

Received 30 May 2023, accepted 23 June 2023, date of publication 27 June 2023, date of current version 6 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3289820

RESEARCH ARTICLE

A Topicality Relevance-Aware Intent Model for Web Search

MENG WANG¹, JIANPING LIU^{1,2}, JIAN WANG³, YINGFEI WANG¹, AND XINTAO CHU¹

¹College of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China

²Key Laboratory of Images and Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan 750021, China

³Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Corresponding author: Jianping Liu (liujianping01@nmu.edu.cn)

This work was supported in part by the Natural Science Foundation Project of Ningxia Province, China, titled “User-Oriented Multi-Criteria Relevance Ranking Algorithm and Its Application” under Grant 2021AAC03205; in part by the Key Research and Development Program for Talent Introduction of Ningxia Province, China, titled “Research on Key Technologies of Scientific Data Retrieval in the Context of Open Science” under Grant 2022YCZX0009 and Grant 61862001; in part by the Starting Project of Scientific Research in the North Minzu University titled “Research of Information Retrieval Model Based on the Decision Process” under Grant 2020KYQD37; and in part by the North Minzu University Postgraduate Innovation Project under Grant YCX22178 and Grant YCX22193.

ABSTRACT To accurately understand user information needs and provide better search experiences, various methods have been proposed to model user search intent from their search logs. Most traditional methods based on query understanding only consider the similarity between the query and documents while ignoring the user’s session interaction sequence. Some researchers also adopt neural network-based methods to model user search intent. However, most neural network-based models mainly focus on the user’s session interaction sequence without considering the role of topic relevance in modeling user search intent. In this paper, we propose a novel topicality relevance-aware intent model (TRIM) for web search. TRIM consists of a topic relevance predictor and a user short-term intent predictor. The topic relevance predictor utilizes the BERT model to predict the topic relevance between the query and documents. The user short-term search intent predictor utilizes session context information to predict the user’s short-term search intent. We further investigate several fusion strategies to integrate the topic relevance and user short-term search intent for user search intent prediction. We conduct our experiments on two public web search datasets named TianGong-QRef and TianGong-ST. The experiments show that TRIM outperforms all baselines in the document ranking task. On TianGong-QRef dataset, TRIM achieves a 15.19% increase over the best-performing baselines M-Match in terms of Mean Average Precision (MAP). On TianGong-ST dataset, TRIM achieves a 5.77% increase over the best-performing baselines CARS in terms of MAP. The experimental results indicate the effectiveness of topic relevance in user search intent modeling.


INDEX TERMS Session search, search intent model, document re-ranking, topic relevance, transformer.

I. INTRODUCTION

Search engine is widely used as a common tool to obtain information from the web. Understanding web search user behavior can help search engines better understand their information needs and improve the performance of retrieval systems. However, according to the four-level theory of information needs [1], user information needs are dynamic. For different users, the same query may represent different search

intents. In order to solve this problem, researchers try to model user search intent and re-rank the search results based on it.

Traditional methods based on query understanding typically assume that the query represents the user’s original search intent. Therefore, these methods focus on analyzing the characteristics of keywords, phrases, and combinations in queries. Several studies [2], [3], [4], [5] have attempted to understand the user search intent through query classification, which may cause some important information loss. When the user’s query is short, using query classification may have

The associate editor coordinating the review of this manuscript and approving it for publication was Ali Kashif Bashir .

a negative impact on search results. To solve this problem, several studies have attempted to expand the query [6], [7], [8], [9], [10], [11]. However, they may introduce irrelevant words to the original query, causing search results to deviate from the user's actual intent.

With the emergence of deep learning, researchers focus on obtaining vector representations of the query and documents [12], [13], [14], [15]. Specifically, they first obtain the interactive or individual representations of the query and documents, then calculate the similarity between the query and the document through the Multi-Layer Perceptron (MLP). Finally, re-ranking the search results. The advantage of these methods is that they can better model the semantics of the queries. However, at the beginning of the search, most user search intent are unclear. So, it is difficult to predict the actual user search intent without considering the user's session interaction sequence.

Recently, some context-aware methods [16], [17], [18], [19], [20], [21] have been proposed to model user search intent. They obtain the representation of queries and documents and then utilize recurrent neural networks [22] or attention mechanism [23] to model the session process. They can better model the changes of user's search behavior, but these methods still face long-term dependency. Furthermore, they ignore the role of topical relevance. To solve this problem, our model aims to utilize the BERT model to predict the topic relevance. The BERT model has powerful language representation and feature extraction capabilities. Previous studies [24], [25] have demonstrated that the BERT performs well in deep semantic matching tasks.

To solve the shortcomings of existing methods, we propose the topicality relevance-aware intent model (TRIM) for web search. The idea is illustrated in Figure 1. We combine topic relevance with users' short-term interests to predict users' real search intent and re-rank the search results. TRIM consists of a topic relevance predictor and a user short-term intent predictor. Firstly, We utilize the BERT model to predict the topic relevance between the query and documents. We input the vector representation of the query and documents into the BERT model to obtain the topic relevance score of the query and documents. Secondly, We use Transformer to model the current session process. We input the user's current session interaction sequence into Transformer to obtain the vector representation of the users' short-term search intent. Thirdly, we further investigate several fusion strategies to integrate the topic relevance and user short-term search intent for user search intent prediction. Then, we get the final document ranking score.

The contributions of this paper can be summarized as follows:

- We proposed a novel topicality relevance-aware intent model named TRIM. It models the users' short-term search intent and topic relevance with an end-to-end neural network and re-ranks the search results.

- TRIM achieves significantly better performance than all baselines in document ranking by utilizing contextual and topic relevance information.
- To explore the importance of topic relevance and user short-term search intent, we further design several fusion strategies to integrate the topic relevance and user short-term search intent for user search intent prediction. Experiment results show that topic relevance and user short-term search intent have different importance in predicting user actual search intent.

