

# 基于点击模型和网络嵌入的查询推荐模型<sup>①</sup>



王奕昕, 范春晓, 吴岳辛

(北京邮电大学 电子工程学院, 北京 100876)

通讯作者: 范春晓, E-mail: cxfan@bupt.edu.cn

**摘要:** 用户在使用现有的搜索引擎时, 常因为无法构造清晰准确的查询词而导致检索效果不佳, 传统的查询推荐方法没有充分考虑用户行为的关联性, 导致了查询推荐的结果不准确. 本文提出了一个新的查询推荐模型, 即基于点击模型和网络嵌入的查询推荐模型. 该模型首先通过点击链式模型嵌入用户的历史检视行为和点击行为, 并通过注意力机制衡量查询和返回文档的相关性; 然后利用属性异构网络来获取复杂异质网络结构中的潜在语义信息; 最后通过多头注意力捕获多个空间的复杂信息, 并利用多任务学习来做评分预测. 在搜狗实验室提供的公开查询日志上的实验结果表明, 我们的模型在查询建议的鉴别式任务和生成式任务中均优于基线模型.

**关键词:** 查询建议; 用户行为; 查询重构; 查询嵌入

引用格式: 王奕昕, 范春晓, 吴岳辛. 基于点击模型和网络嵌入的查询推荐模型. 计算机系统应用, 2021, 30(12): 255-261. <http://www.c-s-a.org.cn/1003-3254/8215.html>

## Query Suggestion Model Based on Click Model and Network Embedding

WANG Yi-Xin, FAN Chun-Xiao, WU Yue-Xin

(School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** When using the existing search engines, users often fail to construct clear and accurate query words, which leads to poor retrieval results. Traditional query recommendation methods do not fully consider the relevance of user behavior, resulting in inaccurate query recommendation results. This study builds a new query recommendation model, which is based on the click model and network embedding. Firstly, the model embeds the user's history view behavior and click behavior through the click chain model and measures the relevance between the query and the returned documents through the attention mechanism; secondly, it uses the attribute heterogeneous network to obtain the potential semantic information in a complex heterogeneous network structure; finally, it captures the complex information in multiple spaces through multi-head attention and uses multi-task learning to make score prediction. The experimental results on the public query log provided by SogouLabs show that our model is superior to the baseline model in both discriminative and generative tasks.

**Key words:** query suggestion; user behavior; query reformulation; query embedding

伴随着互联网的发展越来越迅速, 互联网中的信息越来越繁多, 搜索引擎作为人们获取互联网信息的主要途径, 其在日常生活中扮演的角色也越发重要. 但是在检索过程中, 用户往往难以根据其搜索意图给出简洁准确的查询, 从而导致搜索引擎无法为用户提供

最符合其搜索意图的结果. 搜索引擎如今通过查询推荐为用户构造相关查询的方法来提升用户的搜索满意度.

查询推荐主要包含查询自动补全和查询重构两方面的内容. 查询自动补全指用户在输入少量查询前缀等信息的情境下在搜索框的下拉列表中生成包含输入

<sup>①</sup> 收稿时间: 2021-03-03; 修改时间: 2021-03-29; 采用时间: 2021-04-06

前缀的候选查询;查询重构指用户完成输入后通过对查询词替换、删除等操作来优化前面的查询以获得更令人满意的搜索结果的过程。近些年伴随着查询推荐技术在信息检索系统中发挥着愈发重要的作用,发现查询自动补全相比查询重构具有局限性,故更多的查询推荐研究着重于查询重构推荐。

成功的查询推荐取决于捕获用户的搜索意图、准确的建模用户的信息需求,直观上需要充分利用查询上下文信息,包括同一搜索会话中的历史查询信息等。上下文信息在查询建议和查询自动补全的研究中被广泛利用,传统的方法依赖于查询的相似性如术语复用和查询共现<sup>[1]</sup>等,其在提供相似度较高的候选查询的同时存在数据稀疏的问题。而基于深度学习的方法是如今解决此问题的主流方法,分层递归的编码-解码结构<sup>[2]</sup>在有序编码上下文的同时解决了数据稀疏的问题,反馈记忆网络<sup>[3]</sup>在之前的研究仅将用户反馈信息作为辅助信息进行聚类的基础上解决了数据稀疏的问题且有利于找到更细粒度的搜索意图。查询重构对搜索会话的上下文有极大的影响,重构推理网络<sup>[4]</sup>利用同态查询嵌入系统和异构网络嵌入模型对用户查询和查询重构进行建模,其同时保留了句法和语义属性,解决了由于预定义策略和本体辅助带来的范围外策略和歧义查询的问题。

