

查询推荐研究综述

张晓娟¹, 彭琳², 李倩³

(1. 西南大学计算机与信息科学学院, 重庆 400715; 2. 中国科学院文献情报中心, 北京 100190;
3. 山西大学经济与管理学院, 太原 030006)

摘要 查询推荐是一种提高用户搜索效率的重要技术, 其核心任务是帮助用户构造有效查询并以此准确描述用户信息需求。作为当今搜索引擎的核心技术之一, 查询推荐吸引了学术界和工业界的广泛关注, 一直以来都是信息检索领域中重要的研究主题。本文以国内外会议、期刊发表的有关查询推荐研究的文献为对象, 利用归纳总结方法, 首先详细梳理了查询推荐中主流方法——基于简单共现信息的方法、基于图模型的方法以及融合多种信息的方法, 然后总结评述了评测方法与指标, 最后分析了未来可能的研究方向。

关键词 查询推荐; 查询; 用户意图; 查询会话

A Review of Query Suggestion

Zhang Xiaojuan¹, Peng Lin² and Li Qian³

(1. *School of Computer and Information Science, Southwest University, Chongqing 400715;*
2. *National Science Library, Chinese Academy of Sciences, Beijing 100190;*
3. *School of Economics and Management, Shanxi University, Taiyuan 030006*)

Abstract: Query suggestion is an important technique for improving search efficiency, and its core task is to help users construct effective queries to accurately describe users' information requirements. As a core technology of search engines, query suggestion has attracted wide attention in both academia and industry and has long been considered to be an important research topic in information retrieval. This paper summarizes the recent research progress in query suggestion using papers published in China's and international conferences and journals. On this basis, the mainstream methods—simple occurrence information-based method, graph-based method, and integration of multiple information-based methods—are reviewed in detail in this paper. Then, the related evaluation methods and metrics are summarized and discussed. Finally, the possible future research directions are pointed out.

Key words: query suggestion; query; user intent; query session

1 引言

搜索引擎已经成为人们从海量 Web 数据中获取信息的重要途径。目前搜索引擎采用的主要交互方式是用户输入查询 (query), 即检索系统根据输入的查询提供检索结果。受用户认知水平以及个人习

惯等诸多因素的影响, 用户提交的查询较短且不规则, 使得搜索引擎较难从用户简短的关键词中推测出其真实信息需求并以此返回相关查询结果。为解决此类问题, 大多数商业引擎如 Google、Yahoo!、百度等采用查询推荐 (query suggestion 或 query recommendation) 技术^[1], 即尝试提供各种查询推荐 (发

收稿日期: 2018-09-12; 修回日期: 2018-10-28

基金项目: 国家社会科学基金青年项目“融合用户个性化与实时性意图的查询推荐模型研究”(15CTQ019)。

作者简介: 张晓娟, 女, 1985年生, 博士, 副教授, 硕士生导师, 研究方向为信息检索; E-mail: zxj0614@swu.edu.cn; 彭琳, 女, 1971年生, 硕士, 编辑, 研究方向为信息管理; 李倩, 女, 1983年生, 博士, 讲师, 硕士生导师, 研究方向为网络信息管理。

现或构造一组与原查询 Q 相关的候选查询 $\{Q_1, Q_2, Q_3, \dots\}$, 使得用户查询更可能地与用户需求的文档接近, 以此缓解用户构造查询的压力。

搜索引擎提供的查询推荐服务主要包含查询自动补全(query auto-completion)与查询重构(query reformulation)推荐^[2]。查询自动补全是指导搜索引擎根据用户在搜索框中键入的查询前缀、中缀或者后缀, 通过字符匹配在搜索框中的下拉表中自动生成了与原查询包含部分共同词的候选查询列表^[3]; 而查询重构推荐是指当用户完成查询输入后, 通过对原查询进行替换、删除或添加词等操作, 在检索结果页面后面生成与原查询相关的候选结果列表。其中, 查询自动补全的推荐结果局限于利用与初始查询包含共同词的候选查询来定位用户意图, 限制了可用于解释用户可能意图的查询表达式, 无法提供相关推荐; 而查询重构推荐不受此限制, 其推荐结果可用各种不同形式的查询表达式来描述用户可能意图, 更能体现用户查询语义的多样化。因查询重构推荐具有广泛的应用前景和研究价值, 引起了工业界和学术界的广泛关注。

考虑到搜索引擎运营商为维护竞争地位, 常常将其核心技术作为一级商业机密, 而查询推荐技术作为搜索引擎的一种核心技术, 难以被外界获悉。基于此, 本文主要关注学术界中对查询重构推荐的研究。笔者以检索词“查询推荐”、“查询建议”、“查询重构”、“query suggestion”、“query reformulation”、“query recommendation”等在Web of Science、Springer、CNKI、维普等数据库以及Google学术中不限时间范围地进行标题、主题、关键词的逐步扩展检索, 总共获得相关文献130篇, 通过通篇阅读每篇文献, 最终筛选出与本文研究高度相关的90篇文献。另通过对这些文献内容进行归纳总结后发现, 学术界中对查询重构推荐的已有研究侧重于方法研究。根据方法数据来源与最终所能解决的问题, 本文将主流方法归为以下三类: 基于简单共现信息的方法、基于图模型的方法以及融合多种信息的方法。基于此, 本文首先对这三类方法进行了梳理与概括, 指明它们的主要技术特点以及存在的局限等, 且在此基础上, 也对相关的评测方法与指标进行了总结与评述, 最后指出了未来可能的研究方向。相对李亚楠等^[4]在2010年所进行的查询重构推荐综述工作来说, 本文的主要贡献为: ①对2010年后的相关研究工作进行了综述; ②从不同角度对查询推荐方法进行了全面归类与评述, 并对将来后

续研究进行了多角度方向的展望。需说明的是, 因已有一些研究常常用“查询推荐”直接指代“查询重构推荐”^[1-3], 为表述简洁性, 本文中的“查询推荐”在未特殊说明情况下均指“查询重构推荐”, 且本文所探讨的查询重构推荐主要为搜索引擎的查询重构推荐。

2 基于简单共现的查询推荐研究

最初的查询推荐方法是利用查询或查询词之间的直接共现信息来实现查询间的有效推荐, 并且基于流行度角度, 使得推荐的结果代表群体用户共性的信息需求。其主要思想为: 共现于同一文档或者同一查询会话(session)中的查询是关联的, 可成为彼此的候选查询推荐。根据所依赖的不同数据集, 此类方法又可细分为基于文档共现和基于查询日志共现两类方法。

(1) 基于文档共现方法主要包括全局文档集共现与局部文档集共现两种方法。全局文档集共现方法是指分析利用所有文档中词间的简单共现关系, 以此来获得与原查询关系最紧密的其他查询, 从而构建推荐查询^[4-5]。该方法能避免数据集稀疏问题, 但当文档集合过于庞大时, 计算复杂度较高, 且不利于计算词间的语义关系; 局部文档集方法是通过分析部分文档(典型如查询返回的排名靠前的文档)来实现推荐^[6-9], 该方法能在一定程度上降低计算复杂度, 但如何选取相关文档是一个较难解决的问题。

(2) 基于查询日志共现方法的核心思想为: 根据查询会话中查询的共现信息, 利用互信息^[10]、关联规则^[11-12]、相似度算法等来度量查询或查询词间的相关性^[13], 以此实现查询推荐。根据其推荐单元的不同, 该方法可细分为基于查询级别的推荐和基于查询词级别的推荐, 这两类查询推荐所采用的方法及其主体思想与主要特点如表1所示。

整体来说, 基于查询日志共现是简单共现方法中的主流方法, 此类方法较易实现, 但其缺陷在于难以对间接关系的查询实现推荐, 且难以推荐语义相关的查询; 另查询会话的准确切分是实现此类查询推荐的前提条件, 而基于时间间隔的查询会话切分方法常常影响查询推荐的最终准确度。

