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

Kar Wai Lim

26 November 2012

with Scott Sanner and Shengbo Guo, SIGIR 2012

Outline

- Background and previous works
- How to derive MMR

An example

Full coverage

NAB to customers: you're the voice on security
Sydney Morning Herald - 1 hour ago
National Australia Bank will begin using voice recognition **technology** to identify its phone customers in the latest move towards the use of biometric security among the big banks. The company said that the **technology**, which identifies a person by their speech ...

NAB speaks loud and clear on voice biometrics
Technology Spectator - 2 hours ago
National Australia Bank (NAB) has joined its peer ANZ Banking Group in touting biometrics as a viable replacement to PINs, with the bank's ambitions focused on voice rather than fingerprint recognition. The move comes hot on the heels of ANZ's recent ...

NAB to shift online banking platform
The Australian - 8 hours ago
NATIONAL Australia Bank's popular internet banking platform could have a new home within six months thanks to a significant **technology** upgrade, a senior company executive said. The development comes as the bank announced plans to further cement its ...

Voice recognition **technology** for NAB
Ninemsn - 11 hours ago
Voice recognition **technology** for NAB. 2:07am November 21, 2012. National Australia Bank will become the first major Australian company to roll out voice recognition **technology**, with plans to introduce it next year. Close calls for journalists caught on video ...

Money talks in hi-tech banking
Courier Mail - 7 hours ago
The **technology** is expected to save individual customers three minutes each phone call. NAB executive general manager Adam Bennett said, when fully deployed, Speech Security would save the bank's customers a combined 15 million minutes a year.

NAB deploys customer data aggregator
iT News - 7 hours ago
Chief **technology** officer Denis McGee said the bank had struck "consumption-based" managed services contracts with key suppliers IBM and Telstra. He told iTnews that the vendors typically already had excess capacity – such as bandwidth on existing fibre ...

NAB phone banking will match customers' voices
Banking Day (registration) - 6 hours ago
After first experimenting with the **technology** in 2009, NAB has quietly enrolled 140,000 customers to trial its system. Essentially, the system authenticates the identity of a person calling into NAB's contact centre by matching the person's voice against a voice ...

- Assume current top news is about NAB’s voice recognition technology. We get the search results by querying “technology”.
- Is this desirable?
- We don’t want to get a page full of similar or duplicate news (variant from different sources).

Another example

Apple

- Is this better?

Diversity

- From these examples we can see that diversity is important.
- How can we achieve this?
  – Maximum marginal relevance (MMR)
    - Carbonell & Goldstein, SIGIR 1998
    - Select set S (with K items) from all items set D
    - Choose item greedily until |S| = K

$$s_k^* = \underset{s_k \in D \setminus S_{k-1}^*}{\arg\max} \left[\lambda(\mathrm{Sim}_1(\mathbf{q}, s_k)) - (1-\lambda) \max_{s_i \in S_{k-1}^*} \mathrm{Sim}_2(s_i, s_k)\right]$$

Problem

- MMR is an algorithm, we don't really know what underlying objective that it is optimising.
- There are some previous attempts but full problem remained unsolved for 13 years.

