wsdm 2021
THE 14TH ACM INTERNATIONAL CONFERENCE ON
WEB SEARCH AND DATA MINING

Tutorial
Morning (March, 8)
9:00-12:00, GMT+2

Beyond Probability Ranking Principle: Modeling the Dependencies among Documents

Liang Pang[1], Qingyao Ai[2], Jun Xu[3]

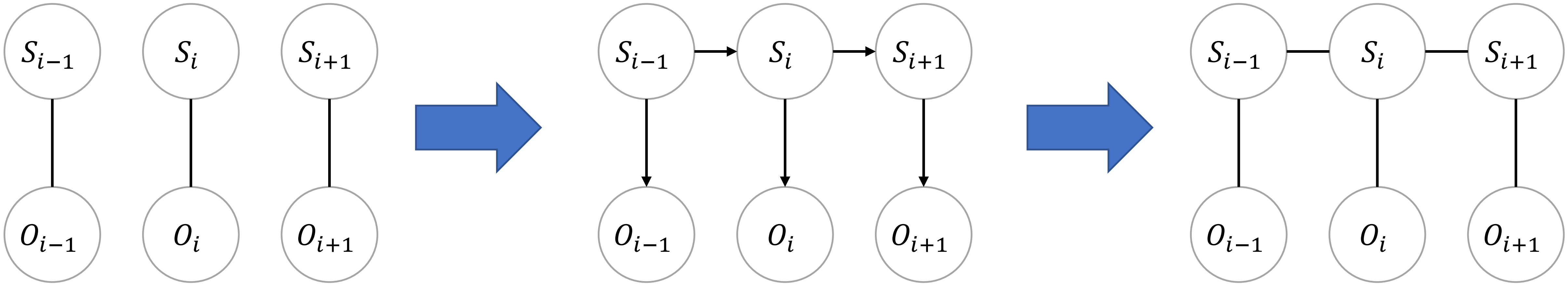
1. CAS Key Lab of Network Data Science and Technology
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
2. The University of Utah, USA
3. Renmin University of China, Beijing, China

Recap: Modeling Dependency in Ranking

No Dependency (between S_i's)

Sequential Dependency

Global Dependency



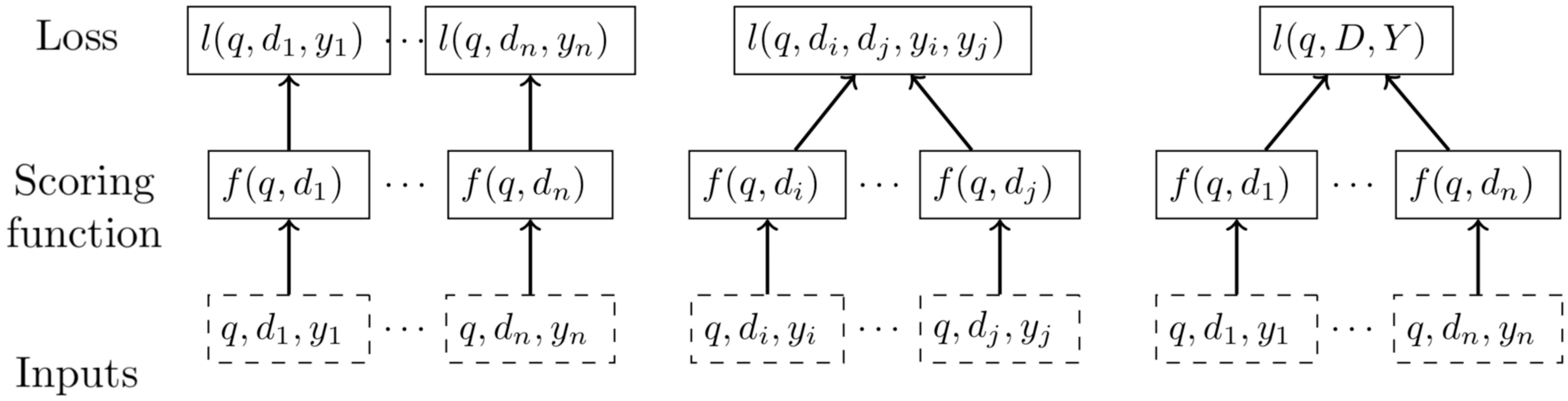
Logistic Regression	HMM	CRF
Probability Ranking Principle	Sequential Decision Making	?

Univariate Ranking with PRP

Uni-variate Scoring Function

Model maps query-doc pair to a score.

$f($ Query, Doc 1 $) \rightarrow$ Score 1 $\quad f: \mathbb{D} \mapsto \mathbb{R}$



(a) Pointwise input
Pointwise loss

(b) Pointwise input
Pairwise loss

(c) Pointwise input
Listwise loss

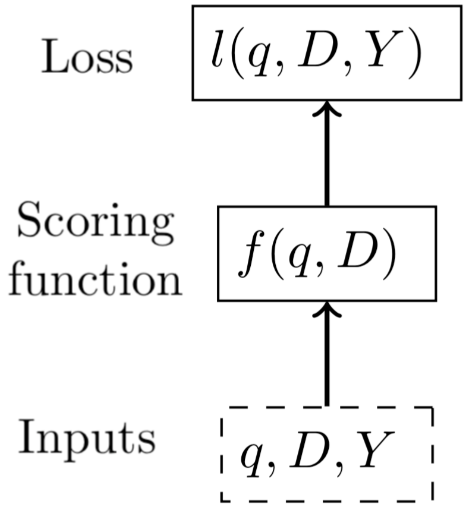
Multivariate Ranking with Global Dependency

Multi-variate Scoring Function

Model maps query-doc pairs in a document set to a list of scores.

$$f\left(\boxed{\begin{array}{c}\text{Query}\\ \text{Doc 1}\end{array}}, \boxed{\begin{array}{c}\text{Query}\\ \text{Doc 2}\end{array}}, \cdots, \boxed{\begin{array}{c}\text{Query}\\ \text{Doc N}\end{array}}\right) \rightarrow \left[\boxed{\text{Score 1}}, \boxed{\text{Score 2}}, \cdots, \boxed{\text{Score N}}\right]$$

$$f\colon \mathbb{D}^N \mapsto \mathbb{R}^N$$



Listwise input
Listwise loss

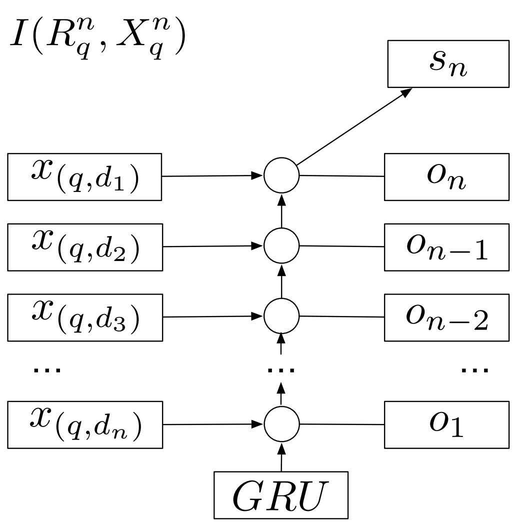
From Univariate to Multivariate Ranking

- Pros
  - From no dependency (PRP) to global dependency
    - More realistic assumptions
    - Context-aware ranking paradigms
    - …

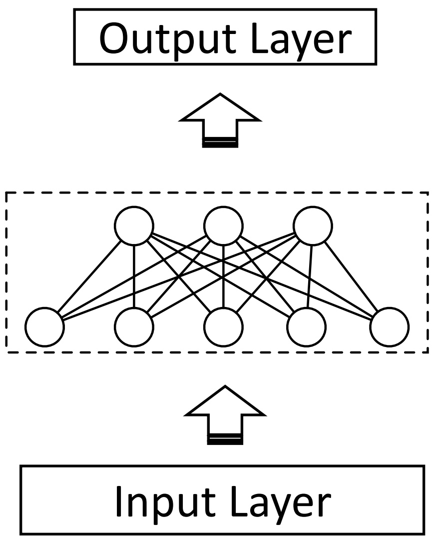
- Cons
  - More complicated input structures
  - More complicated ranking functions
  - Efficiency concerns

Structures for Multivariate Ranking Functions

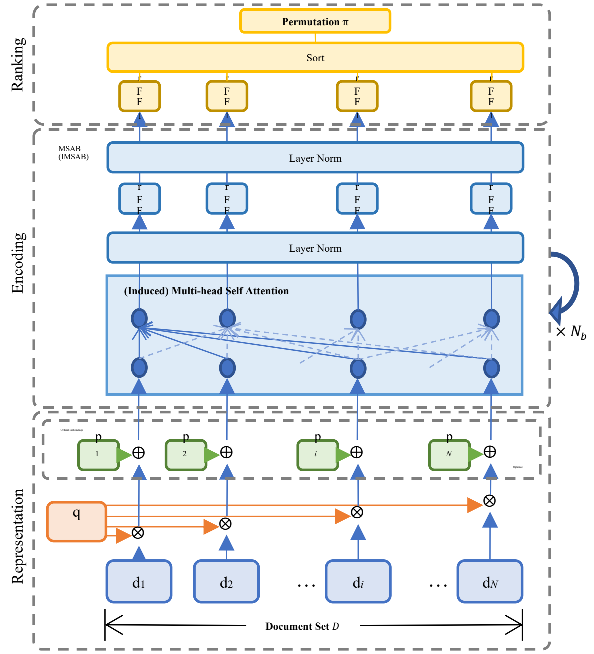
Deep Listwise Context Model



Groupwise Scoring Function



SetRank



	Deep Listwise Context Model	Groupwise Scoring Function	SetRank
Structure	*RNN*	*DNN*	*Self-Attention*
Input	*List*	*List/Set*	*List/Set*

5. Ranking with Global Dependency

5.1 List Inputted Global Ranking Models

5.2 Set Inputted Global Ranking Models

Query-specific learning to rank

- **Learning to rank with local context:**

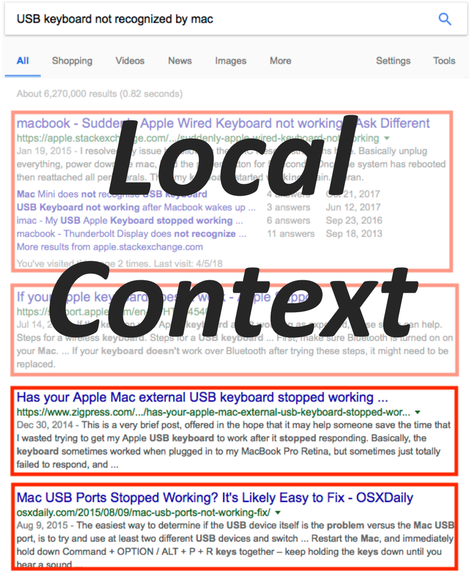
Documents

Features

[0.1, 0.3, …, -0.2]
[-0.3, 0.5, …, -0.8]
[0.4, 0.2, …, -0.1]

Ranking Scores

3.2
-1.4
2.8

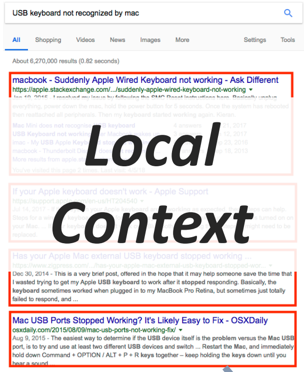


- Powerful and effective
- Memory efficient

[Lavrenko and Croft 2001, Zhai and Lafferty 2001, Diaz 2007, Qin et al. 2008]