前述提到的最新研究中仍然存在一系列问题:在建模用户查询意图的过程中考虑因素比较单一,无法完整表达用户的兴趣;使用的图模型嵌入方法对复杂的查询-URL图无法构造清晰准确的嵌入表达;对同一查询会话内的用户行为及兴趣抽取缺乏相关性,使其缺乏表示完整性。

本文提出了基于点击模型和网络嵌入的查询推荐模型UBM来解决这些问题,帮助用户构建查询。首先通过点击链式模型引入查询时序,将其和用户偏好注意力结合生成反馈意图;利用属性异构网络构建带有标题属性的查询-URL关系图,并以此建模查询重构。引入多头自注意力机制,捕获多个表示子空间的关系和会话行为的内部交互相关性。最后采用多任务学习来训练查询鉴别器和查询生成器来进行查询推荐。

## 1 基本概念

定义1. 点击模型(click model)<sup>[5]</sup>是根据用户的历史点击信息,建模用户的兴趣和行为,从而预测用户的

未来点击行为,提高点击间的相关性。而用户的点击行为为受到结果位置、结果置信度等因素的影响,不能直接利用结果相关性来估计。点击模型包含了用户的整个搜索过程,运用不同的假设去估计不同因素的影响从而更准确地利用用户的隐式反馈来计算用户的点击概率。

定义2. 网络嵌入(network embedding)又叫图嵌入(graph embedding)<sup>[6]</sup>是用来学习网络中节点的低维度表示,并将学习到的特征用来解决基于图的各类任务。其主要思想是找到一种映射函数,从图的拓扑结构、顶点间的关系、子图以及边的性质等信息中挖掘出复杂网络结构中的低纬度信息表示。

## 2 基于点击模型和网络嵌入的查询推荐模型

在搜索引擎平台中,用户不能清晰地表达其查询意图导致查询结果满意度较低,因此需要通过查询推荐策略来帮助用户构建查询词。然而现有的查询推荐策略无法完整建模用户兴趣从而导致推荐结果不佳。本文为了解决查询效率低的问题,设计了关联用户行为的查询推荐方法来帮助用户构造查询从而完成检索过程。如第1节所述,点击模型可以挖掘出用户历史点击行为的相关性从而表达用户兴趣,而网络嵌入可以帮助我们挖掘出复杂的查询-URL二部图中挖掘出用户查询与结果的关系从而得到用户的偏好,通过结合两者的优势,设计了基于点击模型和网络嵌入的查询推荐模型,捕获用户查询行为和重构行为的潜在关系进而为用户生成最符合其查询意图的推荐结果。模型设计结构如图1所示。首先,在嵌入层引入用户点击链式模型,通过对用户点击行为的分析,推断用户对返回文档的兴趣度,综合用户单个会话内针对同一查询的多次点击,得到对于返回文档的点击概率评分矩阵,将其与偏好注意力编码结合得到用户当前查询对于返回文档的反馈意图行为分布;另外将URL的标题信息加入到查询-URL二部图中作为URL的属性信息,并利用属性异构网络对其不同类型的节点与边进行表示,将不同维度不同性质的向量转换到同一纬度,并以此建模用户的查询重构行为;接着在方法层利用多头注意力机制捕获多个表示子空间的信息,在会话级别抽取更多的特征;最后在输出层利用多任务学习来同时训练查询鉴别器和查询生成器来解决不同任务的需求以提高其性能。

