On the Mathematical Relationship between Expected n-call@k and the Relevance vs. Diversity Trade-off

Kar Wai Lim (NICTA & the ANU), Scott Sanner (NICTA & the ANU), Shengbo Guo (Xerox Research Centre Europe)

Highlight

Background: Previous work shows that optimizing *expected 1-call@k* leads to a diverse retrieval algorithm, and relates to MMR with $\lambda = 0.5$. (However, in practice this is not be the best λ). [Sanner, Guo, Graepel, Kharazmi and Karimi, CIKM-11]

Objective: Extend the analysis from *1-call@k* to general *n-call@k*, in order to demonstrate a relationship between MMR's λ and n and k.

$$S_k^* \;=\; \operatorname*{argmax}_{S_k} \text{Exp-}n\text{-Call@}k(S_k, \mathbf{q}) \;=\; \operatorname*{argmax}_{S_k} \mathbb{E}\left[R_k \geq n \mid s_1, \ldots, s_k, \mathbf{q}\right]$$

Questions: 1: How do we optimize this objective?
2: How does the diversification level λ in MMR relate to n and k?

Result: We show that $\lambda = \frac{n}{n+1}$ for Exp-n-call@k.

Set-based Results benefiting from Diversity

- Recommender Systems
  - Books, Music, Movies
  - Real Estate / Apartments
  - Many other products
- Standard IR
  - Search Engine Results
- Text Summarization
- Ad Serving

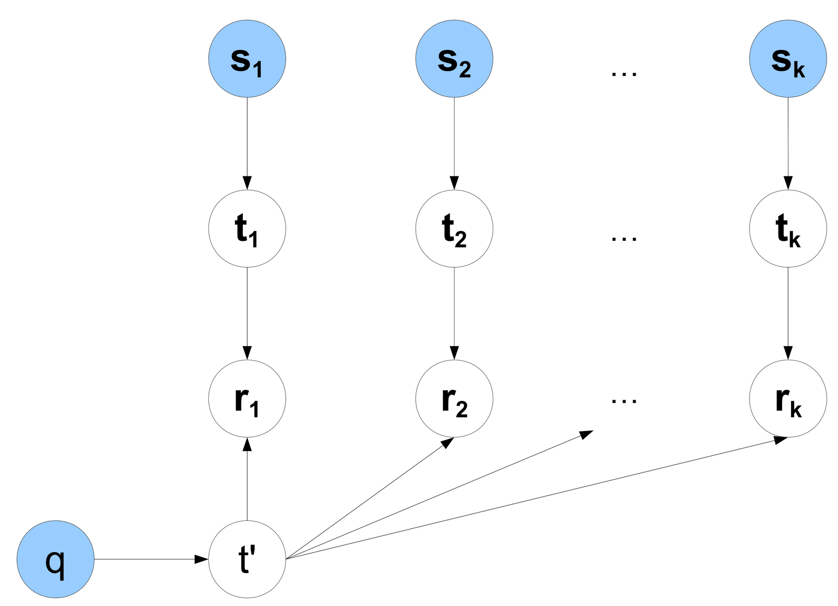
Principle: if one item is irrelevant, similar items may also be irrelevant.

Question: how to define (ir)relevance to account for inter-item similarity?

One answer: *via a latent subtopic relevance model.*

Latent Subtopic Relevance Model (LSRM)

D : item set (e.g., documents)
Q : query set
T : subtopic set (finite)
$s_i \in D$: selected item $i = 1 \ldots k$
$t_i \in T$: subtopic for i-th item
$\mathbf{q} \in Q$: observed query
$t' \in T$: subtopic for $\mathbf{q}$
$r_i \in \{0,1\}$: i-th item relevant?
$R_k = \sum_i^k r_i$: number of relevant items



Subtopics can be manually labeled (facets) or learned (via topic modeling).