Learning to Rank with Local Context

- Problem reformulation:



$$X_q = \{\vec{x}_{(q,d)} | d \in R_q\}$$

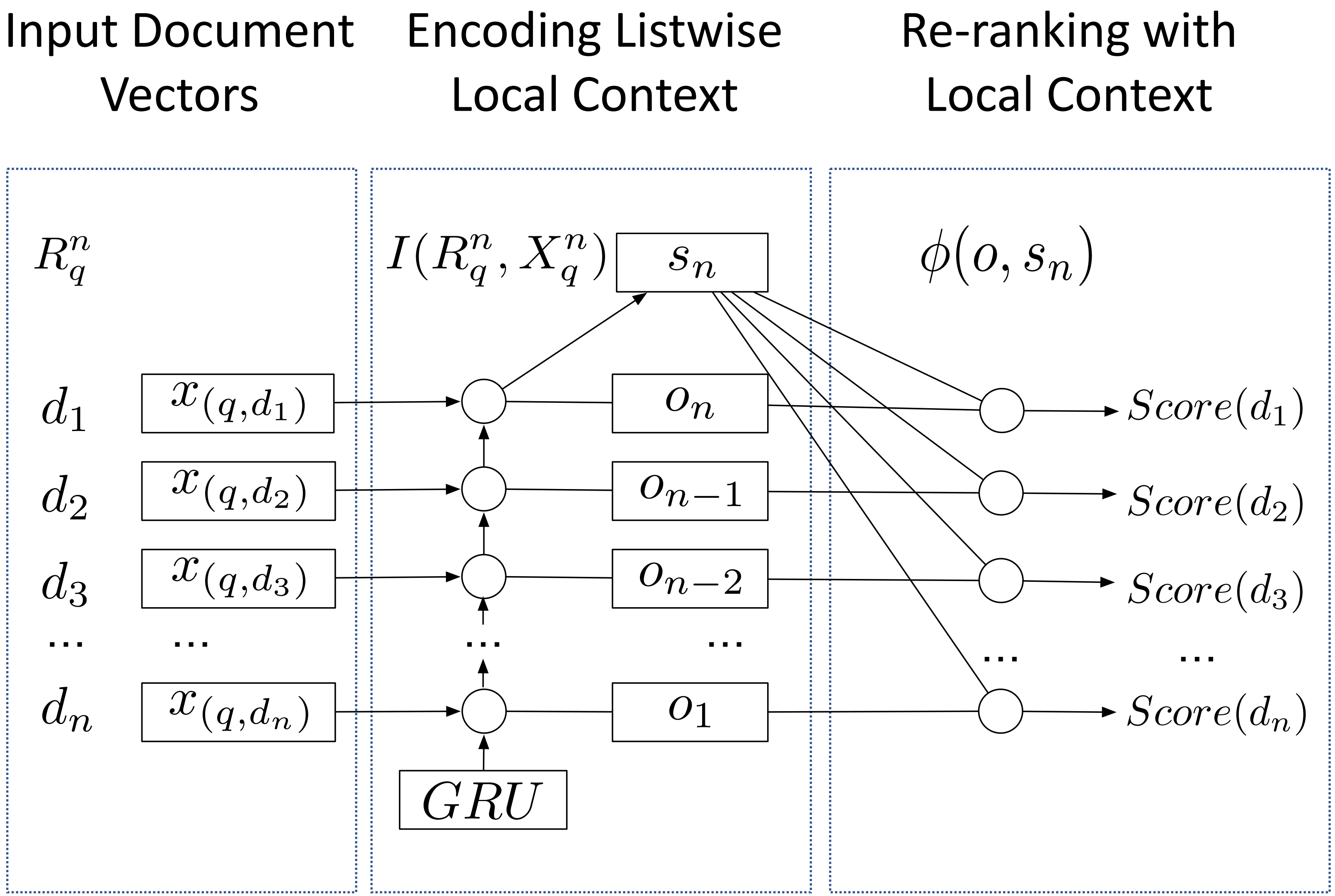
$$\mathcal{L} = \sum_{q \in Q} \ell\Big(\{y_{(q,d)}, \boxed{Score_{(q,d)}} | d \in D\}\Big)$$

$$f(x_{q,d})$$

Scoring function

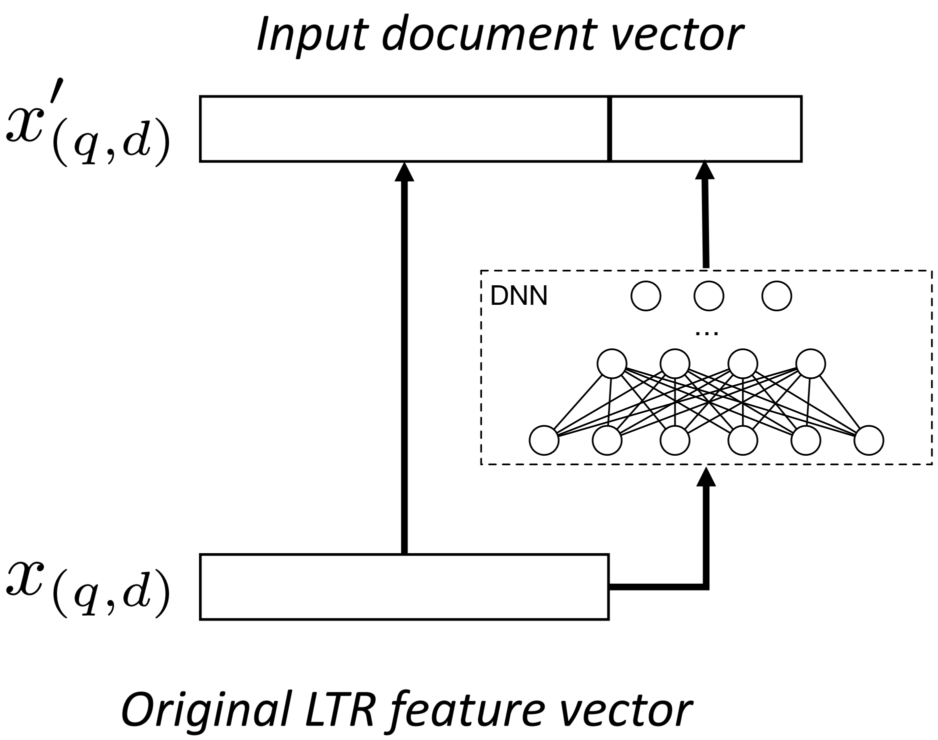
Local context model

Deep Listwise Context Model (DLCM)



Input Document Representation

Feature abstraction and dimensionality increase: [Cheng et al. 2016]



Why?

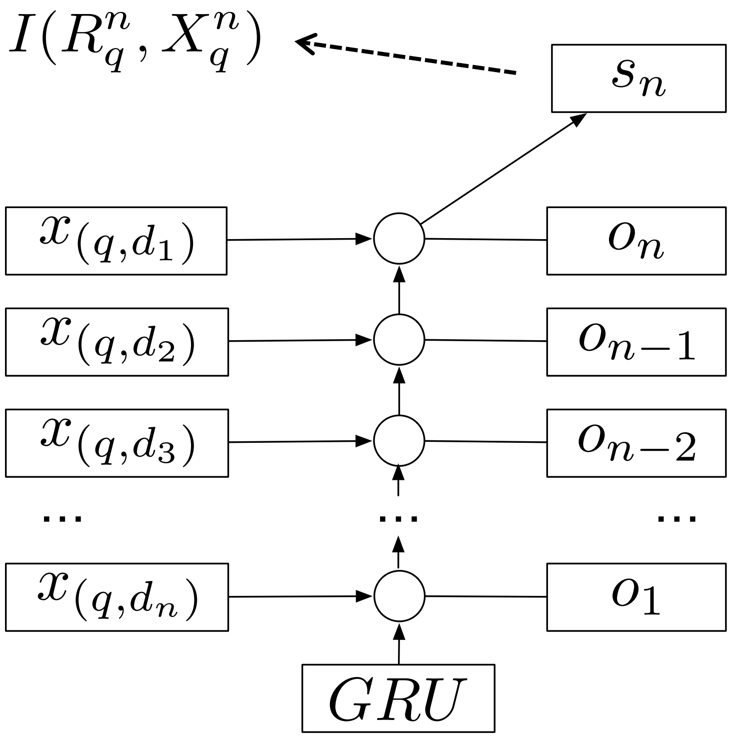
- Flexibility
- Robustness

Encoding Listwise Local Context

Sequentially encode the feature vectors of top results with RNN:

Why?

- Feature interaction
- Inherent structure



Re-ranking with Local Context

- Score documents with an attention function:

Intermediate output of GRU

$$\phi(\vec{o}_{n+1-i}, \vec{s}_n) = \vec{V}_\phi \cdot \left(\vec{o}_{n+1-i} \cdot \tanh(\vec{W}_\phi \cdot \vec{s}_n + \vec{b}_\phi)\right)$$

Final Score

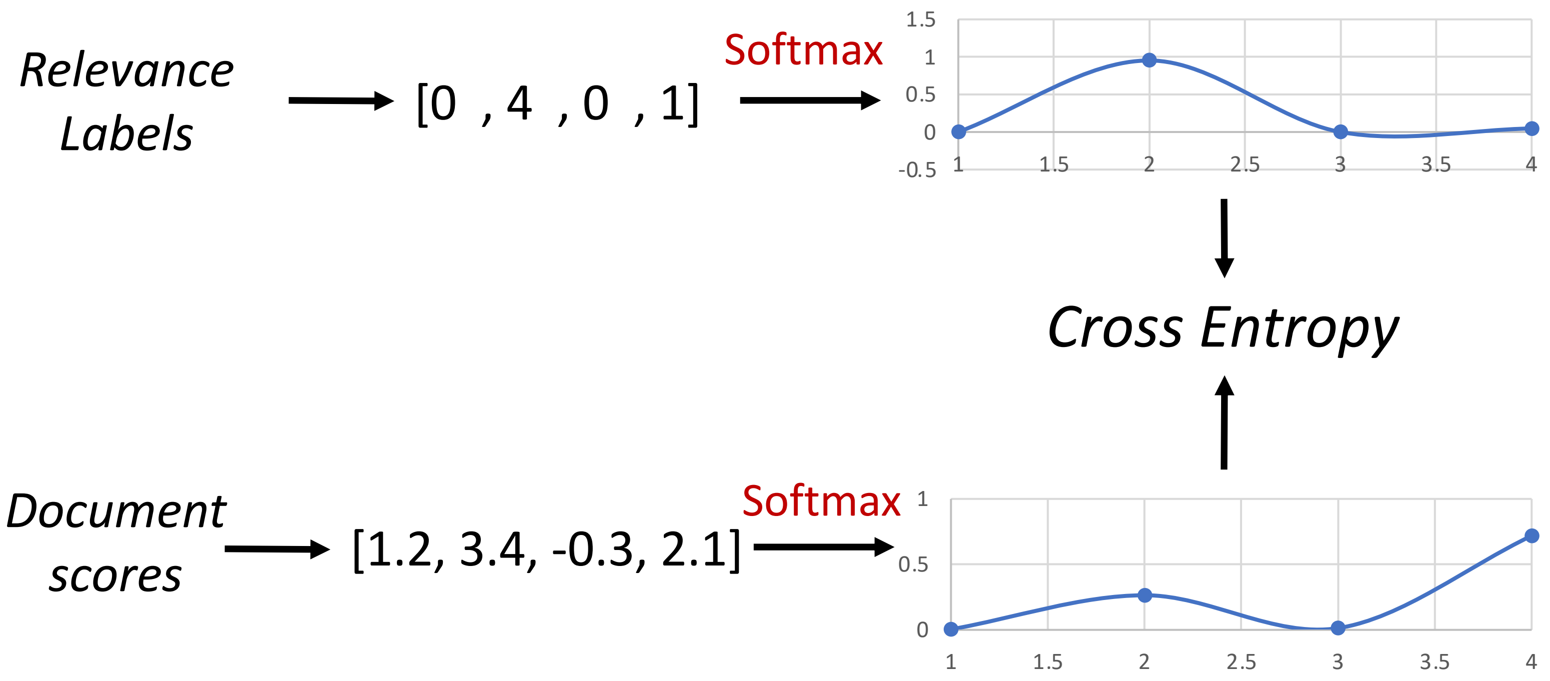
Encoded Local Context

Training

- Listwise ranking loss functions:
  - ListMLE [Xia et al. 2008]
    Optimize the likelihood of the best ranking
  - SoftRank [Taylor et al. 2008]
    Optimize nDCG directly
  - Attention Rank (AttRank)
    Optimize listwise attention allocation

Attention Rank

- Treat ranking as a problem of attention allocation:



Ranking Performance on LambdaMART

* Microsoft 30K

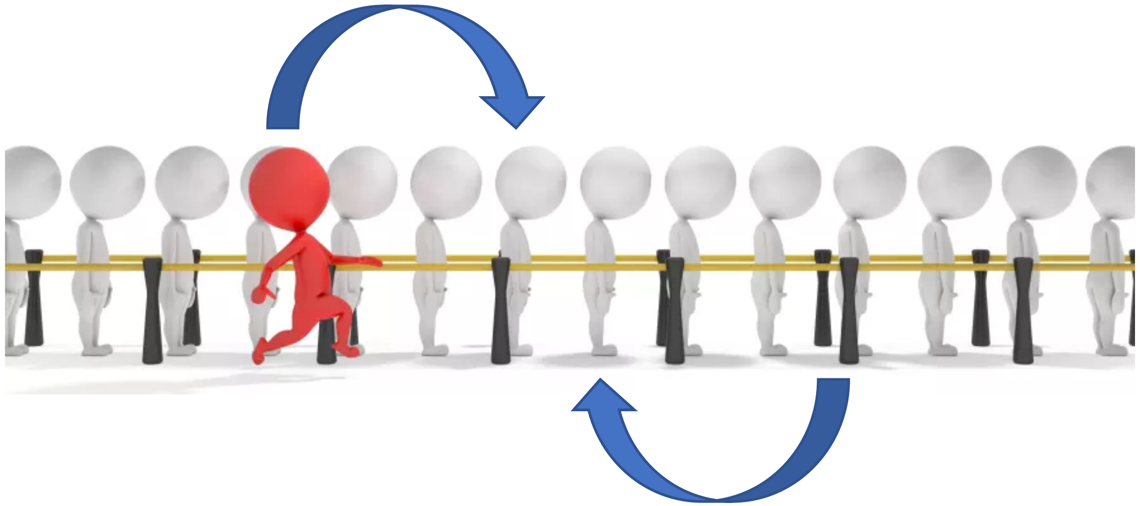
Model	Loss Func	NDCG@1	ERR@1	NDCG@5	ERR@5	NDCG@10	ERR@10
DNN	ListMLE	0.372	0.174	0.386	0.278	0.409	0.299
	SoftRank	0.384	0.209	0.378	0.302	0.397	0.321
	AttRank	0.388	0.199	0.393	0.297	0.416	0.317
LIDNN	ListMLE	0.427	0.219	0.435	0.325	0.455	0.344
	SoftRank	0.457	0.234	0.445	0.336	0.464	0.355
	AttRank	0.455	0.237	0.436	0.334	0.458	0.354
DLCM	ListMLE	0.457	0.235	0.445	0.336	0.464	0.355
	SoftRank	0.463	0.243	0.447	0.342	0.465	0.360
	AttRank	0.463	0.246	0.450	0.344	0.469	0.362
LambdaMART		0.457	0.235	0.445	0.336	0.464	0.355

- Observation
  - DNN damaged the overall ranking performance.
  - DLCM still significantly improved the initial ranker.