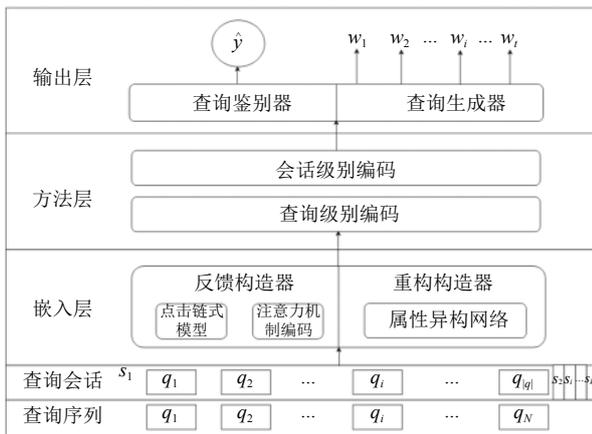


图1 基于点击模型和网络嵌入的查询推荐模型图

### 2.1 嵌入层

嵌入层主要对查询进行嵌入处理,对查询的反馈信息及查询的重构信息进行建模,将其引入模型之中,按功能分为反馈构造器和重构构造器两部分。

#### 2.1.1 反馈构造器

该部分旨在从用户的查询意图中提取反馈信息,这些信息反映在用户的点击顺序、用户的点击文档和略过文档中。

本文利用点击链式模型<sup>[7]</sup>,根据用户在一个查询会

表1 点击概率评分累加例子示意图

点击/文档	$d_1$	$d_2$	$d_3$	$d_4$	$d_5$	$d_6$
点击发生于 $d_3$	$1 - R_i$	$1 - R_i$	1	0	0	0
点击发生于 $d_5$	$\beta(1 - R_i)$	$\beta(1 - R_i)$	$\beta(1 - R_i)$	$\beta(1 - R_i)(\alpha_2(1 - R_i) + \alpha_3 R_i)$	$\beta$	0

注意力机制编码是用来衡量查询和返回文档的相关性,捕获查询中的词对查询文档的重要性,这代表着具有更高注意力得分的文档对应着用户更多的查询意图。

查询  $q = \{w_1^q, \dots, w_j^q, \dots, w_{|q|}^q\}$  被编码为注意力向量  $\vec{a}_q:1$

$$\vec{a}_q = GRU(Emb(q)) \quad (7)$$

其中,  $Emb$  是词嵌入矩阵,  $GRU^{[8]}$  是一种改良的 RNN 模型。

对文档  $D = \{d_1, \dots, d_i, \dots, d_n\}$  中每个文档  $d_i = \{w_1^i, \dots, w_j^i, \dots, w_{|d_i|}^i\}$  用和查询  $q$  同样的方式不同的参数进行编码,编码后的注意力向量为  $\vec{a}_{d_i}$ 。

查询对于每个查询文档的注意力权重是其注意力向量的标准化点积:

话时间内的多次点击行为,计算出每次点击返回的文档列表的点击概率得分:当文档被浏览但被跳过时,得分总是 0,因为它表示用户在查看后没有选择单击;点击的文档得分总是 1;点击文档后的得分根据状态转移方程得到.考虑越往后的点击越能代表用户的搜索意图,而设计倍增系数  $\beta$ ,使后续的每一次点击概率评分都倍增  $\beta$  倍:

$$P(C_i = 1 | E_i = 0) = 0 \quad (1)$$

$$P(C_i = 1 | E_i = 1, R_i) = R_i \quad (2)$$

$$P(E_{i+1} = 1 | E_i = 0) = 0 \quad (3)$$

$$P(E_{i+1} = 1 | E_i = 1, C_i = 0) = \alpha_1 \quad (4)$$

$$P(E_{i+1} = 1 | E_i = 1, C_i = 1, R_i) = \alpha_2(1 - R_i) + \alpha_3 R_i \quad (5)$$

其中,二进制随机变量  $E_i$  表示对位置  $i$  处的文档的检视事件,二进制随机变量  $C_i$  表示对位置  $i$  处的文档的点击事件,  $R_i$  表示用户感知的相关性,  $\alpha_i$  表示用户行为参数。