3 基于图模型的查询推荐研究

基于图模型的查询推荐是为了实现具有间接关

表 1 基于查询日志共现的查询推荐方法

	相关方法及其主体思想	主要特点
基于查询级别的查询推荐: 将查询作为推荐单元,通过 查询之间的相互替换生成候 选查询	基于共现频次的查询推荐:根据查询间共同出现在同一查询会话中的频次来实现查询推荐,即共现频次越高的查询对,越有可能成为彼此的候选查询推荐 ^[12]	简单易于实现,但容易推荐高频的噪声查询,也难以表达查询之间的间接关系
	基于逐点互信息(point-wise mutual information)的查询推荐:将查询日志中相互依赖(共现)程度高的查询作为彼此候选查询推荐 ^[10]	避免高频查询被过多推荐,且能对候选查询进行排序,但对低频查询进行过多推荐
	基于关联规则的查询推荐:从搜索日志中挖掘出与当前查询匹配的搜索模式而实现查询推荐 ^[11-12]	除能获得相关候选查询外,还能对候选查询进行排序,且计算复杂度小
	基于查询邻近算法的查询推荐:利用查询在同一查询会话中前后共现信息进行相关函数评分 ^[2,12,14-15] 而返回候选查询	进一步考虑了查询序列信息,但是非对称函数,即查询对 (q_i, q_j) 与查询对 (q_j, q_i) 计算的结果值不一致
基于词级别的查询推荐:将 查询词作为推荐单元,通过 对初始查询进行添加、替换、 删除词来生成候选查询	基于相似度算法的查询推荐:首先根据查询词在同一查询会话中共现信息而构建共现矩阵,再利用 Jaccard 函数计算两查询词之间的共现频次,或者利用余弦值计算两查询词之间向量相似性来定义查询词的相似性,替换与初始查询相关的词 ^[13]	Jaccard 系数对高频查询有效,但是难以衡量低频查询之间的相关性;余弦值能有效衡量低频查询词之间的相关性,但是构建查询词向量的复杂度比较高
	基于对数似然比例(log likelihood ration, LLR)的查询推荐:利用从假设中抽取的概率来计算二项分布下观察数据的概率,以此对文本元素(如查询词等)的相关度进行衡量 ^[10] ,实现查询中词的替换	等同于互信息的计算,可减少对频次较少查询的过多推荐
	基于转移概率的查询推荐:首先获得每个查询词的上下文词集,然后利用最大似然估计,分别计算每个查询词与其上下文词集中每个词在同一数据集中的共现概率,以此获得每个查询词的上下文分布概率。最后利用 KL 距离(Kullback-Leibler)计算初始查询中词与候选词之间的上下文分布概率相似性,以此实现查询中词替换或添加而生成候选查询 ^[16]	能实现查询词级别的查询推荐,如查询替换、查询添加等,但计算复杂度较高

系查询间的有效推荐。该方法主要通过图中节点关系来表达查询间关系,再利用图遍历等思想来实现查询推荐,其研究主要包括基于查询流动图的推荐与基于 query-URL 点击图的推荐。

3.1 基于查询流动图的查询推荐

查询流动图(query flow graph, QFG)^[14-15]旨在图中利用查询会话中的查询重构信息来表示查询之间的关系,以此实现查询推荐。该方法主要包括查询流动图构建以及候选查询排序两个过程。

(1) 查询流动图构建。根据不同查询在同一查询会话中共现情况(即查询重构行为信息)将其连接起来构成同质有向图 $G=(V,E,w)$ 。其中, $V=Q \cup \{s,t\}$ 表示图中节点集合; Q 表示提交给搜索引擎的不同查询集合; s 与 t 分别表示某搜索任务的起始节点与终止节点;边集合 $E \in V \times V$ 中每个元素表示两查询之间的有效边,表示用户提交查询的序列,若至少有一个用户在同一个查询会话中提交了 q_i 之后又提交了查询 q_j ,则存在着由查询 q_i 指向查询 q_j 的有向边; w 表示每对查询边 $(q_i, q_j) \in E$ 的加

权函数,常用的加权函数主要有基于链接概率的加权^[14]以及基于查询间在同一查询会话中共现频数的加权^[17]。

(2) 候选查询排序。候选查询排序的方法主要包括:①基于权值的排序。根据查询流动图中候选查询节点与初始查询节点之间边权值大小实现候选查询排序^[15]。②基于随机游走的排序。在查询流动图基础上,首先构建查询间转移概率矩阵,通过随机游走迭代后将候选查询与初始查询之间的转移概率作为相关性分数,以此实现查询推荐^[14-15]。

虽 QFG 能实现间接关系查询间的有效推荐,但存在着如下问题:①不能完全真实地表示用户查询重构行为,如在实际行为中,用户会倾向于根据自己偏好来修改查询,也会选择搜索引擎推荐的查询^[18];②难以为用户生成高效用(即满足用户需求)或实时性的查询推荐。于是,学者们尝试通过对 QFG 进行扩展来解决以上问题,如 Anagnostopoulos 等^[18]将查询推荐视为对查询流动图中随机游走转移概率的扰动,提出了为查询流动图中每个查询节点添加查询重构连接,以此来实现最终推荐结果的期

望效能(返回相关查询结果集的概率)最大化; Baraglia等^[19-20]提出了一种增量算法(incremental algorithm)来更新QFG模型,使得最终推荐模型能不断添加数据集中的新信息且能删除时间久远的信息,以此满足用户不断变化的信息需求;李亚楠等^[21]将用户查询或查询之间的间接关系构建关系图,并提出了综合考虑查询关系图全局信息的加权SimRank算法,该算法能挖掘查询间间接关系和语义关系;朱小飞等^[22]提出了一种基于吸收态随机游走的两阶段效用性查询推荐方法对查询重构与查询点击行为进行建模,并根据学习到的各查询效用进行查询推荐;李竞飞等^[23]为了将用户实时性意图融合到查询流动图中,首先利用查询流动图获得候选查询词以及用户对查询推荐结果的满意度,最后基于用户满意状态,利用自适应排序模型对候选查询与初始查询之间的新颖性与相似性加权来对候选查询进行排序。

3.2 基于query-URL点击图的查询推荐

query-URL点击图(或被称为query-URL二部图)是由查询日志中用户查询、查询所对应的点击网页(URL)以及二者之间的连接边构成,此类图表示为 $G=(V,E,w)$ 。其中, $V=V_1\cup V_2$ 表示节点集合,一端节点 V_1 表示用户提交的查询,另一端节点 V_2 表示用户提交查询后点击的URL;边 E 用于连接 V_1 与 V_2 中节点的边集合; w 表示边之间的权重函数,主要是通过点击频次^[24-28]、TF-IDF值^[29]或熵值^[30]确定。在点击图基础上而生成查询推荐的方法主要有:①基于聚类方法的查询推荐。根据查询的点击向量,采用凝聚聚类^[31]、 K 均值^[32]或层次聚类^[33]等算法对query-URL点击图进行聚类,同一类簇中的查询作为彼此的候选查询推荐。②基于随机游走的查询推荐。利用随机游走对点击图进行迭代,获得候选查询到初始查询平均首达时间(hitting time)^[25,34]或者转移概率^[35-36],实现对候选查询的排序。

相对查询流动图来说,query-URL点击图能融合表征用户意图的点击信息,其推荐结果更能满足用户需求,是当前查询推荐研究中最常用的图模型方法。但query-URL点击图存在着如下问题:①忽略了查询日志中的用户信息,难以根据用户兴趣实现查询推荐;②不同查询之间的共同点击URL数有限,难以对包含点击信息少查询间关系进行衡量。为解决以上问题,研究者们尝试对query-URL点击

图进行扩展,如Ma等^[27]为了将用户信息融合到查询流动图中,利用矩阵降维方法,基于点击信息构建了用户-查询二部图与query-URL点击图,通过对点击图进行主题学习,推荐用户感兴趣主题下的查询;Sejal等^[37]为解决查询点击信息稀疏性,构建融合查询点击信息以及查询间文本相似度信息的图模型,即两查询之间的边权值不仅包括查询之间共同点击URL数,也包含了两查询文本相似度,再利用深度优先遍历模型来生成候选查询;张乃洲^[35]为解决点击query-URL点击图中由于数据稀疏所产生的非两同质图现象带来的查询建议失败问题,对query-URL点击图进行非两同质图检测和图合并操作,以此降低或消除图的非连通性,再采用基于随机游走模型的图挖掘算法生成给定查询的查询建议集。除以上研究外,还有学者尝试拓展query-URL点击图来实现对歧义性查询的有效推荐,如Ye等^[38]首先基于查询日志信息构建query-URL点击图,综合考虑点击URL在结果页面中的排序以及点击顺序来计算query节点与URL节点之间边的强度,再通过节点之间边强度和转移概率进行随机游走遍历图来算查询之间的相关性;最后,在基于熵的方法计算每个查询模糊度的基础上,根据查询之间的模糊相似性和相关性来获得候选查询结果列表。

