CIKM 2020
19-23 OCTOBER
Online

PSTIE: Time Information Enhanced Personalized Search

Zhengyi Ma, Zhicheng Dou, Guanyue Bian, Ji-Rong Wen

Renmin University of China

Presented by Zhengyi Ma

Outline

Background

Related Work

Our method
- Problem Formulation
- Time-Sensitive User Interest Modeling
- Time-Sensitive Personalized Ranking

Experiments
- Experiment Settings
- Experiment Results and Analysis

Conclusion

Background

Query: Apple

Apple Company

Apple Fruit

Traditional search engine cannot distinguish different information needs of users.

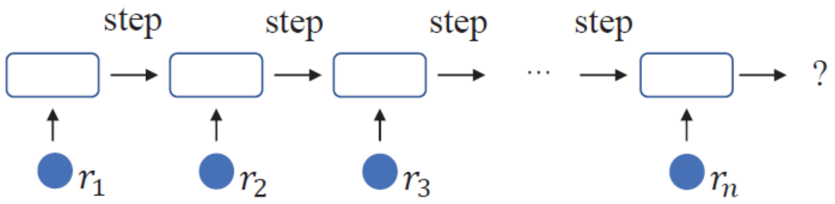
<u>Personalized Search</u>: Re-rank the candidate document list based on user’s history.

Related Work

Traditional methods

P-click: click information

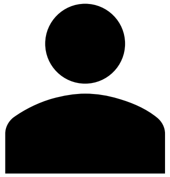
SLTB: click based features + topic based features



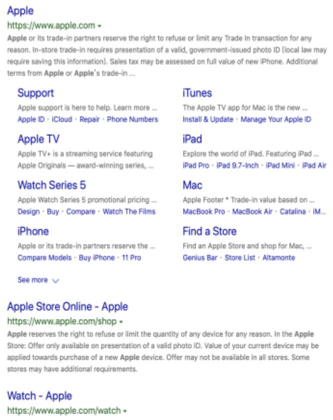
Deep learning methods

HRNN: Hierarchical RNN + Attention mechanism

PSGAN: Generative adversarial network



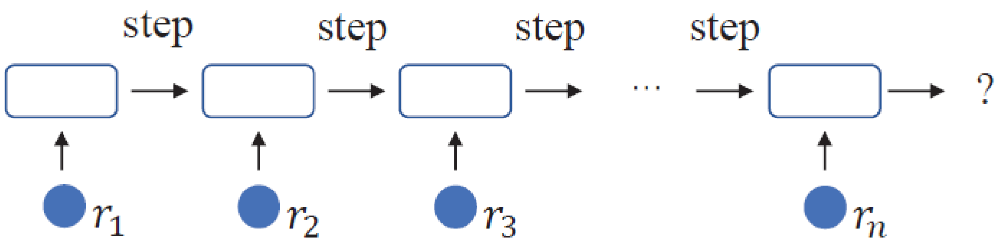
re-rank



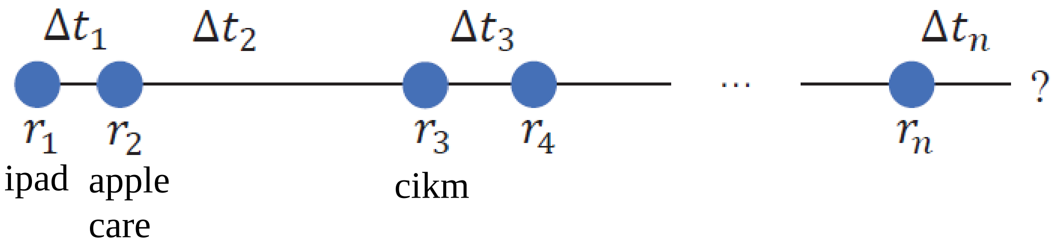
User Interest

Document list

Our Proposal



a) RNN for modeling user profile



b) Behavior of users in personalized search

Existing methods ignored the explicit time information between user's past actions.

Time sensitive interest can be used to enhance personalized search.

Our Proposal

Our contribution includes:

1. We leverage fine-grained time information within user's historical behaviors to improve personalized ranking quality.

2. We track two kinds of time-sensitive evolution of users, including query intent evolution and document interest evolution. We consider both short-term local correlations and long-term re-finding influences between user's search history.

3. We use two matching methods, a representation-based matching and an interaction-based matching, to fuse the time-sensitive interest representations into personalized ranking.

Problem Formulation

For each user u, we have its query log $L = \{(q_1, D_1, t_1), \dots, (q_n, D_n, t_n)\}$

We also calculate the average of the clicked documents under q_i as $\hat{d}_i$.