Challenges in Global Decision

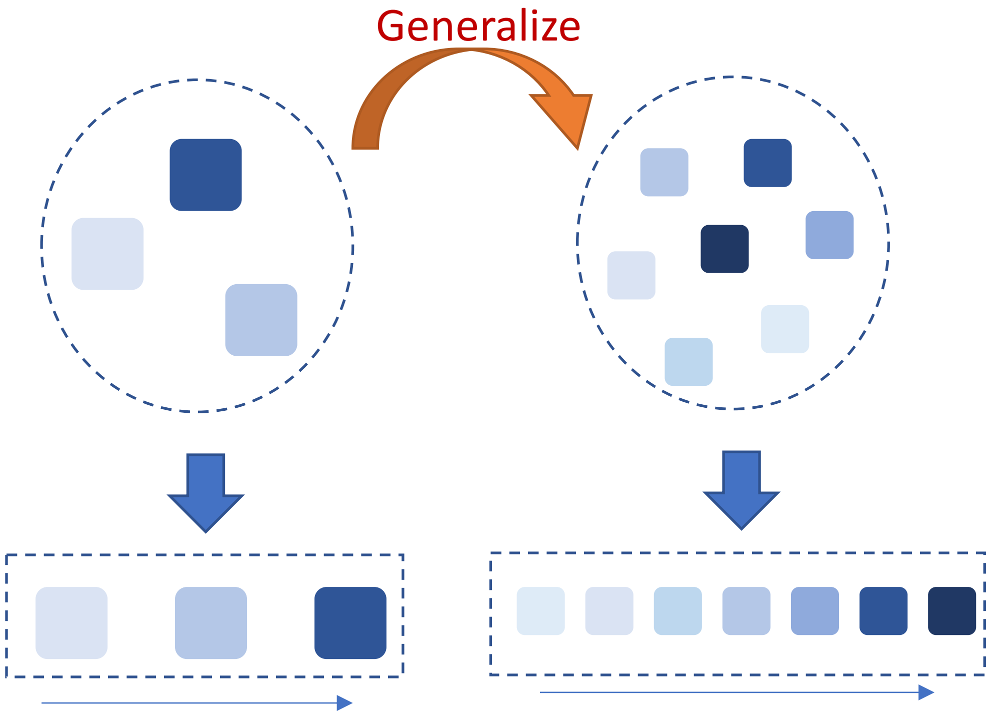
1 Order Sensitive

Change the order of the inputted documents will also change the ranking result.
Strongly rely on the initial document order, e.g. sorted by RankSVM or LambdaMART.

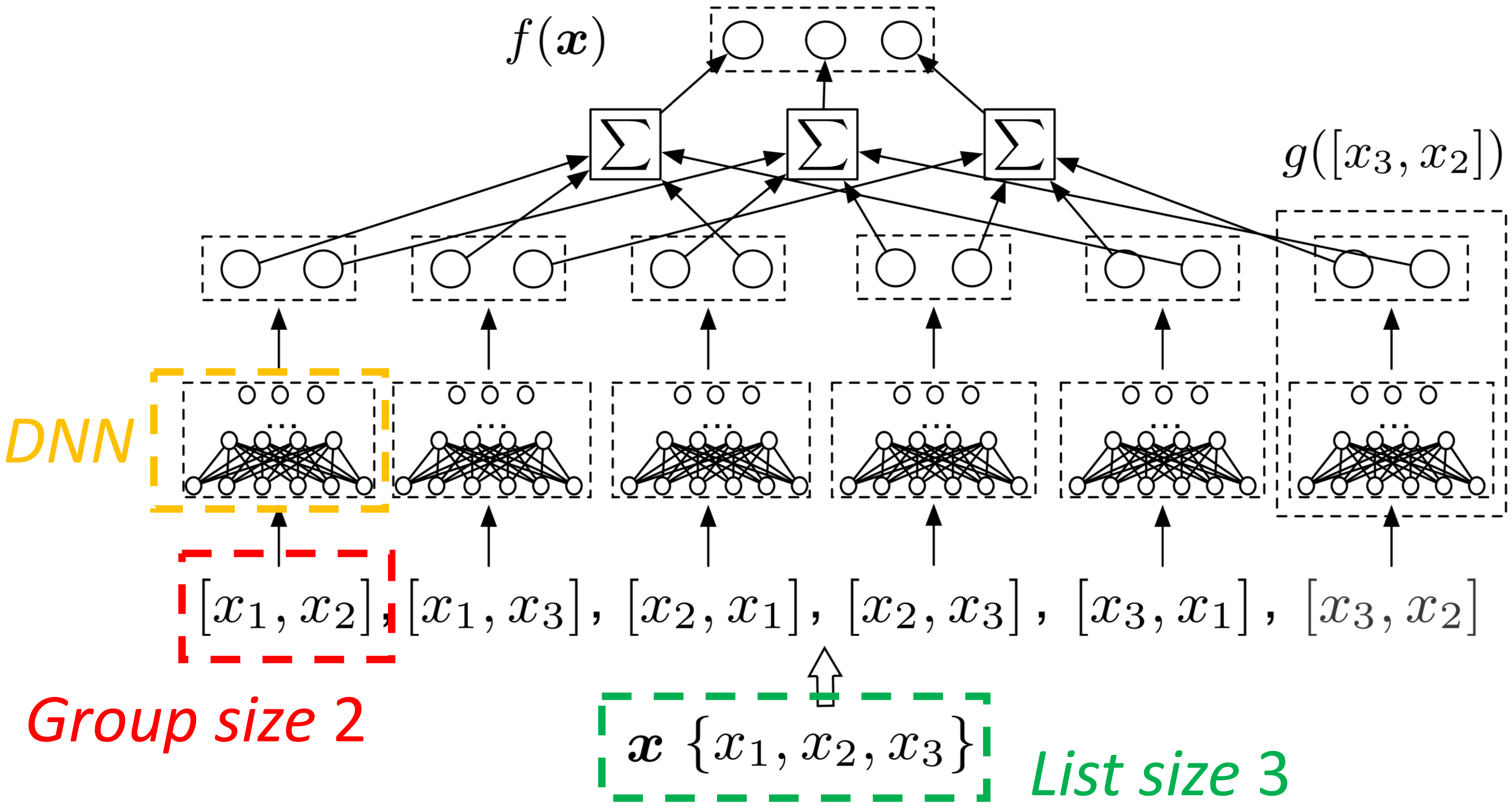


2 Size Adaptation

It is possible that we would have different number of documents for each query.



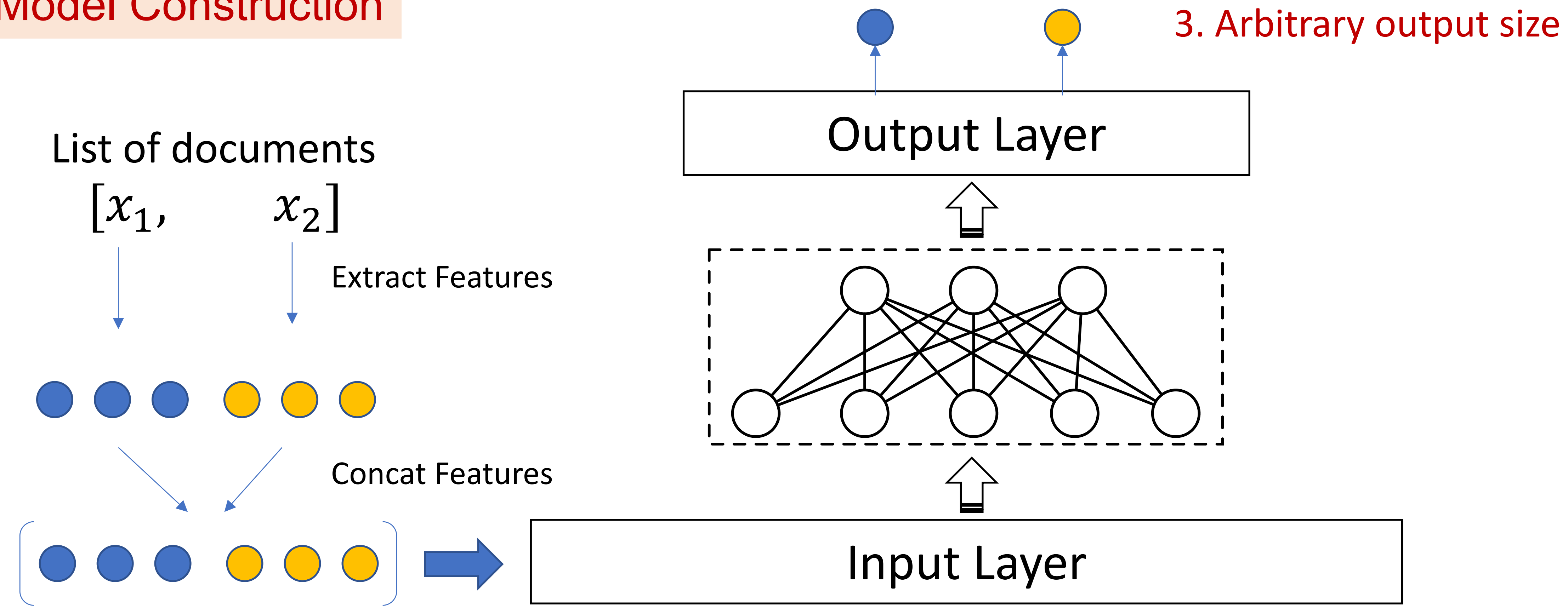
Groupwise Scoring Functions (GSF)



Qingyao Ai, Xuanhui Wang, Sebastian Bruch, Nadav Golbandi, Michael Bendersky, Marc Najork. Learning Groupwise Multivariate Scoring Functions Using Deep Neural Networks . ICTIR' 19

