

Two Alternative Ways of Using LMs

- Classic Probabilistic Model :Doc-Generation as opposed to Query-generation

$$O(R=1|Q,D) \propto \frac{P(D|Q,R=1)}{P(D|Q,R=0)} \approx \frac{P(D|Q,R=1)}{P(D)}$$

- Natural for relevance feedback
- Challenge: Estimate $p(D|Q,R=1)$ without relevance feedback; relevance model [Lavrenko & Croft 01] provides a good solution
- Probabilistic Distance Model :Similar to the vector-space model, but with LMs as opposed to TF-IDF weight vectors
 - A popular distance function: Kullback-Leibler (KL) divergence, covering query likelihood as a special case
 - Retrieval is now to estimate query & doc models and feedback is treated as query LM updating [Lafferty & Zhai 01b; Zhai & Lafferty 01b]

$$score(Q,D) = -D(\theta_Q || \theta_D), \text{ essentially } \sum_{w \in V} p(w | \theta_Q) \log p(w | \theta_D)$$

Both methods outperform the basic LM significantly