Learning For Search Result Diversification

Paper Presentation

Fatimah Al Mubarak

Why use search result diversification ?

To tackle the ambiguous or multi-faceted information needs of different users

Existing methods:

- heuristic predefined ranking function; implicit and explicit methods.

- Usually are “based on greedy approximation, which sequentially select a ‘local-best’ document from the remanent candidate set”

Proposed Solution:

- Relational Learning To Rank framework

- New definitions for ranking function and loss function based on the foundation of sequential selection process for diverse ranking.

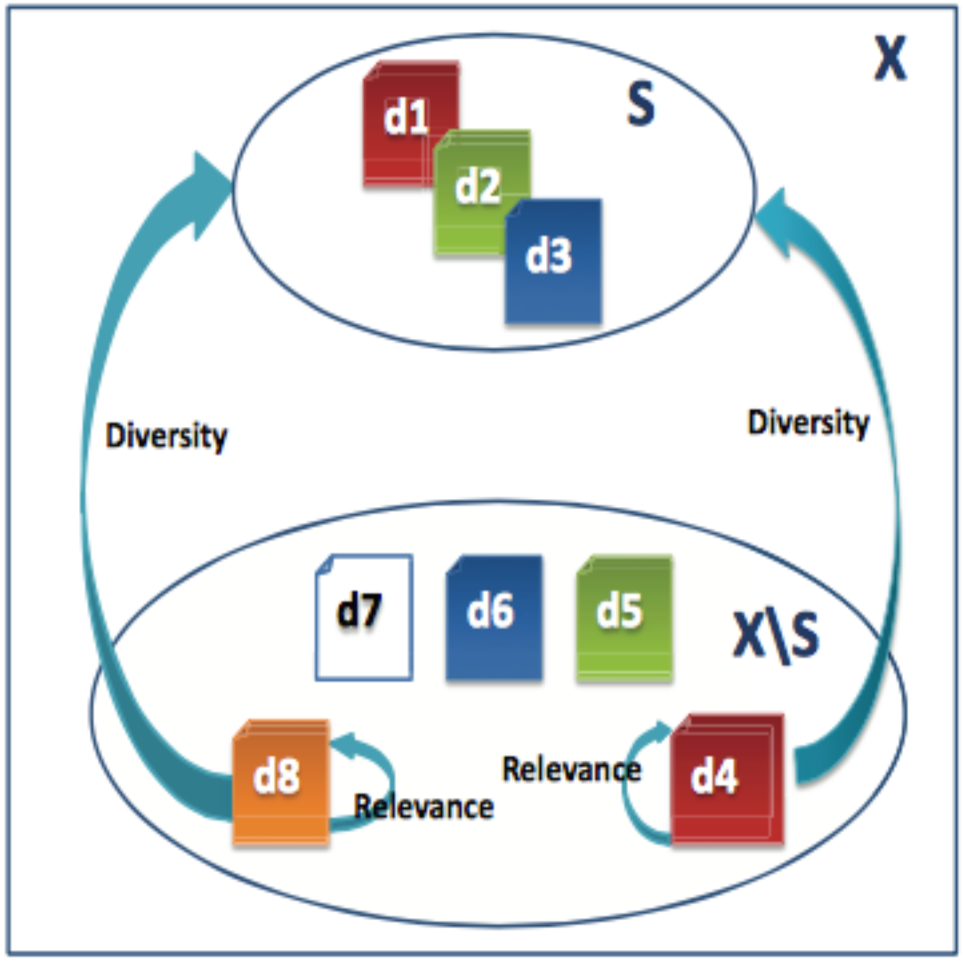
Definitions:

- Ranking Function: “the combination of relevance score and diversity score"
- loss function: “the likelihood loss of the generation probability”

R-LTR

- Documents ranking is obtained by looking to the document’s relevance AND the relations between documents

Adding a new document to the ranking list in a sequential order



R-LTR: Ranking Function

$$f_S(x_i, R_i) = \omega_r^T \mathbf{x}_i + \omega_d^T h_S(R_i), \forall x_i \in X \backslash S, \qquad (2)$$

where :

x_i : the relevance feature vector of the candidate document x_i

R_i : the matrix of relationships between document x_i and other selected documents

h_S (R_i): the relational function on R_i

ω_r^T & ω_d^T : the corresponding relevance and diversity weight vector

R-LTR: Ranking Function

$h_S(R_i)$: represents the diversity relationship between the current document x_i and the previously selected documents in S .

