Learning For Search Result Diversification

Yadong Zhu
Institute of Computing Technology,
Chinese Academy of Sciences
edgewind11@gmail.com

Co-authors: Yanyan Lan, Jiafeng Guo, Shuzi Niu, Xueqi Cheng

Outline

- Motivation
- Our Approach
- Experiments
- Conclusion

Motivation

- Different user needs
  - Ambiguous queries
    - Apple, Jaguar, Band…
  - Multi-faceted needs
    - Britney spears (news, videos, photos…)
- Information redundancy
  - Many duplicate or similar results

Motivation

- Existing approaches

Non-learning

- ✓ MMR
- ✓ IA-Select
- ✓ xQuAD
- ✓ PM-2
- ✓ …

Learning-based

- ✓SVMDIV

Non-learning Methods

- Typical methods
  - MMR: Maximal Marginal Relevance
    - Predefined utility function

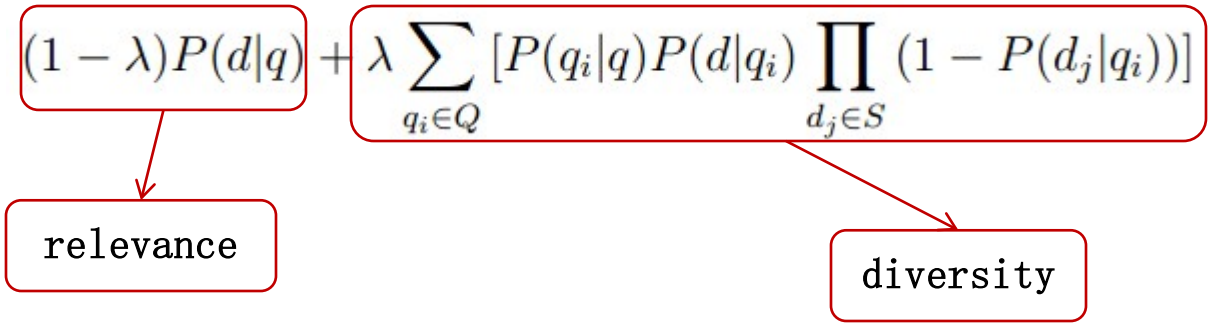
$$MMR \stackrel{def}{=} Arg \max_{d_i \in R \setminus S} [\lambda Sim_1(d_i, q) - (1-\lambda) \max_{d_j \in S} Sim_2(d_i, d_j)]$$

Query Relevance

Similarity with selected documents

Non-learning Methods

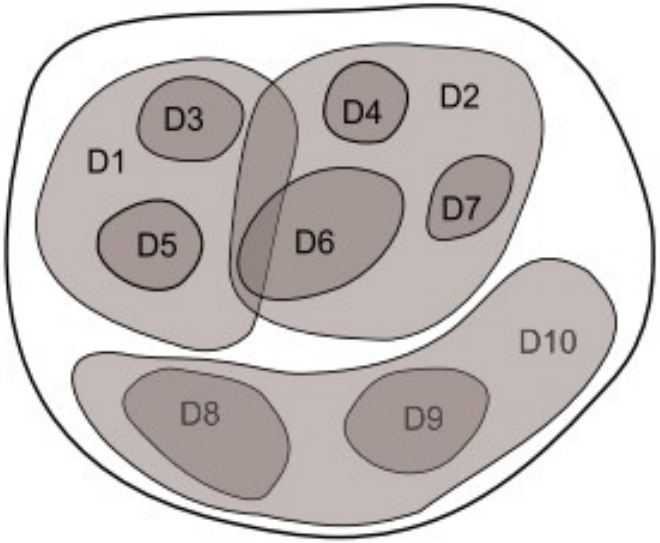
- Typical methods
  - xQuAD: Explicit query aspect diversification
    - Predefined utility function



1) Heuristic predefined utility function;
2) Limited features incorpoated;

Learning-based Methods

- Typical method
  - SVMDIV
    - Only focuses on diversity, discard the relevance
    - Propose to optimize subtopic coverage based on maximizing word coverage



How to model both relevance and diversity?

Outline

- Motivation
- Our Approach
- Experiments
- Conclusion

Our Approach

- Relational Learning-to-rank approach (R-LTR)
  - Considering both ***content*** of individual documents and ***relations*** among documents.