Parameterized with DNN

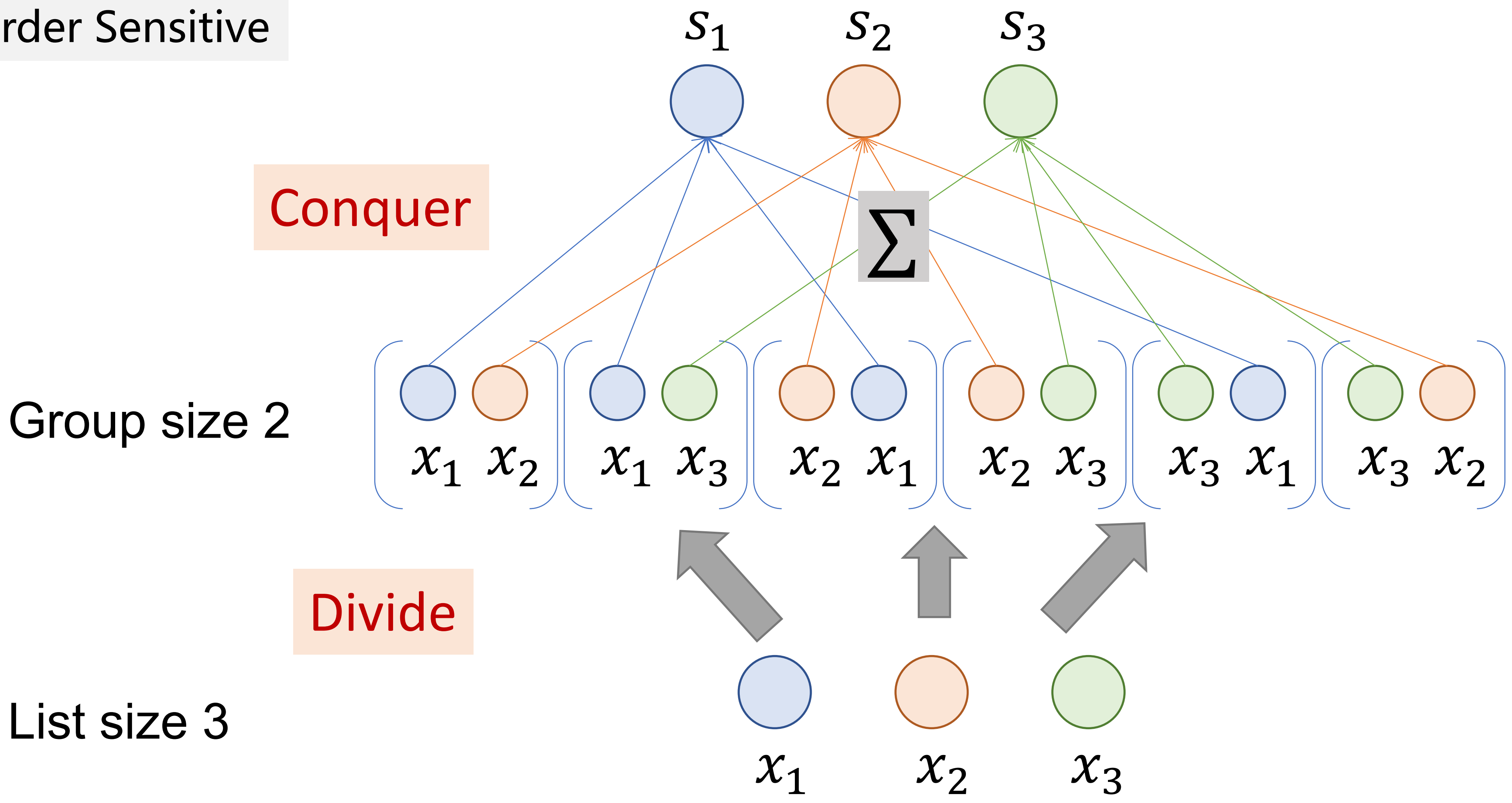
Model Construction



1. Any features: dense features, sparse features, embedding features, …
2. Arbitrary input size: one document, two documents, …

Groupwise Divide & Conquer Strategy

1 Order Sensitive



Training

- Any ranking loss functions:
  - *Sigmoid Cross Entropy (Pointwise)*

$$\hat{\ell}(\boldsymbol{y}, \hat{\boldsymbol{y}}) = -\sum_{j=1}^{n} y_j \log(p_j) + (1 - y_j) \log(1 - p_j), \qquad p_j = \frac{1}{1 + \exp(-\hat{y}_j)}$$

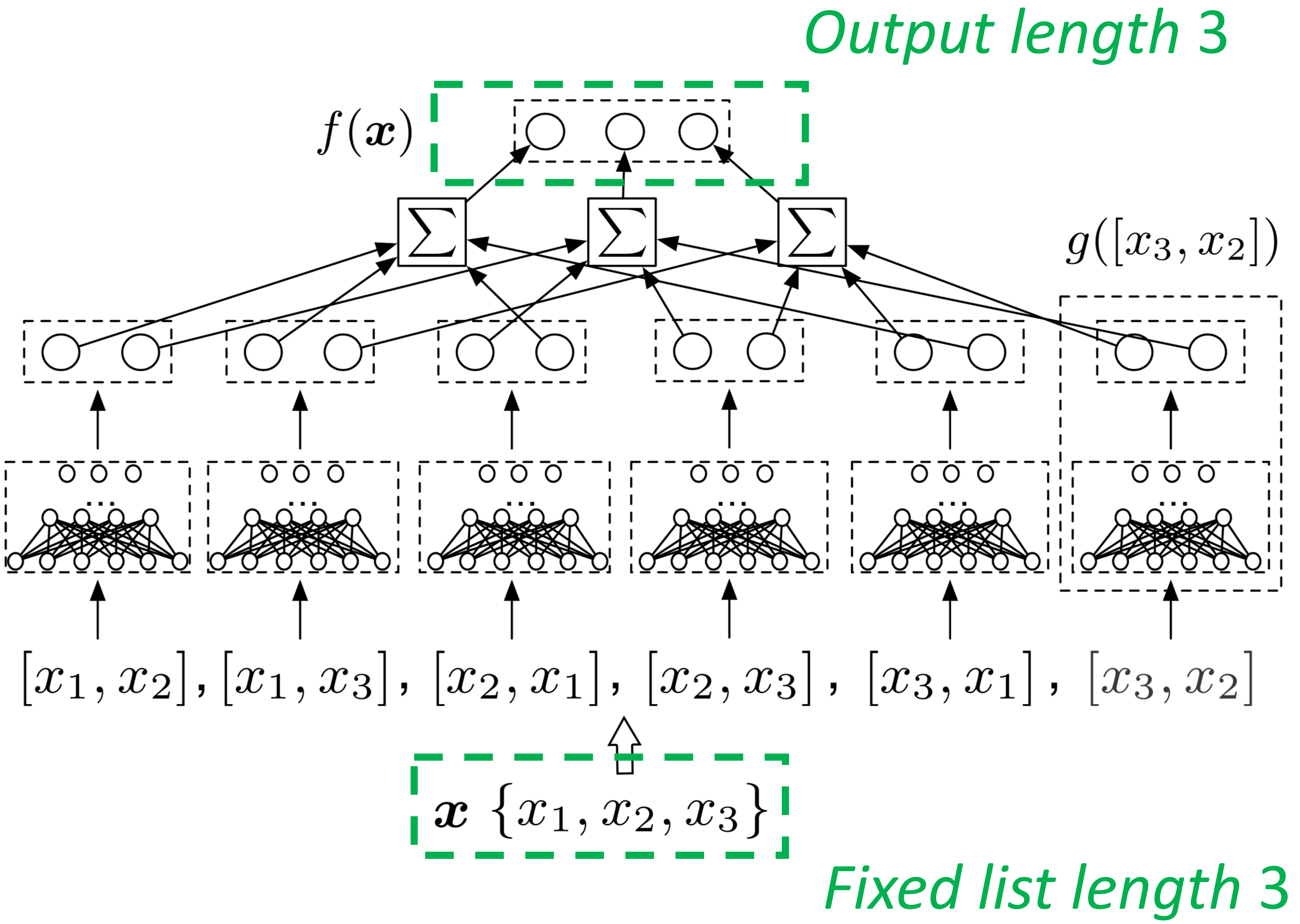
  - *Logistic Loss (Pairwise)*

$$\hat{\ell}(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbb{I}(y_j > y_k) \log(1 + \exp(\hat{y}_k - \hat{y}_j)))$$

  - *Attention Rank (AttRank) (Listwise)*

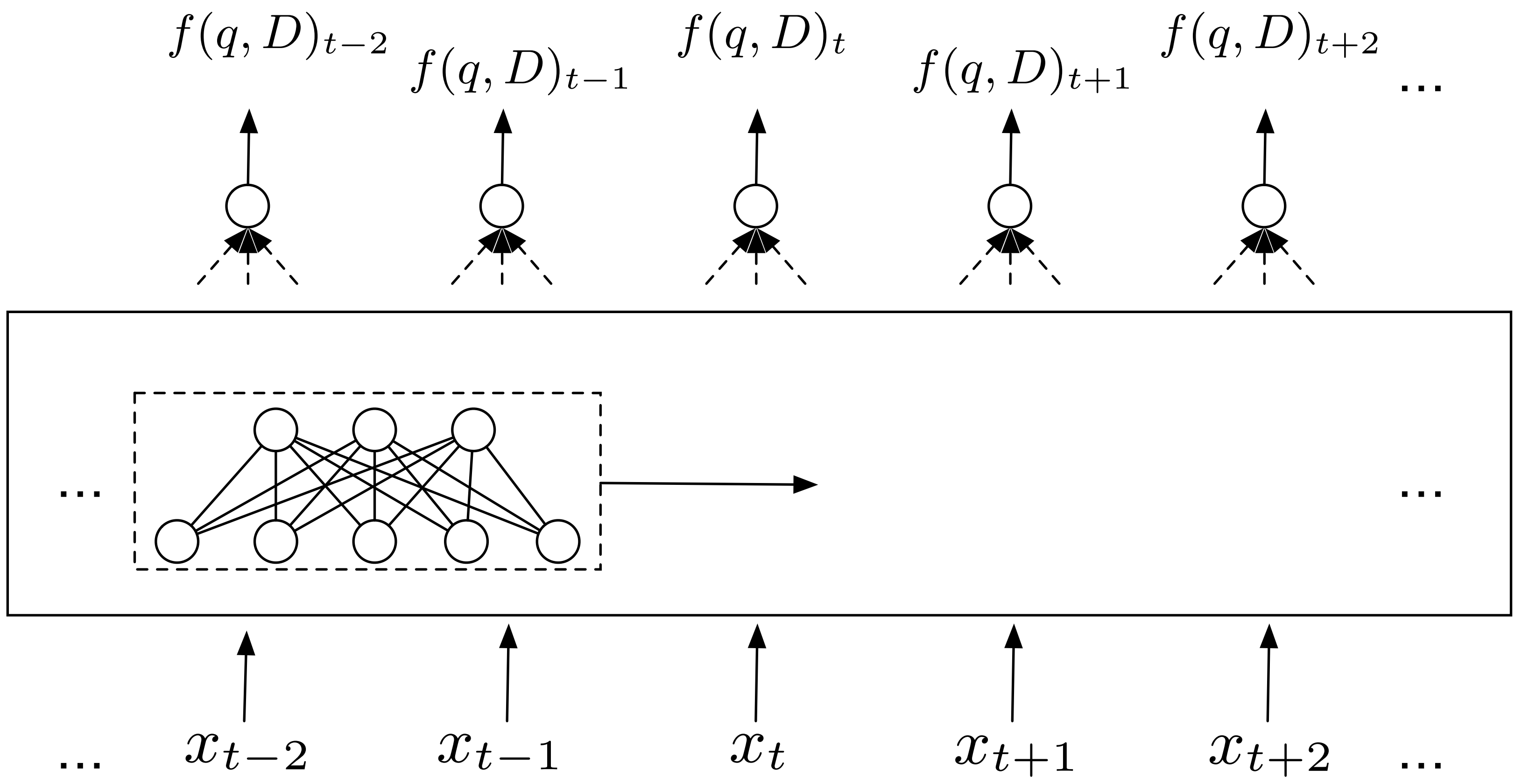
$$\hat{\ell}(\boldsymbol{y}, \hat{\boldsymbol{y}}) = -\sum_{j=1}^{n} y_j \log\left(\frac{\exp(\hat{y}_j)}{\sum_{j=1}^{n} \exp(\hat{y}_j)}\right)$$