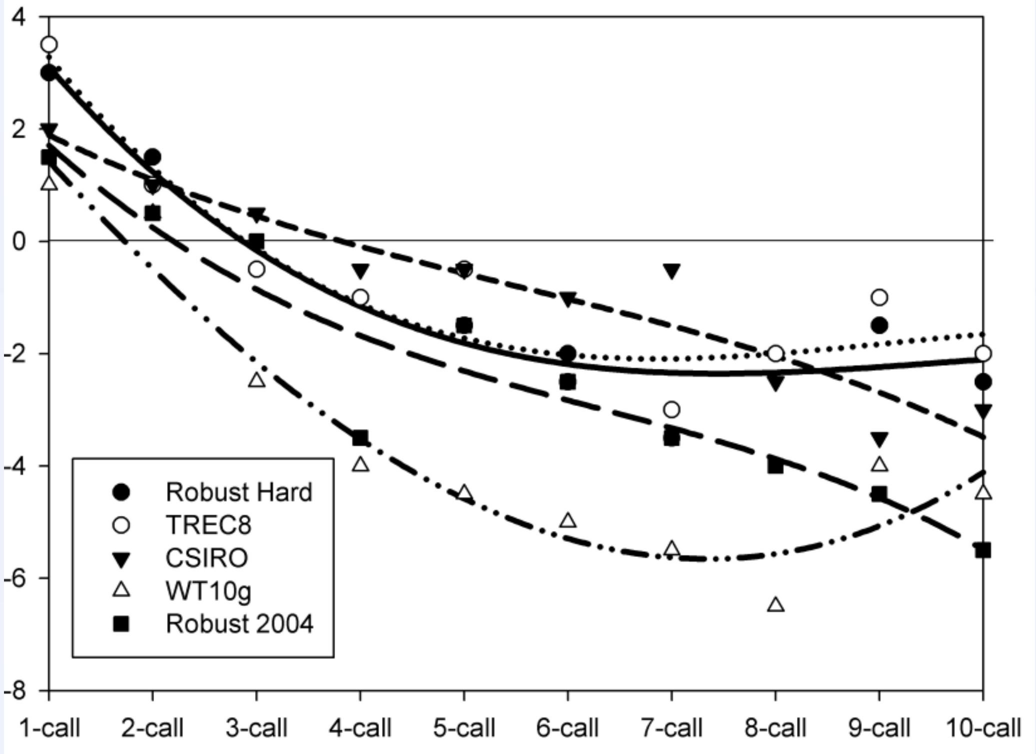
- What objectives would lead to diverse retrieval? (such as MMR)

Problem

- Probability Ranking Principle (PRP)
  - Greedily choose items that are most relevant (potentially gives us the first example before)
- Another extreme is 1-call@k
  - Happy as long as at least 1 item is relevant
  - Diverse!
- Previous work shows that 1-call@k corresponds to MMR with $\lambda = \frac{1}{2}$
  - But in MMR tuning λ is important, is there another objective that leads to tunable λ that modulates diversity ?

Problem

- What about n-call@k?



J. Wang and J. Zhu. Portfolio theory of information retrieval, SIGIR 2009

Hypothesis

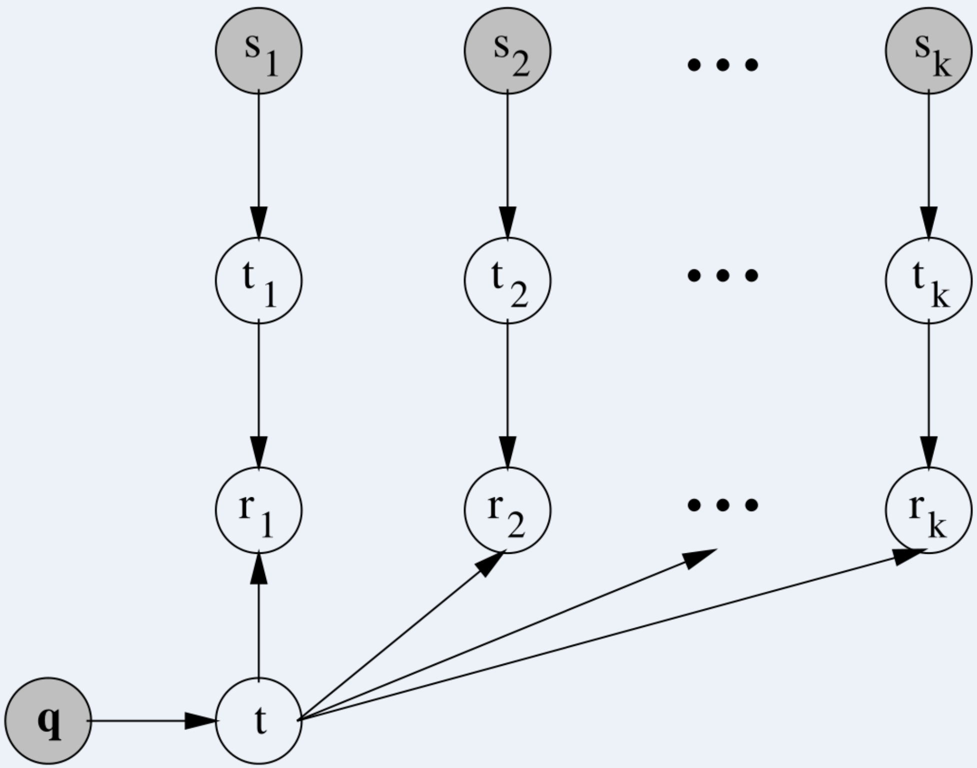
- Start with 2-call@k
  - optimising this leads to MMR with $\lambda = 2/3$
- There seems to be a trend relating λ and n