4 融合多种信息的查询推荐研究

基于简单共现与基于图模型的方法几乎都单独利用查询日志或者文档信息实现查询推荐,最终主要解决的问题是发现与初始查询相关的查询。而融合多种信息的查询推荐方法是综合利用查询日志、词语、文档信息或者外部知识资源(Wikipedia、WordNet等)等实现查询推荐,不仅返回相关查询,也为解决查询推荐中更为复杂的问题,如实现上下文相关的查询推荐、长尾查询的查询推荐、个性化查询推荐以及查询推荐结果多样化。

4.1 上下文相关的查询推荐

上下文相关的查询推荐,旨在推荐与用户提交当前初始查询时之前的搜索或点击行为等相关的候选查询。此类研究的实现主要基于如下假设:上下文信息可以减少查询词的歧义性,提高查询推荐的准确度。根据所需解决的问题,学者们采用了不同来源的上下文信息,主要包括以下三类来源:

(1) 基于历史查询的上下文相关查询推荐。此类研究是将用户提交的历史查询视为上下文信息,

通过对其历史查询进行建模,返回与之相关的候选查询推荐。如Cao等^[39]提出一种基于上下文感知查询建议方法,该方法分为线下和线上两步:在线下,使用户点击图进行聚类,把查询总结成不同概念,然后为查询会话数据序列构造概念后缀树作为查询建议模型;在线上,把用户提交的查询序列映射到概念中,获取用户搜索上下文信息,通过查询概念后缀树得到相关查询;廖振^[2]将用户当前搜索之前的搜索行为作为用户的搜索语境,利用监督式学习方法获取查询之间相似度信息,并基于近邻聚类方法将搜索上下文的查询聚集成搜索任务,再设计基于搜索任务的随机游走推荐算法来生成任务相关的查询;Sordoni等^[40]基于用户已经提交的查询序列提出了生成概率模型(generative probabilistic model)来实现基于上下文的查询推荐,将给定的一序列查询视为前缀,基于层次化的神经网络结构,即层次递归编码器-解码器(hierarchical recurrent encoder-decoder, HRED)预测后续最有可能出现的查询词序列,以此生成候选查询推荐;Dehghani等^[41]提出了通过利用查询感知的注意机制来增强序列-序列(sequence to sequence)以此来对用户会话中的上下文查询进行编码,并能以此控制推断出下一个候选查询所在查询会话的范围;Jiang等^[42]将同一会话中当前搜索之前所提交的查询视为上下文,首先利用异构嵌入(heterogeneous network embeddings)方法将查询以及查询重构以同态隐空间(homomorphic hidden space)来表示,再利用具有注意机制的递归神经网络通过读取同态查询和重构的嵌入来对会话上下文进行编码,最后采用二值分类器和一个基于RNN的解码器作为候选查询鉴别器和生成器。

(2) 基于用户行为的上下文相关查询推荐。即将用户的历史点击行为信息融合到查询推荐模型中实现上下文相关查询推荐。如Liu等^[43]与罗成等^[24]认为搜索结果页面上的摘要为用户进行相关性判断的依据,隐含着迎合用户意图的信息,在利用已有方法生成候选查询的基础上,根据候选查询出现在摘要点击中的情况进行加权而对候选列表进行重排序;石雁等^[44]将用户的点击行为视为上下文信息,基于对用户查询点击日志进行分析与挖掘,首先采用朴素贝叶斯模型预测用户点击URL值,再利用反向点击图将每个URL预测值作为用户意图传播给日志中与其对应的查询项,再结合文本匹配和时间相关因子实现查询推荐;Guo等^[45]为了使得候选推荐

查询更具有高效用性,尝试利用动态贝叶斯网络,通过从用户搜索会话的序列用户行为信息中推测出高效用的候选查询,以此实现推荐。

(3) 基于位置信息的上下文相关查询推荐。此类查询推荐的主要目的是在获得用户位置信息的基础上,获得与用户地理位置相关的查询推荐。如Qi等^[46]将用户所在位置视为上下文,首先构建加权关键词-文档图模型,以此能捕捉查询关键词之间语义相似度以及结果文档集与用户地理位置之间的地理距离,再采用重启式随机游走(random-walk-with-restart)来遍历图,从而选择具有最高分数的关键词查询建议。

以上相关研究大多从与初始查询共现的查询会话中捕捉用户上下文信息。但考虑查询会话中常常包含了多个搜索任务^[47-48],查询会话中某些查询信息可能偏离了初始查询的用户意图,不能作为用户上下文信息。基于此,另有学者尝试在查询会话中识别与初始查询任务相同查询的基础上实现查询推荐。如Field等^[49]将表征用户上下文的查询分为任务查询与非任务查询,首先在时间间隔切分查询会话的基础上,综合根据词汇与语义相似度信息来判断两查询是否包含相同的搜索任务,最后通过随机游走遍历查询流动图实现查询推荐。其实验结果表明,当用户上下文中只包含任务查询时,有利于提高查询推荐的准确度;Ozertem等^[50]在自动标注查询之间是否相关时,考虑到同一查询会话中两查询可能表示不同任务,不能直接进行标注,并未将共现于同一查询会话的查询对直接标注为相关查询,而是在隐式识别用户任务的基础上计算两查询是否属于同一任务,然后再确定查询对之间的标注标签。

4.2 长尾查询的查询推荐

因用户查询的频度分布服从长尾现象(“long tail”)规律,即很大比例的查询被用户提交的频数较低^[51],此类查询具有数据稀疏性。对长尾查询进行推荐的常用方法主要有以下两类:①基于伪相关反馈的长尾查询推荐。即根据初始查询返回的结果文档集来扩展初始查询的相关信息,相关研究有:Song等^[36]借用伪相关反馈思想,在实现查询推荐时,除了考虑用户点击的URL,也考虑了用户跳转的URL信息,分别构建了query-URL点击图与query-URL跳转图,再利用随机游走重启技术分别对两个图进行遍历,最后通过URL之间关系对这两

类图进行组合以此实现查询推荐；Santos等^[52]利用与长尾查询共现于查询会话或者共同点击了同一文档的查询来扩展长尾查询的信息且为其选取分类特征，最后利用学习排序（learning to rank）方法实现候选查询排序；Garigliotti等^[53]为对查询日志中无相关历史数据的查询实现推荐，利用生成概率模型从多种资源（如网络文档片段、网络文档集、Wiki-How等）中获取与初始查询相关的关键词句，再从这些关键词句中生成查询推荐；Liu等^[54]提出基于随机游走和主题概念实现对长尾查询的有效推荐。在该模型中，首先基于两步随机游走来遍历 query-URL 二部图生成初始的候选查询推荐列表，然后利用即时搜索（如百度、Bing等）为初始查询返回排名前10位的结果集，并将结果集中出现频次排名前10位的词用于表征当前初始查询的主题概念集合，最后根据候选查询所包含的主题词在结果集中出现的频次进行累积求和来对初始候选查询进行重排序。②基于实体扩展的长尾查询推荐。即利用模板或本体信息等扩展查询中命名实体信息来对查询进行扩展从而实现查询推荐，如Szpektor等^[17]首先构建查询模板，再通过查询模板之间规则信息进行推理而实现对长尾查询的查询推荐；Huang等^[55-56]为对长尾查询的信息进行扩充，利用知识库如YAGO与Freebase实体之间的联系来扩展与查询中实体相关联的其他实体，扩展后的实体再被用于推荐候选查询。

也有学者尝试在查询流动图或者查询点击图中加入查询词节点信息，通过扩充查询间关系进行查询推荐。如Bonchi等^[57]将长尾查询分解到单词，构建了term-query图，图中节点由查询与查询词构成，图中边由查询间关系以及查询与查询词之间关系构成，且在图中分别以初始查询包含的词作为起始点随机游走遍历图，最后将与初始查询词之间转移概率高的查询作为候选查询；白露等^[58]提出一种关于词项查询图（term-graph model）的概率混合模型，首先将查询意图定义为单词的分布，从单词的角度去预测长尾查询的查询意图，再根据查询意图信息来集成单词的个性化随机游走方法，该方法通过衡量单词在查询中的重要程度对长尾查询进行推荐；刘钰峰等^[59]提出基于查询上下文训练词汇与查询间的语义关系，并结合查询和URL对应的点击图以及查询的序列行为构建 term-query-URL 异构信息网络，采用重启动随机游走算法进行查询推荐，该方法综合了语义和日志信息，提高了稀疏查询的推荐