Inference with Fixed List Length

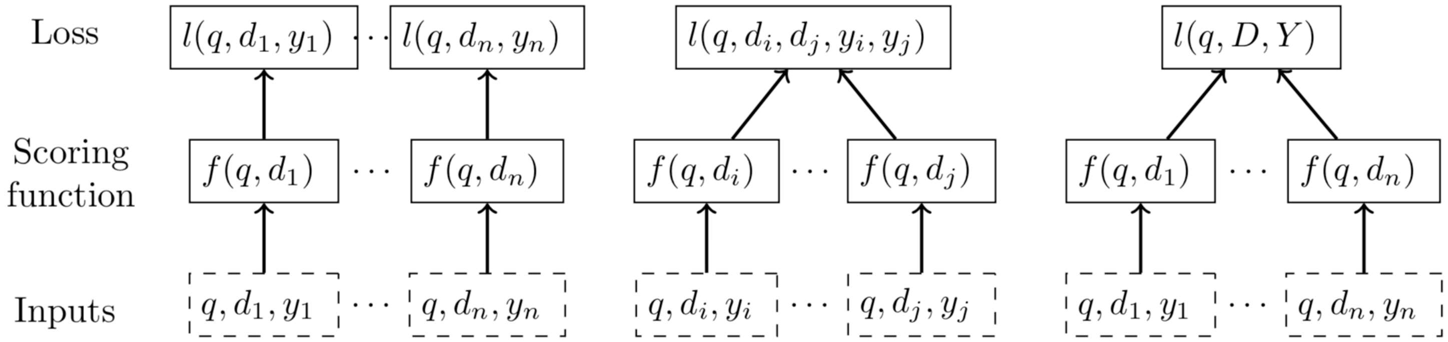


Inference with Arbitrary List Length

2 Size Adaptation



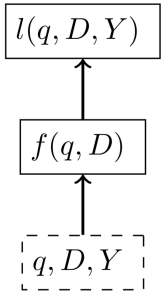
Relationship with Existing Models



(a) Pointwise input
Pointwise loss

(b) Pointwise input
Pairwise loss

(c) Pointwise input
Listwise loss

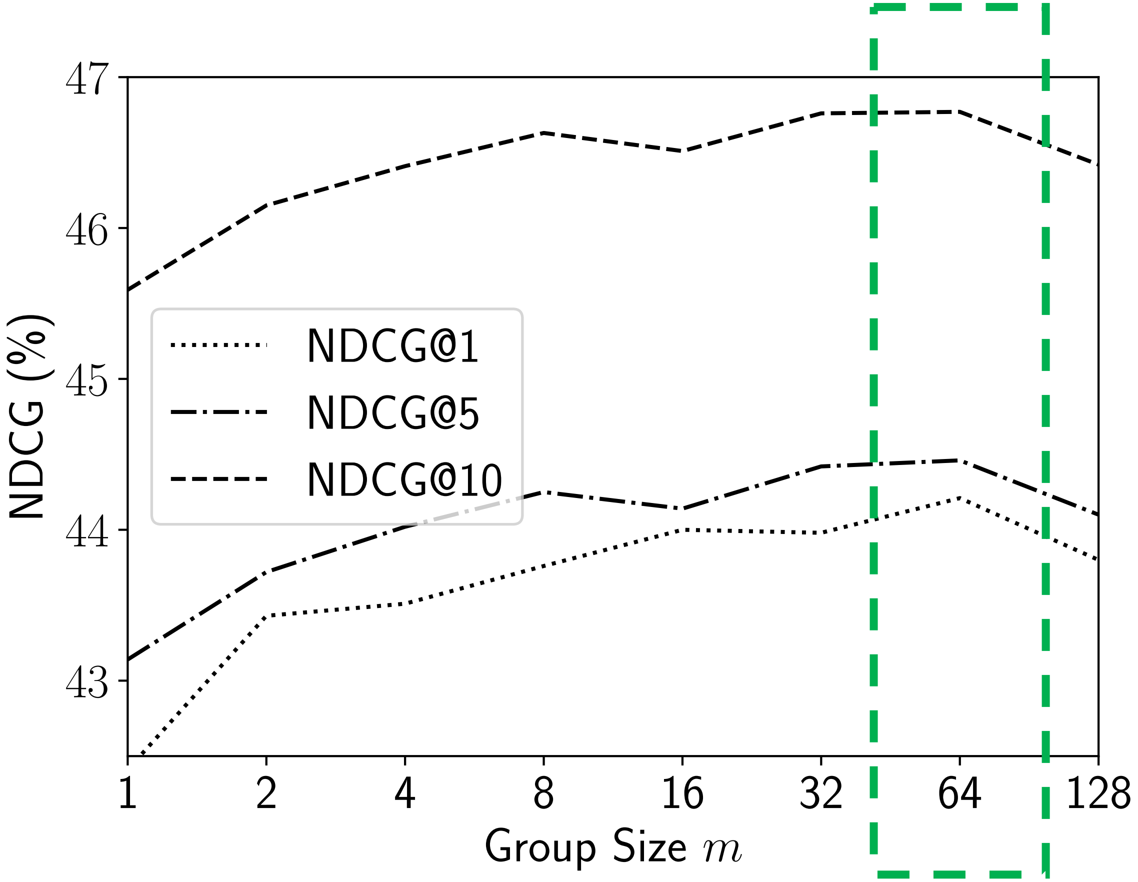


Listwise input
Listwise loss

(a)	(b)	(c)
Group size 1	*Group size* 1	*Group size* 1
List size 1	*List size* 2	*List size* $\|D\|$

Effect of Group Size

- Web30K



Comparison with Baselines

- Web30K

Table 2: A comparison on the Web30K dataset of various GSF flavors and weaker baselines by NDCG@5 (%).

Univariate functions:

	RankNet	*RankSVM*	*PairGSF*	*GSF(1)*
NDCG@5	32.28	34.79	40.40	43.14

Multivariate functions:

	BiGSF	*GSF(2)*	*GSF(64)*
NDCG@5	41.10	43.72	***44.46***

Table 3: A comparison on the Web30K dataset of strong baseline models and the best-performing GSF variant by NDCG at different cut-offs with 95% confidence intervals from 10 trials.

	MART	LambdaMART	GSF(64)
NDCG@1	43.73 ($\pm$0.01)	***45.35*** ($\pm$0.06)	44.21 ($\pm$0.18)
NDCG@5	43.96 ($\pm$0.03)	***44.59*** ($\pm$0.04)	44.46 ($\pm$0.12)
NDCG@10	46.40 ($\pm$0.02)	46.46 ($\pm$0.03)	***46.77*** ($\pm$0.13)

5. Ranking with Global Dependency

5.1 List Inputted Global Ranking Models

5.2 Set Inputted Global Ranking Models

Ranking

Ranking = Set → Permutation

What an Ideal Ranking Model should likes?

An Ideal Ranking Model

from the viewpoint of modeling dependencies

Set → Permutation

A mapping from a document set to a permutation on the set.

Ideal Ranking Model Under Global Dependency

1 Order Sensitive

Ideal Ranking Model

Ranking = Set $\rightarrow$ Permutation

Satisfy Two Critical Requirements

R1: Cross-document Interactions

for modeling the global dependency $s(d_1|\{d_1, d_2, d_3, d_4\})$

R2: Permutation Invariant

because input is a set $f(\{d_1, d_2, d_3, d_4\}) = f(\{d_2, d_1, d_3, d_4\})$

R1: Cross-document Interactions

Uni-variate Scoring Function

Model maps query-doc pair to a score.

$$f\left(\begin{array}{|c|}\hline \text{Query} \\ \hline \text{Doc 1} \\ \hline\end{array}\right) \rightarrow \boxed{\text{Score 1}} \qquad f\colon \mathbb{D} \mapsto \mathbb{R}$$

Multi-variate Scoring Function

Model maps query-doc pairs in a document set to a list of scores.

$$f\left(\begin{array}{|c|}\hline \text{Query} \\ \hline \text{Doc 1} \\ \hline\end{array}, \begin{array}{|c|}\hline \text{Query} \\ \hline \text{Doc 2} \\ \hline\end{array}, \cdots, \begin{array}{|c|}\hline \text{Query} \\ \hline \text{Doc N} \\ \hline\end{array}\right) \rightarrow \left[\boxed{\text{Score 1}}, \boxed{\text{Score 2}}, \cdots, \boxed{\text{Score N}}\right]$$

$$f\colon \mathbb{D}^N \mapsto \mathbb{R}^N$$

R2: Permutation Invariant

1 Order Sensitive

A permutation-invariant ranking model is assigning relevance scores to the documents with a permutation equivariant scoring function.

Definition of permutation equivariant:

A function $F: X^N \mapsto Y^N$ is permutation equivariant

$$F([d_{\pi(1)}, \cdots, d_{\pi(N)}]) = [F(D)|_{\pi(1)}, \cdots, F(D)|_{\pi(N)}],$$

for any permutations of indices π.

Note that: The uni-variate scoring function is a permutation equivariant function.

Existing Work

Following PRP

X **R1**: Cross-document Interactions
√ **R2**: Permutation Invariant

Point-wise Approaches

- SVM, GBT

Pair-wise Approaches

- RankSVM, RankBoost

List-wise Approaches

- MART, LambdaMART

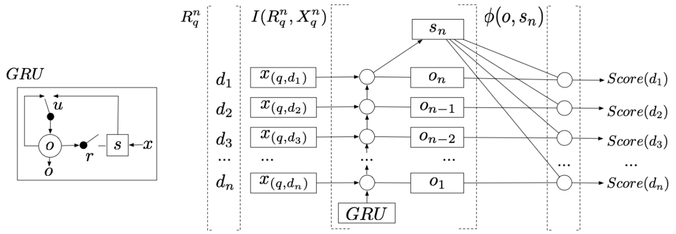
Uni-variate scoring functions: $\mathbb{D} \mapsto \mathbb{R}$

Do not modeling cross-document interactions and capturing local context information

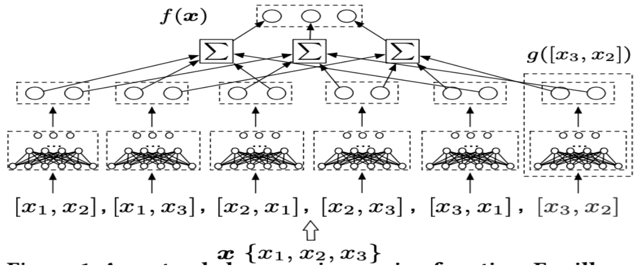
Break PRP

√ **R1**: Cross-document Interactions
X **R2**: Permutation Invariant

Deep Listwise Context Model (DLCM)



Group-wise Multivariate Scoring Func. (GSF)



Multi-variate scoring functions: $\mathbb{D}^N \mapsto \mathbb{R}^N$

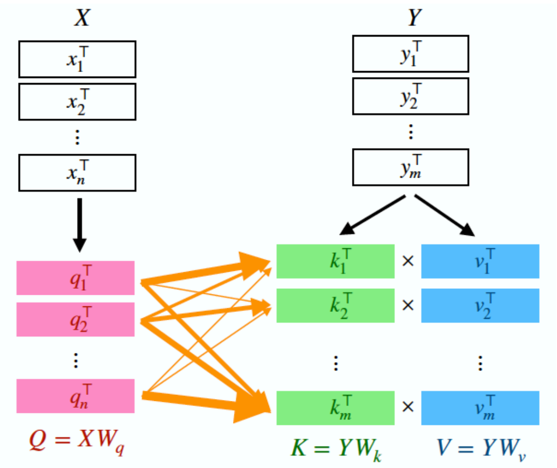
Input is a sequence of documents and heavily biased to its initial ranking.

How to design a Multi-variate & Permutation equivariant function?

Set Transformer (Lee et al. ICML 2019)

- Follow the work Set-input problems and Deep Sets [Zaheer et al., 2017]
- Deep Sets: a simple way to construct permutation invariant set-input neural networks, but does not effectively modeling interactions between elements in sets.

$$f(X) = \rho(\sum_{x \in X} \phi(x))$$



- Note that a self-attention is permutation equivariant

$$SelfAtt(\pi \cdot X) = \pi \cdot SelfAtt(X)$$

- Propose two types of building blocks, e.g. self-attention block and induced self-attention block

Proof

Self-Attention Definition

Note that the multi-head parameters $\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V$, the row-wise feedforward function rFF$(\cdot)$, and layer normalization LayerNorm$(\cdot)$ are all element wise functions, which do not affect the permutation-equivariant property. Therefore, we only need to consider the permutation-equivariant of the self-attention function:

$$\mathrm{SelfAttn}(\mathbf{X}) = \mathrm{Attn}(\mathbf{X}, \mathbf{X}, \mathbf{X}) = \mathrm{softmax}\left(\frac{\mathbf{X}\mathbf{X}^T}{\sqrt{E}}\right)\mathbf{X}.$$

Proof

Proof. Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$. For simplicity and with no loss of generality we discard the scalar factor $\sqrt{E}$, we have

$$\mathrm{SelfAttn}(\mathbf{X}) = \mathrm{softmax}\left(\mathbf{X}\mathbf{X}^T\right)\mathbf{X} = \mathrm{softmax}\left(\left[\mathbf{x}_i\mathbf{x}_j^T\right]_{ij}\right)\mathbf{X}$$

$$= \left[\frac{e^{\mathbf{x}_i\mathbf{x}_j}}{\sum_k e^{\mathbf{x}_i\mathbf{x}_k}}\right]_{ij}\mathbf{X} = \left[\frac{\sum_j e^{\mathbf{x}_i\mathbf{x}_j} \cdot \mathbf{x}_j}{\sum_k e^{\mathbf{x}_i\mathbf{x}_k}}\right]_i,$$

For any permutation $\pi \in \Pi_N$, since sum is permutation-invariant,

$$\mathrm{SelfAttn}(\pi\mathbf{X}) = \mathrm{SelfAttn}\left([\mathbf{x}_{\pi(i)}]_i\right) = \left[\frac{\sum_j e^{\mathbf{x}_{\pi(i)}\mathbf{x}_{\pi(j)}} \cdot \mathbf{x}_{\pi(j)}}{\sum_k e^{\mathbf{x}_{\pi(i)}\mathbf{x}_{\pi(k)}}}\right]_{\pi(i)}$$

$$= \left[\frac{\sum_j e^{\mathbf{x}_{\pi(i)}\mathbf{x}_j} \cdot \mathbf{x}_j}{\sum_k e^{\mathbf{x}_{\pi(i)}\mathbf{x}_k}}\right]_{\pi(i)} = \pi\left(\mathrm{SelfAttn}(\mathbf{X})\right).$$

