Geometric Perspectives of the BM25

Giorgio Maria Di Nunzio
Dept. of Information Engineering
University of Padua

IIR 2015 - 6th Italian Information Retrieval Workshop

May 26th, 2015, Cagliari, Italy

Outline

- Probabilistic Retrieval Models
- Binary Independence Model (BIM)
- BestMatch 25 (BM25)
- Likelihood Spaces
- Conclusions

Probability Ranking Principle (PRP)

- The objective of a retrieval function is to rank the documents of a collection according to a specific information need or query.
- Optimal retrieval can be achieved by ranking documents in order of decreasing probability of relevance (PRP).

$$P(R = r|D)$$

S.E. ROBERTSON, (1977) "The Probability Ranking Principle in IR", Journal of Documentation, Vol. 33 Iss: 4, pp.294 - 304

Probability Ranking Principle (PRP)

- The probability of a document being relevant is rank equivalent (given a query) to the odds

$$P(R = r|D) \propto_{rank} \frac{P(R{=}r|D)}{P(R{=}\overline{r}|D)}$$

$$= \frac{P(D|R{=}r)P(R{=}r)}{P(D|R{=}\overline{r})P(R{=}\overline{r})}$$

$$\propto_{rank} \frac{P(D|R{=}r)}{P(D|R{=}\overline{r})}$$

Bayesian Decision Theory 1/2

- The identification of relevant documents can be achieved by selecting the decision that minimizes the conditional risk (under zero-one loss):

$$P(R = r|D) > P(R = \overline{r}|D)$$

$$\frac{P(R = r|D)}{P(R = \overline{r}|D)} > 1$$

Bayesian Decision Theory 2/2

- The identification of relevant documents can be achieved by selecting the decision that minimizes the conditional risk (under zero-one loss):

$$P(D|R=r)P(R=r) > P(D|R=\overline{r})P(R=\overline{r})$$

$$\frac{P(D|R=r)}{P(D|R=\overline{r})} > \frac{P(R=\overline{r})}{P(R=r)}$$

Fabio Crestani, Mounia Lalmas, Cornelis J. Van Rijsbergen, and Iain Campbell. 1998. “Is this document relevant?…probably”: a survey of probabilistic models in information retrieval. ACM Comput. Surv. 30, 4.

Binary Independence Model (BIM)

- Documents as binary vectors. Each word is distributed as a Bernoulli function.

$$\frac{P(D = d_i | R = r)}{P(D = d_i | R = \overline{r})} = \prod_{w_t \in \mathcal{V}} \frac{\theta_{w_t|r}^{x_{it}} (1 - \theta_{w_t|r})^{1 - x_{it}}}{\theta_{w_t|\overline{r}}^{x_{it}} (1 - \theta_{w_t|\overline{r}})^{1 - x_{it}}}$$

$$\log \left(\frac{P(D = d_i | R = r)}{P(D = d_i | R = \overline{r})} \right) = \sum_{w_t \in \mathcal{V}} x_{it} \log \left(\frac{\theta_{w_t|r} (1 - \theta_{w_t|r})}{\theta_{w_t|\overline{r}} (1 - \theta_{w_t|\overline{r}})} \right) + \sum_{w_t \in \mathcal{V}} \log \left(\frac{(1 - \theta_{w_t|r})}{(1 - \theta_{w_t|\overline{r}})} \right)$$

S.E. Robertson and K. Spärck Jones, Relevance weighting of search terms.
Journal of the American Society for Information Science 27, 129-46 (1976)

Binary Independence Model (BIM)

- Documents ranked according to the sum of relevance weights RW.

$$\log\left(\frac{P(D|R=r)}{P(D|R=\overline{r})}\right) \propto_{rank} \sum_{w_t \in d_i} \log\left(\frac{\theta_{w_t|r}(1-\theta_{w_t|r})}{\theta_{w_t|\overline{r}}(1-\theta_{w_t|\overline{r}})}\right)$$

$$RW(w_t) = \log\left(\frac{p(1-q)}{q(1-p)}\right)$$

BM25 (exact formula)

- Each word is a mixture of 2 Poisson distributions
  - tf = 0 , $w_t = 0$
  - monotonic in tf
  - asymptote approximates BIM

$$w_t = \log\left(\frac{(p'\lambda^{tf}e^{-\lambda} + (1-p')\mu^{tf}e^{-\mu})(q'e^{\lambda} + (1-q')e^{\mu})}{(q'\lambda^{tf}e^{-\lambda} + (1-q')\mu^{tf}e^{-\mu})(p'e^{\lambda} + (1-p')e^{\mu})}\right)$$

Robertson, S. E., & Walker, S. (1994). Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval. In Proceedings of SIGIR'94, pp. 232–241.