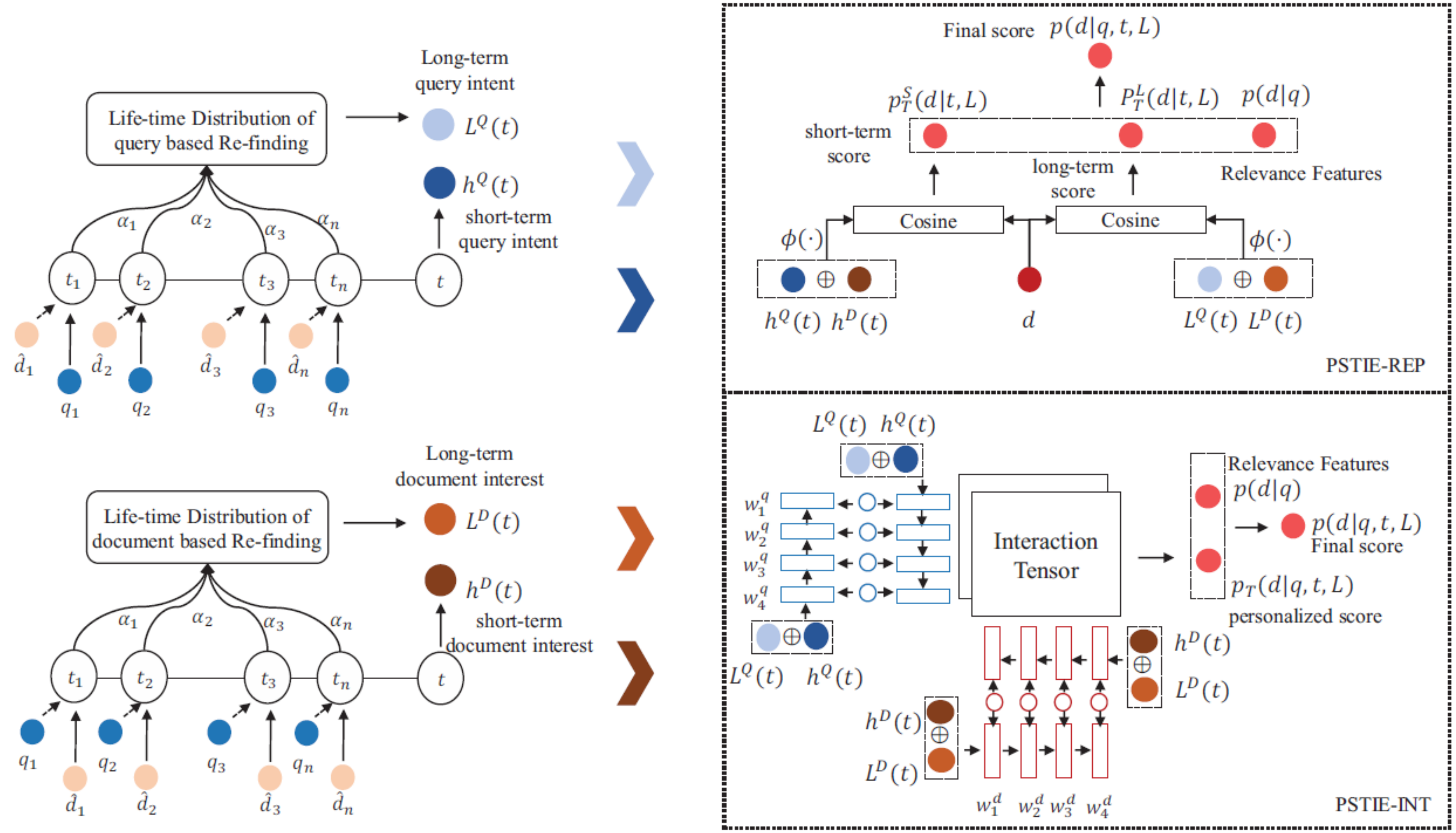
We define the output of our model as $p(d|q,t,L) = \phi(p_T(d|q,t,L), p(d|q))$

- q: the current query, t: the current time, L: the user's query log, d: the candidate document
- $p_T(d|q,t,L)$: the time-sensitive personalized score of document d at time t.
- $p(d|q)$: the ad-hoc relevance between q and d

Overall Framework

Document-driven
Time-aware LSTM

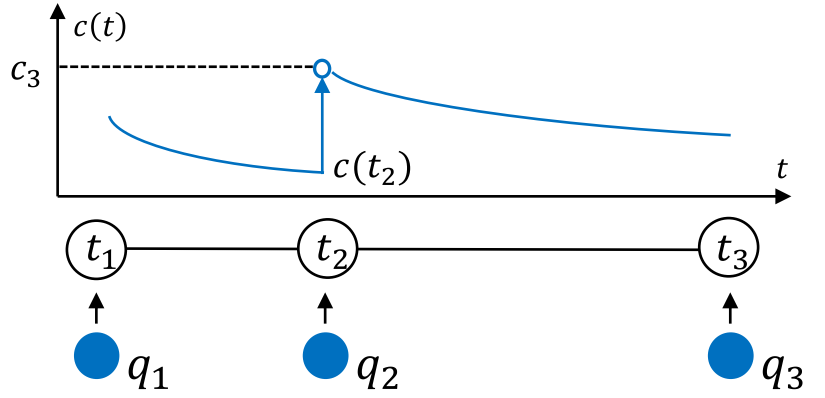
Query-driven
Time-aware LSTM



Time-Sensitive Interest Modeling

We design document-driven time-aware LSTM for modeling short-term query intent

$$h(t) = o_k \odot \left(2\sigma\left(2c(t)\right) - 1\right)$$



Query intent shows self-exciting characteristic and decays with time.

$$c(t) = \bar{c}_{i+1} + (c_{i+1} - \bar{c}_{i+1}) \exp\left(-\delta_{i+1}(t - t_i)\right), t \in (t_i, t_{i+1}]$$

$c_{i+1} \leftarrow f_{i+1} \odot c(t_i) + i_{i+1} \odot z_{i+1}$ *# the jump of $c(t)$ to a specific value c_{i+1} at t_i*

The clicked documents $\hat{d}_i$ can satisfy the user's information need after the issued query q_i

$\bar{c}_{i+1} \leftarrow \bar{f}_{i+1} \odot \bar{c}_i + \bar{i}_{i+1} \odot z_{i+1} + \bar{d}_{i+1} \odot \hat{d}_i$ *# the $c(t)$ will decays towards target $\bar{c}_{i+1}$*