对于查询返回文档列表  $D$ ,计算文档概率评分  $G_n = \{g_1, \dots, g_i, \dots, g_n\}$ ,其中  $g_i$  为每次点击计算出的点击概率评分的累加(如表 1),之后对文档概率评分计算归一化操作得到文档归一化评分  $G$ :

$$G = softmax\{G_n | d_i \in D\} \quad (6)$$

$$A_i = \vec{a}_q \cdot \vec{a}_{d_i} \quad (8)$$

$$A = softmax\{A_i | d_i \in D\} \quad (9)$$

根据文档概率评分和每个查询的注意力权重相加后经过 Softmax 标准化后的值,通过加权求和得到用户的反馈行为分布  $F$ :

$$F = \sum_{d_i \in D} softmax(A + G) \cdot \vec{a}_{d_i} \quad (10)$$

#### 2.1.2 重构构造器

查询重构<sup>[9,10]</sup>是用户对以前的查询不满意,通过对查询进行一系列的优化得到满意结果的过程.重构构造器在图嵌入的基础上,参考词嵌入的句法和语义信息获取方法,构造重构嵌入.从查询日志出发,结合每个查询的特点,提取查询词、URL 和标题,构造查询

URL 二部图。

图2具有以下特点:第一其查询及URL均与其附加特征相连,第二在图中节点和边的类型不完全相同。考虑这些特点,选用类似属性异构网络<sup>[11]</sup>的方式来构建嵌入。首先对图中不同类型的顶点使用不同空间的向量表示;之后使用非线性网络函数对顶点进行向量变换,转换到同一维度;最后利用连接边的邻居顶点相似来构造目标函数。

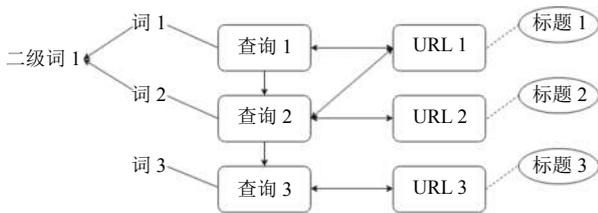


图2 包含标题信息的查询-URL 二部图

基于词嵌入的定义与图嵌入的补充,对于查询 $q_i$ 和 $q_{i+1}$ 间的差异可以定义为查询重构 $r_i$ :

$$r_i = V_{q_{i+1}} - V_{q_i} \quad (11)$$

式中, $r_i$ 可以从句法和语义两个方面捕获查询间相关性,建模查询重构行为。

## 2.2 方法层

方法层主要对嵌入层的输出进行两层结构的编码:查询级别的编码和会话级别的编码;查询级别的编码主要是将嵌入层的输出映射到固定长度的向量,会话级别的编码主要对顺序及位置敏感的信息进行编码<sup>[12]</sup>。

### 2.2.1 查询级别编码

查询级别编码是对用户的查询 $q$ 使用和式(7)相同的形式与参数将其构造为标准的嵌入向量 $q_k$ 。之后将用户查询 $q_k$ 、用户反馈 $F_k$ 及用户重构 $r_{k-1}$ 三个部分,将其经过组合拼接线性变换得到用户输入序列 $X = \{X_1, \dots, X_i, \dots, X_n\}$ ,其中 $W_r$ 为可学习到的线性变换矩阵。

### 2.2.2 会话级别编码

会话级别的编码使用Transformer<sup>[13,14]</sup>中的编码器对输入序列 $X$ 进行编码处理。最终输出的隐层向量为 $X = \{X_{hidden1}, \dots, X_{hiddeni}, \dots, X_{hiddenK}\}$ ,上下文向量 $c$ 可以表示为编码器隐层向量的加权和:

$$c = \sum_{X_{hiddeni} \in X} \gamma_i X_{hiddeni}$$

其中, $\gamma_i$ 为隐层权值。

## 2.3 输出层

输出层的设计思路来源于查询推荐问题的场景描述,其存在两个场景:根据查询相似度等的候选查询列

表的排序问题和查询生成问题;本文根据不同的场景设计了多任务学习框架,同时训练两个不同的模块:查询鉴别器和查询生成器。

### 2.3.1 查询鉴别器

查询鉴别器的任务是评估候选查询列表中的每个候选查询成为符合查询意图的查询的可能性。

### 2.3.2 查询生成器

查询生成器的任务是在没有任何候选查询的基础上,利用方法层编码器的输出生成查询序列。

交叉熵被用作该层两个模块的损失函数,在模型训练和优化过程中,将两个损失函数结合起来,共同解决任务。

## 3 实验分析

### 3.1 实验数据

本实验的实验数据来源于搜狗实验室提供的搜狗搜索引擎在2006年8月中31天的查询日志<sup>[15]</sup>。查询日志包括查询时间、用户ID、查询词、该URL在结果中的排名、用户点击的序号、用户点击的URL等信息。运用标准的30分钟间隔<sup>[16]</sup>划分数据集,将查询分为多个会话。而仅包含一个查询的会话将被放弃,因为必须有先前查询的会话作为上下文信息。把会话中的最后一个查询视为正确的建议。由于网页链接的时效性等原因无法从链接得到文档的更多信息,只能从搜狗本身提供的互联网语料库中找到链接对应的标题信息,将其作为文档的内容信息。将会话随机分为3个部分:60%为训练集、30%为开发集、10%为测试集。训练集训练神经基线模型和我们的模型;开发集是在训练集训练出来的模型基础上进行测试,通过测试结果不断优化模型;测试集评估所有方法。数据集的划分结果信息如表2所示。

表2 数据集统计信息表

数据	训练集	开发集	测试集
会话数	7978441	3989220	36519
查询数	27191564	13551903	117225

### 3.2 评价指标

本实验选用MRR指标作为评价重排序结果的指标,选用PER指标作为评估生成结果好坏的评价指标。MRR是国际上通用的对搜索算法进行评价的机制,利用平均倒数排名来评估推荐列表返回结果的可靠性。PER代表与位置无关的错误率,它基于生成的查询和

目标查询之间的单词覆盖率,代表的是忽略单词顺序将生成的查询转换为目标查询所需的最小单词插入和删除次数的长度归一化表示。

### 3.3 方法对比

本实验目的在于测试深度学习环境下的不同查询推荐策略对推荐结果的影响,因此在此对比以下5个模型:

MPS是一种依靠上下文中的查询共现来排序查询的最大似然方法,其常作为深度学习方法中生成候选查询列表的方法。

HRED是利用用户已经提交的查询基于层次递归编码-解码器预测后续查询,进而得到候选查询推荐序列。

ACG<sup>[17]</sup>是利用查询感知的注意力机制增强序列到序列模型以编码用户查询会话的上下文信息,并提出了复制和生成两种机制来确定后续推荐词。

FMN将用户与系统的交互信息加入到了模型之中,通过用户反馈在提供相同查询前缀时给出符合用户需求的查询推荐。

RIN是利用异构嵌入网络 and 同态隐空间建模查询以及查询重构,捕获重构信息对查询推荐的影响。

### 3.4 实验结果及分析

这部分将首先利用评估指标对两类任务进行评估,之后对文章所提出的新思路及方法进行分析和讨论。如表3所示为各个模型对于两类不同任务在测试集上的结果。

表3 不同方法在鉴别式和生成式任务上的表现

方法/评价指标	MRR	PER
MPS	0.4639	—
HRED	0.5370	0.8284
ACG	0.5625	0.7152
FMN	0.5812	0.7026
RIN	0.7318	0.6865
UBM	0.8090	0.6252

#### 3.4.1 基于鉴别式任务的评估

对于鉴别式任务使用MRR指标来进行评估,MRR指标越高表明推荐列表的结果越符合用户查询意图。基于查询共现的MPS方法仅仅根据查询间的相似度及流行度等消息进行排序,效果相对较差。选取的基线模型都是深度学习模型,主要在于深度学习模型其对于上下文的关系以及顺序关系等相对而言有更好的效果。ACG由于其门控机制控制复制和生成的选择使模型效果有所提升;FMN建模了用户的正负反馈行为,

使模型更能学习到用户兴趣相关的信息;RIN建模了用户的查询重构行为,使其对在一个会话中有多次查询的场景有较好的效果。本文同时建模了用户的查询点击行为、检视行为及查询重构行为等,尽可能地利用了查询日志信息来模拟了用户在一次查询过程中所面对的场景,使模型在鉴别式任务的指标上相比之前效果最好的RIN有了7.72%的提升。