$P(t_i|s_i) =$ distribution over latent subtopics t_i of selected item (document) s_i
$P(t'|\mathbf{q}) =$ distribution over latent subtopics t' of query

marginalizing latent subtopics gives LSI-like relevance:

$$P(r_i|t', t_i) = \mathbb{I}[t_i = t'] \;\rightarrow\; P(r_i|s_i, \mathbf{q}) = \sum_{t'} \sum_{t_i} \underbrace{P(r_i|t, t_i)}_{\mathbb{I}[t_i = t']} P(t_i|s_i) P(t'|\mathbf{q}) = \underbrace{\sum_t P(t|s_i) P(t|\mathbf{q})}_{[\vdots]\cdot[\vdots]}$$

Theoretical Contribution

Optimizing Expected n-call@k

Greedy optimization: Choose s_k assuming $S_{k-1}^* = \{s_1^*, \ldots, s_{k-1}^*\}$ (with topics $T_{k-1} = \{t_1, \ldots, t_{k-1}\}$) have been selected. Following Chen and Karger [SIGIR-06]:

$$\begin{aligned}
s_k^* &= \operatorname*{argmax}_{s_k} \text{Exp-}n\text{-Call@}k(S_{k-1}^* \cup \{s_k\}, \mathbf{q}) \\
&= \operatorname*{argmax}_{s_k} P(R_k \geq n | S_{k-1}^*, s_k, \mathbf{q}) \\
&= \operatorname*{argmax}_{s_k} \sum_{T_k} \left(P(t|\mathbf{q})\, P(t_k|s_k) \prod_{i=1}^{k-1} P(t_i|s_i^*) \cdot P(R_k \geq n | T_k, S_{k-1}^*, s_k, \mathbf{q}) \right) \\
&= \operatorname*{argmax}_{s_k} \sum_{T_k} P(t|\mathbf{q})\, P(t_k|s_k) \prod_{i=1}^{k-1} P(t_i|s_i^*) \cdot \Big(\underbrace{P(r_k \geq 0 | R_{k-1} \geq n, t_k, t)}_{1} P(R_{k-1} \geq n | T_{k\text{-}1}) \\
&\qquad + P(r_k = 1 | R_{k-1} = n-1, t_k, t) P(R_{k-1} = n-1 | T_{k\text{-}1}) \Big) \quad \text{[all } r_i \text{ are D-separated]} \\
&= \operatorname*{argmax}_{s_k} \Bigg(\sum_{T_{k\text{-}1}} \underbrace{\left[\sum_{t_k} P(t_k|s_k) \right]}_{1} P(R_{k-1} \geq n | T_{k\text{-}1}) P(t|\mathbf{q}) \prod_{i=1}^{k-1} P(t_i|s_i^*) \\
&\qquad + \sum_t P(t|\mathbf{q}) P(t_k = t|s_k) \sum_{t_1, \ldots, t_{k-1}} P(R_{k-1} = n-1 | T_{k\text{-}1}) \prod_{i=1}^{k-1} P(t_i|s_i^*) \Bigg) \quad \text{[sum over } t_k \text{]} \\
&= \operatorname*{argmax}_{s_k} \sum_t P(t|\mathbf{q}) P(t_k = t|s_k) P(R_{k-1} = n-1 | S_{k-1}^*, t) \quad \text{[dropping constant term w.r.t. } s_k \text{]}
\end{aligned}$$

The last probability $P(R_{k-1} = n-1 | S_{k-1}^*, t)$ is recursively defined:

$$P(R_k = n | S_k, t) = \begin{cases}
n \geq 1, k > 1: & \big(1 - P(t_k = t|s_k)\big) P(R_{k-1} = n | S_{k-1}, t) \\
& \quad + P(t_k = t|s_k) P(R_{k-1} = n-1 | S_{k-1}, t) \\
n = 0, k > 1: & \big(1 - P(t_k = t|s_k)\big) P(R_{k-1} = 0 | S_{k-1}, t) \\
n = 1, k = 1: & P(t_1 = t|s_1) \\
n = 0, k = 1: & 1 - P(t_1 = t|s_1)
\end{cases}$$

(derived *via* similar approach described above)

Unrolling the objective recursively we get (for $n < k/2$):

$$s_k^* = \operatorname*{argmax}_{s_k} \sum_t \left(P(t|\mathbf{q})\, P(t_k = t|s_k) \cdot \sum_{j_1, \ldots, j_{n-1}} \prod_{l \in \{j_1, \ldots, j_{n-1}\}} P(t_l = t|s_l^*) \prod_{\substack{i=1 \\ i \notin \{j_1, \ldots, j_{n-1}\}}}^{k-1} \left(1 - P(t_i = t|s_i^*)\right) \right)$$

(a symmetrical result holds for $n > k/2$)

Interpreting the Result as a Variant of MMR

- **Deterministic:** Subtopic probabilities are deterministic, *i.e.* each document covers only a single subtopic of the query.