- Hypothesis
  - Optimising n-call@k leads to MMR with $\lambda = n/(n+1)$

Outline

- ~~Background and previous work~~
- How to derive MMR

Graphical model of Relevance



s = selected docs

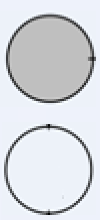
t = subtopics ∈ T

r = relevance ∈ {0, 1}

q = observed query

T = discrete subtopic set

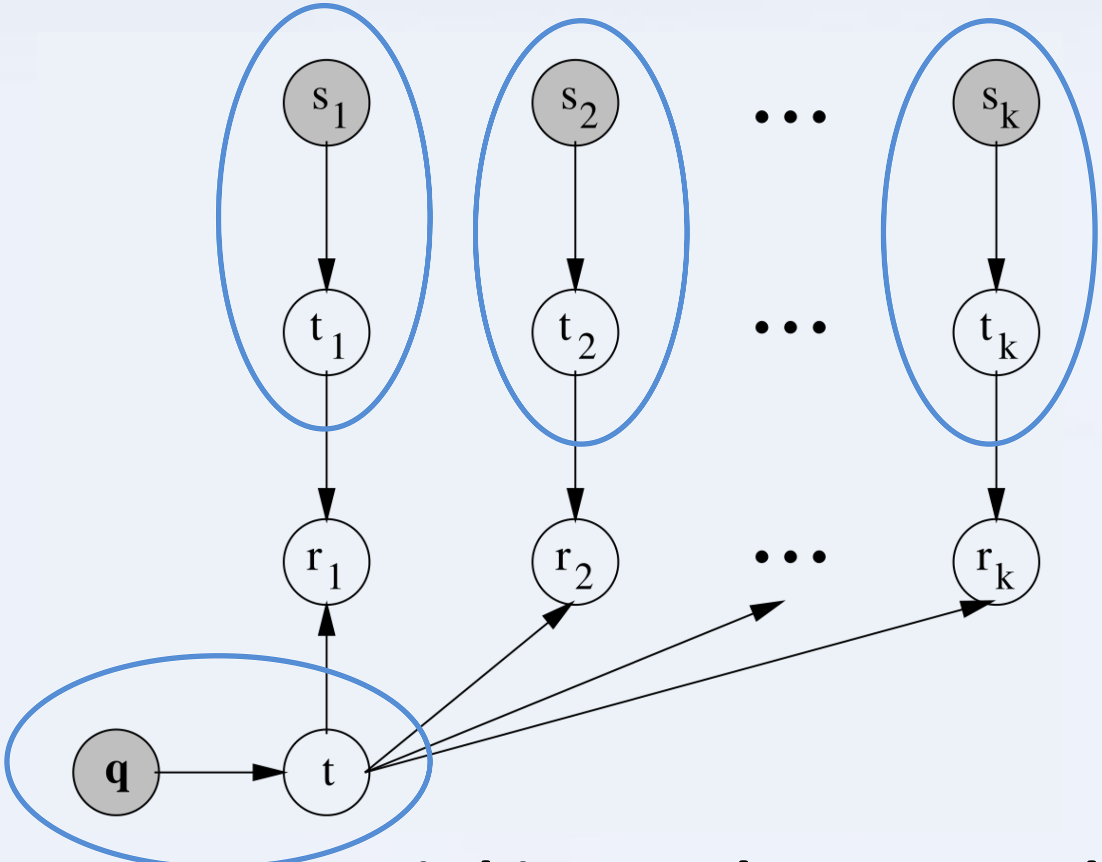
Latent subtopic binary relevance model



Observed

Latent (unobserved)