We calculate historical query intent representations $H_q = \{h^Q(t_1), h^Q(t_2), \ldots h^Q(t_n)\}$ and short-term query intent $h^Q(t)$ using query-driven time-aware LSTM

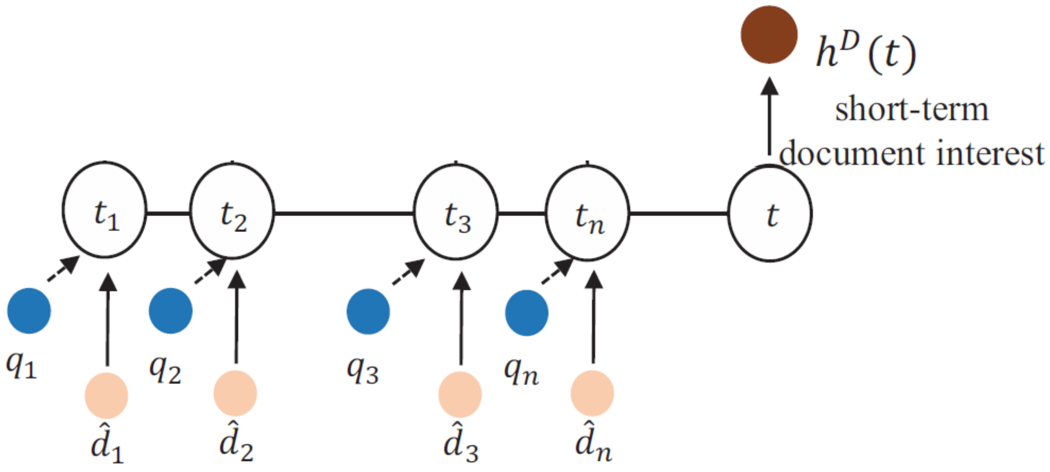
Time-Sensitive Interest Modeling

For document sequences, we design query-driven time-aware LSTM since the query straightly reflects the information need of users.

$c_{i+1} \longleftarrow f_{i+1} \odot c(t_i) + i_{i+1} \odot z_{i+1} + \bar{q}_{i+1} \odot q_i$

$\bar{c}_{i+1} \leftarrow \bar{f}_{i+1} \odot \bar{c}_i + \bar{\imath}_{i+1} \odot z_{i+1}$

We calculate historical document interest representations $H_d = \{h^d(t_1), h^d(t_2), \ldots h^d(t_n)\}$ and short-term document interest $h^{\mathrm{D}}(t)$

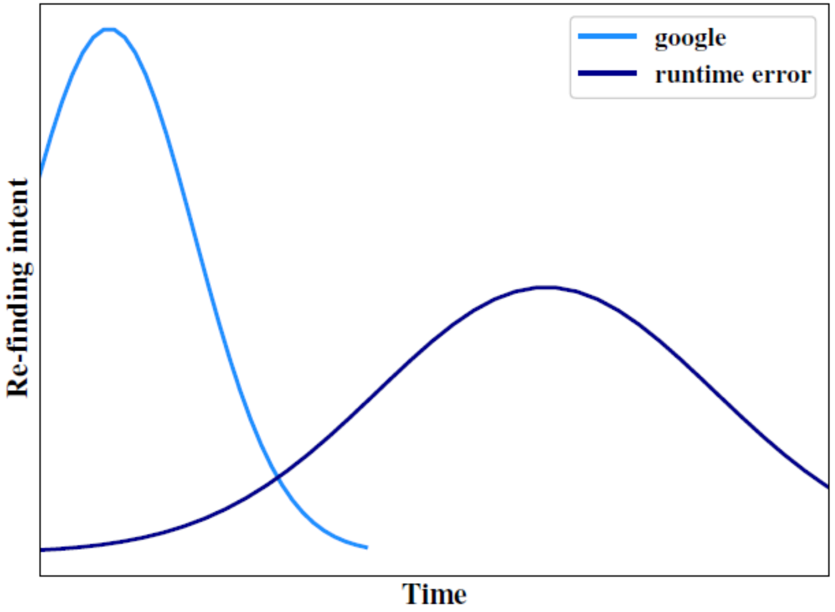


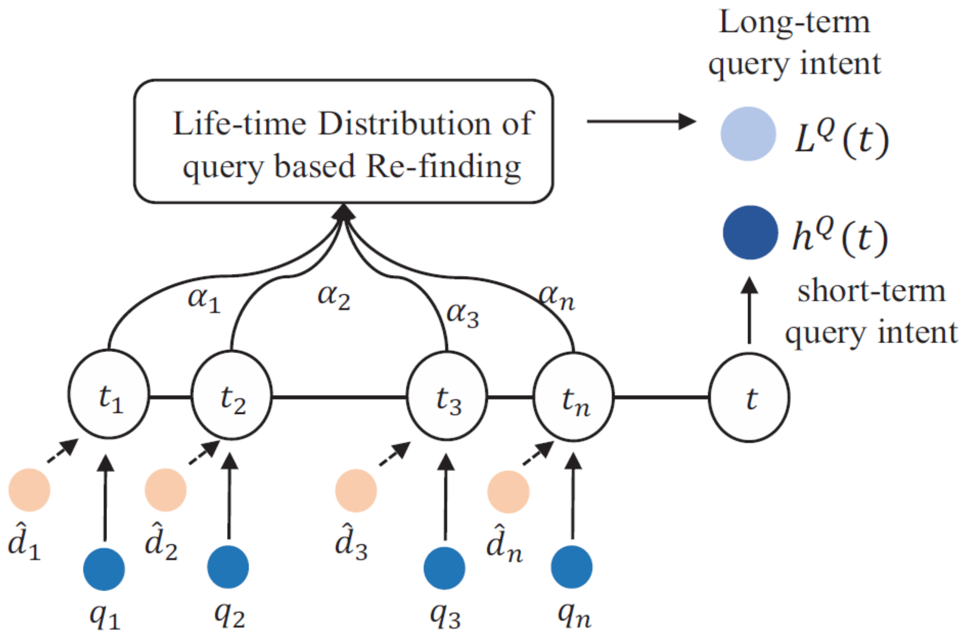
Query-driven time-aware LSTM

Time-Sensitive Interest Modeling

Time-aware LSTM can hardly capture the influence a long time ago.