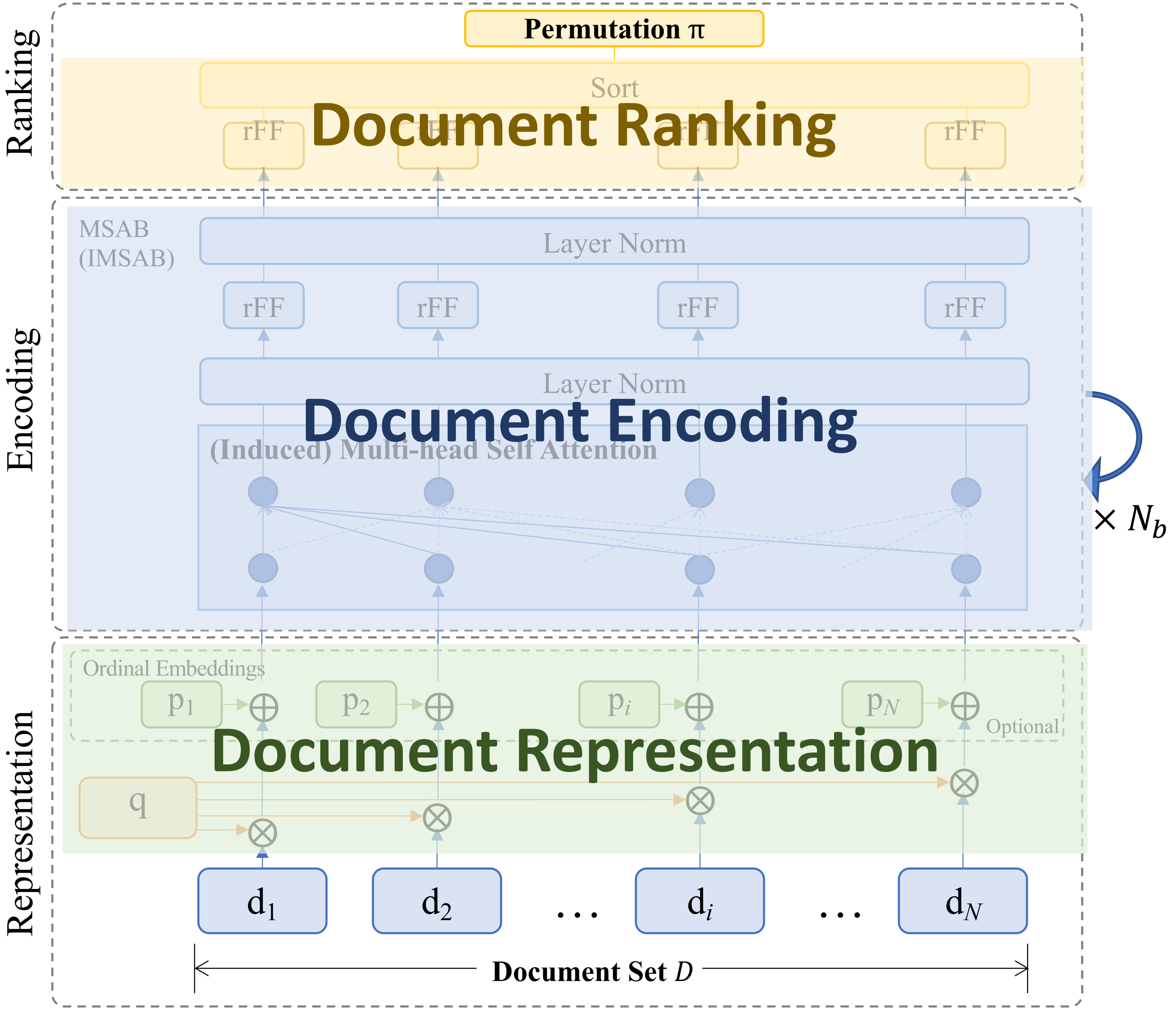
□

SetRank
[Pang et al., SIGIR 2020]

√ **R1**: Cross-document Interactions
√ **R2**: Permutation Invariant

Three components:

1. Document Representation
2. Document Encoding
3. Document Ranking



Liang Pang, Jun Xu, Qingyao Ai, Yanyan Lan, Xueqi Cheng, Jirong Wen. SetRank: Learning a Permutation-Invariant Ranking Model for Information Retrieval . SIGIR’20

Document Representation

- Represent each document with query-document pair:

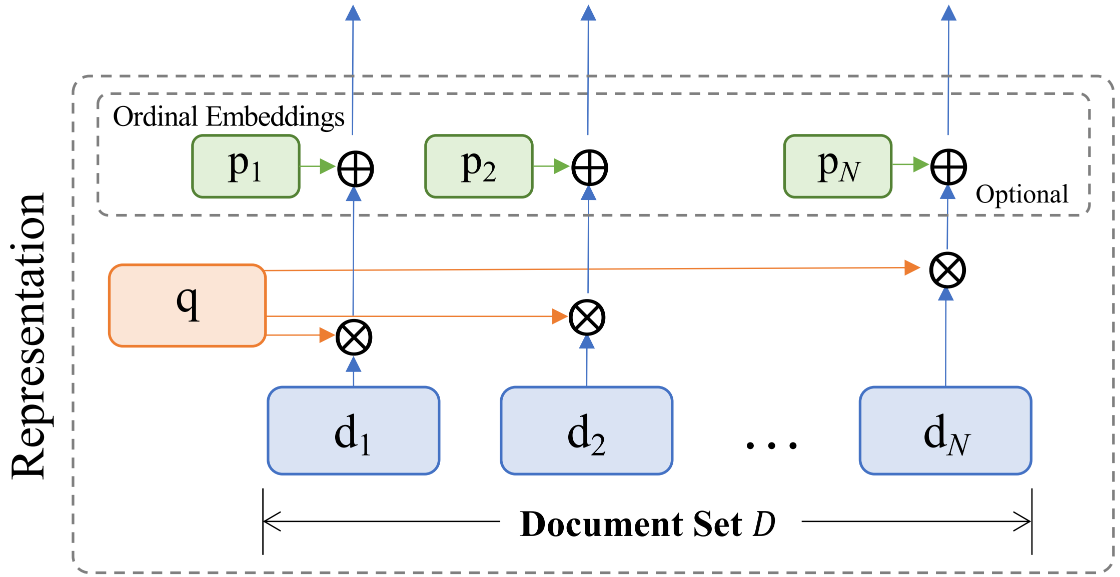
$$\mathbf{d}_i = \phi(q, d_i), \text{ where } \mathbf{d}_i \in \mathbb{R}^E,$$

- Append ordinal embeddings:

$$\mathbf{p}_i = P(\text{rank}(d_i)), \text{ where } \mathbf{p}_i \in \mathbb{R}^E,$$

$$\mathbf{X} = [\mathbf{d}_1 + \mathbf{p}_1, \mathbf{d}_2 + \mathbf{p}_2, \cdots, \mathbf{d}_N + \mathbf{p}_N]^T.$$

Different from positional embeddings!



Document Encoding

- Scaled Dot-Product Attention

$$\mathrm{Attn}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{E}}\right)\mathbf{V},$$

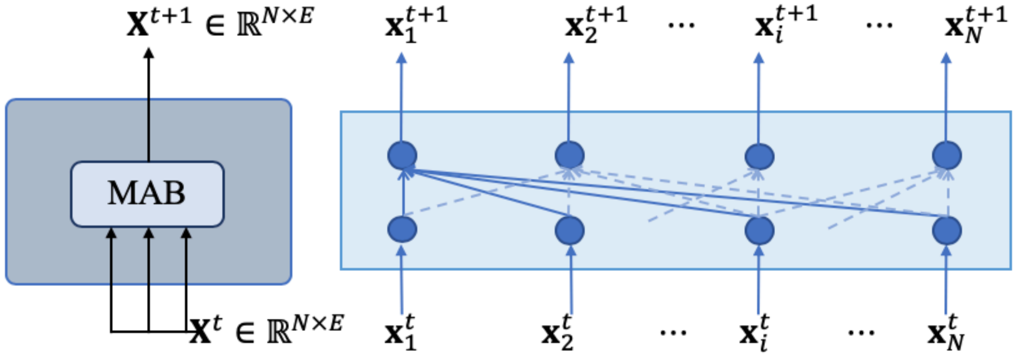
- Multi-head Self Attention Block

$$\mathrm{MSAB}(\mathbf{X}) = \mathrm{MAB}(\mathbf{X},\mathbf{X},\mathbf{X}).$$

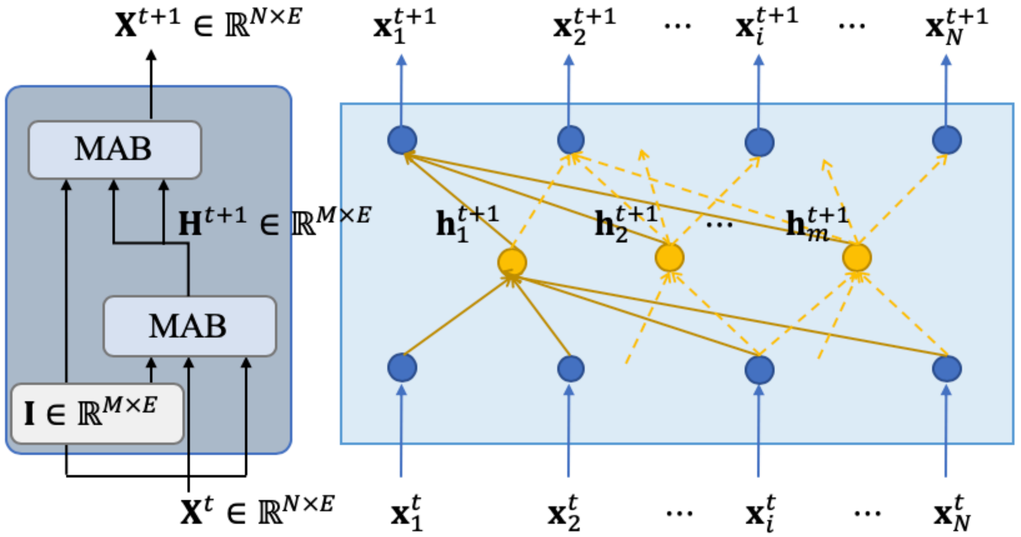
- Induced Multi-head Self Attention Block

$$\mathrm{IMSAB}_M(\mathbf{X}) = \mathrm{MAB}(\mathbf{X},\mathbf{H},\mathbf{H}),$$
$$\text{where } \mathbf{H} = \mathrm{MAB}(\mathbf{I},\mathbf{X},\mathbf{X}).$$

MSAB and IMSAB are multi-variate scoring functions and prove to be permutation equivariant.



2 Size Adaptation



Document Ranking

- Stack multiple blocks:

$$\mathbf{X}^{N_b}_{\text{MSAB}} = \underbrace{\text{MSAB}(\text{MSAB}\ldots(\text{MSAB}(\mathbf{X}^0))}_{N_b}),$$

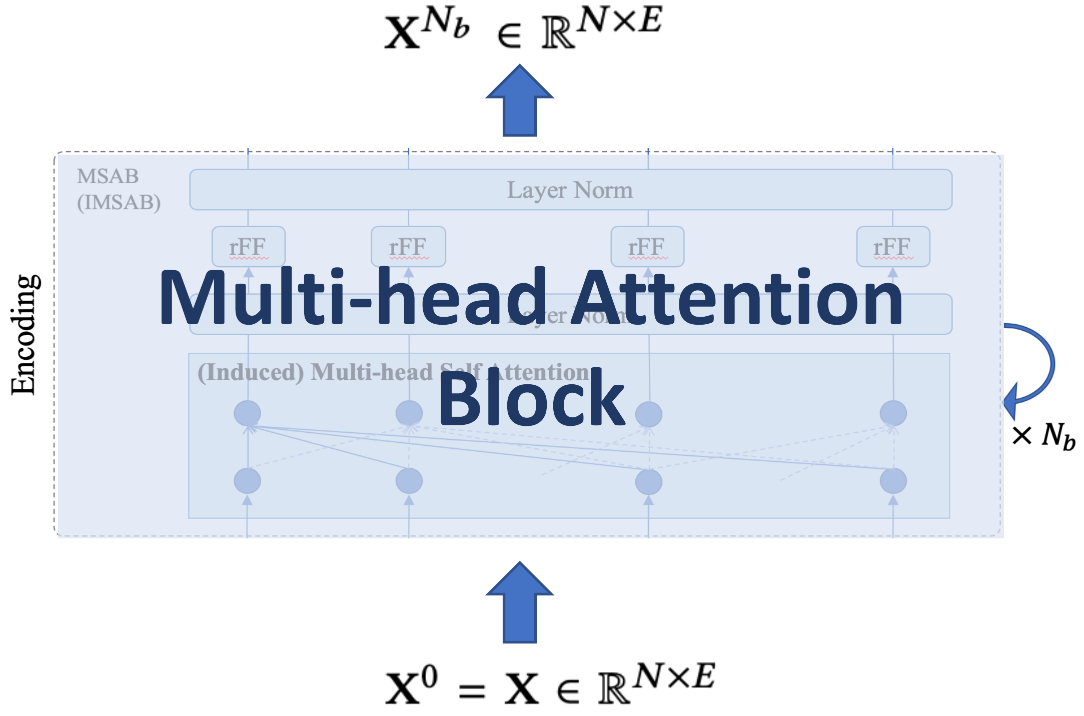
$$\text{SetRank}_{\text{MSAB}}(D) = \text{rFF}_1\left(\mathbf{X}^{N_b}_{\text{MSAB}}\right);$$

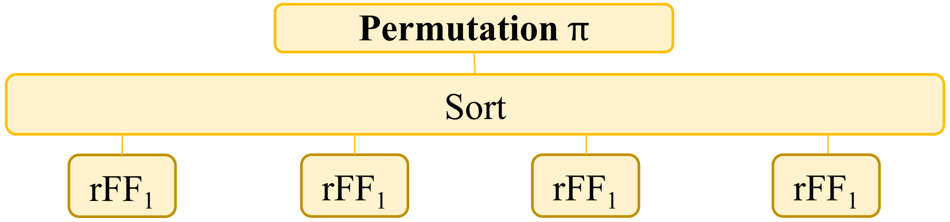
Stack of permutation equivariant functions is also a permutation equivariant function.