The rest of the paper is arranged as follows. Related works are summarized in section II. The proposed method is shown in section III. We introduce the experimental settings in section IV, and analyze the results in section V. The conclusion is drawn in section VI.

II. RELATED WORK

Overall, through an in-depth investigation of existing research, we divide the study of modeling user search intent into two research directions: (1) Search intent modeling based on query understanding. (2) Search intent modeling based on session sequence.

A. SEARCH INTENT MODELING BASED ON QUERY UNDERSTANDING

Some works are already trying to understand the query to model the user search intent. Border systematically expounded the classification of query keywords, which laid a solid foundation for the later research on query classification [2]. Recently, many neural query intent classification models have been proposed [3], [4], [5]. Specifically, Xu et al. [3] proposed a hybrid deep neural network model for query intent classification. They encoded the query representations using the recurrent and recursive neural networks, respectively. Wang et al. [4] used a nature network algorithm to classify different building information-related queries. Yuan et al. [5] proposed a Multi-granularity Matching Attention Network to comprehensively extract features from the query and a query-category interaction matrix. However, query classification may cause the loss of important information, which could result in inaccurate predictions of user search intent.

The majority of queries issued by users are short and ambiguous [26], [27]. In this case, it is necessary to expand the query. Savitha et al. [6] fused statistical information from the news corpus and topic diversity of news articles to expand the query. Wang et al. [7] generated words from local word embeddings to expand the original query. Alqahtani et al. [8] proposed a novel approach of hybrid COOT-based Cat and Mouse Optimization algorithm to select optimal candidate terms in the automatic query expansion process. The advantage of these methods is that they improve the recall rate of search engines and the diversity of search results. However, they lack feedback information from users.

Rocchio et al. [9] used explicit feedback from users to expand the query. However, it is difficult to obtain active

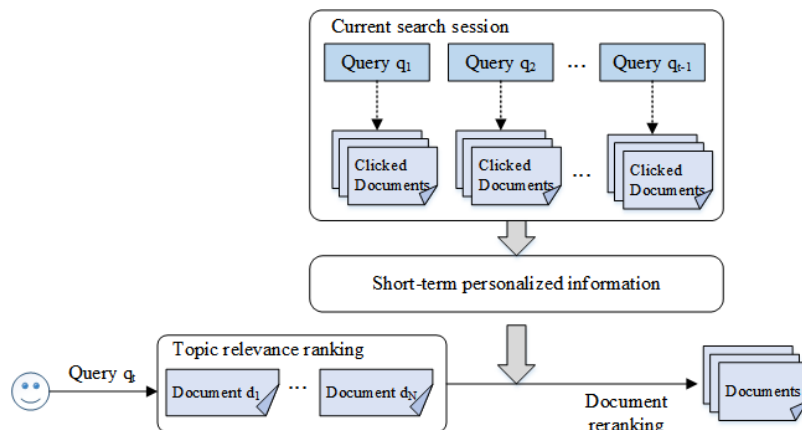


FIGURE 1. The idea of our paper. When the user issues a query, the search engine will return a list of results based on relevance. On this basis, the user will choose the most relevant documents according to their search intent.

feedback from users during retrieval. Thus, some studies attempt to expand queries through pseudo relevance feedback. Specifically, Dev and Balasubramanian [10] utilized the semantic properties of context phrases that occur within the top-ranked retrieved documents to generate diversified query expansions. Nasir et al. [11] used information from a knowledge base to improve the pseudo-relevance feedback process and further expand the query. In addition, some of the pseudo-relevant documents obtained by the user's initial search may not be relevant to the user's query, which leads to a deviation in understanding the user's search intent.

With the emergence of deep learning, Most existing models use deep neural networks to obtain vector representations of queries and documents. They then re-rank search results based on the similarity between the query and document vectors [12], [13], [14]. Specifically, Hu et al. [12] proposed ARC-I model, which obtains the text representation by two layers of 1-D convolution network. Then, the MLP is used to compute their matching score. Compared to the ARC-I model, the ARC-II model [12] considers the order of words and interaction information between sentences. It utilizes 2D-convolution network to obtain a global vector representation. Xiong et al. [13] used the kernel pooling method to obtain the ranking features on the interaction matrix between the query and documents. Mitra et al. [14] simultaneously considered the interactive information and distributed representation of queries and documents. Li et al. [15] selected key blocks of a long document as the input of the BERT model.

However, users usually put forward a series of queries to solve a search task or multiple similar search tasks [28]. Therefore, the representation of the query and documents may not be able to encode the user's actual search intent. On the basis of considering the representation of queries and documents, our model utilizes the user's session behavior sequence to enhance user search intent. By this, we attempt accurately model the actual search intent.

B. SEARCH INTENT MODELING BASED ON SESSION SEARCH

Some traditional approaches already utilize session context to infer search intent [29], [30], [31], [32], [33]. Specifically, Carterette et al. [29] proved that using session data can improve retrieval effectiveness. Van Gysel et al. [30] explored the viability of lexical query matching in session search. They found that specialized models can make better use of long session history than naive term weighting methods.

With the emergence of deep learning, researchers have focused on designing neural context-aware ranking models [16], [17], [18], [19], [20], [21]. Specifically, Ahmad et al. [16], [17] encoded the session contextual information using RNNs and attention mechanism. They jointly optimized the ranking task and the next query prediction task. Cheng et al. [18] learned user search intent from their long-term and short-term behavior. They utilized a multi-hop memory network to infer the users' long-term search intent. Deng et al. [19] modeled multi-granularity user feedback information. Zuo et al. [20] modeled the historical query change in the user's current session. Chang et al. [21] molded user latent intent by a probabilistic modeling approach. They incorporated the latent intention model into the RNN-based sequential recommendation model.

However, these methods still face long-term dependency. The appearance of Transformer [23] technology effectively alleviates this problem. BERT4Rec [34] have proved the effectiveness of Transformer in sequence problems. Due to its powerful ability to leveraging contextual information, we apply it to encode user current session. It is worth mentioning that these methods all ignore the role of topical relevance. Our model attempt to apply the BERT model to predict the topic relevance between the query and documents. By this, we aim to predict user search intent more comprehensively.