效果。

从以上相关研究中可以看出，已有研究大多基于对查询或者查询词进行扩展而实现查询推荐。但因长尾查询包含表征用户意图的用户行为信息稀疏而难以被利用，则此类查询的查询推荐结果常常难以定位到用户意图。因此，长尾查询的查询推荐研究可考虑与长尾查询类似查询的查询重构行为以及点击行为来优化推荐结果；另外，在一般情况下，长尾查询频率低，是因为用户的个性化比较强，提交这些查询词的用户较少，也可纵向获取用户更多信息，改进查询推荐结果。

4.3 个性化查询推荐

考虑到用户背景不同会带来搜索需求的差异，为使得个体用户尤其是弱势群体用户的信息需求不被强势群体淹没，学者们提出了满足用户个性化需求的查询推荐，即针对不同用户背景推荐不同的候选查询。此研究主要包括以下两个方面：

(1) 基于单个用户信息的个性化查询推荐。其主流方法是根据单个用户历史提交查询或者点击文档来获得用户偏好信息，以此实现查询推荐。Chen等^[60]首先构建了查询-概念二部图（query-concept bipartite graph）和概念关系树（concept relation tree），再基于用户当前最新提交的查询以及点击URL，动态更新查询-概念二部图以及概念关系树来进一步实现个性化查询推荐；Jiang等^[34]将查询日志中每个用户条目视为一个伪文档，且假设每个文档是由一组潜在主题构成，在生成用户伪文档时，将同一查询会话中的查询词与URL限定在同一主题下，利用Beta分布去捕捉每个主题的时间突出性（temporal prominence），且在每个特定主题下为每个用户定制了查询词与点击URL（即每个用户针对每个特定主题所偏好的查询词与URL），最后根据查询与用户伪文档相似性实现个性化推荐；Bing等^[61]根据用户点击文档与所提交的历史查询构建伪文档，再利用LDA（latent Dirichlet allocation）分别对其进行潜在主题分析，以此获得每个词的潜在主题以及用户对每个主题的偏好概率，最后利用隐马尔可夫模型，根据词之间依存性、主题依存性以及用户偏好来建模，以此实现个性化查询推荐；Chen等^[62]将用户提交的历史查询表征用户个性化信息，通过计算候选查询与用户历史查询之间的语义相似度的平均值来对查询进行候选查询的个性化排序；张晓娟^[63]首先利用用户提交的历史查询来构建表征用户兴趣的个

个性化偏好,然后利用查询词嵌入技术为每个查询获得该词上下文信息的词向量,最后利用词向量进一步构建表征用户偏好的向量,从而基于词向量与用户向量实现根据用户偏好生成候选查询;Chen等^[64]认为用户长期和短期搜索行为能表征其偏好,通过线性方式将两类行为信息结合起来,再在概率模型中融合查询之间语义相似度以及用户行为信息来实现个性化查询推荐。另一些研究者也尝试利用用户历史检索文档信息来获得偏好信息。如王卫国等^[65]首先对用户检索文档集和用户历史查阅文档分别进行聚类而得到全局文档簇和用户兴趣簇,再对全局簇进行LSA(latent semantic analysis)分析,构建基于概念群的潜在语义空间,然后从用户兴趣簇提取前 n 个代表性关键词而形成用户向量,最后基于查询之间的潜在语义空间以及用户向量相似性来生成与用户兴趣相关的候选查询。

(2) 基于多个相似用户的个性化查询推荐。为了克服个性化推荐中获取单个用户信息的困难,一些学者^[66-67]认为具有相似搜索行为的搜索偏好相似,将相似用户历史行为数据扩充到单个用户数据中,以此实现个性化查询推荐。如石雁等^[66]提出将每个用户的查询日志作为文档,利用空间向量模型计算文档间的相似度,故可通过将用户在历史数据中对链接的点击频率作为对链接的偏好评分,采用改进的欧氏距离计算用户的最近邻居,以此将相似用户历史行为数据扩充到单个用户数据中,然后基于朴素贝叶斯模型,训练数据并预测查询-链接的点击率,且将其作为权重用于query-URL点击图中,最后利用点击传播产生查询推荐;孙达明等^[67]在查询流动图的基础上,构造密集行为块来表示用户特征,从而构建用户行为模型以此区分用户背景,且提出了面向多样化搜索背景的查询推荐方法。

当前个性化查询推荐研究大多基于“所有查询都适合采用个性化技术”的假设。但Dou等^[68]研究发现,个性化技术并非适合所有查询,某些查询采用个性化技术反而会降低用户满意度,如为查询“百度”返回查询结果时,因用户提交此查询更多关注百度网站主页,故此类查询更适合基于大众访问的排序方法。因此,个性化查询推荐也应考虑这一问题,需要深入理解某查询是否具有个性化意图基础上再进行选择性的个性化查询推荐^[69]。

4.4 多样化查询推荐

多样化查询推荐是满足用户大众化意图与个性

化意图的一种折中方法,通过为多含义查询提供尽量多的涵盖含义类别的推荐集合,以降低推荐失败的风险。其中,当前一些搜索引擎的查询推荐结果体现了多样化,如图1中百度为模糊查询“苹果”生成的候选查询推荐,既包括了与电子产品子主题相关,也包括了与水果主题相关的查询;同样,图2中Google为模糊查询“earthquake”生成了包含“earthquake”多个潜在主题(如“earthquake california”、“earthquake philippines”等)的查询推荐。



图1 百度为查询“苹果”生成的查询推荐



图2 Google为查询“earthquake”生成的查询推荐

多样化查询推荐的主要思路是采用不同迭代方法,发现既与原始查询相关且与已有候选查询存在差异的候选查询。实际上,大多数研究者通过迭代选择能最大限度地覆盖原查询潜在子主题,且尽可能地选取与已有候选查询中子主题冗余度最小的候选查询。其中,根据考虑初始查询潜在子主题的方法,查询推荐多样化方法主要分为以下两种:

(1) 隐式多样化方法。此类方法基于相似查询包含相似子主题的假设,隐式地考虑初始查询中的潜在子主题,通过比较待排序查询与已选择候选查询的差异度来减少候选查询之间的语义冗余度。其中,MMR(maximal marginal relevance)是常用的迭代算法,即在每次迭代中,通过比较待排序查询与已有候选查询相似度来判断当前待排序查询是否存在差异以及差异程度,并据此对当前待排序查询进行选择。其中,如何计算查询之间相异性是此类方法的研究重点。主要相关研究有:Ma等^[70]首先基于构建query-URL点击图为初始查询返回第一个相关查询并构建候选查询集合 S ,然后利用点击时

间 (hitting time) 来选择其他多样化的候选查询, 即每次将到 S 的点击时间最大的查询作为候选查询并添加到 S , 以此实现多样化查询; Bordino 等^[71]首先利用查询日志中信息构建查询流动图, 利用谱投影技术 (spectral projection) 将初始 QFG 映射到低维欧式空间 (即利用广度优先原则为每个查询抽取子图), 通过计算投影点之间的距离来计算图中两查询之间的差异性, 最后利用贪婪算法 (greed algorithm) 实现查询结果多样化; Song 等^[72]首先利用随机游走从查询日志中为初始查询返回系列候选查询, 再构建多样化函数对候选查询进行重排序, 以此返回多样化结果。其中, 该多样化函数是在综合查询相似性以及如 URL 的差异性、域名多样性等特征的基础上, 利用学习排序算法来优化 MMR 多样化函数而生成; Zhu 等^[73]假设查询空间是流行的 (manifold), 提出在查询流行排序 (manifold ranking) 中引入停止点信息来计算查询之间的相关度以及相异度, 以此实现多样化查询推荐; Hu 等^[74]提出了对词级别查询推荐进行的方法, 首先将每个查询词映射到 Wikipedia 主题中, 再基于 MMR 算法来对查询词之间的相关性以及主题相异性建模, 最后利用贪婪算法实现词级别查询推荐的多样性; Jiang 等^[34]提出了具有多样化感知的个性化查询建议 (personalized query suggestion with diversity awareness) 方法。在该方法中, 首先对查询日志进行多-二项 (multi-bipartite) 表示, 再利用规则化框架为原查询识别最相关的候选查询推荐, 然后再利用交叉二项首达时间 (cross-bipartite hitting time) 模型来选择相关且彼此之间存在差异的候选查询。除以上基于贪婪算法的研究外, 也有学者尝试利用其他算法来实现查询推荐结果的多样化, 如 Ding 等^[75]分别尝试利用流行后缀 (popular suffix)、神经语言模型以及序列-序列模型生成候选查询, 以此能描述用户初始查询中可能包含的潜在子任务, 其实验结果表明序列-序列模型取得了最好的实验效果。