#### 3.4.2 基于生成式任务的评估

对于生成式任务使用PER指标来进行评估,PER指标越低代表生成的查询和目标查询差异越小。MPS不是生成模型所以不考虑。ACG由于其门控的生成器及RIN中的查询生成器与生成式任务相贴近,其相比其它模型在生成式任务上有一定优势。本文利用了序列到序列生成模型效果比较好的Transformer中的编码器-解码器结构,在生成模型上提升了6.13%的覆盖率。

#### 3.4.3 分析与讨论

为了进一步验证并深入了解本文所提出的模型,本文进行了消融研究去比较所提出的关键技术的作用。

点击链式模型的影响。点击模型的原理是用户点击行为导致的不同位置的文档的不同点击可能性。由于反馈构造器部分由点击链式模型与注意力机制编码构成,图3展现了反馈构造器不同组成对MRR性能的影响。通过观察可知,同时排除点击链式模型与注意力机制编码的模型UBM(-a-c)与RIN的结果基本相同。而单独考虑注意力机制编码UBM(-c)仅能带来接近2%的提升,仅考虑点击链式模型UBM(-a)的效果提升了接近5%说明在反馈构造器中起主要作用的是点击链式模型。

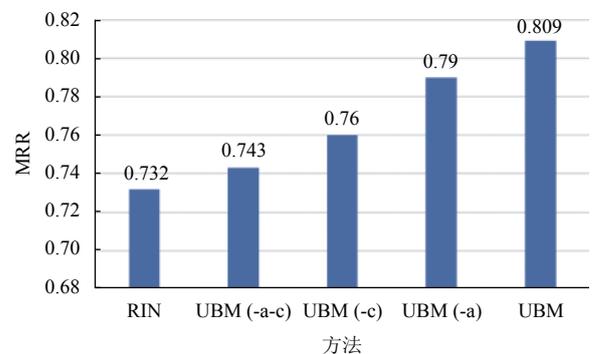


图3 反馈构造器不同组成的UBM的MRR性能

属性异构网络的影响。属性异构网络的原型是查询-URL二部图,其常规方法是使用随机游走<sup>[18]</sup>或Node2Vec

等图嵌入方法建立嵌入。本文引入 URL 的标题作为辅助信息,使图中存在了性质不同的边和节点。图 4 展现了不同图嵌入模型及辅助信息影响下的 MRR 性能影响。RIN 中使用的是 Node2Vec 模型但其没有考虑标题信息,其相比添加了标题信息的模型 RIN(+title) 效果相差不大,分析原因主要在于 Node2Vec 对于不同类型节点的关系无法正确建模,导致其存在一些差错使得整体效果提升不大。将属性异构网络模型嵌入作为图嵌入模型使用时,其不添加标题 UBM(-title) 的效果相比使用 Node2Vec 的效果提升了 5%,体现了属性异构网络本身的优越性;而添加标题后 UBM 因为其附加的辅助信息与属性异构网络的设计目的相同,使得模型效果提升了 2%。

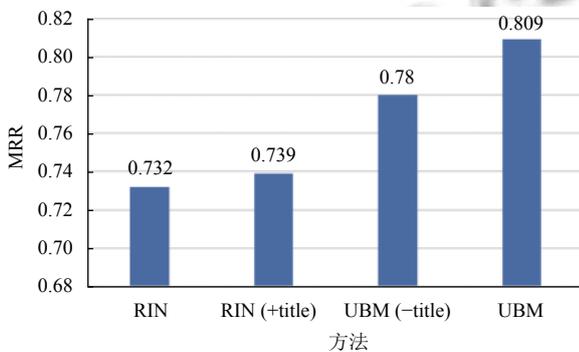


图 4 不同网络嵌入与辅助信息的 UBM 的 MRR 性能

#### 4 总结与展望

在本文中,我们通过建模用户行为包括其点击行为及查询重构行为,对其查询意图进行了分析,提出了一种基于用户行为的查询建议方法。以用户检视行为和点击行为作为基准设计了点击模型,从而借助注意力机制捕获了用户查询与返回结果的相关性,使得从会话内查询结果中可以得到更多的用户查询意图信息;接着受日常生活启发,了解到标题作为用户第一时间检视的信息,其极大的影响了用户的查询行为,于是利用属性异构网络将标题作为辅助属性信息引入我们的模型,并以此来构建我们的查询重构,使得重构行为更具有可解释性;最后利用多头注意力机制并借助多任务学习,使得该模型在查询鉴别和查询生成两类任务中都相比于基线模型有了不同程度的提升效果。此外,对结果的分析也证明了我们所提出的方法的有效性和鲁棒性。在未来的工作中,我们试图通过在不同场景中