III. METHOD

In this section, we describe the architecture of TRIM. We first make a definition of the session search task, and then we introduce three components in the model: topic relevance predictor, user short-term intent predictor and fusion layer.

A. PROBLEM DEFINITION

In session search, the search engine retrieves and sorts candidate documents according to the query and session context submitted by the user. The session context contains the user's historical queries and the clicks for each query. We assume that for every query, there is at least one document clicked by a user. Suppose the given query is q_t , the search session S is represented as a sequence of query-clicks documents pairs, which is defined as: $S = [(q_1, d_1), \dots, (q_{t-1}, d_{t-1})]$. q_i is the i -th query in the session and d_i is the clicked document for query q_i . The task is to rank the candidate documents in D_t for query q_t .

B. TRIM: TOPICALITY RELEVANCE-AWARE INTENT MODEL

The overall structure of TRIM is shown in Figure 2. The model can be divided into three parts: (1) topic relevance predictor. We use the BERT model to obtain a topic relevance score between the query and each candidate document. (2) user short-term intent predictor. Based on users' feedback information within the current session, we use Transformer to obtain the user's short-term search intent. Then calculate the user's short-term interest relevance score. (3) fusion layer. We fuse the topic relevance score and the user's short-term interest relevance score to calculate the final relevance score of each candidate document. Then, we re-rank candidate documents based on their relevance scores.

1) TOPIC RELEVANCE PREDICTOR

Suppose the given query is q_t , the set of candidate documents D_t returned for this query can be obtained, which is defined as: $D_t = [d_1, d_2, \dots, d_t]$. To obtain the joint representation of query and candidate documents, we use the BERT model (bert-base-chinese) for learning. The BERT model sends every word in the input text to the token embedding layer to map each word into a low-dimensional vector space, and then convert it into a text representation vector. The Embedding layer consists of the following three parts: 1) Token embedding. It transforms words into unified dimensions. 2) Segment embedding. It enables the model to distinguish between two texts. 3) Position embedding. It enables the model to understand the word order of words.

For the current query q_t , we combine the query and candidate documents as the input of BERT, and the output of BERT is as follows:

$$s_t = \text{BERT} ([CLS] q_t [SEP] d_1 [SEP], \dots, [CLS] q_t [SEP] d_t [SEP]) \quad (1)$$

where s_t refers to the vector representation of CLS.

We apply a linear layer to calculate the relevance score from the obtained representation. The relevance score can be regarded as the aggregation of local relevance information, and we take the final relevance score as the topic relevance score O_t .

$$O_t = \text{Linear} (s_t) \quad (2)$$

2) USER SHORT-TERM INTENT PREDICTOR

In this part, we use Transformer to model the user's search and click behavior in the current session. The Transformer layer contains a Multi-head Self-attention layer and a Position-wise Feed-forward (FFN) layer.

For the current query q_t and search session $S = [(q_1, d_1), \dots, (q_{t-1}, d_{t-1})]$, we join the query with its clicked documents to obtain their vector representation.

$$E = \text{Emb} (q_i \oplus d_i) \quad i \in [1, t] \quad (3)$$

For the current query q_t , there is no document clicked by the user, so d_t is initialized by zero vector.

The embedding vector E is calculated by multi-head self-attention. Specifically, It converts E into query vector $Q = EW_i^Q$, keyword vector $K = EW_i^K$, and value vector $V = EW_i^V$ through W_i^Q , W_i^K and W_i^V , respectively. The formula is as follows.

$$\text{head}_i = \text{Attention} (EW_i^Q, EW_i^K, EW_i^V) \quad (4)$$

$$\text{Attention} (Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d}} \right) V \quad (5)$$

$$H = \text{Concat} (\text{head}_1, \dots, \text{head}_h) W^O \quad (6)$$

where head_i represents the i -th self-attention. The final output is the connection of all heads.

We incorporate Point-wise Feed-Forward Networks (FFN) to further enhance the model with non-linearity, which is defined as follows.

$$I_t = \phi (LN (H + D (FFN (H)))) \quad (7)$$

where I_t is the output vector of the FFN layer as user's short-term search intent. The function $\phi(\cdot)$ is the multilayer perceptron (MLP) with LeakyRelu(\cdot) as the activation function. LN(\cdot) is layer normalization to stabilize the output. D(\cdot) is a dropout layer with 0.1 probability. FFN is a Position-wise Feed-forward layer.

Finally, we connect the current query q_t with the short-term search intention I_t through non-linear transformation, and apply a sigmoid function to generate the user's short-term interest relevance score U_t .

$$U_{it} = \text{Sigmoid} \left(d_i \cdot \tanh \left(W^P [q_t; I_t] + b^P \right) \right) \quad i \in [1, t] \quad (8)$$

where W^P and b^P are parameters of our model, N is the number of candidate documents, d_i is the i -th candidate document for query q_t .

