1),(Q,D_2,0)\}) = \frac{1}{1+\exp(-\beta_0 - 0.7\beta_1 - 0.11\beta_2 - 0.65\beta_3)} * (1 - \frac{1}{1+\exp(-\beta_0 - 0.3\beta_1 - 0.05\beta_2 - 0.4\beta_3)})$$

$$\vec{\beta}^* = \arg\max_{\vec{\beta}} p(\{(Q_1,D_{11},R_{11}),(Q_1,D_{12},R_{12}),\ldots,(Q_n,D_{m1},R_{m1}),\ldots\})$$

Once β's are known, we can take Xi(Q,D) computed based on a new query and a new document to generate a score for D w.r.t. Q.

More Advanced Learning Algorithms

- Attempt to directly optimize a retrieval measure (e.g. MAP, nDCG)
  - More difficult as an optimization problem
  - Many solutions were proposed [Liu 09]
- Can be applied to many other ranking problems beyond search
  - Recommender systems
  - Computational advertising
  - Summarization
  - …

Summary

- Machine learning has been applied to text retrieval since many decades ago (e.g., Rocchio feedback)
- Recent use of machine learning is driven by
  - Large-scale training data available
  - Need for combining many features 用于优化结果.
  - Need for robust ranking (again spams)
- Modern Web search engines all use some kind of ML technique to combine many features to optimize ranking
- Learning to rank is still an active research topic

Additional Readings

- Tie-Yan Liu. Learning to Rank for Information Retrieval. Foundations and Trends in Information Retrieval 3, 3 (2009): 225-331.
- Hang Li. A Short Introduction to Learning to Rank, IEICE Trans. Inf. & Syst. E94-D, 10 (Oct. 2011): n.p.