Graphical model of Relevance



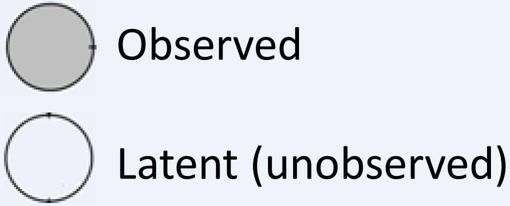
Latent subtopic binary relevance model

$P(t_i = C|s_i)$

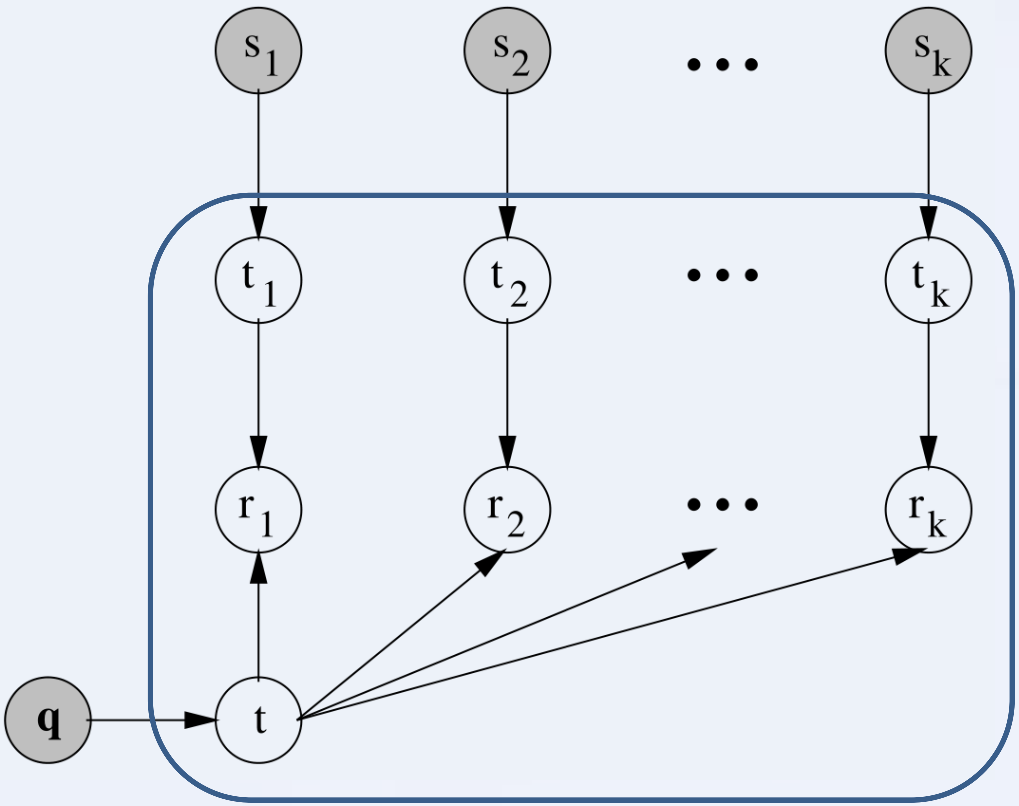
= prob. of document s belongs to subtopic C

$P(t = C|\mathbf{q})$

= prob. of query **q** refer to subtopic C



Graphical model of Relevance



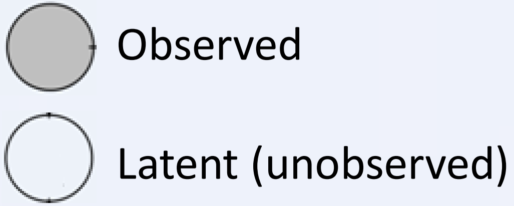
Latent subtopic binary relevance model

If $t_i = t$:

$P(r_i=1|t_i,t) = 1$

Else:

$P(r_i=1|t_i,t) = 0$



Optimising Objective

- Expected n-call@k objective:

$$\text{Exp-}n\text{-Call@}k(S_k, \mathbf{q}) = \mathbb{E}[R_k \geq n | s_1, \ldots, s_k, \mathbf{q}]$$

$$R_k = \sum_{i=1}^{k} r_i$$

- We want at least n out of the chosen k documents to be relevant, by choosing **s** that maximises the objective.
- Note that jointly optimise **s** is NP-hard.