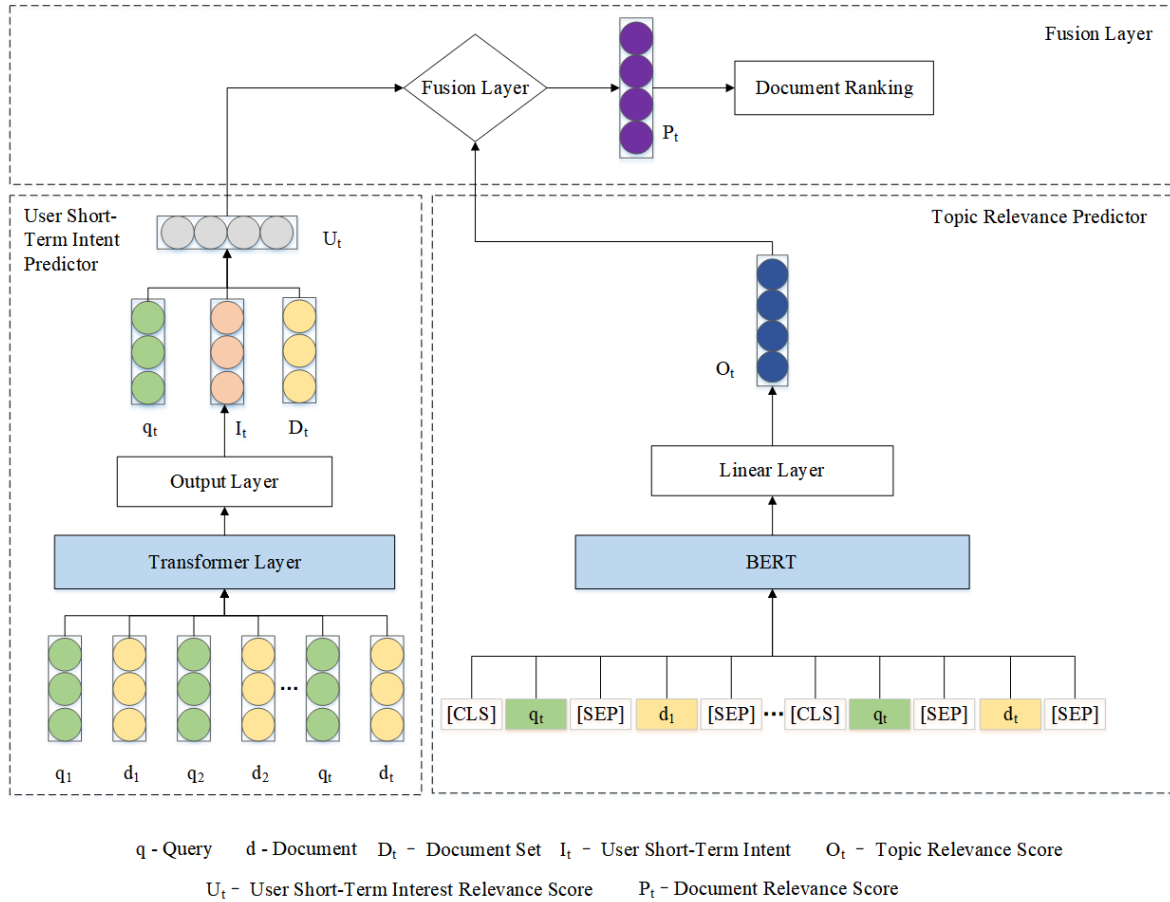


FIGURE 2. The overall structure of TRIM. TRIM consists of a topic relevance predictor and a user short-term intent predictor. The topic relevance predictor is used to estimate the topic relevance score O_t . The user short-term intent predictor is used to obtain the user short-term interest relevance score U_t . TRIM integrates O_t and U_t through a fusion layer to predict users' search intent.

3) FUSION LAYER

In this part, our task is to fuse the topic relevance score O_t and the user's short-term interest relevance score U_t and re-rank the documents according to the fused score P_t . Suppose the given query is q_t . There are 10 candidate documents under the query. We have formulated the following three integration strategies:

Strategy 1: When checking search results, users are often influenced by the position of the results [35], [36]. Users tend to check the top-ranked documents and ignore the results with lower rankings. The user's clicking behavior is usually complex, often influenced by various factors such as presentation form and credibility. Therefore, it is difficult for users to accurately express the document's relevance by clicking. To this end, we have formulated the following strategy.

Firstly, the relevance scores U_t and O_t are sorted in descending order, and the descending sorting lists Lu and Lo of the two relevance scores are obtained. Secondly, we reorder the candidate documents according to Lu and Lo , and get the ranking lists $L1$ and $L2$ of the candidate documents. The relevance score of any candidate document is calculated as

follows.

$$P_{ii} = \begin{cases} Lu_j + Lo_k, & d_i \in \{L1_j, L2_k\} \quad j, k \in (1, 2, 3) \\ Lo_k, & otherwise \end{cases} \quad (9)$$

where $i \in \{1, 2, \dots, 10\}$, j, k represent the position of the document d_i in the list $Lu, Lo, L1$ and $L2$.

Strategy 2: According to the data fusion algorithm CombSum [37], we add the relevance scores U_t and O_t to represent the final relevance score of the document. This strategy holds that topic relevance and short-term user feedback are equally important.

$$P_t = U_t + O_t \quad (10)$$

Strategy 3: We use the linear combination method to obtain the final relevance score of candidate documents. This strategy holds that topic relevance and short-term user feedback have different importance to document relevance scores. Therefore, we use a trainable parameter to adjust the weights of U_t and O_t .

$$P_t = m * U_t + (1 - m) * O_t \quad (11)$$

TABLE 1. Configuration of TRIM.

Configuration of TRIM			
embedding size	64	batch size	1
epoch	16	learning rate	0.05
head number	8	dropout	0.2
Transformer block	1	weight decay	10^{-5}

We re-rank the candidate documents according to the relevance score P_t . To learn the weights and parameters of the model, we use the binary cross entropy loss function as the objective function of the model.

$$L_{rank} = -\frac{1}{N} \sum_{i=1}^N C_i \log(\bar{C}_i) + (1 - C_i) \log(1 - \bar{C}_i) \quad (12)$$

where N is the number of candidate documents, and C_i is the real label of the i candidate document under the current query q_t . \bar{C}_i is the prediction label.

IV. EXPERIMENT

In this section, we conduct experiments to answer the following questions:

RQ1: Compared with baseline models, Does TRIM achieve the best performance in the document ranking task on TianGong-QRef and TianGong-ST datasets?

RQ2: What is the effectiveness of the topic relevance prediction component in modeling user intent?

RQ3: What strategy performs best in integrating topic relevance and users' short-term feedback?