In the paper, $h_S(R_i)$ is presented as the distance of x_i to the set S, thus $h_S(R_i)$ can be represented by three ways:

- **Minimal Distance:** the distance between a document x_i and a set S is defined as the minimal distance of all the document pairs $(x_i,x_j), x_j \in S$.

- **Average Distance:** The distance between a document x_i and a set S is defined as the average distance of all the document pairs $(x_i,x_j), x_j \in S$.

- **Maximal Distance:** The distance between a document x_i and a set S is defined as the maximal distance of all the document pairs $(x_i,x_j), x_j \in S$.

How to capture Diversity

- **Semantic Diversity:**

- Subtopic Diversity

-Text Diversity

-title diversity

-anchor text diversity

-ODP-based diversity; using existing ODP taxonomy, “the distance between documents on similar topics in the taxonomy is likely to be small”
”

- **Structural Diversity:**

linked-based diversity, and URL-based diversity

Constructing Training Data

Algorithm 1 **Construction of Approximate Ideal Ranking List**

Input:
$(q_i, X^{(i)}, \mathbf{T}_i, P(x_j^{(i)}|t)), t \in \mathbf{T}_i, x_j^{(i)} \in X^{(i)}$

Output: $\mathbf{y}^{(i)}$

1: Initialize $S_0 \leftarrow \emptyset, \mathbf{y}^{(i)} = (1, \cdots, n_i)$
2: **for** $k = 1, ..., n_i$ **do**
3: bestDoc $\leftarrow \mathrm{argmax}_{x \in X^{(i)} \setminus S_{k-1}} ODM(S_{k-1} \cup x)$
4: $S_k \leftarrow S_{k-1} \cup$ bestDoc
5: $y^{(i)}(k) =$ the *index* of bestDoc
6: **end for**
7: **return** $\mathbf{y}^{(i)} = (y^{(i)}(1), \cdots, y^{(i)}(n_i))$.

Optimization Process

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

Ranking Prediction

Algorithm 3 **Ranking Prediction via Sequential Selection**

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

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

Feature Importance Analysis

Table 6: Order list of diversity features with corresponding weight value.

feature	weight
R_{ij1}(topic)	3.71635
R_{ij3}(title)	1.53026
R_{ij4}(anchor)	1.34293
R_{ij2}(text)	0.98912
R_{ij5}(ODP)	0.52627
R_{ij6}(Link)	0.04683
R_{ij7}(URL)	0.01514

Evaluation:

- TREC dataset.
- Intent-aware measures: Precision-IA & Subtopic recall
- Performance comparison with state-of-art approaches

Experiment Setup

Implementation:

- Retrieval platform: Indri toolkit
- Data processing: porter stemmer and stopwords removing for both indexing and query processing.

Evaluation Results

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

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%)
$PM\text{-}2_{list}$	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%)
$R\text{-}LTR_{min}$	**0.3647** (+84.19%)	**0.4924** (+62.83%)	**0.3293** (+112.59%)
$R\text{-}LTR_{avg}$	0.3587 (+81.16%)	0.4781 (+58.10%)	0.3125 (+101.74%)
$R\text{-}LTR_{max}$	0.3639 (+83.79%)	0.4836 (+59.92%)	0.3218 (+107.74%)
TREC-Best	0.2981	0.4178	0.2616

Some Highlights of the Comparison

- explicit methods(xQuAD & PM-2) outperformed implicit methods (MMR) in the test
- “SVMDIV simply uses weighted word coverage as a proxy for explicitly covering subtopics, while our R-LTR approach directly models the generation probability of the diverse ranking based on the sequential ranking formulation.”
- R-LTR outperformance is significantly different (p-value <0.01)