(2) 显式多样化方法。在抽取查询子主题的基础上, 通过显式地判断当前待排序查询对未覆盖子主题的覆盖程度来对其进行选择。其中, 如何挖掘查询的潜在主题以及如何计算查询对子主题的覆盖度是此类研究的重点。主要相关研究有: Kim 等^[76]首先从初始查询返回结果文档中抽取关键词, 再基于关键词之间的检索性能和主题相似性来对关键词进行聚类, 以此生成查询分面, 再基于查询返回伪文档训练决策树为每个查询分面生成候选查询推荐, 最后利用 xQuAD (explicit query aspect diversifi-

cation) 多样化模型^[52] (类似 MMR 算法) 来筛选与查询文档相关且与已有候选查询主题不同的查询; Zheng 等^[77]考虑到贪婪式方法的计算复杂度较高, 提出了一种基于 WordNet 本体的查询推荐多样化方法, 即首先基于图模型生成候选查询, 再根据初始查询在 WordNet 中的不同含义来对候选查询进行多样化; Chen 等^[62]提出了融合用户个性化与多样化意图的查询推荐模型, 在基于 ODP (open directory project) 结合点击文档识别查询分面的基础上, 将用户的长时间搜索行为融合到贪婪模型, 计算待排序查询对未覆盖子主题的覆盖度来实现多样化查询推荐。和隐式查询推荐多样化方法相比, 显式查询推荐多样化方法的结果中必然包括涵盖初始查询潜在子主题的候选查询; 但相对隐式查询推荐多样化方法, 显式查询推荐多样化方法在挖掘识别查询的潜在子主题以及计算候选查询对每个子主题的覆盖程度方面会付出更大代价。

查询推荐结果多样化是当前比较新的研究任务, 虽然目前研究主要采用了在主题覆盖度或信息多样性等与相关度之间寻求一种平衡的语义化策略。而实际上, 多样化是个复杂性的问题, 还需考虑许多其他策略如新颖性、价值、质量等^[78]。例如, 网络是一个动态空间, 用户信息需求随时间而变化, 需要考虑用户提交查询发生的特定场景才能返回精确结果^[79-80], 故在对具有时态意图的语义模糊性查询时, 其多样化结果排序还需考虑到用户的实时意图^[81]。

5 数据集与结果评测

5.1 数据集

数据集是查询推荐评价的重要基础。目前, 查询推荐研究领域还无权威的数据集, 各学者在进行研究时往往根据自己需要选取不同的数据集。综合已有研究, 查询推荐的数据集主要来源于搜索日志, 且可公开获取的主要有 AOL、MSN 与 Sogou 查询日志数据集, 其具体统计信息可参见文献[3]。

5.2 查询推荐评测方法

查询推荐评测主要有人工评价和自动评价两种方法。

(1) 人工评测方法。针对每个评测查询, 取出某个推荐方法为其返回的前 n 个推荐结果, 邀请评测者对这些推荐结果的相关度 (如相关、不相关或部分相关) 进行评价^[40,62]。

(2) 自动评价方法。利用查询日志或者其他资源判断查询之间的相关性。包括：①基于查询日志的相关性判定。将每个查询会话中查询分为满意和满意两类^[27]，若某查询位于某查询会话的结束处且用户提交该查询后至少点击了一个URL地址，则该查询将被视为满意查询；而在同一查询会话中位于满意查询之前的查询被视为不满意查询；将不满意查询作为测试查询，利用相关查询推荐为其返回推荐结果，最后根据推荐结果集中包含满意查询的情况来衡量该荐方法的效果^[61]。②基于其他资源的相关性判定。利用人工编辑目录(ODP)、WordNet与Google Directory通过计算候选查询与初始查询之间语义相似度来衡量查询间的相关度^[27,30,70,82]。

5.3 查询推荐评测指标

查询推荐的最初目标是为用户推荐几个最相关的查询，主要的评价指标为信息检索领域被广泛使

用的准确度(precision)、MAP(mean average precision)、NDCG(normalized discounted cumulative gain)等。随着后期满足用户意图的查询推荐以及多样化查询推荐等研究工作的开展，一些研究者^[19,22,83-85]在评价某查询推荐结果时，除了考虑推荐结果是否与初始结果相关，也分别从效用性(utility)^[19,22,83]或多样化角度^[84-85]进行评价。其中，效用性角度与多样化角度旨在判断推荐结果是否与初始查询相关的基础上，进一步分别判断查询是否满足用户信息需求或候选查询之间存在着主题差异性。表2中总结了基于以上三个角度提出的主要评测指标及其说明。

6 结语

6.1 研究总结

纵览国内外相关研究，查询推荐研究近年来受

表2 查询推荐的主要评测指标

评价指标 的类别	评价指标	具体解释以及使用情况	说明
相关度：候 选查询与初 始查询之间 是否主题相 似	准确率(precision): $\text{precision}@N = \frac{\sum_{i=1}^N \text{Rel}(i)}{N}$	排名前N个查询推荐的准确度。Rel(i)是二值函数,若第i个推荐为相关,则其值为1,否则为0。该指标被用于文献[44,60,83,86-88]	较多使用,有时被称为quality或者accuracy
	召回率(recall): $\text{recall}@N = \frac{\sum_{i=1}^N \text{Rel}(i)}{S}$	召回相关查询的比例。Rel(i)是二值函数,若第i个推荐为相关,则其值为1,否则为0;S表示所有可能潜在的查询推荐。该指标被用于文献[60,83,86,88]	因很难获得所有的相关查询数,故较少使用
	平均准确率(MAP): $\text{MAP} = \frac{1}{K} \sum_{i=1}^K \frac{1}{ R_i } \sum_{k=1}^{ R_i } P(R_i[k])$	所有测试查询下的平均准确率。K为查询测试集; R _i 为测试查询q _i 对应的相关查询推荐集合;P(R _i [k])是在查询q _i 的排序队列中观察到查询R _i [k]的概率,如查询相关则为1,否则为0。该指标被用于文献[44,87]	MAP指标能解决准确度与召回率指标中的单点值局限性,同时能考虑推荐结果的排名问题
	正规化的累积收益折扣(NDCG)指标: $\text{NDCG}@N = \frac{1}{\text{IDCG}(q)} \sum_i \frac{2^{r(i)} - 1}{\log_2(i + 1)}$	该指标使用等级相关性分数,根据查询在结果列表中的排序来衡量查询的有用性或者增益。r(i)表示排序为i的查询推荐的相关等级;IDCG(q)是归一化因子,对应最佳情况下的排序结果。该指标被用于文献[7-8,28]	最佳的推荐结果难以获取,一般采用一些近似方法来获取如推荐方法中前50个结果里的最佳排序来替代
	MRR(mean reciprocal rank)指标: $\text{MRR}(A) = \frac{1}{ Q } \sum_{i=1}^{ Q } \frac{1}{\text{rank}_i}$	多个测试查询所返回查询推荐中第一个相关查询排名倒数的均值。rank _i 表示第i个测试查询所返回查询推荐中第一个相关查询的排序数;Q表示测试查询集。该指标被用于文献[40,49,62,88]	计算简单,是常用的评测指标
	similarity 指标:计算查询对的语义相似度,具体的similarity方法依赖于所采用的语义资源,其定义往往不统一	一种基于Google Directory、ODP、Wikipedia的相似度。如文献[27,30,70,79];另一种为基于点击向量或上下文向量的查询相似度,如文献[83]	一般在无基准实验结果集的情况下使用;但如何获取相应的语义资源是采用此类指标的关键问题