A. EXPERIMENTAL SETUP

1) IMPLEMENTATION DETAILS

The operating system used is Windows, the GPU is NVIDIA TITAN V. We train our model with Adam optimizer. We give detailed model parameters in Table 1.

2) DATASET

We evaluate our proposed method on two public datasets: TianGong-QRef search logs [38] and Tiangong-ST query logs [39]. The statistics of the datasets can be found in Table 2. TianGong-QRef has a longer session length than TianGong-ST, so the user's interaction behavior is more complicated than TianGong-ST. TianGong-ST has a shorter user query, which is more difficult than a long query in predicting the user search intent.

- TianGong-QRef¹: The TianGong-QRef search logs is a Chinese web search log data set contains the user's search behavior log (queries, candid documents, mouse movements, clicks and scrolling, URLs, timestamps) and the user's usefulness evaluation of each search result.

¹<http://www.thuir.cn/tiangong-qref/>

TABLE 2. Statistics of datasets.

Dataset	TianGong-QRef	TianGong-ST
Session	2356	147,155
Query	7464	356,252
Average Session Length	3.47	2.42
Average Query Length	7.73	2.88
Average Click per Query	1.04	0.98
Average Document Length	18.73	8.24
Relevance Judgments	7464	2000

- TianGong-ST²: The Tiangong-ST is collected from a Chinese commercial search engine contains web search session data extracted from an 18-day search log. It contains the query text and query identifier. URL, document identifier, title, and six click-based relevance tags are given for each document in the result list.

3) EVALUATION METRIC

We use three widely used evaluation measures for document ranking: Mean Average Precision (MAP), Mean Reciprocal Ranking (MRR), and Normalized Discounted Cumulative Gain (NDCG). Defined as follows:

$$MAP = \frac{\sum_{q=1}^Q AveP(q)}{Q} \quad (13)$$

where Q represents the number of queries, $AveP$ represents the average accuracy of each query, and the calculation formula of $AveP$ is as follows:

$$AveP = \frac{\sum_{k=1}^n (P(k) * rel(k))}{N} \quad (14)$$

where k represents the ranking position in the search result list, $P(k)$ represents the accuracy of the first k results, and $rel(k)$ represents whether the documents in the position are relevant, with correlation being 1 and irrelevance being 0. N represents the total number of related documents.

$$NDCG@k = \frac{DCG@k}{IDCG@k} \quad (15)$$

$$DCG@K = \sum_{i=1}^K \frac{g_i}{\log_2(i+1)} \quad (16)$$

where g_i is the relevance label of the i -th document in the result list, with relevance of 1 and irrelevance of 0. $IDCG@K$ rank the retrieved K documents from high to low in relevance (that is, the most ideal ranking order), and then calculate the value DCG in this order.

$$MRR = \frac{1}{N} \sum_{i=1}^N \frac{1}{rank(i)} \quad (17)$$

where N is the number of queries, and $rank(i)$ is the ranking of the first related document in the retrieval results under the k -th query.

²<http://www.thuir.cn/tiangong-st/>

4) BASELINES

To evaluate the effectiveness of our proposed model TRIM, we compare it with two kinds of baselines:

(1) Ad-hoc ranking models. These models only use the information of query and document to get the ranking score.

1) ARC-I [12] obtains the representation of query and candidate documents through a two-layer one-dimensional convolutional network. 2) ARC-II [12], it uses a 2-D convolution network to obtain the deep features of text on the interaction matrix between query and candidate documents. 3) KNRM [13], it uses the kernel pooling method to obtain the ranking features on the interaction matrix between the query and document. 4) DUET [14], the model integrates the representation-based method and the interaction-based method.

(2) Context-aware ranking models. These models attempt to understand the search intent by modeling session context.

1) M-NSRF [16] and 2) M-Match [16]. The two models combine current query and session information to solve document ranking task. They joint query suggestion and document ranking tasks. 3) CARS [17]. It introduces the user's click and attention mechanism to obtain a better representation of session context. 4) LostNet [18]. It use hierarchical session-based attention mechanism and multi-hop memory network to infer the user's search intent.

V. RESULTS AND ANALYSIS

A. OVERALL PERFORMANCE COMPARISON (ANSWER RQ1)

We compared our model with other baseline models on TianGong-QRef and TianGong-ST datasets, and the results are shown in Table 3. We can observe:

(1) Compared with all ad-hoc models, our model shows the best effect on both datasets. Compared with the best ad-hoc baseline model DUET, our model has significant improvements in all evaluation metrics. Concretely, for TianGong-QRef dataset, TRIM outperforms DUET by over 21.5% improvement on MAP, while the improvement percentage is 57.6% for TianGong-ST dataset. These results show the importance of modeling the session context containing the historical behavior of users. However, we find some session search models perform worse than some ad-hoc search, e.g. M-NSRF and CARS are worse than DUET on TianGong-QRef dataset. The reason may be that DUET combines the two methods of representation-based and interaction-based, which is more useful than the user's context information on the TianGong-QRef dataset.

(2) Our model shows the best effect compared with all context-aware Ranking models. Compared with the context-aware ranking models, our model significantly improved in all evaluation metrics on both datasets. This proves the effectiveness of modeling the topic relevance between the query and documents. Modeling topic relevance can better understand users' search intent. Specifically, compared with the best context-aware baseline model CARS on

TianGong-ST dataset, our models have significant improvements in all evaluation metrics. TRIM outperforms CARS by over 5.7% improvement on MAP, and on the NDCG@1 Metric, it increased by 9.3%. Compared with the best session search baseline model M-Match on TianGong-QRef dataset. TRIM outperforms M-Match by over 15.1% improvement on MAP, and on the NDCG@1, it increased by 27.0%. In particular, because TianGong-ST dataset does not contain user ID information, it can't obtain the long-term historical information of users, so we compare the LostNet model's variant LostNet-Short [18] proposed by the author on TianGong-ST dataset. Compared with the LostNet-short on TianGong-ST dataset, TRIM outperforms LostNet-short by over 7.8% improvement on MAP.