- Formalization
  - Four key components: input space, out space, ranking function *f*, loss function *L*

Difference

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}\in\mathcal{F}} \sum_{i=1}^{N} L(\mathbf{f}(X^{(i)}, R^{(i)}), \mathbf{y}^{(i)}).$$

Challenges for R-LTR

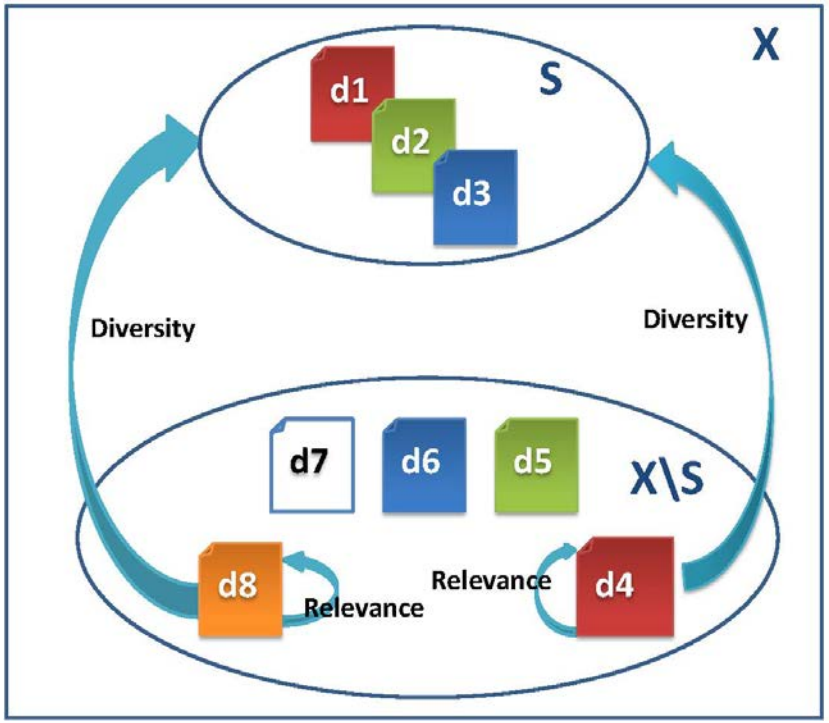
→How to define ranking function

→How to define loss function

Definition of Ranking Function

Sequential Ranking Process

1) Top-down user browsing behavior;
2) NP-hard, greedy sequential approximation



Definition of Ranking Function

- Definition

Relevance Score

Diversity Score

$$f_S(x_i, R_i) = \omega_r^T \mathbf{x}_i + \omega_d^T h_S(R_i), \forall x_i \in X \setminus S$$

- Relational function $h_S(R_i)$
  - Minimal Distance

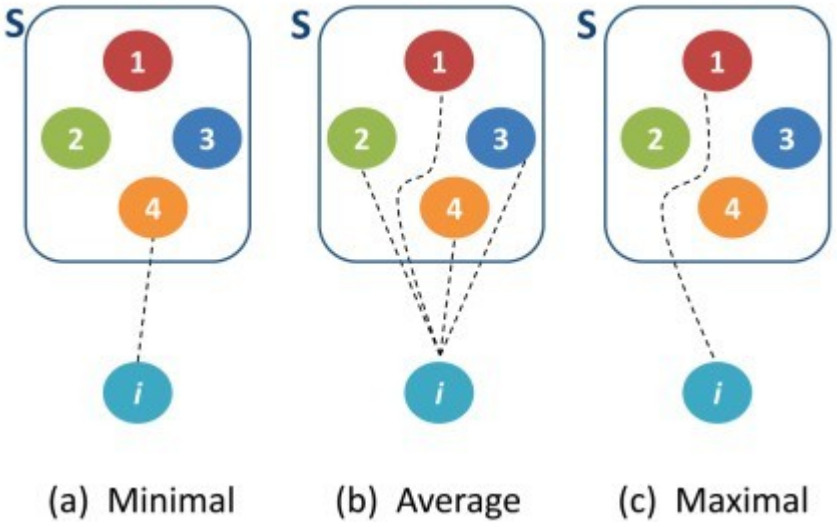
$$h_S(R_i) = (\min_{x_j \in S} R_{ij1}, \cdots, \min_{x_j \in S} R_{ijl}).$$

  - Average Distance

$$h_S(R_i) = (\frac{1}{|S|} \sum_{x_j \in S} R_{ij1}, \cdots, \frac{1}{|S|} \sum_{x_j \in S} R_{ijl}).$$