Users tend to show similar interest near the end of the information's lifetime so it is natural to use Gaussian mixture distribution to model the life-time evolution of re-finding.





Time-Sensitive Interest Modeling

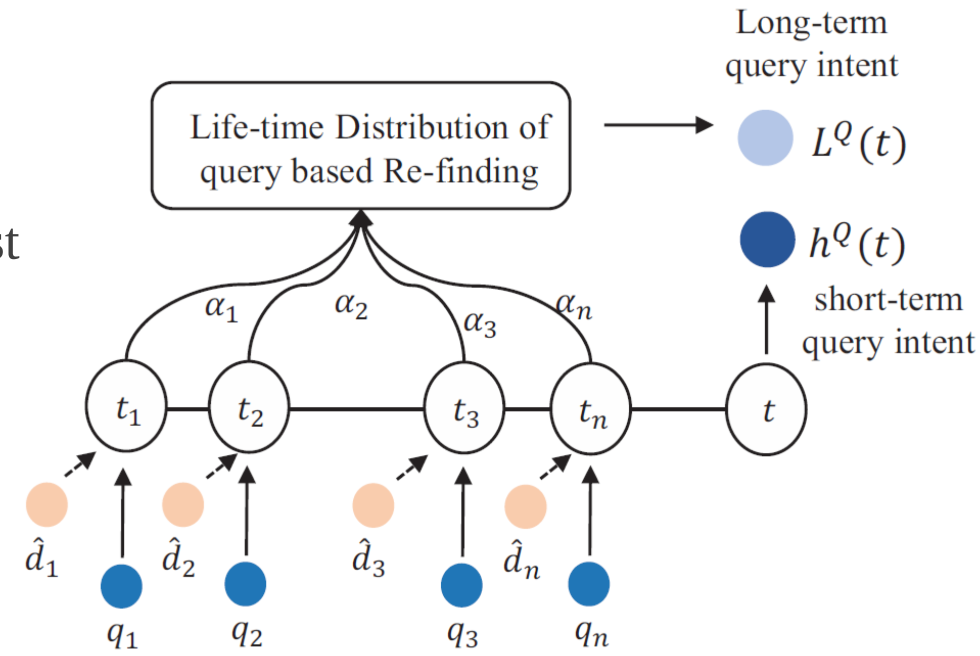
We calculate the re-finding possibility based on the query-specific Gaussian mixture distribution

$$\alpha_i = N(\delta t_i | \mu_i, \sigma_i) \text{ , } \delta t_i = t - t_i$$

And calculate the long-term query intent and document interest

$$L^Q(t) = \sum_{i=1}^{n} \frac{\exp(\alpha_i)}{\sum_{j=1}^{n} \exp(\alpha_j)} h^Q(t_i)$$

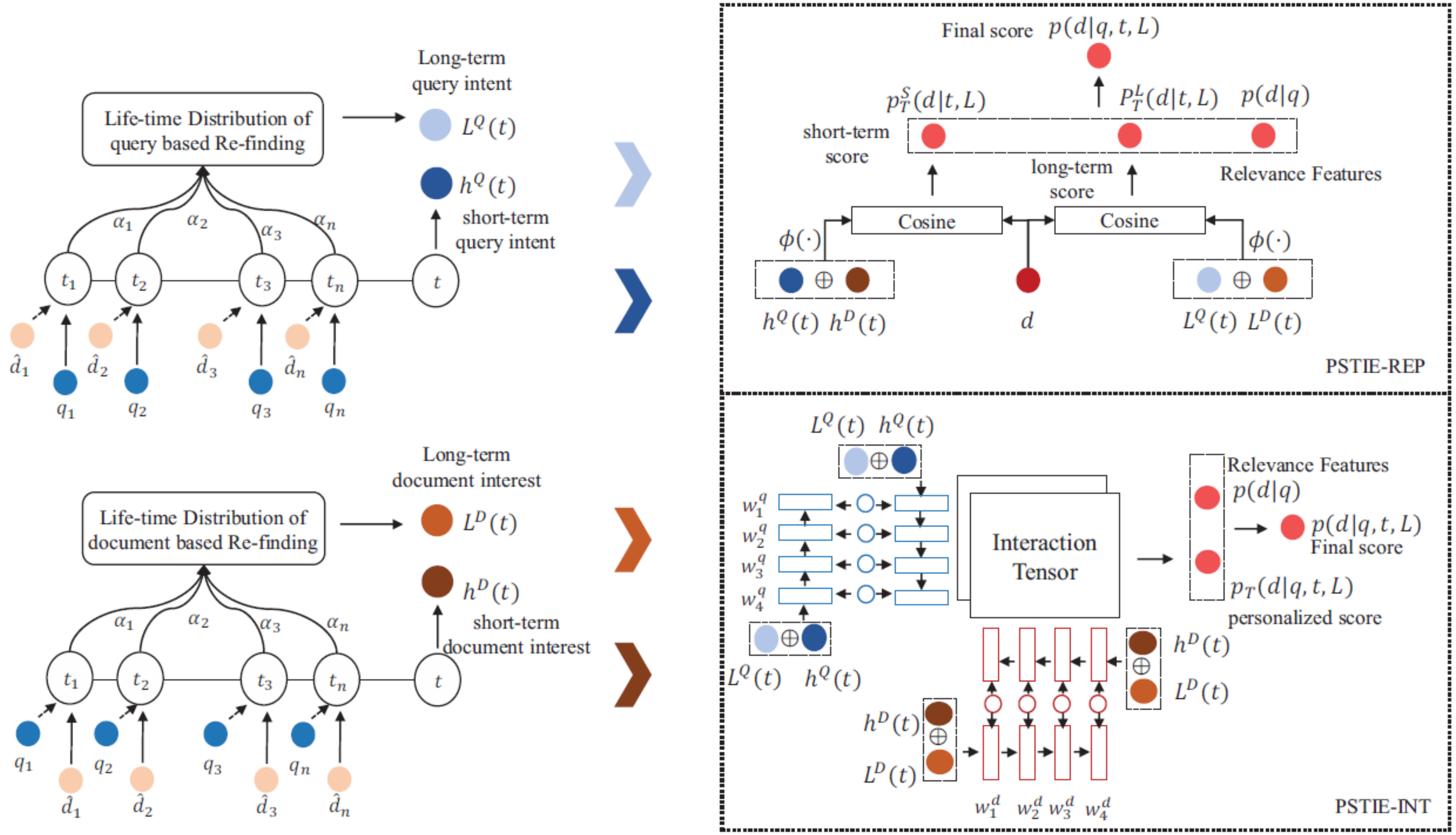
$$L^D(t) = \sum_{i=1}^{n} \frac{\exp(\alpha_i)}{\sum_{j=1}^{n} \exp(\alpha_j)} h^D(t_i)$$