Greedy approach

- We choose the documents consecutively with a greedy approach.
  – select the next document given all previously chosen documents.

$$s_k^* = \arg\max_{s_k} \mathbb{E}[R_k \geq n | S_{k-1}^*, s_k, \mathbf{q}]$$

Derivation

- Nontrivial
  - I will explain at high level and highlight the main mathematical tricks that are used.
  - Rather than going through the details step by step.

Derivation

$$
\begin{aligned}
s_k^* &= \arg\max_{s_k} \mathbb{E}[R_k \geq n | S_{k-1}^*, s_k, \mathbf{q}] \\
&= \arg\max_{s_k} P(R_k \geq n | S_{k-1}^*, s_k, \mathbf{q})
\end{aligned}
$$

Derivation

$$
\begin{aligned}
s_k^* &= \arg\max_{s_k} \mathbb{E}[R_k \geq n | S_{k-1}^*, s_k, \mathbf{q}] \\
&= \arg\max_{s_k} P(R_k \geq n | S_{k-1}^*, s_k, \mathbf{q}) \\
&= \arg\max_{s_k} \sum_{T_k} \Big(P(t|\mathbf{q})\, P(t_k|s_k) \prod_{i=1}^{k-1} P(t_i|s_i^*) \\
&\qquad \cdot P(R_k \geq n | T_k, S_{k-1}^*, s_k, \mathbf{q}) \Big)
\end{aligned}
$$

Marginalise out all subtopics (using conditional probability)

$T_k = \{t, t_1, \ldots, t_k\}$ and $\sum_{T_k} \circ = \sum_t \sum_{t_1} \cdots \sum_{t_k} \circ$

Derivation

$$
\begin{aligned}
s_k^* &= \arg\max_{s_k} \mathbb{E}[R_k \geq n | S_{k-1}^*, s_k, \mathbf{q}] \\
&= \arg\max_{s_k} P(R_k \geq n | S_{k-1}^*, s_k, \mathbf{q}) \\
&= \arg\max_{s_k} \sum_{T_k} \Big(P(t|\mathbf{q})\, P(t_k|s_k) \prod_{i=1}^{k-1} P(t_i|s_i^*) \\
&\qquad \cdot P(R_k \geq n | T_k, S_{k-1}^*, s_k, \mathbf{q}) \Big) \\
&= \arg\max_{s_k} \sum_{T_k} P(t|\mathbf{q})\, P(t_k|s_k) \prod_{i=1}^{k-1} P(t_i|s_i^*) \\
&\qquad \cdot \Big(\underbrace{P(r_k \geq 0 | R_{k-1} \geq n, t_k, t)}_{1} P(R_{k-1} \geq n | T_{k\text{-}1}) \\
&\qquad + P(r_k = 1 | R_{k-1} = n-1, t_k, t) P(R_{k-1} = n-1 | T_{k\text{-}1}) \Big)
\end{aligned}
$$

We write r_k as conditioned on R_{k-1}.

Note that relevance **r** are independent given the subtopics **t**.

Derivation

$$
\begin{aligned}
s_k^* =& \arg\max_{s_k} \mathbb{E}[R_k \geq n | S_{k-1}^*, s_k, \mathbf{q}] \\
=& \arg\max_{s_k} P(R_k \geq n | S_{k-1}^*, s_k, \mathbf{q}) \\
=& \arg\max_{s_k} \sum_{T_k} \Big(P(t|\mathbf{q})\, P(t_k|s_k) \prod_{i=1}^{k-1} P(t_i|s_i^*) \\
& \qquad \cdot P(R_k \geq n | T_k, S_{k-1}^*, s_k, \mathbf{q}) \Big) \\
=& \arg\max_{s_k} \sum_{T_k} P(t|\mathbf{q})\, P(t_k|s_k) \prod_{i=1}^{k-1} P(t_i|s_i^*) \\
& \quad \cdot \Big(\underbrace{P(r_k \geq 0 | R_{k-1} \geq n, t_k, t)}_{1} P(R_{k-1} \geq n | T_{k\text{-}1}) \\
& \qquad + P(r_k = 1 | R_{k-1} = n-1, t_k, t) P(R_{k-1} = n-1 | T_{k\text{-}1}) \Big) \\
=& \arg\max_{s_k} \Big(\sum_{T_{k\text{-}1}} \underbrace{\Big[\sum_{t_k} P(t_k|s_k) \Big]}_{1} P(R_{k-1} \geq n | T_{k\text{-}1}) P(t|\mathbf{q}) \prod_{i=1}^{k-1} P(t_i|s_i^*) + \\
& \sum_{t} P(t|\mathbf{q}) P(t_k = t | s_k) \sum_{t_1,\ldots,t_{k-1}} P(R_{k-1} = n-1 | T_{k\text{-}1}) \prod_{i=1}^{k-1} P(t_i|s_i^*) \Big)
\end{aligned}
$$