  - Maximal Distance

$$h_S(R_i) = (\max_{x_j \in S} R_{ij1}, \cdots, \max_{x_j \in S} R_{ijl}).$$



(a) Minimal (b) Average (c) Maximal

Definition of Ranking Function

- Relevance features
  - Traditional LTR relevance features, such as: TFIDF, bm25, LM, Proximity……
- Diversity features
  - Subtopic diversity: semantic distance based on topic distribution.
  - Text, title, anchor diversity based on cosine similarity;
  - ODP-based: existing ODP taxonomy
  - Link-based
  - url-based
  - …

Definition of Loss Function

Sequential Ranking Process

↓

Model the generation of the result list
in a Sequential way

↓

Loss function:
likelihood loss of the generation probability

$$L(\mathbf{f}(X,R),\mathbf{y}) = -\log P(\mathbf{y} \mid X)$$

$$P(\mathbf{y} \mid X) = P(x_{y(1)}, x_{y(2)}, \cdots, x_{y(n)} \mid X)$$

$$= P(x_{y(1)} \mid X) P(x_{y(2)} \mid X \setminus S_1) \cdots P(x_{y(n-1)} \mid X \setminus S_{n-2})$$

How to define properly?

Definition of Loss Function

- Plackett-Luce Model

$$\mathbf{P}(\pi \mid \boldsymbol{v}) = \prod_{i=1}^{M} \frac{v_{\pi(i)}}{v_{\pi(i)} + v_{\pi(i+1)} + \cdots + v_{\pi(M)}}$$

- Detailed definition

$$P(x_{y(1)} \mid X) = \frac{\exp\{f_\phi(x_{y(1)})\}}{\sum_{k=1}^{n} \exp\{f_\phi(x_{y(k)})\}}, \quad P(x_{y(j)} \mid X \setminus S_{j-1}) = \frac{\exp\{f_{S_{j-1}}(x_{y(j)}, R_{y(j)})\}}{\sum_{k=j}^{n} \exp\{f_{S_{k-1}}(x_{y(k)}, R_{y(k)})\}}.$$

- maximize the sum of the likelihood function

$$-\sum_{i=1}^{N} \sum_{j=1}^{n_i} \log\left\{ \frac{\exp\{\omega_r^T \mathbf{x}_{y(j)}^{(i)} + \omega_d^T h_{S_{j-1}^{(i)}}(R_{y(j)}^{(i)})\}}{\sum_{k=j}^{n_i} \exp\{\omega_r^T \mathbf{x}_{y(k)}^{(i)} + \omega_d^T h_{S_{k-1}^{(i)}}(R_{y(k)}^{(i)})\}} \right\}$$

Learning

- Unconstrained optimization problem
  - Stochastic Gradient Descent

Algorithm 2 Optimization Algorithm

Input: training data $\{(X^{(i)}, R^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{N}$,
parameter: learning rate η, tolerance rate ϵ
Output: model vector: ω_r, ω_d
Initialize parameter value ω_r, ω_d
repeat
  Shuffle the training data
  for $i = 1, ..., N$ **do**
    Compute gradient $\Delta{\omega_r}^{(i)}$ and $\Delta{\omega_d}^{(i)}$
    Update model: $\omega_r = \omega_r - \eta \times \Delta{\omega_r}^{(i)}$,
    $\omega_d = \omega_d - \eta \times \Delta{\omega_d}^{(i)}$
  end for
  Calculate likelihood loss on the training set
until the change of likelihood loss is below ϵ

Prediction

- Sequential Prediction Process

Algorithm 3 Ranking Prediction via Sequential Selection

Input: $X^{(t)}, R^{(t)}, \omega_r, \omega_d$
Output: $\mathbf{y}^{(t)}$

1: Initialize $S_0 \leftarrow \emptyset, \mathbf{y}^{(t)} = (1, \cdots, n_t)$
2: **for** $k = 1, ..., n_t$ **do**
3: $\quad$ bestDoc $\leftarrow \text{argmax}_{x \in X_t} f_{S_{k-1}}(x, R)$
4: $\quad$ $S_k \leftarrow S_{k-1} \cup$ bestDoc
5: $\quad$ $y^{(t)}(k) \leftarrow$ the *index* of bestDoc
6: **end for**
7: **return** $\mathbf{y}^{(t)} = (y^{(t)}(1), \cdots, y^{(t)}(n_t))$

Outline

- Motivation
- Our Approach
- Experiments
- Conclusion

Experiments

- Dataset:
  - Diversity task in TREC Web track 2009, 2010, 2011
- Evaluation measures (K=20):
  - TREC official evaluation measures: ERR-IA, a-NDCG, NRBP;
  - Traditional diversity measures: Precision-IA, Subtopic Recall;
- Baseline methods:
  - QL, MMR, xQuAD, PM-2, ListMLE, SVMDIV
- Platform:
  - Indri toolkit (version 5.2)

TREC Official Measures

Table 2: Performance comparison of all methods in official TREC diversity measures for WT2009.