Overall Framework

Document-driven
Time-aware LSTM

Query-driven
Time-aware LSTM



Time-Sensitive Personalized Ranking

PSTIE-REP: Representation-based Similarity

Short-term score:

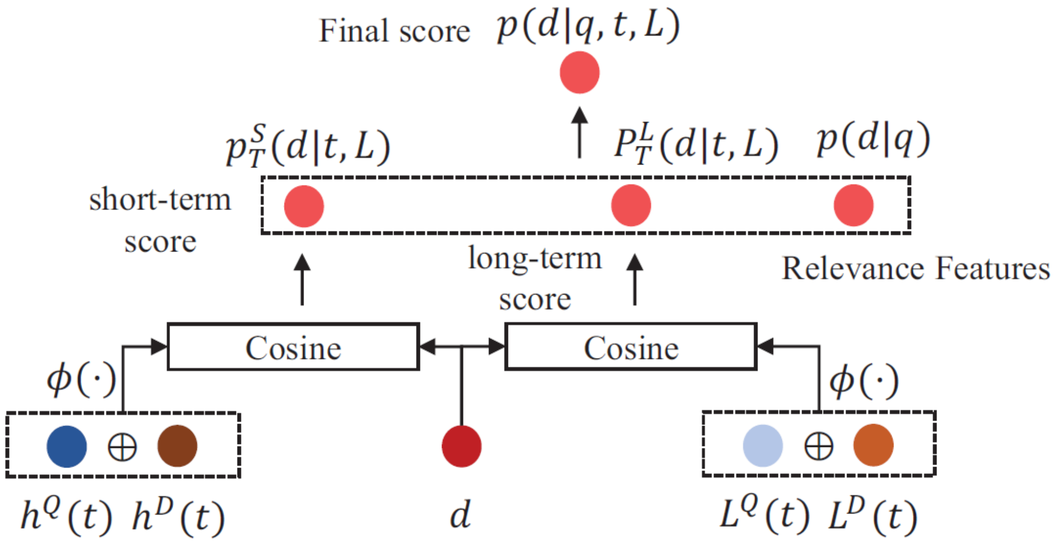
- $p_T^S(d|q,t,L) = \mathrm{sim}(\phi([h^Q(t); h^D(t)]), d)$

Long-term score:

- $p_T^L(d|q,t,L) = \mathrm{sim}(\phi([L^Q(t); L^D(t)]), d)$

Personalized score:

- $p_T(d|q,t,L) = \phi(p_T^S(d|q,t,L), p_T^L(d|q,t,L))$



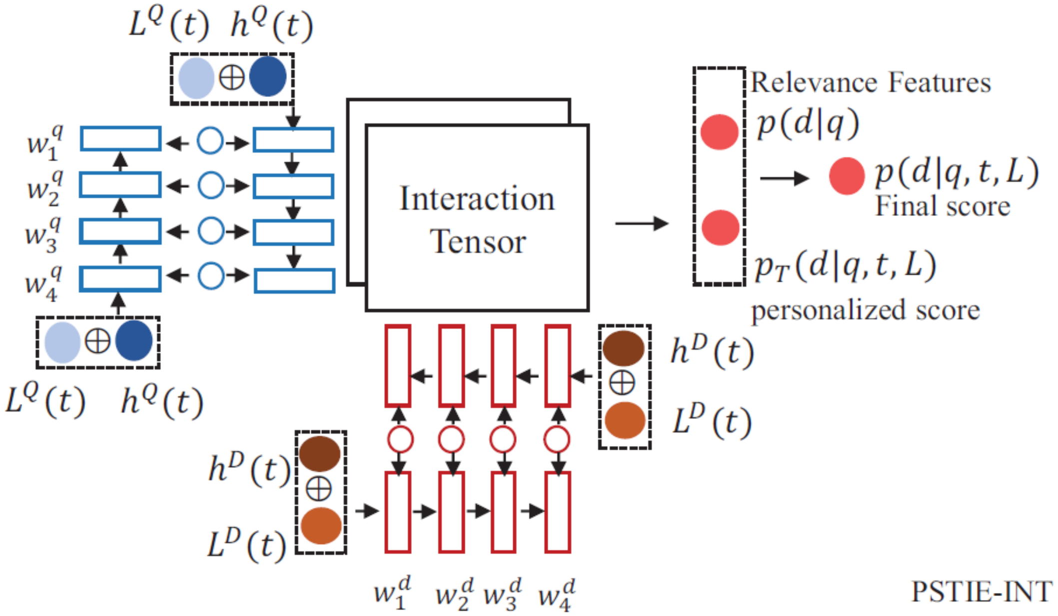
PSTIE-REP

Time-Sensitive Personalized Ranking

PSTIE-ITE: Interaction-based Matching

- Using query intent and document interest to initialize the bi-directional LSTM of interactive matching model (*MV-LSTM*)

PSTIE-ITE can naturally incorporate query into ranking module, and capture word-level matching feature.



PSTIE-INT

Experiment Setting

Dataset

- AOL Dataset: 1st March 2006 to 31st May 2006
- Commercial Dataset: 1st Jan. 2013 to 28th Feb. 2013

AOL		Commercial	
Item	Statistic	Item	Statistic
days	92	days	58
users	118,067	users	7,648
queries	3,461,636	queries	694,837
distinct queries	1,555,829	distinct queries	278,661
SAT-clicks	4,701,531	SAT-clicks	443,428
Co-queries	8,184,227	Co-query	4,109,396
Re-queries	84.70%	Re-query	80.75%

Baselines

- Ad-hoc ranking models: BM25, KNRM, MV-LSTM
- Personalized models: P-Click, SLTB, HRNN, PSGAN,
- Personalized models that exploit time information: HRNN+Time, HTime-LSTM, H-CTLSTM