$$
\begin{aligned}
& \textstyle\sum_{t_k} P(t_k|s_k) P(r_k{=}1 | t_k, t) \\
=& \textstyle\sum_{t_k} P(t_k|s_k) \mathbb{I}[t_k{=}t] = P(t_k{=}t | s_k)
\end{aligned}
$$

Sum over t_k

Derivation

$$
\begin{aligned}
s_k^* &= \arg\max_{s_k} \mathbb{E}[R_k \geq n | S_{k-1}^*, s_k, \mathbf{q}] \\
&= \arg\max_{s_k} P(R_k \geq n | S_{k-1}^*, s_k, \mathbf{q}) \\
&= \arg\max_{s_k} \sum_{T_k} \Big(P(t|\mathbf{q})\, P(t_k|s_k) \prod_{i=1}^{k-1} P(t_i|s_i^*) \\
&\qquad \cdot P(R_k \geq n | T_k, S_{k-1}^*, s_k, \mathbf{q}) \Big) \\
&= \arg\max_{s_k} \sum_{T_k} P(t|\mathbf{q})\, P(t_k|s_k) \prod_{i=1}^{k-1} P(t_i|s_i^*) \\
&\qquad \cdot \Big(\underbrace{P(r_k \geq 0 | R_{k-1} \geq n, t_k, t)}_{1} P(R_{k-1} \geq n | T_{k\text{-}1}) \\
&\qquad + P(r_k = 1 | R_{k-1} = n-1, t_k, t) P(R_{k-1} = n-1 | T_{k\text{-}1}) \Big) \\
&= \arg\max_{s_k} \Big(\sum_{T_{k\text{-}1}} \underbrace{\Big[\sum_{t_k} P(t_k|s_k) \Big]}_{1} P(R_{k-1} \geq n | T_{k\text{-}1}) P(t|\mathbf{q}) \prod_{i=1}^{k-1} P(t_i|s_i^*) + \\
&\qquad \sum_t P(t|\mathbf{q}) P(t_k = t|s_k) \sum_{t_1,\ldots,t_{k-1}} P(R_{k-1} = n-1 | T_{k\text{-}1}) \prod_{i=1}^{k-1} P(t_i|s_i^*) \Big) \\
&= \arg\max_{s_k} \sum_t P(t|\mathbf{q}) P(t_k = t|s_k) P(R_{k-1} = n-1 | S_{k-1}^*)
\end{aligned}
$$

dropping the first line

Derivation

- We arrive at

$$= \arg\max_{s_k} \sum_t P(t|\mathbf{q}) P(t_k = t | s_k) P(R_{k-1} = n-1 | S^*_{k-1})$$

- This is still a complicated term, but it can be expressed recursively.

Recursion

$$P(R_k = n|S_k, t) =$$

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

This is derived using method that are very similar to previous derivation.

Explicit expression

- We then unroll the optimising objective recursively to arrive at the explicit expression

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

$$n \leq k/2$$

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

$$n > k/2$$

where $j_1, \ldots, j_{n-1} \in \{1, \ldots, k-1\}$ satisfy that $j_i < j_{i+1}$

Trick 1: Convert to max

- To further simplify the objective, we assume that the subtopics of each document are known (deterministic), hence:

$$P(t_i|s_i) \in \{0, 1\}$$

- where in general the probability is between 0 and 1.
- Example next slide.