续表

评价指标 的类别	评价指标	具体解释以及使用情况	说明
效用性：候 选查询是否 满足用户需 求	MM-AMAP 指标： 最大匹配均值准确度(maximum matching averaged mean average precision)	根据查询推荐所返回文档覆盖查询潜在子主题 的能力来对其推荐结果进行评价；若候选查询 中包含了初始查询的某个子主题，则说明该候 选查询是有效的。该指标被用于文献[89]	如何获得查询子主题是采用该 指标的难点问题，故较少使用
	QRR 指标： $QRR(q) = \frac{RQ(q)}{N(q)}$	RQ(q) 表示候选查询推荐 q 作为重构时，用户 点击到相关文档的总查询次数；N(q) 表示查询 q 作为查询重构的指数。该指标被用于文献 [22,45]	因难以获得查询的点击文档信 息，故较少使用
	MRD 指标： $MRD(q) = \frac{RD(q)}{N(q)}$	RD(q) 表示查询 q 作为重构时，用户点击到的 相关文档数目。该指标被用于文献[22,45]	因需查询相关文档，评测工作 较复杂，较少使用
多样化：候 选查询之间 是否存在着 主题差异性	AMDR 指标： $AMDR(q) = \frac{\sum_{q_i, q_j \in SQ} MDR(q_i, q_j)}{C_{ SQ }^2}$	查询推荐结合中两两查询之间的相关文档的相 异度。MDR(q _i , q _j) 表示两个查询在同一搜索 引擎下返回相关文档的相异比值；SQ 表示查询 推荐集合。该指标被用于文献[35]	因需查询相关文档，评测工作 较复杂，较少使用
	ERR-IA 指标： $ERR-IA = \sum_{r=1}^n \frac{1}{r} \prod_{r=1}^{r-1} (1 - R_r) R_r$	该指标表示用户的需求被满足时停止的位置的 倒数的期望。R _r 表示查询相关度等级的函数。 单独计算每个主题的 ERR 值，然后计算子主题 的加权平均值。该指标被用于文献[53,90]	计算相对简单，查询推荐结果 多样化评测中常采用的指标
	SD 指标： $SD(S_q) = \frac{\sum_{i=1}^K \sum_{j=1, i \neq j}^K D(q_i, q_j)}{K \times (K - 1)}$	根据查询之间点击文档差异性衡量查询之间差 异性。K 表示查询集合 S _q 中包含的查询个数； D(q _i , q _j) 表示两查询所点击文档的差异性。该 指标被用于文献[70]	如何为每个查询获取相关的点 击文档是关键问题，较少使用
	α - NDCG@N 指标： $\alpha - NDCG@N = \frac{\sum_{i=1}^N NG(i/\log(i+1))}{\sum_{i=1}^k NG \times (i) / \log(i+1)}$	该指标是 NDCG 指标的变形，其中新发现的查 询主题将会被奖励，而多余的主题将会被惩罚。 NG(i) 表示排名列表中排在 i 位置的新颖性收 益。该指标被用于文献[52,62]	计算较为复杂，但是是查询推 荐结果多样化评测中常采用的 指标

到了信息检索领域的广泛关注且取得了较大进展，其研究主要集中在查询推荐方法以及查询推荐结果评测两方面。根据所依赖的数据集与所能解决的最终问题，查询推荐方法可分为三类：基于简单共现、基于图模型和融合多种信息的方法。其中，基于简单共现方法主要分为基于文档与基于查询日志两类方法，基于图模型细分为基于查询流动图与 query-URL 点击图两类方法，而融合多种信息方法根据其解决的问题又可细分为上下文相关的查询推荐、长尾查询的查询推荐、个性化查询推荐以及多样化查询推荐四类研究。另外，查询推荐评测方法主要包括人工评测与自动评价，而评测指标除了信息检索领域被广泛使用的准确度等指标外，还包括对推荐结果效用性以及多样性进行评价。

6.2 研究展望

未来查询推荐的研究可能主要体现在以下几个

方面：

(1) 基于知识图谱技术的查询推荐研究。当前大多研究仅停留在基于用户点击信息或者用户提交的历史查询信息，只是浅层次地理解用户意图，难以返回完全满足用户意图的查询推荐结果。后续的研究可从查询中的实体或者自然语言要素（如动词、形容词等）等角度，借用知识图谱等技术对查询中命名实体、实体属性以及包含的其他语义信息进行挖掘，以此深入理解查询中的用户意图，以此实现满足用户意图的查询推荐。比如，借助知识图谱技术表达查询中实体、关系、属性等知识要素之间关系，再通过知识融合消除实体之间的歧义性或并通过知识推理等技术进一步挖掘查询中隐含知识，有助于返回满足用户意图的查询推荐结果；也可借助知识图谱技术对多源异构数（如结合用户的浏览日志信息、用户的位置信息或者社交网络信息等）进行整合，实现大数据环境下互联网上的数据

进行知识抽取, 得到更加细粒度的用户和查询的特征信息, 从而更精准地计算用户与用户、用户与查询以及查询之间的相关性, 在查询推荐结果中进一步考虑用户个性化信息; 另也可考虑将知识图谱中丰富的语义信息融合到现有查询推荐方法(如随机游走、神经网络方法)中来增强数据的语义信息以进一步提高推荐准确度。

(2) 查询推荐结果的多维度多样化研究。已有的查询推荐结果多维度多样化探讨研究, 仅局限于将多样化与个性化融合到查询推荐中, 以此尽可能地为用户罗列初始查询不同主题的查询推荐结果外, 也将会为用户返回个性化查询推荐结果。后续相关研究应尝试从多种不同维度进行推荐结果多样化, 比如, 目前检索结果多样化研究中存在着如何将多样化与时间维度融合的热点研究, 而此类研究在查询推荐领域仍未得以探讨。而对于时态模糊性查询(temporal ambiguous query)如查询“CIKM”包含了多个潜在时间意图, 此类查询的理想化多样化推荐结果应该是来源于不同时间段的子主题, 如与“CIKM 2018”、“CIKM2017”、“CIKM2016”等相关的子主题。为实现对此类查询结果信息多样化, 首先需探讨如何借助相应外部知识资源获得查询的潜在时间段信息, 再在此基础上, 探讨如何将时间分面和主题分面融合到隐式或显式多样化查询推荐模型中。

(3) 查询推荐结果的可解释性研究。目前, 结果可解释性已经成为推荐系统与文本自动分类研究领域的热点之一, 如在推荐系统应用实践中, 向用户推荐某商品时, 给出如“推荐A因为您曾经购买了与A类似的某产品”的解释; 在文本分类领域, 将某文本归为A类时, 给出如“将此文本归为A类主要是因为它具有XX特征”。已有查询推荐研究忽略了查询推荐结果的可解释性, 直接影响到用户对推荐结果的满意度与信任度。因此, 未来研究可考虑如何将推荐系统中的推荐可解释性方法应用于查询推荐结果的解释研究, 提高用户对推荐结果的置信度: 如基于协同过滤推荐的解释; 或者利用主题模型、深度学习以及矩阵分解, 从而将主题与矩阵分解中的隐变量进行映射和解释、借助知识图谱连接用户的历史记录和推荐结果等。

(4) 查询推荐的性能评测研究。虽当前已有研究从查询点击文档或者相关文档集中选取特征信息来判断查询是否满足用户意图, 以此对查询推荐性能进行评价。但实际应用中, 用户更在关注所生成

的候选查询推荐能否为其返回满意的查询结果, 故可考虑候选查询推荐的搜索结果来对查询推荐进行评价, 如可通过分析用户针对不同候选查询推荐所返回结果的搜索结果页面之间的互动(如利用眼动仪视线跟踪技术对结果页面的停留时间以及眼睛关注位置), 或者利用动态贝叶斯网络、学习型排序等模型预测查询可能点击的文档对其性能进行评价。除以上方面外, 为建立健全的查询推荐评价体系, 还需从借鉴信息检索等相关领域评价指标和发展适应需求跟踪查询推荐领域自身需要的评价指标两个方面同时开展研究, 从而使得构建的评价体系不仅能评估当前方法的优劣, 而且还能为查询推荐方法的改进指出方向。