- Sort to get the documents permutation:

$$\hat{\pi}_{\text{MSAB}} = sort \circ \text{SetRank}_{\text{MSAB}}(D);$$

Sort the result of permutation equivariant functions yield Permutation-Invariant function.





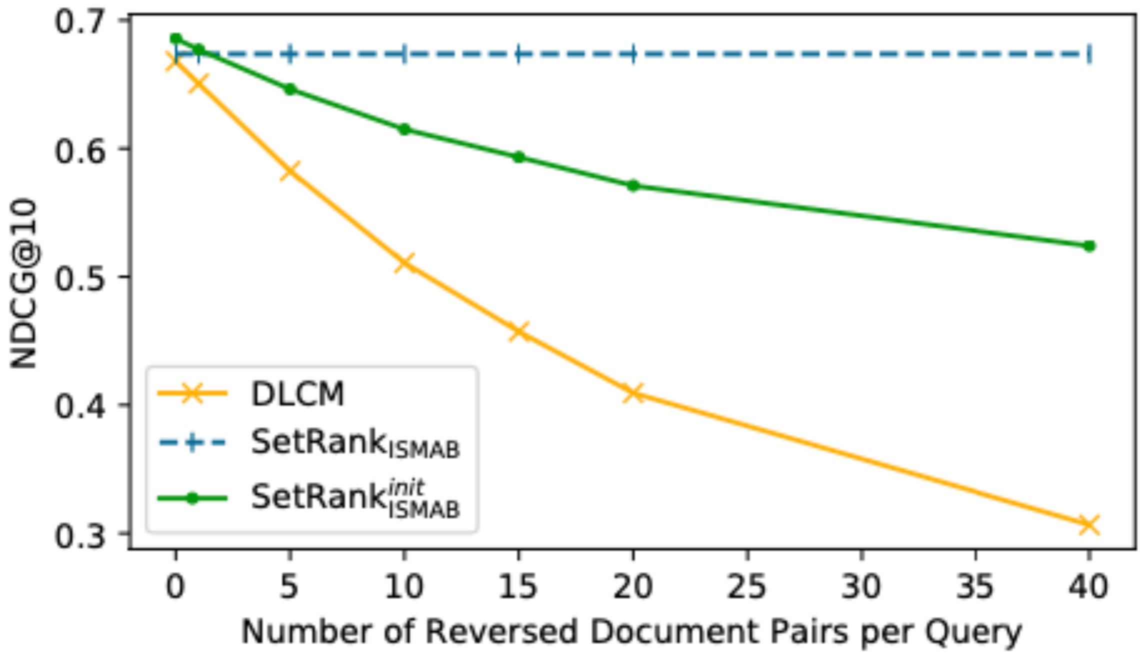
Experiment – Performance

(a) Ranking accuracies on Istella LETOR dataset

		Model	NDCG @1	NDCG @3	NDCG @5	NDCG @10
Obey PRP		RankSVM	0.5269	0.4867	0.5041	0.5529
		RankBoost	0.4457	0.3977	0.4097	0.4511
		Mart	0.6185	0.5633	0.5801	0.6285
		LambdaMart	0.6571	0.5982	0.6118	0.6591
		Without initial rankings				
Beyond PRP	**No Perm-Inv**	$\text{DLCM}^{w/o\ init}$	0.6272	0.5717	0.5848	0.6310
		GFS	0.6224	0.5796	0.5968	0.6508
	SetRank	$\text{SetRank}_{\text{MSAB}}$	$0.6702^{+\dagger}$	$0.6150^{+\dagger}$	$0.6282^{+\dagger}$	$0.6766^{+\dagger}$
		$\text{SetRank}_{\text{IMSAB}}$	$0.6733^{+\dagger}$	$0.6136^{+\dagger}$	$0.6278^{+\dagger}$	$0.6737^{+\dagger}$
		With initial rankings generated by LambdaMart				
	No Perm-Inv	DLCM	0.6558	0.6030^{+}	0.6194^{+}	0.6680^{+}
	SetRank	$\text{SetRank}^{init}_{\text{MSAB}}$	$0.6745^{+\dagger}$	$0.6201^{+\dagger}$	$\mathbf{0.6350}^{+\dagger}$	$0.6819^{+\dagger}$
		$\text{SetRank}^{init}_{\text{IMSAB}}$	$\mathbf{0.6760}^{+\dagger}$	$\mathbf{0.6202}^{+\dagger}$	$0.6345^{+\dagger}$	$\mathbf{0.6834}^{+\dagger}$

SetRank outperformed the traditional L2R baselines (obey PRP) and DLCM & GFS which break PRP but not fit permutation invariant requirement.

Experiment – Robustness



- Randomly reversed some document pairs in the ranking lists generated by LambdaMart for each query.
- Feed them to the models in the test phase.

SetRank which satisfied permutation invariance is more robust than **DLCM**.

The effects of set size adaptation

2 Size Adaptation

Model	# docs associated per query training set (N_0)	test set (N_1)	NDCG@1 (Δ)	NDCG@3 (Δ)	NDCG@5 (Δ)	NDCG@10 (Δ)
$\text{DLCM}^{w/o\,init}$	40	40	0.6233	0.5684	0.5825	0.6298
		240	0.5943 (-0.0290)	0.5394 (-0.0290)	0.5518 (-0.0307)	0.5964 (-0.0334)
		500	0.5844 (-0.0389)	0.5300 (-0.0384)	0.5428 (-0.0397)	0.5868 (-0.0430)
	240	240	0.6199	0.5708	0.5836	0.6309
	500	500	0.6258	0.5700	0.5833	0.6298
$\text{SetRank}_{\text{MSAB}}$	40	40	0.6702	0.6150	0.6282	0.6766
		240	0.6578 (-0.0124)	0.6026 (-0.0124)	0.6155 (-0.0127)	0.6602 (-0.0164)
		500	0.6533 (-0.0169)	0.5969 (-0.0181)	0.6102 (-0.0180)	0.6547 (-0.0219)
	240	240	0.6736	0.6141	0.6295	0.6777
	500	500	0.6712	0.6170	0.6316	0.6816
$\text{SetRank}_{\text{IMSAB}}$	40	40	0.6733	0.6136	0.6278	0.6737
		240	0.6674 (-0.0059)	0.6104 (-0.0032)	0.6244 (-0.0034)	0.6688 (-0.0049)
		500	0.6665 (-0.0068)	0.6082 (-0.0054)	0.6220 (-0.0058)	0.6662 (-0.0075)
	240	240	0.6699	0.6115	0.6264	0.6750
	500	500	0.6696	0.6117	0.6247	0.6727

Large Decrease

Small Decrease

SetRank is set size adaption.

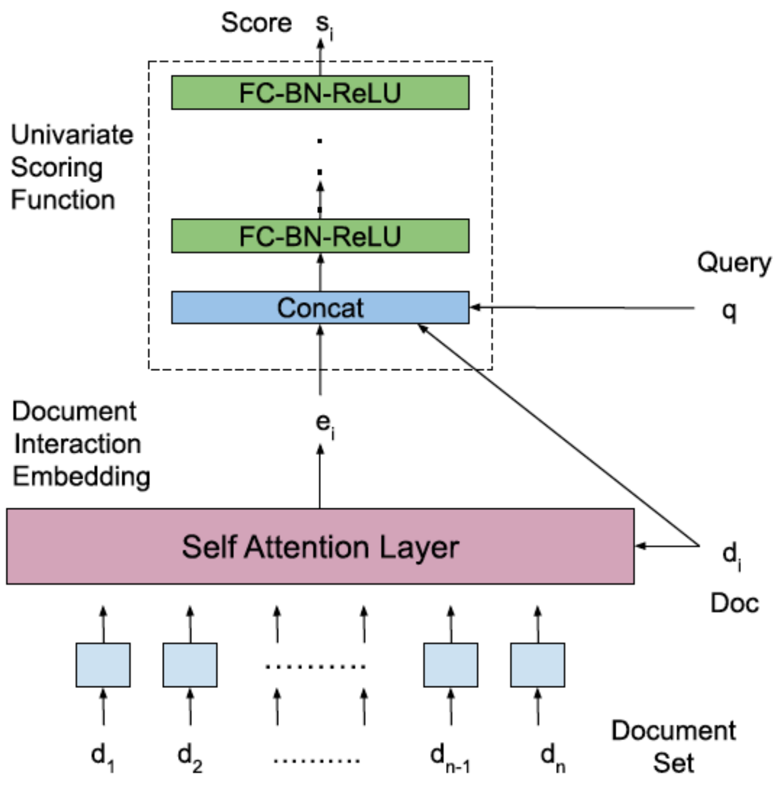
Attn-DIN (Pasumarthi et.al ICTIR 2020)

- Self-Attentive Document Interaction Network

- Propose another permutation invariant scoring function also based on Transformer.

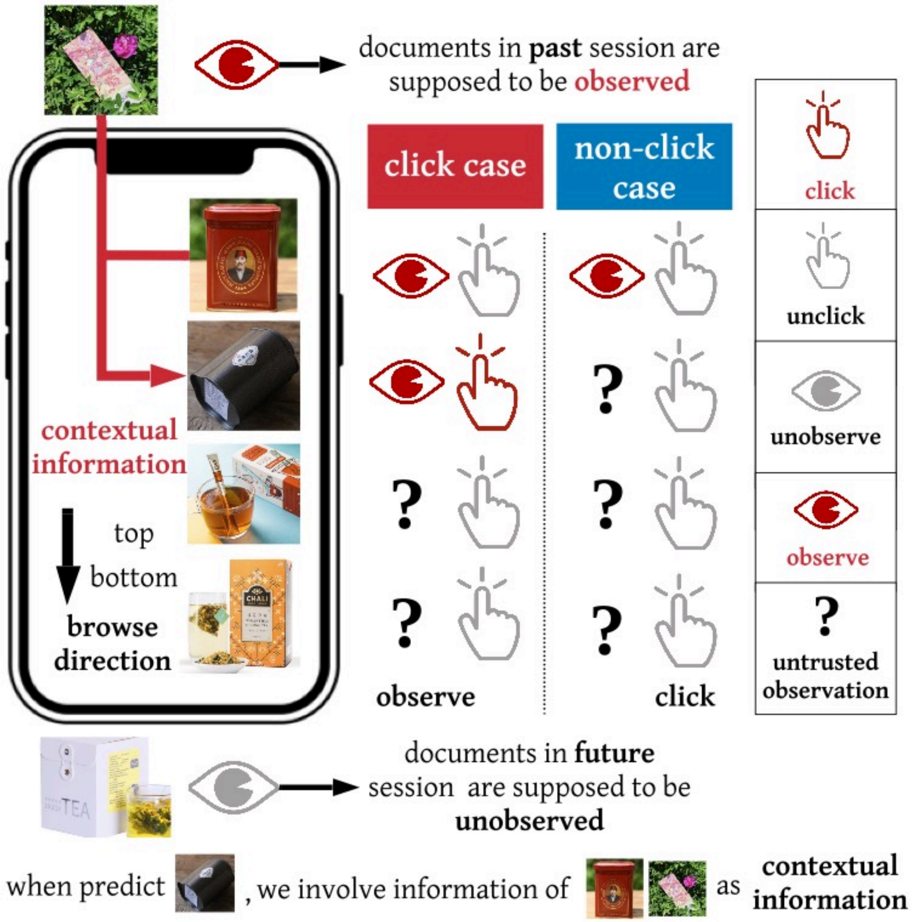
$$s_{DIN}(q, \boldsymbol{d}) = s(q, concat(\boldsymbol{d}, MHSA(\boldsymbol{d})))$$

- A univariate scoring function s_{DIN} to combine self-attention output with query and document features through concatenation.



Pasumarthi, Rama Kumar, Xuanhui Wang, Michael Bendersky, and Marc Najork. "Self-Attentive Document Interaction Networks for Permutation Equivariant Ranking." ICTIR 2020.