Trick 1: Convert to max

- Generally:

$$\begin{bmatrix} P(t_i = C_1|s_i) \\ P(t_i = C_2|s_i) \\ \vdots \\ P(t_i = C_{|T|}|s_i) \end{bmatrix} = \begin{bmatrix} 0.24 \\ 0.62 \\ \vdots \\ 0.01 \end{bmatrix}$$

- Deterministic:

$$\begin{bmatrix} P(t_i = C_1|s_i) \\ P(t_i = C_2|s_i) \\ \vdots \\ P(t_i = C_{|T|}|s_i) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

Trick 1: Convert to max

- This assumption allows us to convert a product $\prod$ to a max:

$$x_i \in \{0,1\}$$
$$\prod x_i = 0 \text{ iff at least } 1\ x_i = 0$$
$$\prod(1-x_i) = 0 \text{ iff at least } 1\ x_i = 1$$
$$1 - \prod(1-x_i) = 1 \text{ iff at least } 1\ x_i = 1$$

$$\text{also } \max x_i = 1 \text{ iff at least } 1\ x_i = 1$$
$$\text{hence they are equivalent (when } x_i \in \{0,1\})$$

Trick 1: Convert to max

- From the optimising objective when $n \leq k/2$, we can write

$$\prod_{\substack{i=1 \\ i \notin \{j_1,\ldots,j_{n-1}\}}}^{k-1} \big(1 - P(t_i = t | s_i^*)\big) = 1 - \left(1 - \prod_{\substack{i=1 \\ i \notin \{j_1,\ldots,j_{n-1}\}}}^{k-1} \big(1 - P(t_i = t | s_i^*)\big)\right)$$

$$= 1 - \Big(\max_{\substack{i \in [1,k-1] \\ i \notin \{j_1,\ldots,j_{n-1}\}}} P(t_i = t | s_i^*)\Big)$$

After Trick 1

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

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

Trick 2: combinatory simplification

- Assuming that m documents out of the chosen (k-1) are relevant, then $\sum_{j_1,\ldots,j_{n-1}} \prod_{l\in\{j_1,\ldots,j_{n-1}\}} P(t_l=t|s_l^*)$ (the top term) are non-zero $\binom{m}{n-1}$ times.
- $\sum_{j_1,\ldots,j_{n-1}} \prod_{l\in\{j_1,\ldots,j_{n-1}\}} P(t_l=t|s_l^*) \max_{\substack{i\in[1,k-1] \\ i\notin\{j_1,\ldots,j_{n-1}\}}} P(t_i=t|s_i^*)$ (bottom term) are non-zero $\binom{m}{n}$ times.

Final form

- After applying trick 2 and some manipulation, we derive the objective

$$= \arg\max_{s_k} \binom{m}{n-1} \underbrace{\sum_t P(t|\mathbf{q}) P(t_k = t|s_k)}_{\text{relevance: } \mathrm{Sim}_1(s_k, \mathbf{q})} - \binom{m}{n} \max_{s_i \in S^*_{k-1}} \underbrace{\sum_t P(t_i = t|s_i) P(t|\mathbf{q}) P(t_k = t|s_k)}_{\text{diversity: } \mathrm{Sim}_2(s_k, s_i, \mathbf{q})}$$

$$= \arg\max_{s_k} \frac{n}{m+1} \mathrm{Sim}_1(s_k, \mathbf{q}) - \frac{m-n+1}{m+1} \max_{s_i \in S^*_{k-1}} \mathrm{Sim}_2(s_k, s_i, \mathbf{q})$$

Using Pascal rule to normalise: $\binom{m}{n-1} + \binom{m}{n} = \binom{m+1}{n}$

Comparison to MMR

- The optimising objective used in MMR is

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

- We note that the optimising objective for expected n-call@k has the same form as MMR, with $\lambda = \frac{n}{m+1}$.
  - but m is unknown

Expected value for m

- Note that under expected n-call@k's greedy algorithm, we would expect m to be approximately equal to n after choosing k-1 documents (note that k >> n).
- Hence replacing m by n gives us $\lambda = \frac{n}{n+1}$.
  - Our hypothesis!

Our contributions

- We show the first derivation of MMR from first principle.
  - MMR optimises expected n-call@k
  - Analyse if MMR is appropriate for a given problem

- This framework can be used to derive new diversification algorithms by changing
  - the model
  - the objective
  - the assumptions

Under certain assumptions,
MMR optimises
expected n-call@k