捕获的不同辅助信息,将其集成入我们的模型中,观察其是否对这类任务具有普适性。

#### 参考文献

- 1 Fonseca BM, Golgher PB, De Moura ES, *et al.* Using association rules to discover search engines related queries. Proceedings of the IEEE/LEOS 3rd International Conference on Numerical Simulation of Semiconductor Optoelectronic Devices. Santiago: IEEE, 2003. 66–71.
- 2 Sordoni A, Bengio Y, Vahabi H, *et al.* A hierarchical recurrent encoder-decoder for generative context-aware query suggestion. Proceedings of the 24th ACM International Conference on Information and Knowledge Management. Melbourne: ACM, 2015. 553–562.
- 3 Wu B, Xiong CY, Sun MS, *et al.* Query suggestion with feedback memory network. Proceedings of the 2018 World Wide Web Conference. Lyon: ACM, 2018. 1563–1571.
- 4 Jiang JY, Wang W. RIN: Reformulation inference network for context-aware query suggestion. Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino: ACM, 2018. 197–206.
- 5 Borisov A, Markov I, De Rijke M, *et al.* A neural click model for web search. Proceedings of the 25th International Conference on World Wide Web. Montreal: ACM, 2016. 531–541.
- 6 Goyal P, Ferrara E. Graph embedding techniques, applications, and performance: A survey. Knowledge-Based Systems, 2018, 151: 78–94. [doi: 10.1016/j.knosys.2018.03.022]
- 7 Guo F, Liu C, Kannan A, *et al.* Click chain model in web search. Proceedings of the 18th International Conference on World Wide Web. Madrid: ACM, 2009. 11–20.
- 8 Cho K, van Merriënboer B, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: ACL, 2014. 1724–1734.
- 9 Jiang JY, Ke YY, Chien PY, *et al.* Learning user reformulation behavior for query auto-completion. Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval. Gold Coast: ACM, 2014. 445–454.
- 10 Chen J, Mao JX, Liu YQ, *et al.* Investigating query reformulation behavior of search users. 25th China Conference on Information Retrieval. Fuzhou: Springer, 2019. 39–51.

- 11 Chang SY, Han W, Tang JL, *et al.* Heterogeneous network embedding via deep architectures. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney: ACM, 2015. 119–128.
- 12 Feng YF, Lv FY, Shen WC, *et al.* Deep session interest network for click-through rate prediction. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao: International Joint Conferences on Artificial Intelligence Organization, 2019. 2301–2307.
- 13 Chen QW, Zhao H, Li W, *et al.* Behavior sequence transformer for e-commerce recommendation in Alibaba. Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data. Anchorage: ACM, 2019. 12.
- 14 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: NIPS, 2017. 6000–6010.
- 15 Liu YQ, Miao JW, Zhang M, *et al.* How do users describe their information need: Query recommendation based on snippet click model. Expert Systems with Applications, 2011, 38(11): 13847–13856.
- 16 Jansen BJ, Spink A, Blakely C, *et al.* Defining a session on web search engines. Journal of the American Society for Information Science and Technology, 2007, 58(6): 862–871. [doi: 10.1002/asi.20564]
- 17 Dehghani M, Rothe S, Alfonseca E, *et al.* Learning to attend, copy, and generate for session-based query suggestion. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. Singapore: ACM, 2017. 1747–1756.
- 18 Liu JW, Li QS, Lin YS, *et al.* A query suggestion method based on random walk and topic concepts. 2017 IEEE/ACIS 16th International Conference on Computer and Information Science. Wuhan: IEEE, 2017. 251–256.