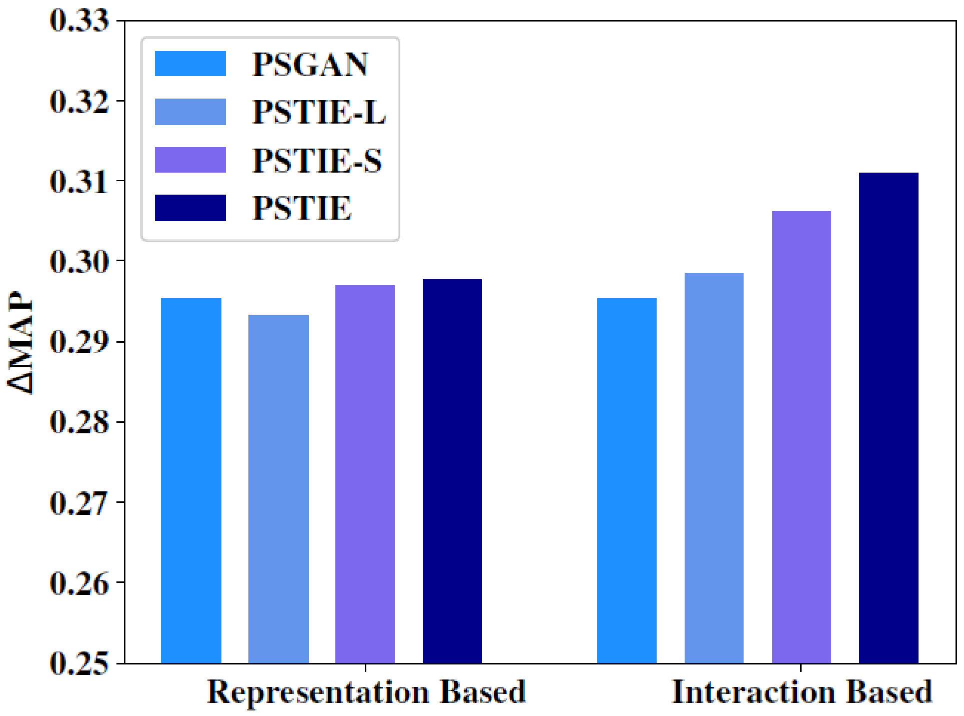
Evaluation

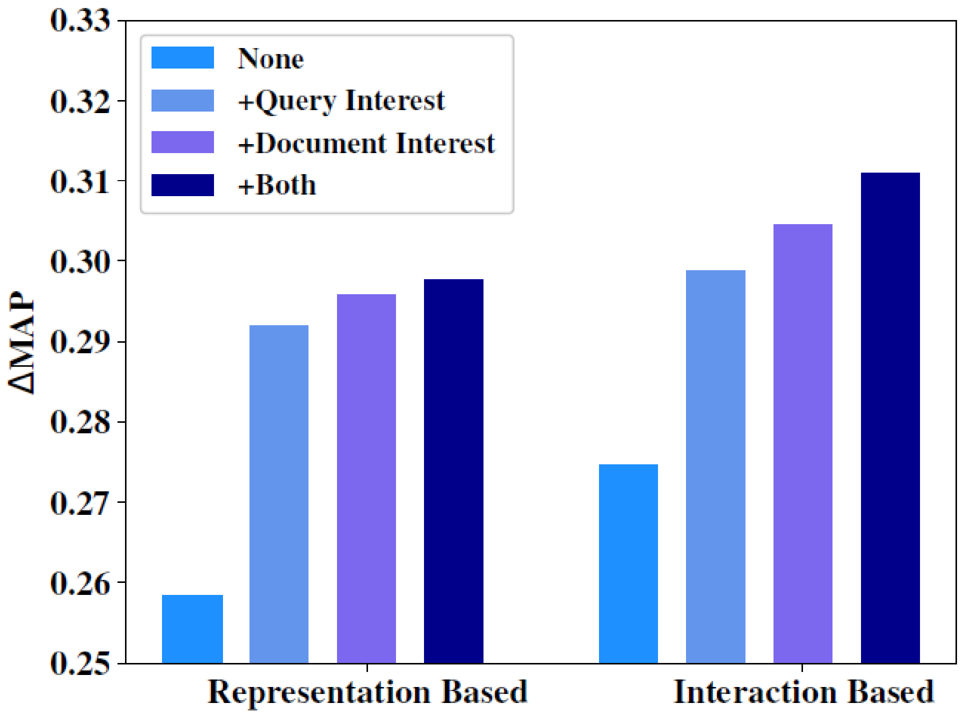
- MAP, MRR, P-Improve, P@1, P@3, P@5

Overall Performance

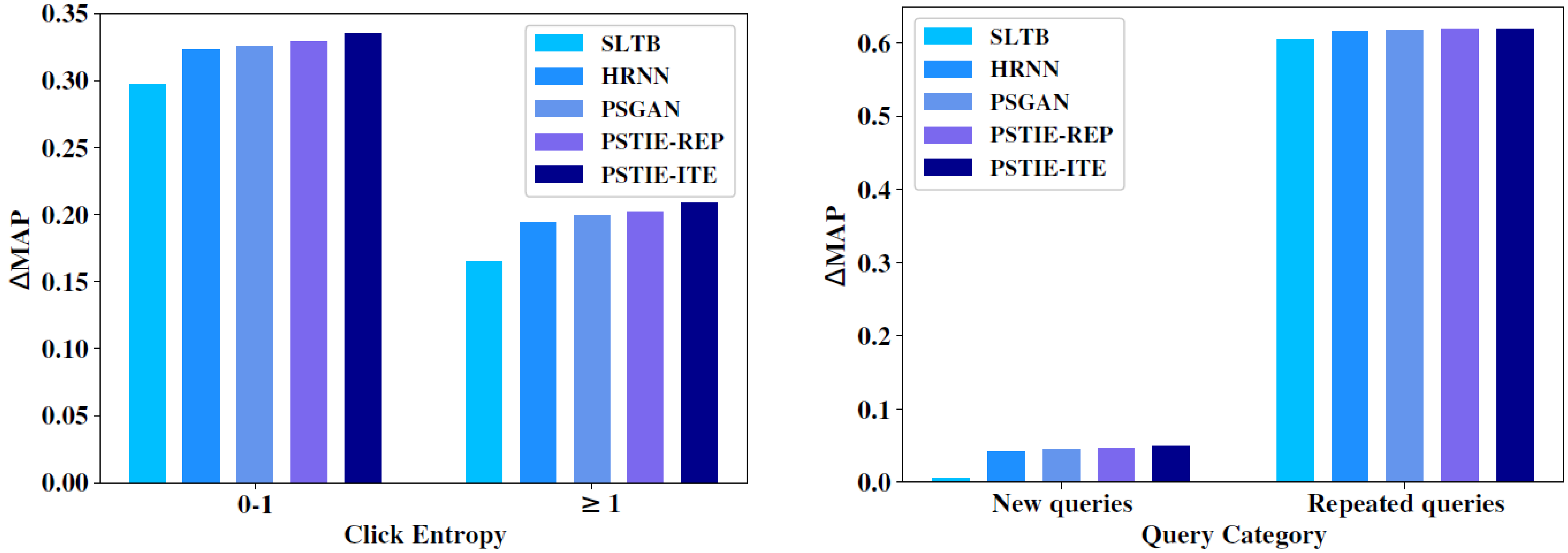
Models	AOL Query Log						Commercial Query Log					
	MAP	MRR	P-Imp	P@1	P@3	P@5	MAP	MRR	P-Imp	P@1	P@3	P@5
Ori.R	.2529	.2640	-	.1531	.2769	.3492	.7399	.7506	-	.6162	.8459	.9394
K-NRM	.4298	.4399	.6633	.2718	.5130	.6089	.4927	.5007	.0665	.2855	.3391	.3646
MV-LSTM	.4315	.4452	.6605	.2762	.5186	.6131	.4893	.4966	.0624	.2816	.3348	.3592
P-Click	.4221	.4305	.1657	.3780	.4128	.4431	.7509	.7634	.0611	.6260	.8823	.9598
SLTB	.5113	.5237	.3374	.4693	.5244	.5507	.7921	.7998	.1184	.6901	.9016	.9573
HRNN	.5438	.5565	.5927	.4841	.5663	.6042	.8065	.8191	.2401	.7127	.9061	.9590
PSGAN	.5482	.5609	.5985	.4898	.5741	.6190	.8135	.8234	.2494	.7174	.9114	.9658
HRNN+time	.5452	.5554	.5934	.4861	.5623	.6076	.8017	.8136	.2324	.7097	.9012	.9526
HTime-LSTM	.5476	.5578	.5975	.4896	.5677	.6097	.8077	.8210	.2413	.7156	.9131	.9610
H-CTLSTM	.5479	.5574	.5984	.4875	.5687	.6127	.8094	.8231	.2386	.7199	.9165	.9645
PSTIE-REP	.5506	.5610	.6042	.4929	.5734	.6261	.8105	.8238	.2445	.7210	.9181	.9680
PSTIE-ITE	**.5639**†	**.5769**†	**.6847**†	**.5033**†	**.5965**†	**.6413**†	**.8211**†	**.8301**†	**.2636**†	**.7295**†	**.9274**†	**.9766**†