BM25 (approximated)

- TF and document length normalisation
- Pivoted normalisation

$$w(t, d_i) = \frac{TF(t, d_i)}{TF(t, d_i) + K} \log \left(\frac{p(1-q)}{q(1-p)} \right)$$

$$K = k_1((1-b) + b\ dl(d_i)/\Delta)$$

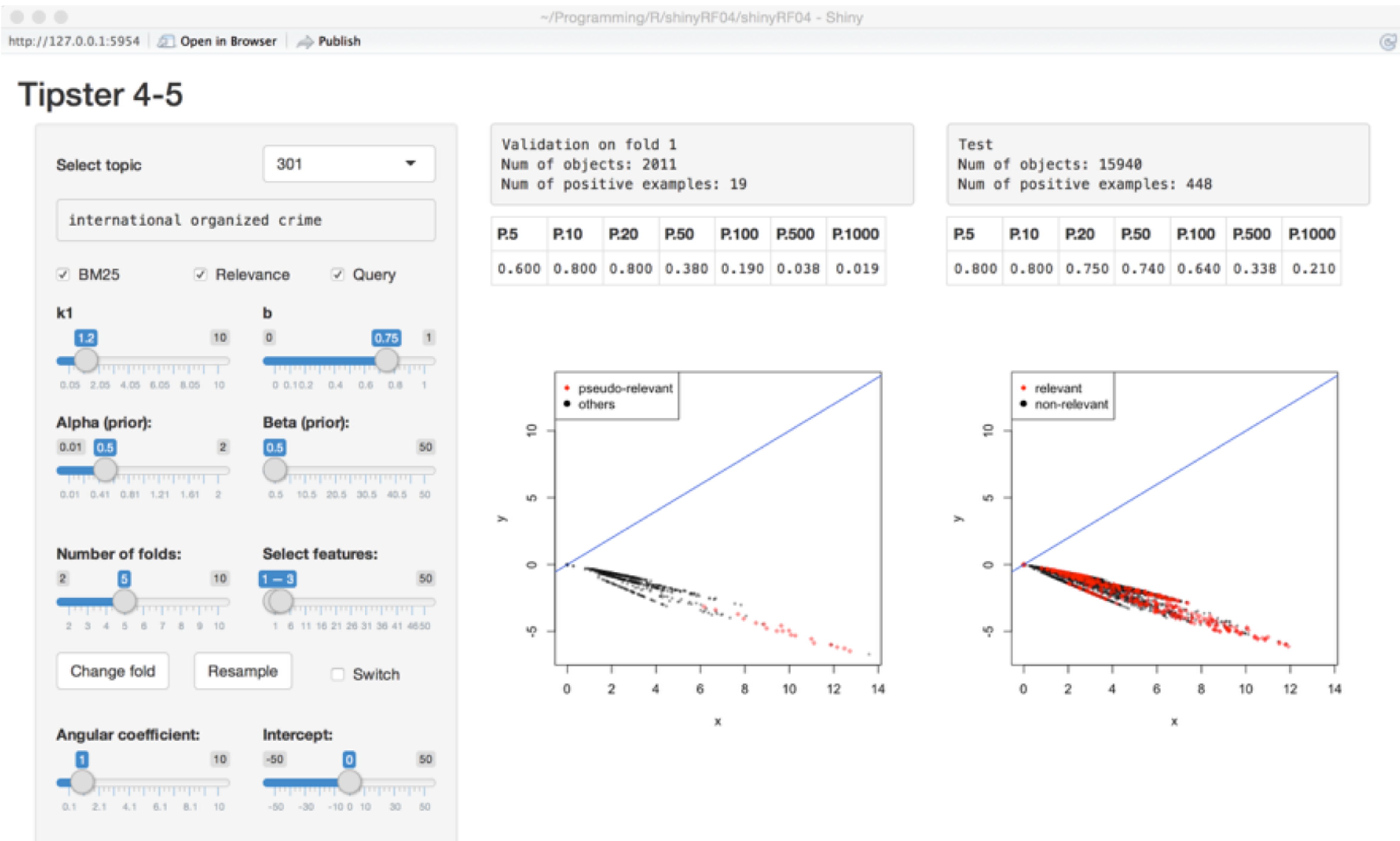
Singhal, A., Buckley, C., & Mitra, M. (1996). Pivoted document length normalization.
In Proceedings of SIGIR '96 (pp. 21–29).