Evaluation on Traditional Diversity Metrics

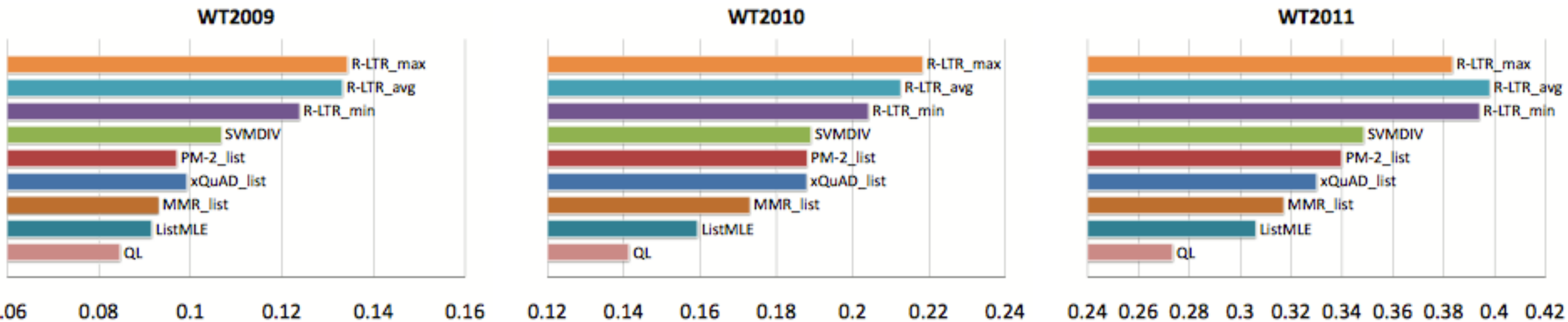
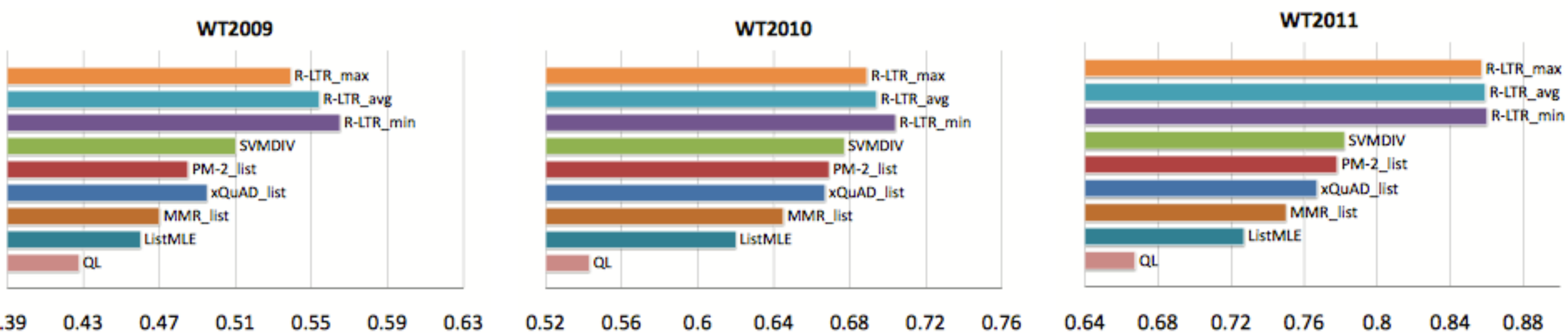


Figure 2: Performance comparison of all methods in Precision-IA for WT2009, WT2010, WT2011.



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

Robustness Analysis

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

Future improvements

- Algorithms running time : $O(n * K)$

- Average training time:

ListMLE ($\sim$ 1.5h) < SVMDIV ($\sim$ 2h) < R-LTR ($\sim$ 3h)

Discussion

- where would this approach be most useful?
- what features that you think should be considered in this approach ?
- what parts of the paper needs more clarification ?

Source Paper

Learning for Search Result Diversification

Yadong Zhu Yanyan Lan Jiafeng Guo Xueqi Cheng
Shuzi Niu Institute of Computing Technology, Chinese
Academy of Sciences, Beijing 100190, China