Method	ERR-IA	α-NDCG	NRBP
QL	0.1637	0.2691	0.1382
ListMLE	0.1913 (+16.86%)	0.3074 (+14.23%)	0.1681 (+21.64%)
MMR_{list}	0.2022 (+23.52%)	0.3083 (+14.57%)	0.1715 (+24.09%)
$xQuAD_{list}$	0.2316 (+41.48%)	0.3437 (+27.72%)	0.1956 (+41.53%)
$PM\text{-}2_{list}$	0.2294 (+40.13%)	0.3369 (+25.20%)	0.1788 (+29.38%)
SVMDIV	0.2408 (+47.10%)	0.3526 (+31.03%)	0.2073 (+50.00%)
$R\text{-}LTR_{min}$	**0.2714** (+65.79%)	0.3915 (+45.48%)	**0.2339** (+69.25%)
$R\text{-}LTR_{avg}$	0.2671 (+63.16%)	**0.3964** (+47.31%)	0.2268 (+64.11%)
$R\text{-}LTR_{max}$	0.2683 (+63.90%)	0.3933 (+46.15%)	0.2281 (+65.05%)
TREC-Best	0.1922	0.3081	0.1617

R-LTR approaches show better performance!

Table 3: Performance comparison of all methods in official TREC diversity measures for WT2010.

Method	ERR-IA	α-NDCG	NRBP
QL	0.1980	0.3024	0.1549
ListMLE	0.2436 (+23.03%)	0.3755 (+24.17%)	0.1949 (+25.82%)
MMR_{list}	0.2735 (+38.13%)	0.4036 (+33.47%)	0.2252 (+45.38%)
xQuAD_{list}	0.3278 (+65.56%)	0.4445 (+46.99%)	0.2872 (+85.41%)
PM-2_{list}	0.3296 (+66.46%)	0.4478 (+48.08%)	0.2901 (+87.28%)
SVMDIV	0.3331 (+68.23%)	0.4593 (+51.88%)	0.2934 (+89.41%)
R-LTR_{min}	**0.3647** (+84.19%)	**0.4924** (+62.83%)	**0.3293** (+112.59%)
R-LTR_{avg}	0.3587 (+81.16%)	0.4781 (+58.10%)	0.3125 (+101.74%)
R-LTR_{max}	0.3639 (+83.79%)	0.4836 (+59.92%)	0.3218 (+107.74%)
TREC-Best	0.2981	0.4178	0.2616

Table 4: Performance comparison of all methods in official TREC diversity measures for WT2011.

Method	ERR-IA	α-NDCG	NRBP
QL	0.3520	0.4531	0.3123
ListMLE	0.4172 (+18.52%)	0.5169 (+14.08%)	0.3887 (+24.46%)
MMR_{list}	0.4284 (+21.70%)	0.5302 (+17.02%)	0.3913 (+25.30%)
xQuAD_{list}	0.4753 (+35.03%)	0.5645 (+24.59%)	0.4274 (+36.86%)
PM-2_{list}	0.4873 (+38.44%)	0.5786 (+27.70%)	0.4318 (+38.26%)
SVMDIV	0.4898 (+39.15%)	0.5910 (+30.43%)	0.4475 (+43.29%)
R-LTR_{min}	**0.5389** (+53.10%)	**0.6297** (+38.98%)	**0.4982** (+59.53%)
R-LTR_{avg}	0.5276 (+49.89%)	0.6219 (+37.25%)	0.4724 (+51.26%)
R-LTR_{max}	0.5285 (+50.14%)	0.6223 (+37.34%)	0.4741 (+51.81%)
TREC-Best	0.4380	0.5220	0.4070