(3) In all session search models, the RNN network is mainly used to process the sequence information of session context. Our model is superior to the model using the RNN network, which shows that Transformer can better learn the deeper representation of each feature in the behavior sequence.

B. ABLATION ANALYSIS(ANSWER RQ2)

To prove the influence of topic relevance on the modeling of users' search intentions, we conducted ablation experiments on TianGong-QRef and TianGong-ST datasets. Specifically, we remove one component at a time for performance comparison in the following.

- w/o. TPR. We removed the topic relevance predictor and only modeled users' search intent through user short-term search intent predictor.
- w/o. SFM. We removed the user short-term search intent predictor and only used topic relevance predictor to model the user's search intent.

The experimental results are shown in Table 4 and Table 5, and we can get the following conclusions through the ablation analysis:

(1) The introduction of topic relevance predictor can improve the model's performance. When the topic relevance predictor in the model is deleted, model's performance on all evaluation metrics will be greatly reduced. Specially, for TianGong-QRef dataset, the evaluation of MAP and MRR decreased by over 40% and NDCG@1 metrics decreased by more than 70%. For TianGong-ST dataset, the evaluation of MAP and MRR decreased by over 55% and NDCG@1 metrics decreased by more than 70%. By analyzing the model structure, we can see the importance of topic relevance predictor in modeling user search intent.

(2) The user short-term search intent predictor effectively models the user's intention. When the user short-term search intent predictor is removed, the model's performance on all evaluation metrics will be reduced. Specially, for TianGong-QRef dataset, the MAP and MRR metrics dropped by over 6%. The NDCG@1 metric dropped by over 14%. For TianGong-ST dataset, the MAP and MRR metrics drop by over 2%. The NDCG@1 metric dropped by over 3%. These

TABLE 3. Overall performance of all models on TianGong-QRef and TianGong-ST dataset. The best results are shown in bold.

Model Type	Model name	TianGong-QRef						TianGong-ST					
		MAP	MRR	NDCG				MAP	MRR	NDCG			
				@1	@3	@5	@10			@1	@3	@5	@10
Ad-hoc Ranking Models	ARC-I	0.4656	0.4869	0.3107	0.4167	0.4971	0.5884	0.4641	0.4977	0.3116	0.408	0.4795	0.5753
	ARC-II	0.4687	0.4941	0.3164	0.4045	0.4824	0.6051	0.4610	0.4876	0.3021	0.4076	0.4740	0.5716
	KNRM	0.4341	0.4501	0.2671	0.3661	0.4404	0.5700	0.4349	0.4776	0.2797	0.3786	0.4515	0.5533
	DUET	0.4893	0.5102	0.3310	0.4370	0.5078	0.6149	0.4651	0.5020	0.3289	0.4097	0.4696	0.5718
Context-aware Ranking Models	M-NSRF	0.4824	0.5002	0.3281	0.4302	0.4869	0.6134	0.6754	0.6921	0.5591	0.6513	0.6995	0.7601
	M-Match	0.5161	0.5346	0.3672	0.4632	0.5444	0.6424	0.6567	0.6729	0.5380	0.6259	0.6784	0.7456
	CARS	0.4800	0.5023	0.3288	0.4298	0.5025	0.6119	0.6932	0.7094	0.5805	0.6717	0.7170	0.7736
	LostNet	0.5105	0.5285	0.3678	0.4551	0.5146	0.6337	0.6801	0.7001	0.5686	0.6549	0.7076	0.7647
	TRIM	0.5945	0.6198	0.4667	0.5593	0.6126	0.7010	0.7332	0.7512	0.6348	0.7163	0.7573	0.8046

TABLE 4. Ablation study of TRIM on the TianGong-QRef dataset. We sequentially perform the following steps: I. remove the topic relevance predictor; II. remove the user short-term search intent predictor. The best results are shown in bold.

Model	MAP	MRR	NDCG			
			@1	@3	@5	@10
I w/o. TPR	0.3278	0.3406	0.1363	0.2474	0.3249	0.4913
II w/o. SFM	0.5584	0.5781	0.4013	0.5119	0.5792	0.6717
TRIM	0.5945	0.6198	0.4667	0.5593	0.6126	0.7010

TABLE 5. Ablation study of TRIM on the TianGong-ST dataset. We sequentially perform the following steps: I. remove the topic relevance predictor; II. remove the user short-term search intent predictor. The best results are shown in bold.

Model	MAP	MRR	NDCG			
			@1	@3	@5	@10
I w/o. TPR	0.3213	0.3322	0.1272	0.2380	0.3240	0.4852
II w/o. SFM	0.7157	0.7334	0.6153	0.6929	0.7382	0.7910
TRIM	0.7332	0.7512	0.6348	0.7163	0.7573	0.8046

TABLE 6. Performances of different fusion strategies on TianGong-QRef. The best results are shown in bold. Strategy 1 is to re-rank to obtain relevance scores. Strategy 2 is to add the topic relevance score and the user's short-term interest relevance score. Strategy 3 is to assign different weights to the topic relevance score and the user's short-term interest relevance score.

Method	MAP	MRR	NDCG			
			@1	@3	@5	@10
strategy1	0.5684	0.5901	0.4290	0.5269	0.5878	0.6803
strategy2	0.5765	0.5995	0.4401	0.5366	0.5978	0.6867
strategy3	0.5945	0.6198	0.4667	0.5593	0.6126	0.7010

results show the effectiveness of the user's short-term intent modeling. We should consider both topic relevance and users' short-term feedback.