Likelihood Spaces

- Coordinates of a two-dimensional space

$$\underbrace{\log(P(D|R=r))}_{X} > \underbrace{\log(P(D|R=\overline{r}))}_{Y} + \log\left(\frac{P(R=\overline{r})}{P(R=r)}\right)$$

Ranking in Likelihood spaces



Ranking in Likelihood Spaces

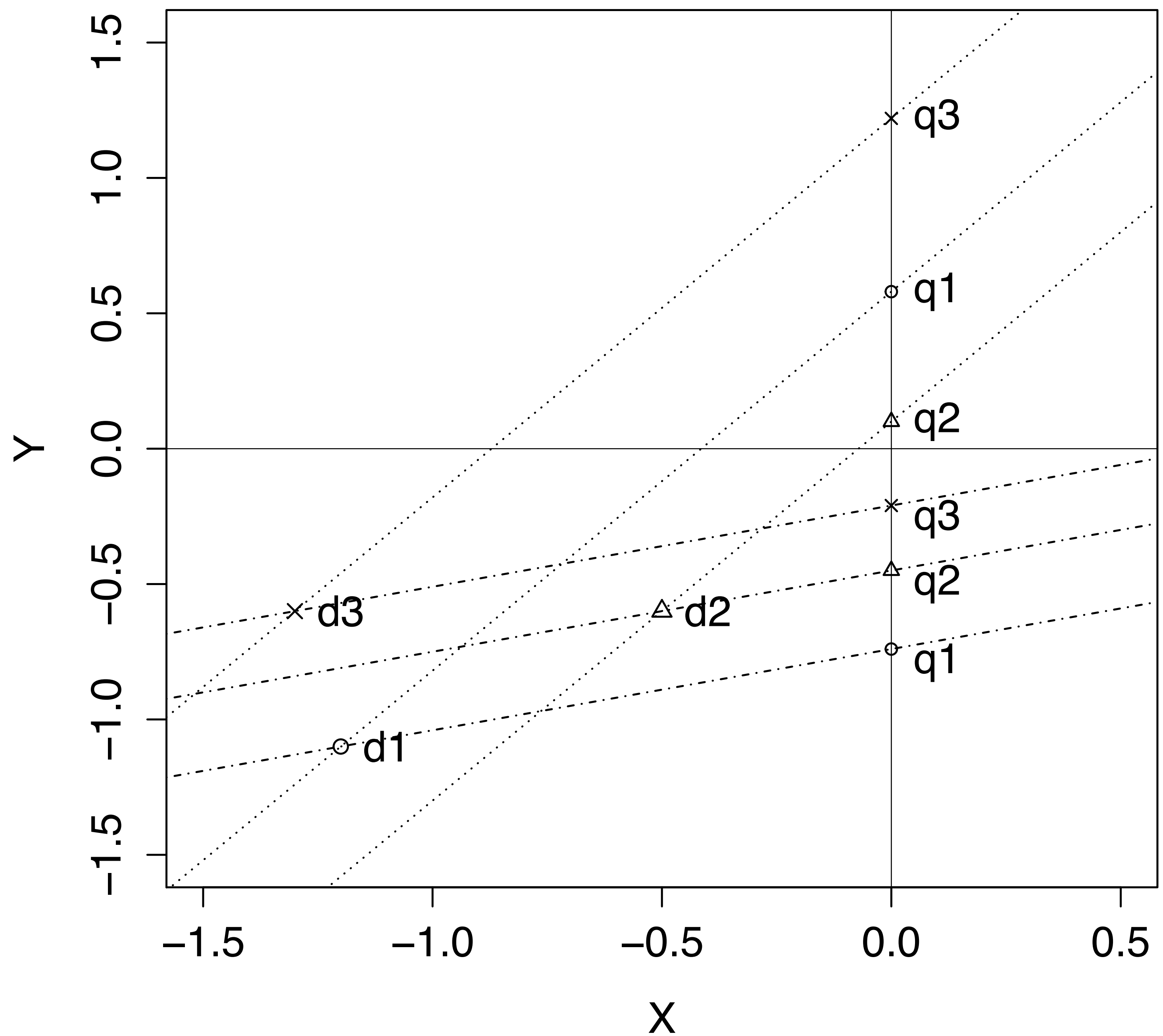
- Change slope M of the decision line

$$\underbrace{\log(P(D|R=r)}_{X} > \underbrace{\log(P(D|R=\overline{r})}_{Y}$$

$$M \underbrace{\log(P(D|R=r)}_{X} + Q > \underbrace{\log(P(D|R=\overline{r})}_{Y}$$

Giorgio Maria Di Nunzio. 2014. A new decision to take for cost-sensitive Naïve Bayes classifiers. *Inf. Process. Manage.* 50, 5 (September 2014), 653-674.

Ranking in Likelihood Spaces



Conclusions

- Study probabilistic models on two dimensions
- BM25 vs BIM
- BM25 and Bayesian Decision Theory
- Ranking Line and BDT