(5) 构建更为真实的实验环境。查询推荐研究的实验环境与企业的应用环境之间存在着一定差异, 从而导致学界中探讨的模型难以被应用于企业实践中。例如, 学界对查询推荐研究所采用的查询日志数据集大多是从真实查询日志数据集中抽取的部分数据, 是静态的且非大量的数据集, 而真实的搜索日志数据是一种大数据(Big Data), 具有海量、复杂以及变化速度快等特征。在两种不同环境下, 数据处理、模型参数调整以及算法执行时间等因素会存在着差异, 故在实验环境中能取得较好实验效果的模型, 未必能在企业实验中也取得好的结果。因此, 后续研究将会进一步与企业建立合作, 构建与企业实际应用相符的实验环境, 以便提升学界中所探讨的查询推荐模型的实践价值。

参 考 文 献

- [1] 李亚楠, 王斌, 李锦涛. 搜索引擎查询推荐技术综述[J]. 中文信息学报, 2010, 24(6): 75-84.
- [2] 廖振. 基于查询点击核心图的查询推荐问题研究[D]. 天津: 南开大学, 2013.
- [3] Cai F, de Rijke M. A survey of query auto completion in information retrieval[J]. Foundations and Trends® in Information Retrieval, 2016, 10(4): 273-363.
- [4] Deerwester S. Indexing by latent semantic indexing[J]. Journal of the American Society for Information Science, 1990, 41(6): 391-407.
- [5] Jing Y, Croft W B. An association thesaurus for information retrieval[C]// Proceedings of the Conference on Intelligent Text and Image Handling. New York: ACM Press, 1994: 146-160.
- [6] Xu J, Croft W B. Query expansion using local and global docu-

- ment analysis[C]// Proceedings of the 19th ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 1996: 4-11.
- [7] Plansangket S. New weighting schemes for document ranking and ranked query suggestion[M]. University of Essex, 2017.
- [8] Nogueira R, Cho K. Task-oriented query reformulation with reinforcement learning[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2017: 574-583.
- [9] Singh V, Garg S, Kaur P. Efficient algorithm for web search query reformulation using genetic algorithm[C]// Proceedings of the Conference on Computational Intelligence in Data Mining—Volume 1, Advances in Intelligent Systems and Computing. New Delhi: Springer, 2016, 410: 459-470.
- [10] Jones R, Rey B, Madani O, et al. Generating query substitutions [C]// Proceedings of the 15th International Conference on World Wide Web. New York: ACM Press, 2006: 387-396.
- [11] Shi X, Yang C C. Mining related queries from Web search engine query logs using an improved association rule mining model[J]. Journal of the American Society for Information Science and Technology, 2007, 58(12): 1871-1883.
- [12] Fonseca B M, Golgher P B, De Moura E S, et al. Discovering search engine related queries using association rules[J]. Journal of Web Engineering, 2003, 2(4): 215-227.
- [13] Huang C K, Chien L F, Oyang Y J. Relevant term suggestion in interactive web search based on contextual information in query session logs[J]. Journal of the Association for Information Science and Technology, 2003, 54(7): 638-649.
- [14] Boldi P, Bonchi F, Castillo C, et al. The query-flow graph: model and applications[C]// Proceedings of the 17th ACM Conference on Information and Knowledge Management. New York: ACM Press, 2008: 609-618.
- [15] Boldi P, Bonchi F, Castillo C, et al. Query suggestions using query-flow graphs[C]// Proceedings of the 2009 Workshop on Web Search Click Data. New York: ACM Press, 2009: 56-63.
- [16] Wang X, Zhai C X. Mining term association patterns from search logs for effective query reformulation[C]// Proceedings of the 17th ACM Conference on Information and Knowledge Management. New York: ACM Press, 2008: 479-488.
- [17] Szpektor I, Gionis A, Maarek Y. Improving recommendation for long-tail queries via templates[C]// Proceedings of the 20th International Conference on World Wide Web. New York: ACM Press, 2011: 47-56.
- [18] Anagnostopoulos A, Becchetti L, Castillo C, et al. An optimization framework for query recommendation[C]// Proceedings of the Third ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2010: 161-170.
- [19] Baraglia R, Nardini F M, Castillo C, et al. The effects of time on query flow graph-based models for query suggestion[C]// Proceedings of Adaptivity, Personalization and Fusion of Heterogeneous Information. New York: ACM Press, 2010: 182-189.
- [20] Baraglia R, Castillo C, Donato D, et al. Aging effects on query flow graphs for query suggestion[C]// Proceedings of the 18th ACM Conference on Information and Knowledge Management. New York: ACM Press, 2009: 1947-1950.
- [21] 李亚楠, 许晟, 王斌. 基于加权 SimRank 的中文查询推荐研究[J]. 中文信息学报, 2010, 24(3): 3-10.
- [22] 朱小飞, 郭嘉丰, 程学旗, 等. 基于吸收态随机行走的两阶段效用性查询推荐方法[J]. 计算机研究与发展, 2013, 50(12): 2603-2611.
- [23] 李竞飞, 商振国, 张鹏, 等. 融合用户实时搜索状态的自适应查询推荐模型[J]. 计算机科学与探索, 2016, 10(9): 1290-1298.
- [24] 罗成, 刘奕群, 张敏, 等. 基于用户意图识别的查询推荐研究[J]. 中文信息学报, 2014, 28(1): 64-72.
- [25] Mei Q Z, Zhou D Y, Church K. Query suggestion using hitting time[C]// Proceedings of 17th ACM Conference on Information and Knowledge Management. New York: ACM Press, 2008: 469-478.
- [26] Craswell N, Szummer M. Random walks on the click graph[C]// Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2007: 23-27.
- [27] Ma H, Yang H X, King I, et al. Learning latent semantic relations from click through data for query suggestion[C]// Proceedings of the 17th ACM Conference on Information and Knowledge Management. New York: ACM Press, 2008: 709-718.
- [28] Liu Y, Song R H, Chen Y, et al. Adaptive query suggestion for difficult queries[C]// Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2012: 15-24.
- [29] Li L, Yang Z L, Liu L, et al. Query-URL bipartite based approach to personalized query recommendation[C]// Proceedings of the 23rd National Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2008: 1189-1194.
- [30] Deng H B, King I, Lyu M R. Entropy-biased models for query representation on the click graph[C]// Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2009: 339-346.
- [31] Beeferman D, Berger A. Agglomerative clustering of a search engine query log[C]// Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2000: 407-416.