C. COMPARISON OF FUSION LAYER STRATEGIES (ANSWER RQ3)

We study the influence of different fusion strategies on the experimental results. The fusion strategy is described in section III. The performance of different fusion strategies on TianGong-QRef and TianGong-ST is shown in Table 6 and Table 7, respectively. From Table 6 and Table 7, we can observe that strategy 3 can achieve the best results in the fusion layer. The result of using strategy 2 is worse than strategy 3. This shows that the topic relevance and the users'

TABLE 7. Performances of different fusion strategies on TianGong-ST. The best results are shown in bold. Strategy 1 is to re-rank to obtain relevance scores. Strategy 2 is to add the topic relevance score and the user's short-term interest relevance score. Strategy 3 is to assign different weights to the topic relevance score and the user's short-term interest relevance score.

Method	MAP	MRR	NDCG			
			@1	@3	@5	@10
strategy1	0.7301	0.7504	0.6262	0.7165	0.7300	0.8000
strategy2	0.7212	0.7379	0.6187	0.7015	0.7460	0.7950
strategy3	0.7332	0.7512	0.6348	0.7163	0.7573	0.8046

short-term feedback have different influences on modeling user search intent.

VI. CONCLUSION

In this paper, we propose a topicality relevance-aware intent model (TRIM) for web search. TRIM consists of a topic relevance predictor and a users' short-term intent predictor. We further investigate the performances of several fusion strategies which integrate topic relevance and user short-term intent for user actual search intent prediction.

We conduct extensive experiments on two open Web search datasets. We can find that: 1) TRIM significantly outperforms all baselines on the document ranking task. On TianGong-QRef dataset, TRIM achieves a 15.19% increase over the best-performing baselines M-Match in terms of MAP. On TianGong-ST dataset, TRIM achieves a 5.77% increase over the best performing baselines CARS in terms of MAP. 2) The ablation experiment indicates the effectiveness of topic relevance information in modeling user search intent. 3) The fusion strategy 3 has the best overall performance among all the fusion strategies. The reason is that using a learnable method to assign weights is more flexible than manually assigning them.

Modeling user search intent is a challenging task due to the complex and dynamic nature of user search intent. We combine topic relevance and users' short-term intent in modeling user search intent. However, our work only considers the current session information of users. Compared with the user's current session information, the longer-term session information contains the user's more stable interests. In future work, we plan to apply user's long-term sessions so as to enhance the current session search. Also, External

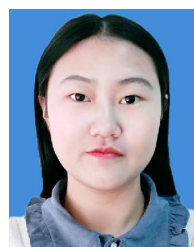
domain knowledge can supplement the information learned by the model. So, we would like to integrate external domain knowledge to further improve the performance of TRIM.

ACKNOWLEDGMENT

The authors would like to thank the data providers used in this article.

