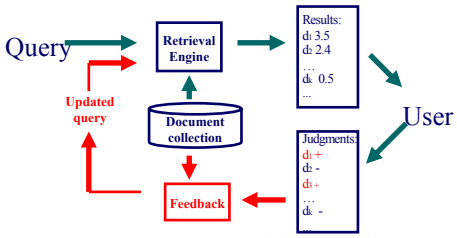
Feedback

(Search and Data Mining)

Hui Fang
Department of Electrical and Computer Engineering
University of Delaware

Relevance Feedback

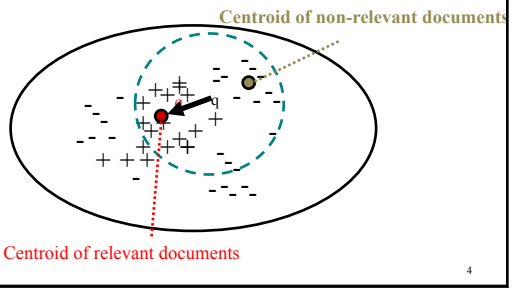


- **Implicit Feedback:** Use clickthrough data
- **Psuedo Feedback:** Use top-K ranked results

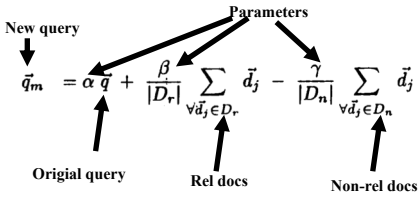
Feedback in Vector Space Model

- How to learn from examples to improve retrieval performance?
  - Positive examples: documents that are known to be relevant
  - Negative examples: documents that known to be non-relevant
- General method: query modification
  - Adding new terms
  - Adjusting weights of old terms

Illustration of Rocchio Feedback



Rocchio Feedback: Formula



Rocchio in Practice

- Negative examples are not very important.
- Need to keep relatively high weight on the original query weights.
- Can be used for relevance feedback and pseudo feedback
- Usually robust and effective

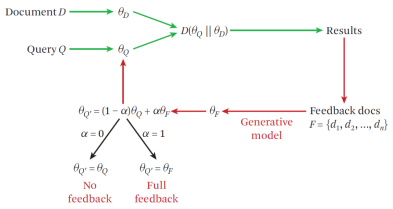
Kullback-Leibler (KL) Divergence

Query likelihood $f(q,d) = \sum_{w \in d, w \in q} c(w,q) [\log \frac{p_{seen}(w|d)}{\alpha_d p(w|C)}] + n \log \alpha_d$

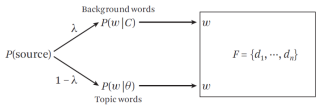
KL-divergence (cross entropy) $f(q,d) = \sum_{w \in d, p(w|\theta_Q)>0} [p(w|\hat{\theta}_Q)] \log \frac{p_{seen}(w|d)}{\alpha_d p(w|C)} + \log \alpha_d$

Query LM $p(w|\hat{\theta}_Q) = \frac{c(w,Q)}{|Q|}$

Model-based Feedback



Mixture Model for Feedback



$\log p(F|\theta) = \sum_i \sum_w c(w; d_i) \log[(1-\lambda)p(w|\theta) + \lambda p(w|C)]$

Maximum likelihood $\theta_F = \arg\max_\theta \log p(F|\theta)$

λ = noise in feedback documents

Example of learned query model

Query: "airport security"

Mixture model approach

$\lambda = 0.9$

w	$P(w\|\theta_F)$
security	0.0558
airport	0.0546
beverage	0.0488
alcohol	0.0474
bomb	0.0236
terrorist	0.0217
author	0.0206
license	0.0188
bond	0.0186
counter-terror	0.0173
terror	0.0142
newsnet	0.0129
attack	0.0124
operation	0.0121
headline	0.0121

Web database

Top 10 docs

$\lambda = 0.7$

w	$P(w\|\theta_F)$
the	0.0405
security	0.0377
airport	0.0342
beverage	0.0305
alcohol	0.0304
to	0.0268
of	0.0241
and	0.0214
author	0.0156
bomb	0.0150
terrorist	0.0137
in	0.0135
license	0.0127
state	0.0127
by	0.0125