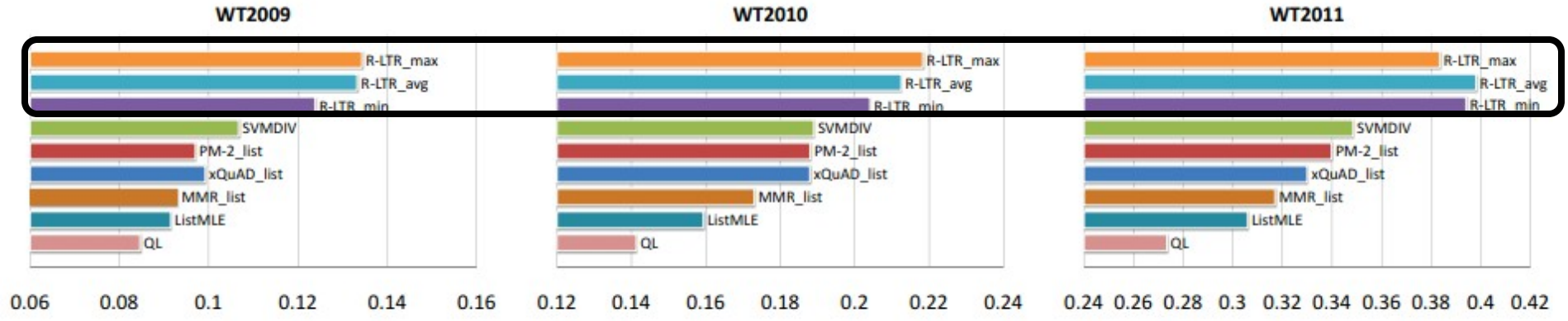


Figure 2: Performance comparison of all methods in Precision-IA for WT2009, WT2010, WT2011.

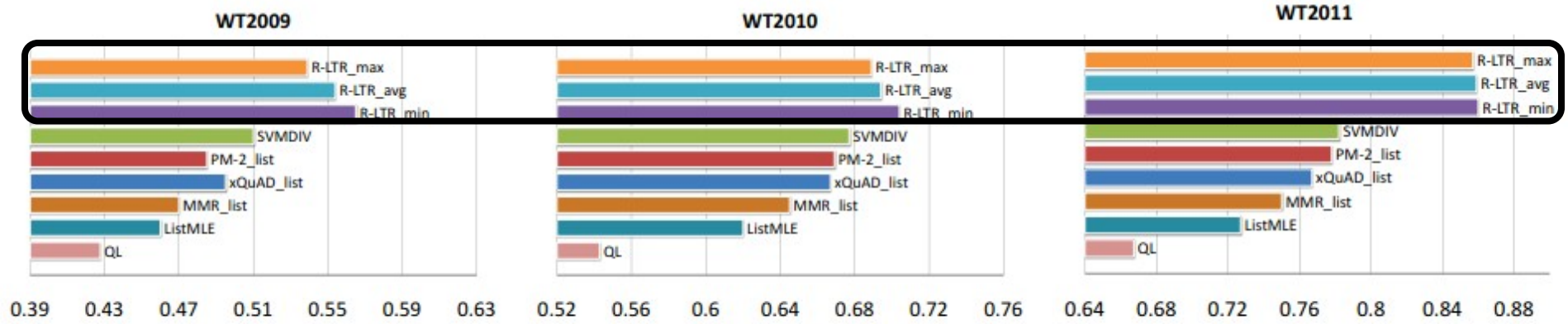


Figure 3: Performance comparison of all methods in Subtopic Recall for WT2009, WT2010, WT2011.

R-LTR approaches also shown better performance!

Experiments

- Robustness Analysis
  - Win/Loss Ratio (vs QL)

Table 5: The robustness of the performance of all diversity methods in Win/Loss ratio

	WT2009	WT2010	WT2011	*Total*
ListMLE	20/18	27/16	26/11	73/45
MMR_{list}	22/15	29/13	29/10	80/38
xQuAD_{list}	28/11	31/12	31/12	90/35
PM-2_{list}	26/15	32/12	32/11	90/38
SVMDIV	30/12	32/11	32/11	94/34
R-LTR_{min}	**34/9**	**35/10**	**35/9**	**104/28**
R-LTR_{avg}	33/9	34/11	34/10	101/30
R-LTR_{max}	33/10	35/10	34/10	102/30

Experiments

- Offline Training Time

$$\text{ListMLE } (\sim 1.5h) \prec \text{SVMDIV } (\sim 2h) \prec \text{R-LTR } (\sim 3h)$$

Outline

- Motivation
- Our Approach
- Experiments
- Conclusion

Conclusion

- Contributions
  - Propose a ***novel relational learning-to-rank*** framework for search results diversification
  - The R-LTR is very ***general*** and can be easily extended to other fields such as summarization or recommendation.
  - Extensive experimental evaluation

http://www.yadongzhu.com

SIGIR Student Travel Grants☺

Q&A