Ablation Performance

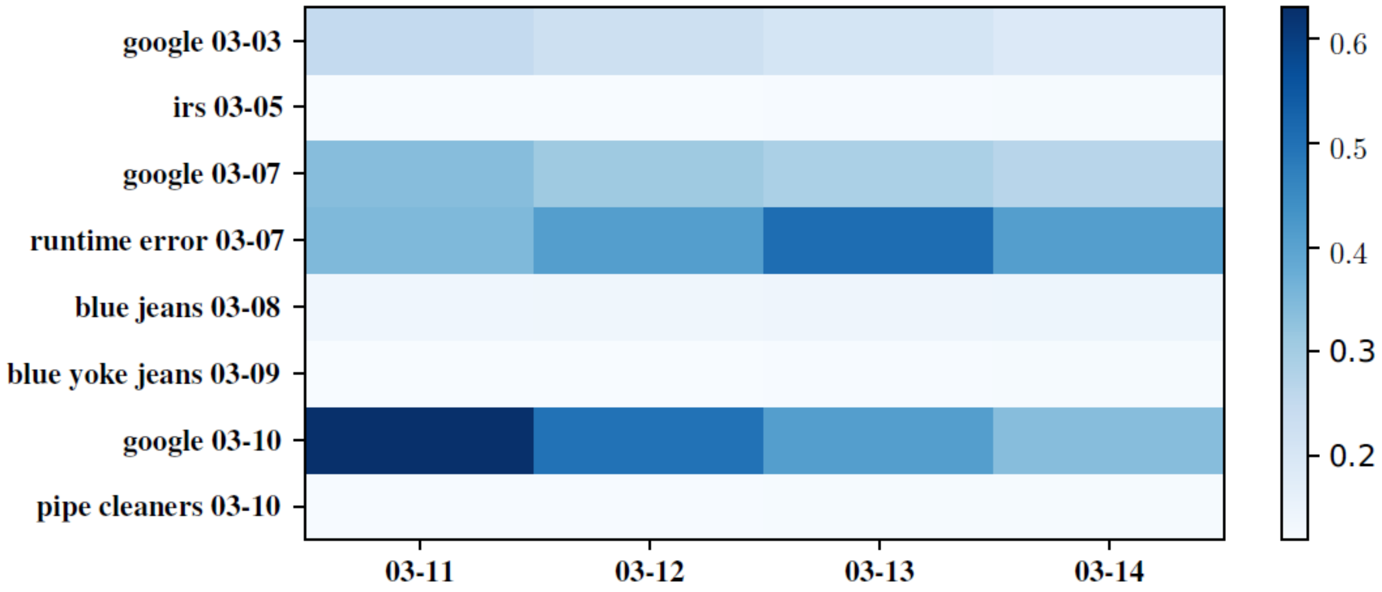




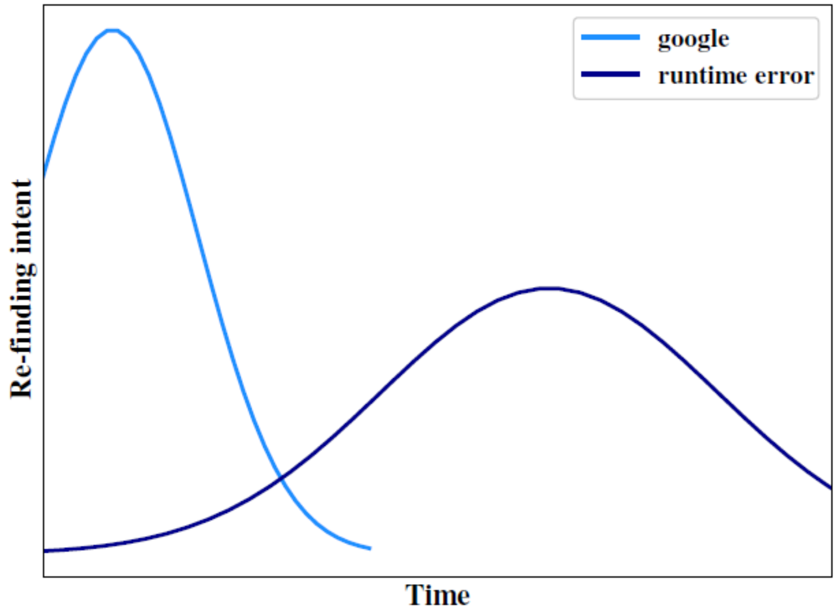
Results on Different Query Sets



Case Study



Visualization



Query: “Google” VS “runtime error”

Take-away Today

1. We leverage fine-grained time information within user's historical behaviors to improve personalized ranking quality.

2. We track two kinds of time-sensitive evolution of users, including query intent evolution and document interest evolution. We consider both short-term local correlations and long-term re-finding influences between user's search history.

3. We use two matching methods, a representation-based matching and an interaction-based matching, to fuse the time-sensitive interest representations into personalized ranking.

Thanks For Your Attention!

PSTIE: Time Information Enhanced Personalized Search

Email: zymaa@ruc.edu.cn

Zhengyi Ma, Zhicheng Dou, Guanyue Bian, Ji-Rong Wen from Renmin University of China