Probabilistic Latent Maximal Marginal Relevance

NICTA

Shengbo Guo, Scott Sanner (Statistical Machine Learning Group, NICTA)

Highlight

Objective The first fully principled derivation of a criterion that balances relevance versus diversity of search results in information retrieval.

Given a query $\vec{q}$ and a corpus D, we want to find a set of documents S^*,

$$S^* = \operatorname*{argmax}_S P(\bigvee_{i=1}^{k} r_i = 1 | s_1, \ldots, s_k, \vec{q})$$

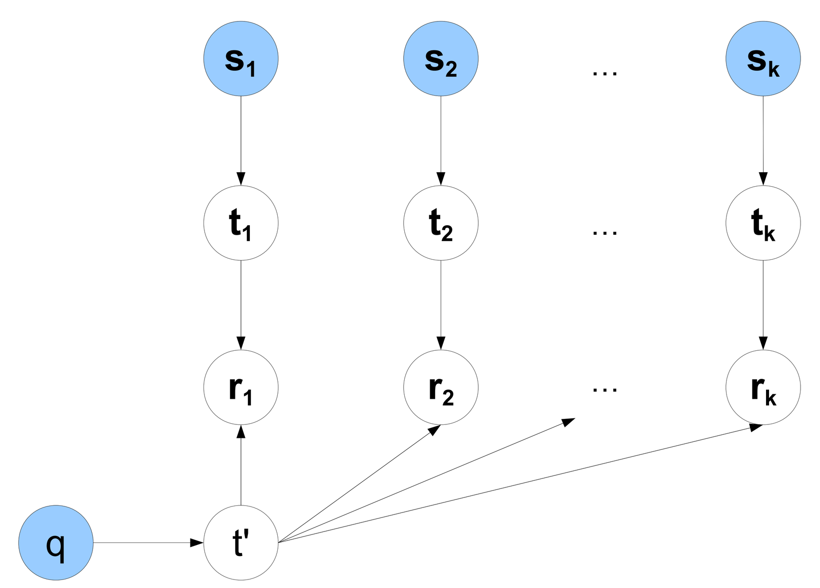
Approach 1: Set relevance objective to optimize,
2: Take a greedy approach like maximal marginal relevance (MMR).

Set-based Results benefiting from Diversity

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

Graphical Model of Relevance

s_i : document selection $i = 1, \ldots, k$
t_i : topic for i-th document
r_i : i-th document relevant?
$\vec{q}$: query
t' : topic for $\vec{q}$



Conditional probability tables for topic distributions of document and query, $P(t_i|s_i)$ and $P(t'|\vec{q})$, can be learnt via latent Dirichlet allocation.

Relevance Definition

The relevance of each document depends on the topics for this document and query. Specifically, we define

$$P(r_i|t', t_i) = \begin{cases} 1 & \text{if } t_i = t', \\ 0 & \text{otherwise} \end{cases}$$

Objective to Optimize

Set Relevance Objective to Optimize

The optimization objective is to maximize the set relevance of the query $\vec{q}$ with a set of documents S, given as below:

$$\begin{aligned} S^* &= \operatorname*{argmax}_{S=\{s_1,\ldots,s_k\}} P(\bigvee_{i=1}^{k} r_i = 1 | s_1, \ldots, s_k, \vec{q}) \\ &= P(r_1 = 1 \vee [r_1 = 0 \wedge r_2 = 1] \vee [r_1 = 0 \wedge r_2 = 0 \wedge r_3 = 1] \vee | s_1, \ldots, s_k, \vec{q}) \\ &= \sum_{i=1}^{k} P(r_i = 1, r_1 = 0, \ldots, r_{i-1} = 0 | s_1, \ldots, s_k, \vec{q}) \\ &= \sum_{i=1}^{k} P(r_i = 1 | r_1 = 0, \ldots, r_{i-1} = 0 | s_1, \ldots, s_k, \vec{q}) \underbrace{P(r_1 = 0, \ldots, r_{i-1} = 0 | s_1, \ldots, s_k, \vec{q})}_{s_k \text{ D-separated from } r_1,\ldots,r_{k-1}} \end{aligned}$$

We thus can be greedy when selecting s_i^* using

$$s_i^* = \operatorname*{argmax}_{s_i} P(r_i = 1 | r_1 = 0, \ldots, r_{i-1} = 0, s_1^*, \ldots, s_{i-1}^*, s_i, \vec{q})$$

Objective to Optimize: s_1^*

- Take a greedy approach (like MMR)
- Choose s_1^* first

$$\begin{aligned} s_1^* &= P(r_1 | s_1, \vec{q}) \\ &= \operatorname*{argmax}_{s_1} \sum_{t_1, t'} I[t' = t_1] P(t'|\vec{q}) P(t_1|s_1) \\ &= \operatorname*{argmax}_{s_1} \sum_{t'} P(t'|\vec{q}) P(t_1 = t'|s_1) \end{aligned}$$

- Note that $\sum_{t'} P(t'|\vec{q}) P(t_1 = t'|s_1)$ is actually in the form of a dot product, $[\vdots] \cdot [\vdots]$
- Binary relevance derivation of LSI kernel

Objective to Optimize: s_2^*

Choose s_2^* via Accumulated Relevance (AccRel) next, conditioning on s_1 and $r_1 = 0$:

$$\begin{aligned} s_2^* &= \operatorname*{argmax}_{s_2} P(r_2 | r_1 = 0, s_1^*, s_2, \vec{q}) \\ &= \operatorname*{argmax}_{s_2} \sum_{t_1, t_2, t'} I[t_2 = t'] I[t_1 \neq t'] P(t_1|s_1^*) P(t_2|s_2) P(t'|\vec{q}) \\ &= \operatorname*{argmax}_{s_2} \sum_{t_2, t'} I[t_2 = t'] P(t_2|s_2) P(t'|\vec{q}) \underbrace{\sum_{t_1} I[t_1 \neq t'] P(t_1|s_1^*)}_{1 - P(t_1 = t'|s_1^*)} \\ &= \operatorname*{argmax}_{s_2} \sum_{t'} (1 - P(t_1 = t'|s_1^*)) P(t'|\vec{q}) \underbrace{\sum_{t_2} I[t_2 = t'] P(t_2|s_2)}_{P(t_2 = t'|s_2)} \\ &= \operatorname*{argmax}_{s_2} \underbrace{\left[\sum_{t'} P(t'|\vec{q}) P(t_2 = t'|s_2)\right]}_{\text{Relevance}} - \underbrace{\left[\sum_{t'} P(t'|\vec{q}) P(t_1 = t'|s_1^*) P(t_2 = t'|s_2)\right]}_{\text{Query topic-weighted diversity}} \end{aligned}$$

which suggests the title *Probabilistic Latent Maximal Marginal Relevance*.

Experimental Results

Evaluate using weighted subtopic loss (WSL) of three methods using all words and first 10 words on a subset of TREC 6-8 data focusing on diversity. Standard error estimates are shown for PLMMR-LDA.

Method	WSL (first 10 words)	WSL (all words)
MMR-TF	0.555	0.534
MMR-TFIDF	0.549	0.493
PLMMR-LDA	**0.458** $\pm$ **0.0058**	**0.468** $\pm$ **0.0019**

Summary

We proposed a binary set-relevance model of diversity, which derived

- LSI kernel
- LSI diversity kernel
- Probabilistic variant and justification of MMR

Acknowledgement

We thank Thore Graepel for important derivations and interesting discussions, and thank the anomalous reviewers for their comments.