Unbiased Learning to Rank



AutoULTR framework: Dual Learning Algorithm

$$P(c_q^i = 1) = P(o_q^i = 1)P(r_q^i = 1)$$

$$\tilde{l}_{IRW}(E, q) = -\sum_{i=1, c_q^i=1}^{i=|\pi_q|} \frac{\mathcal{F}_\theta^1(\mathbf{X}_q)}{\mathcal{F}_\theta^i(\mathbf{X}_q)} \times \log \mathcal{G}_\phi^i(\pi_q)$$

$$\tilde{l}_{IPW}(S, q) = -\sum_{i=1, c_q^i=1}^{i=|\pi_q|} \frac{\mathcal{G}_\phi^1(\pi_q)}{\mathcal{G}_\phi^i(\pi_q)} \times \log \mathcal{F}_\theta^i(\mathbf{X}_q)$$

A multivariant relevance estimation function $\mathcal{F}$ parameterized by θ for ranking system S and a propensity estimation function $\mathcal{G}$ parameterized by ϕ for propensity model E .

Permutation invariance of function *F* a necessary condition and a sufficient condition for the convergence of Dual Learning Algorithm (DLA).

Yang, Tao, Shikai Fang, Shibo Li, Yulan Wang, and Qingyao Ai. Analysis of Multivariate Scoring Functions for Automatic Unbiased Learning to Rank. CIKM’ 20

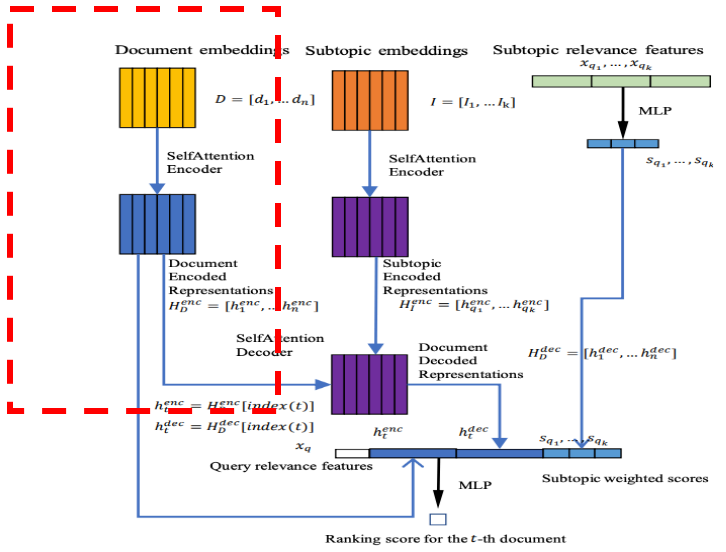
Diverse Ranking (Qin et.al CIKM2020 & Yan et.al WWW2021)

SetRank → DESA

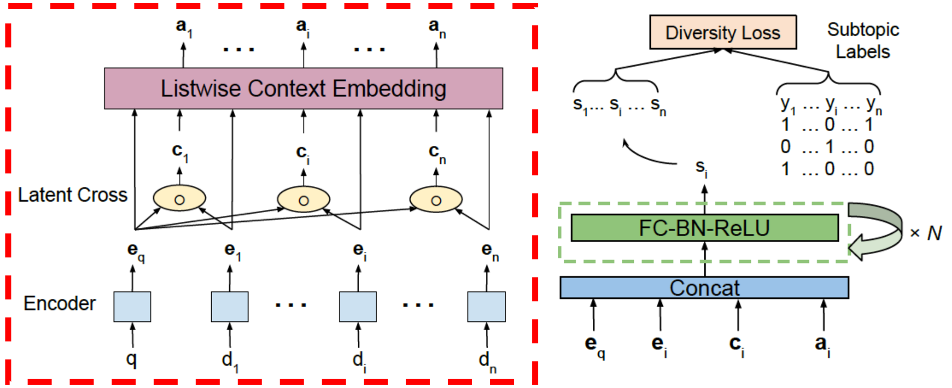
- DESA is inspired by SetRank model in IR with subtopic representations.

DIN → DALETOR

- Document Interaction Network (DIN) is incorporated into DALETOR framework to enhance scoring function.



DESA Model



DALETOR Model

6. Conclusion and Open Discussions

Summary

- Theoretical foundation of ad hoc retrieval paradigms is the probability ranking principle (PRP)
  - Most existing ranking models obey PRP: ranking = scoring + sorting
  - However, PRP is suboptimal, goes beyond PRP !
  - Key idea: breaking the independency of document relevance assumption
- Modeling the dependency sequentially
  - Heuristic sequential ranking models
  - Learning sequential ranking models
- Modeling the dependency globally
  - List inputted global ranking models
  - Set inputted global ranking models

Future Research Opportunities

- **Model**: Explore Contextual Ranking Functions
  - Other properties of contextual ranking function
- **Data**: Learn from Biased Data
  - Hard to annotate
  - Noise label
- **Algorithm**: Optimize with Richer Supervision Signals
  - Multiple retrieval metrics
  - Session-level retrieval metrics
- **Task**: From 1D Ranking to 2D Representation
  - Diverse representative forms beyond ranking list

OP1 – Explore Contextual Ranking Functions

Multi-variate Scoring Function

Model maps query-doc pairs in a document set to a list of scores.

$$f\left(\begin{bmatrix}\text{Query}\\ \text{Doc 1}\end{bmatrix}, \begin{bmatrix}\text{Query}\\ \text{Doc 2}\end{bmatrix}, \cdots, \begin{bmatrix}\text{Query}\\ \text{Doc N}\end{bmatrix}\right) \rightarrow [\,\boxed{\text{Score 1}}, \boxed{\text{Score 2}}, \cdots, \boxed{\text{Score N}}\,]$$

$$f: \mathbb{D}^N \mapsto \mathbb{R}^N$$

Permutation Invariant
Scoring Function

$$f(\{d_1, d_2, d_3, d_4\}) = f(\{d_2, d_1, d_3, d_4\})$$

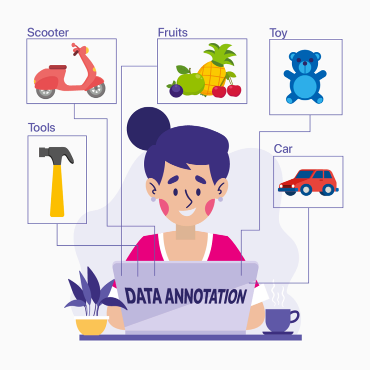
Size Adaptation
Scoring Function

$$f(\{d_1, d_2\})$$
$$f(\{d_1, d_2, d_3, d_4, d_5 \ldots\})$$

Other Properties ?

Low Dimension ?
Local Sensitive ?

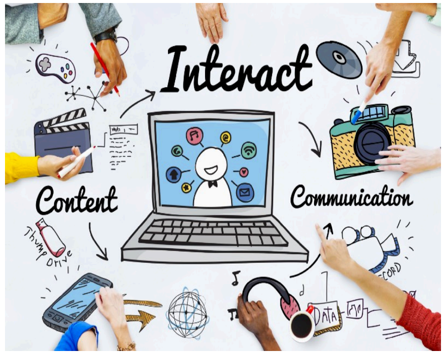
OP2 – Learn from Biased Data



Annotated Relevance Data

- Small Dataset
- Low Diversity
- Accurate Label

VS



Observed Relevance Data

- Huge Dataset
- High Diversity
- Noisy Label

Observed Relevance Data:

- Click-through data
- Mouse dwell data
- Like / Follow / Retweet data
- ……

Noisy Label:

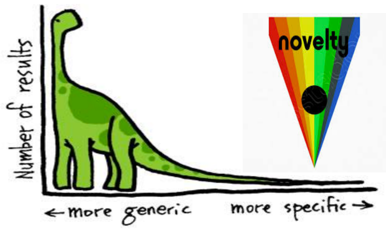
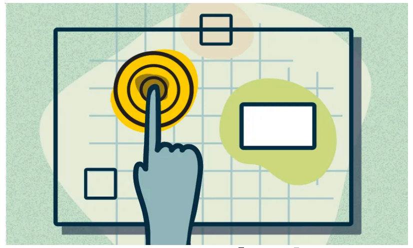
- Position bias
- Selection bias
- Click noise
- Exposure bias
- ……

Weak Supervision Learning

OR

Few-shot Learning

OP3 – Optimize with Richer Supervision Signals

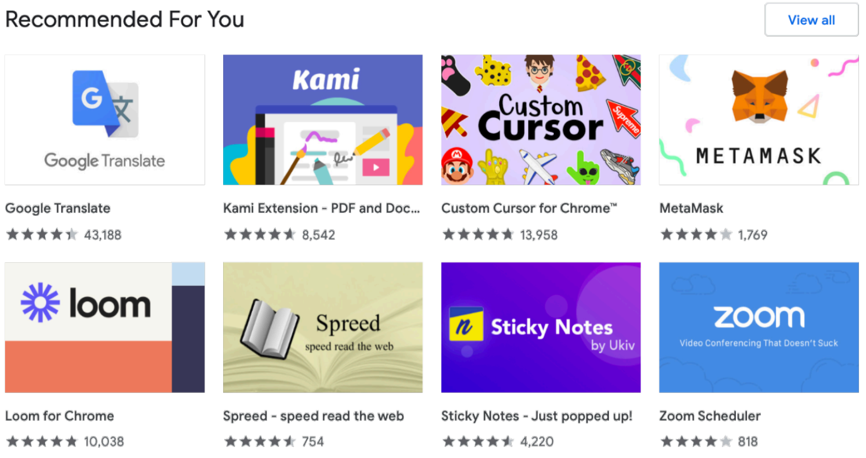
	Item	User	Session
Instant Feedback	 Product Novelty	 User Click	 Temporal Diversity
Delayed Feedback	 Conversion Rate	 User Engagement	 Platform Profits
Incomplete Feedback	 Like & Share	 User Dwell Time	

OP4: From 1D Ranking to 2D Representation (Zhuang et.al WWW2021)

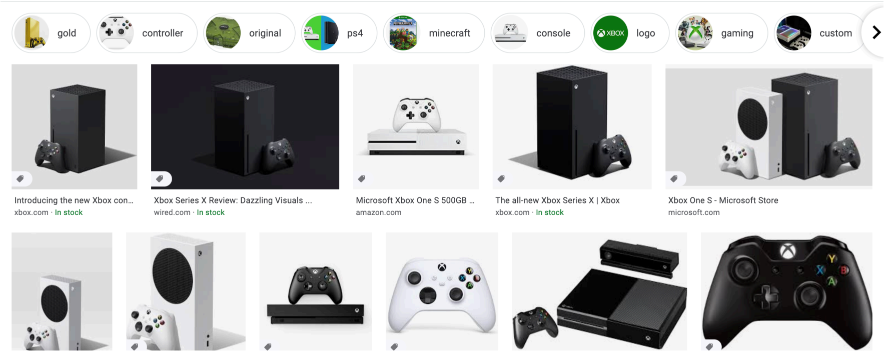
- Assuming a simple position-based bias or enforcing a sequential order in user examination behaviors.

- Insufficient to capture complex real-world user behaviors and hardly generalize to modern user interfaces (UI) in web applications.

Results shown in a **Grid View**



Results shown in a **Complex View**



Reference

- Qingyao Ai, Keping Bi, Jiafeng Guo and W. Bruce Croft. Learning a Deep Listwise Context Model for Ranking Refinement. The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, Ann Arbor Michigan, U.S.A. (SIGIR 2018).
- Qingyao Ai, Xuanhui Wang, Sebastian Bruch, Nadav Golbandi, Michael Bendersky, and Marc Najork. 2019. Learning Groupwise Multivariate Scoring Functions Using Deep Neural Networks. In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval (ICTIR '19). Association for Computing Machinery, New York, NY, USA, 85–92.
- Liang Pang, Jun Xu, Qingyao Ai, Yanyan Lan, Xueqi Cheng, and Jirong Wen. SetRank: Learning a Permutation-Invariant Ranking Model for Information Retrieval. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'20)*
- Lee, Juho, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. "Set transformer: A framework for attention-based permutation-invariant neural networks." In International Conference on Machine Learning, pp. 3744-3753. PMLR, 2019.
- Zaheer, Manzil, Satwik Kottur, Siamak Ravanbhakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. "Deep Sets." In Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 3394-3404. 2017.
- Yang, Tao, Shikai Fang, Shibo Li, Yulan Wang, and Qingyao Ai. Analysis of Multivariate Scoring Functions for Automatic Unbiased Learning to Rank. CIKM 2020.
- Pasumarthi, Rama Kumar, Xuanhui Wang, Michael Bendersky, and Marc Najork. "Self-Attentive Document Interaction Networks for Permutation Equivariant Ranking." ICTIR 2020.
- Qin, Xubo, Zhicheng Dou, and Ji-Rong Wen. "Diversifying Search Results using Self-Attention Network." In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 1265-1274. 2020.
- Zhuang, Honglei, Zhen Qin, Xuanhui Wang, Mike Bendersky, Xinyu Qian, Po Hu, and Chary Chen. "Cross-Positional Attention for Debiasing Clicks." (2021).

Thanks Q & A

Name: Liang Pang | Email: pangliang@ict.ac.cn