REFERENCES

- [1] R. S. Taylor, "Question-negotiation and information seeking in libraries," *College Res. Libraries*, vol. 29, no. 3, pp. 178–194, 1968.
- [2] A. Broder, "A taxonomy of web search," *ACM SIGIR Forum*, vol. 36, no. 2, pp. 3–10, 2002.
- [3] B. Xu, Y. Ma, and H. Lin, "A hybrid deep neural network model for query intent classification," *J. Intell. Fuzzy Syst.*, vol. 36, no. 6, pp. 6413–6423, Jun. 2019.
- [4] N. Wang, R. R. A. Issa, and C. J. Anumba, "Transfer learning-based query classification for intelligent building information spoken dialogue," *Autom. Construct.*, vol. 141, Sep. 2022, Art. no. 104403.
- [5] C. Yuan, Y. Qiu, M. Li, H. Hu, S. Wang, and S. Xu, "A multi-granularity matching attention network for query intent classification in e-commerce retrieval," in *Proc. Companion ACM Web Conf.*, Apr. 2023, pp. 416–420.
- [6] G. Savitha, S. Thamizharasan, and R. Prasath, "Query expansion using information fusion for effective retrieval of news articles," in *Pattern Recognition and Machine Intelligence*. Tezpur, India: Springer, Dec. 2019, pp. 275–284.
- [7] Y. Wang, H. Huang, and C. Feng, "Query expansion with local conceptual word embeddings in microblog retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 4, pp. 1737–1749, Apr. 2021.
- [8] A. S. Alqahtani, P. Saravanan, M. Maheswari, and S. Alshmrany, "An automatic query expansion based on hybrid CMO-COOT algorithm for optimized information retrieval," *J. Supercomput.*, vol. 78, pp. 8625–8643, Jan. 2022.
- [9] J. J. Rocchio, "Relevance feedback in information retrieval," in *The SMART Retrieval System: Experiments in Automatic Document Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1971, pp. 313–323.
- [10] R. E. Dev and V. Balasubramanian, "An implicit aspect modelling framework for diversity focused query expansion," *J. Intell. Inf. Syst.*, vol. 55, no. 2, pp. 207–231, Oct. 2020.
- [11] J. A. Nasir, I. Varlamis, and S. Ishfaq, "A knowledge-based semantic framework for query expansion," *Inf. Process. Manage.*, vol. 56, no. 5, pp. 1605–1617, Sep. 2019.
- [12] B. Hu, Z. Lu, H. Li, and Q. Chen, "Convolutional neural network architectures for matching natural language sentences," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, p. 1–9.
- [13] C. Xiong, Z. Dai, J. Callan, Z. Liu, and R. Power, "End-to-end neural ad-hoc ranking with kernel pooling," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2017, pp. 55–64.
- [14] B. Mitra, F. Diaz, and N. Craswell, "Learning to match using local and distributed representations of text for web search," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 1291–1299.
- [15] M. Li, D. N. Popa, J. Chagnon, Y. G. Cinar, and E. Gaussier, "The power of selecting key blocks with local pre-ranking for long document information retrieval," *ACM Trans. Inf. Syst.*, vol. 41, no. 3, pp. 1–35, Jul. 2023.
- [16] W. U. Ahmad and K.-W. Chang, "Multi-task learning for document ranking and query suggestion," in *Proc. 6th Int. Conf. Learn. Represent.*, 2018, pp. 1–14.
- [17] W. U. Ahmad, K.-W. Chang, and H. Wang, "Context attentive document ranking and query suggestion," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2019, pp. 385–394.
- [18] Q. Cheng, Z. Ren, Y. Lin, P. Ren, Z. Chen, X. Liu, and M. D. de Rijke, "Long short-term session search: Joint personalized reranking and next query prediction," in *Proc. Web Conf.*, Apr. 2021, pp. 239–248.
- [19] C. Deng, Y. Zhou, and Z. Dou, "Improving personalized search with dual-feedback network," in *Proc. 15th ACM Int. Conf. Web Search Data Mining*, Feb. 2022, pp. 210–218.
- [20] X. Zuo, S. Dou, and J.-R. Wen, "Improving session search by modeling multi-granularity historical query change," in *Proc. 15th ACM Int. Conf. Web Search Data Mining*, Feb. 2022, pp. 1534–1542.
- [21] B. Chang, A. Karatzoglou, Y. Wang, C. Xu, E. H. Chi, and M. Chen, "Latent user intent modeling for sequential recommenders," in *Proc. Companion ACM Web Conf.* New York, NY, USA: Association for Computing Machinery, Apr. 2023, pp. 427–431.
- [22] A. Graves, "Long short-term memory," in *Supervised Sequence Labelling With Recurrent Neural Networks (Studies in Computational Intelligence)*, vol. 385. Berlin, Germany: Springer, 2012, pp. 37–45, doi: 10.1007/978-3-642-24797-2_4.
- [23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
- [24] S. Choudhary, H. Guttikonda, D. R. Chowdhury, and G. P. Learmonth, "Document retrieval using deep learning," in *Proc. Syst. Inf. Eng. Design Symp. (SIEDS)*, Apr. 2020, pp. 1–6.
- [25] Y. Wang, M. Xu, Y. Yan, T. Zhao, Y. Chen, and J. Yang, "Exploring topic supervision with BERT for text matching," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2022, pp. 1–7.
- [26] S. Cronen-Townsend and W. B. Croft, "Quantifying query ambiguity," in *Proc. HLT*, vol. 2, 2002, pp. 94–98.
- [27] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, "Analysis of a very large web search engine query log," *ACM SIGIR Forum*, vol. 33, no. 1, pp. 6–12, 1999.
- [28] E. Agichtein, R. W. White, S. T. Dumais, and P. N. Bennet, "Search, interrupted: Understanding and predicting search task continuation," in *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2012, pp. 315–324.
- [29] B. Carterette, P. Clough, M. Hall, E. Kanoulas, and M. Sanderson, "Evaluating retrieval over sessions: The TREC session track 2011–2014," in *Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2016, pp. 685–688.
- [30] C. Van Gysel, E. Kanoulas, and M. de Rijke, "Lexical query modeling in session search," in *Proc. ACM Int. Conf. Theory Inf. Retr.*, Sep. 2016, pp. 69–72.
- [31] P. N. Bennett, R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisjuk, and X. Cui, "Modeling the impact of short- and long-term behavior on search personalization," in *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2012, pp. 185–194.
- [32] X. Shen, B. Tan, and C. Zhai, "Context-sensitive information retrieval using implicit feedback," in *Proc. 28th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2005, pp. 43–50.
- [33] R. W. White, W. Chu, A. Hassan, X. He, Y. Song, and H. Wang, "Enhancing personalized search by mining and modeling task behavior," in *Proc. 22nd Int. Conf. World Wide Web*, May 2013, pp. 1411–1420.
- [34] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, and P. Jiang, "BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 1441–1450.
- [35] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey, "An experimental comparison of click position-bias models," in *Proc. Int. Conf. Web Search Web Data Mining (WSDM)*, 2008, pp. 87–94.
- [36] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately interpreting clickthrough data as implicit feedback," *ACM SIGIR Forum*, vol. 51, no. 1, pp. 4–11, 2017.
- [37] S. Wu, "Linear combination of component results in information retrieval," *Data Knowl. Eng.*, vol. 71, no. 1, pp. 114–126, Jan. 2012.
- [38] J. Chen, J. Mao, Y. Liu, F. Zhang, M. Zhang, and S. Ma, "Towards a better understanding of query reformulation behavior in web search," in *Proc. Web Conf.*, Apr. 2021, pp. 743–755.
- [39] J. Chen, J. Mao, Y. Liu, M. Zhang, and S. Ma, "TianGong-ST: A new dataset with large-scale refined real-world web search sessions," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 2485–2488.



MENG WANG is currently pursuing the master's degree with the College of Computer Science and Engineering, North Minzu University. Her research interest includes interactive information retrieval. Her research has been published in the International Conference on Cloud Computing and Intelligent Systems (CCIS 2022).



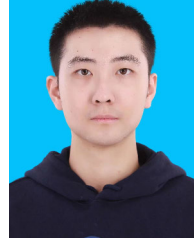
JIANPING LIU received the Ph.D. degree in information technology and digital agriculture from the Chinese Academy of Agricultural Sciences. He is currently a Lecturer in information sciences with the College of Computer Science and Engineering, North Minzu University. His research has been published in *Library and Information Science Research*, in 2019, and *Data Science Journal*, in 2020.



YINGFEI WANG is currently pursuing the master's degree with the College of Computer Science and Engineering, North Minzu University. Her research interests include interactive information retrieval and click model. Her research has been published in the International Conference on Cloud Computing and Intelligent Systems (CCIS 2022).



JIAN WANG received the Ph.D. degree in geoinformatics from the Chinese Academy of Sciences. He is currently a Professor in information sciences with the Agricultural Information Institute, Chinese Academy of Agricultural Sciences. His research was published in *Sensor Letters*, in 2010, *Data Science Journal*, in 2020, and *Library and Information Science Research*, in 2019.



XINTAO CHU is currently pursuing the master's degree with the College of Computer Science and Engineering, North Minzu University. His research interests include interactive information retrieval and deep learning. His research has been published in the International Conference on Cloud Computing and Intelligent Systems (CCIS 2022).

...