- [32] Baeza-Yates R, Hurtado C, Mendoza M. Query recommendation using query logs in search engines[C]// Proceedings of International Conference on Extending Database Technology. Heidelberg: Springer, 2004: 588-596.
- [33] 吴家丽. 基于用户意图识别的查询重构研究[D]. 哈尔滨: 哈尔滨工程大学, 2015.
- [34] Jiang D, Leung K W T, Vosecky J, et al. Personalized Query Suggestion with Diversity Awareness[C]// Proceedings of the IEEE 30th International Conference on Data Engineering. IEEE, 2014: 400-411.
- [35] 张乃洲. 基于时间点击图挖掘的查询建议方法[J]. 计算机工程, 2015, 41(5): 191-196.
- [36] Song Y, He L W. Optimal rare query suggestion with implicit user feedback[C]// Proceedings of the 19th International Conference on World Wide Web. New York: ACM Press, 2010: 901-910.
- [37] Sejal D, Shailesh K G, Tejaswi V, et al. Query click and text similarity graph for query suggestions[M]// Machine Learning and Data Mining in Pattern Recognition. Heidelberg: Springer, 2015: 328-341.
- [38] Ye F Y, Sun J. Combining query ambiguity and query-URL strength for log-based query suggestion[C]// Proceedings of International Conference on Swarm Intelligence. Heidelberg: Springer, 2016: 590-597.
- [39] Cao H H, Jiang D X, Pei J, et al. Context-aware query suggestion by mining click-through and session data[C]// Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2008: 875-883.
- [40] Sordoni A, Bengio Y, Vahabi H, et al. A hierarchical recurrent encoder-decoder for generative context-aware query suggestion[C]// Proceedings of the 24th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2015: 553-562.
- [41] Dehghani M, Rothe S, Alfonseca E, et al. Learning to attend, copy, and generate for session-based query suggestion[C]// Proceedings of the 26th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2017: 1747-1756.
- [42] Jiang J Y, Wang W. RIN: Reformulation Inference Network for context-aware query suggestion[C]// Proceedings of the 27th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2018: 197-206.
- [43] Liu Y Q, Miao J W, Zhang M, et al. How do users describe their information need: Query recommendation based on snippet click model[J]. Expert Systems with Applications, 2011, 38(11): 13847-13856.
- [44] 石雁, 李朝锋. 基于朴素贝叶斯点击预测的查询推荐方法[J]. 计算机应用与软件, 2016, 33(10): 19-23.
- [45] Guo J F, Zhu X F, Lan Y Y, et al. Modeling users' search sessions for high utility query recommendation[J]. Information Retrieval Journal, 2017, 20(1): 4-24.
- [46] Qi S Y, Wu D M, Mamoulis N. Location aware keyword query suggestion based on document proximity[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(1): 82-97.
- [47] Lucchese C, Orlando S, Perego R, et al. Identifying task-based sessions in search engine query logs[C]// Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2011: 277-286.
- [48] Liao Z, Song Y, He L W, et al. Evaluating the effectiveness of search task trails[C]// Proceedings of the 21st International Conference on World Wide Web. New York: ACM Press, 2012: 489-498.
- [49] Feild H, Allan J. Task-aware query recommendation[C]// Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2013: 83-92.
- [50] Ozertem U, Chapelle O, Donmez P, et al. Learning to suggest: a machine learning framework for ranking query suggestions[C]// Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2012: 25-34.
- [51] Goel S, Broder A, Gabrilovich E, et al. Anatomy of the long tail: ordinary people with extraordinary tastes[C]// Proceedings of the Third ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2010: 201-210.
- [52] Santos R L T, Macdonald C, Ounis I. Learning to rank query suggestions for adhoc and diversity search[J]. Information Retrieval, 2013, 16(4): 429-451.
- [53] Garigliotti D, Balog K. Generating query suggestions to support task-based search[C]// Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2017: 1153-1156.
- [54] Liu J W, Li Q S, Lin Y S, et al. A query suggestion method based on random walk and topic concepts[C]// Proceedings of IEEE/ACIS 16th International Conference on Computer and Information Science. IEEE, 2017: 251-256.
- [55] Huang Z P, Cautis B, Cheng R, et al. KB-enabled query recommendation for long-tail queries[C]// Proceedings of the 25th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2016: 2107-2112.
- [56] Huang Z P, Cautis B, Cheng R, et al. Entity-based query recommendation for long-tail Queries[J]. ACM Transactions on Knowledge Discovery from Data, 2018, 12(6): Article No. 64.

- [57] Bonchi F, Perego R, Silvestri F, et al. Recommendations for the long tail by term-query graph[C]// Proceedings of the 20th International Conference Companion on World Wide Web. New York: ACM Press, 2011: 15-16.
- [58] 白露, 郭嘉丰, 曹雷, 等. 基于查询意图的长尾查询推荐[J]. 计算机学报, 2013, 36(3): 636-642.
- [59] 刘钰锋, 李仁发. 基于 Term-Query-URL 异构信息网络的查询推荐[J]. 湖南大学学报(自然科学版), 2014, 41(5): 106-112.
- [60] Chen Y, Zhang Y Q. A personalised query suggestion agent based on query-concept bipartite graphs and Concept Relation Trees[J]. International Journal of Advanced Intelligence Paradigms, 2009, 1(4): 398-417.
- [61] Bing L D, Lam W, Wong T L, et al. Web query reformulation via joint modeling of latent topic dependency and term context[J]. ACM Transactions on Information Systems, 2015, 33(2): Article No. 6.
- [62] Chen W Y, Cai F, Chen H H, et al. Personalized query suggestion diversification[C]// Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2017: 817-820.
- [63] 张晓娟. 利用嵌入方法实现个性化查询重构[J]. 情报学报, 2018, 37(6): 621-630.
- [64] Chen W Y, Hao Z P, Shao T H, et al. Personalized query suggestion based on user behavior[J]. International Journal of Modern Physics C, 2018, 29(4): 1850036.
- [65] 王卫国, 徐炜民. 基于潜在语义分析的个性化查询扩展模型[J]. 计算机工程, 2010, 36(21): 43-45.
- [66] 石雁, 李朝锋. 基于协同相似计算的查询推荐[J]. 计算机工程, 2016, 42(8): 188-193.
- [67] 孙达明, 张斌, 张书波, 等. 面向差异化搜索背景的查询推荐方法[J]. 计算机工程, 2016, 42(11): 202-206.
- [68] Dou Z C, Song R H, Wen J R. A large-scale evaluation and analysis of personalized search strategies[C]// Proceedings of the 16th International Conference on World Wide Web. New York: ACM Press, 2007: 581-590.
- [69] Cai F, de Rijke M. Selectively personalizing query auto-completion[C]// Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2016: 993-996.
- [70] Ma H, Lyu M R, King I. Diversifying query suggestion results [C]// Proceedings of the 24th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2010: 1399-1404.
- [71] Bordino I, Castillo C, Donato D, et al. Query similarity by projecting the query-flow graph[C]// Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2010: 515-522.
- [72] Song Y, Zhou D Y, He L W. Post-ranking query suggestion by diversifying search results[C]// Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2011: 815-824.
- [73] Zhu X F, Guo J F, Cheng X Q, et al. A unified framework for recommending diverse and relevant queries[C]// Proceedings of the 20th International Conference on World Wide Web. New York: ACM Press, 2011: 37-46.
- [74] Hu H, Zhang M X, He Z Y, et al. Diversifying query suggestions by using topics from Wikipedia[C]// Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies. Washington, DC: IEEE Computer Society, 2013: 139-146.
- [75] Ding H, Zhang S, Garigliotti D, et al. Generating high-quality query suggestion candidates for task-based search[C]// Proceedings of the 40th European Conference on Information Retrieval. Heidelberg: Springer, 2018: 625-631.
- [76] Kim Y, Croft W B. Diversifying query suggestions based on query documents[C]// Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2014: 891-894.
- [77] Zheng H T, Zhao J, Zhang Y C, et al. An ontology-based approach to query suggestion diversification[C]// Proceedings of the International Conference on Neural Information Processing. Heidelberg: Springer, 2014: 437-444.
- [78] 任鹏杰, 陈竹敏, 马军, 等. 一种综合语义和时效性意图的检索结果多样化方法[J]. 计算机学报, 2015, 38(10): 2076-2091.
- [79] Gupta D, Berberich K. Diversifying search results using time[C]// Proceedings of the European Conference on Information Retrieval. Heidelberg: Springer, 2016: 789-795.
- [80] Nguyen T N, Kanhabua N. Leveraging dynamic query subtopics for time-aware search result diversification[C]// Proceedings of the European Conference on Information Retrieval. Heidelberg: Springer, 2014: 222-234.
- [81] Zhang X J, Peng L. Time-aware diversified query suggestion[C]// Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries. New York: ACM Press, 2018: 399-400.
- [82] 朱小飞, 郭嘉丰, 程学旗, 等. 基于流形排序的查询推荐方法[J]. 中文信息学报, 2011, 25(2): 38-44.
- [83] Jain A, Ozertem U, Velipasaoglu E. Synthesizing high utility suggestions for rare web search queries[C]// Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2011: 805-814.
- [84] Bhatia S, Majumdar D, Mitra P. Query suggestions in the absence of query logs[C]// Proceedings of the 34th International ACM SI-

- GIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2011: 795-804.
- [85] Ma Z R, Chen Y, Song R H, et al. New assessment criteria for query suggestion[C]// Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2012: 1109-1110.
- [86] Yan X H, Guo J F, Cheng X Q. Context-aware query recommendation by learning high-order relation in query logs[C]// Proceedings of the 20th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2011: 2073-2076.
- [87] Song Y, Zhou D Y, He L W. Query suggestion by constructing term-transition graphs[C]// Proceedings of the Fifth ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2012: 353-362.
- [88] Chen W Y, Cai F, Chen H H, et al. Attention-based hierarchical neural query suggestion[C]// Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2018: 1093-1096.
- [89] Nallapati R, Shah C. Evaluating the quality of query refinement suggestions in information retrieval[EB/OL]. [2018-09-01]. <http://maroo.cs.umass.edu/getpdf.php?id=663>.
- [90] Miyanishi T, Sakai T. Time-aware structured query suggestion [C]// Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2013: 809-812.

(责任编辑 车 尧)