  $\forall i\; P(t_i|s_i) \in \{0,1\}$ and $P(t|\mathbf{q}) \in \{0,1\}$

  This allows us to convert a $\prod$ to a $\max$. (see paper for details)
- **Number of relevant documents selected:** Denoting this as m, we assume that m is large and $m > n$.

$$\begin{aligned}
s_k &= \operatorname*{argmax}_{s_k} \sum_t \Bigg(P(t|\mathbf{q}) P(t_k = t|s_k) \sum_{j_1, \ldots, j_{n-1}} \prod_{l \in \{j_1, \ldots, j_{n-1}\}} P(t_l = t|s_l^*) \\
&\qquad - P(t|\mathbf{q}) P(t_k = t|s_k) \sum_{j_1, \ldots, j_{n-1}} \prod_{l \in \{j_1, \ldots, j_{n-1}\}} P(t_l = t|s_l^*) \max_{\substack{i \in [1, k-1] \\ i \notin \{j_1, \ldots, j_{n-1}\}}} P(t_i = t|s_i^*) \Bigg) \\
&= \operatorname*{argmax}_{s_k} \binom{m}{n-1} \underbrace{\sum_t P(t|\mathbf{q}) P(t_k = t|s_k)}_{\textbf{relevance}:\ \text{Sim}_1(s_k, \mathbf{q})} - \binom{m}{n} \max_{s_i \in S_{k-1}^*} \underbrace{\sum_t P(t_i = t|s_i) P(t|\mathbf{q}) P(t_k = t|s_k)}_{\textbf{diversity}:\ \text{Sim}_2(s_k, s_i, \mathbf{q})} \\
&= \operatorname*{argmax}_{s_k} \underbrace{\left(\frac{n}{m+1}\right)}_{\lambda} \text{Sim}_1(s_k, \mathbf{q}) - \underbrace{\left(\frac{m-n+1}{m+1}\right)}_{1-\lambda} \max_{s_i \in S_{k-1}^*} \text{Sim}_2(s_k, s_i, \mathbf{q}) \quad \text{[after normalized]}
\end{aligned}$$

[Refer to paper's appendix for a full derivation]

Comparison to Maximal Marginal Relevance

Maximal Marginal Relevance (MMR) [Carbonell and Goldstein, SIGIR-98]:

$$s_k^* = \operatorname*{argmax}_{s_k \in D \setminus S_{k-1}^*} \left[\lambda(\text{Sim}_1(\mathbf{q}, s_k)) - (1-\lambda) \max_{s_i \in S_{k-1}^*} \text{Sim}_2(s_i, s_k)\right]$$

Sim_1, Sim_2 kernels for query/item similarity; $\lambda \in [0,1]$ trades off relevance & diversity.

- **Set diversity function:** MMR uses $\max$, Exp-n-Call@k uses Π — in the special case that subtopic probabilities are deterministic then Π equivalent to $\max$.
- **Similarity and diversity kernels:** Exp-n-Call@k supports popular MMR kernels: LSI directly; L_1 normalized TF and TFIDF if words are equated to subtopics.
- **Query re-weighted diversity:** Exp-n-Call@k *reweights* diversity by $P(t'|\mathbf{q})$.
- **λ relevance diversity tradeoff:** Assuming $m \approx n$ since Exp-n-Call@k optimizes for the case where n relevant documents are selected, $\lambda = \frac{n}{n+1}$.

Summary

- **Derived an MMR-like algorithm** by optimizing Exp-n-call@k in an LSRM.
- **Relevance vs. diversity tradeoff** expressed by a function of n: $\lambda = \frac{n}{n+1}$.
- **Diversification level** decreases linearly as n increases.