

面向用户任务的查询推荐研究*

张晓娟 唐祥彬

(武汉大学信息资源研究中心 武汉 430072)

摘要:【目的】基于 AOL 查询日志数据集,从 Session 级别实现面向用户任务的查询推荐。【方法】从用户任务级别衡量查询间关系,再通过随机游走遍历图的思想为查询构建向量,以此实现候选查询推荐。【结果】本文方法的推荐效果优于基于查询共现来衡量查询间关系的推荐效果。【局限】未对拼写错误的候选查询进行拼写纠错;未从查询级别来实现面向用户任务的查询推荐;稀有查询和模糊性查询的推荐效果不佳。【结论】基于用户任务来衡量查询之间相关关系,能提高查询推荐的实验效果。

关键词: 查询推荐 用户任务 查询日志

分类号: G353.4

1 引言

伴随着网络资源的不断增长,用户的检索兴趣已从与查询相关的文档过渡到与用户信息需求相关的文档。而在大多数情况下,用户所提交的有限关键词常常不能正确地表达其信息需求,很难构造成功的查询,对搜索引擎来说,理解用户信息需求是一项比较困难的任務。鉴于此,搜索引擎服务提供商如 Google 等努力尝试采用多种方法去探测用户的信息需求,并将生成查询推荐作为其重要的环节。其中,基于查询日志方法是当前查询推荐的主流方法,其主要思想为,从 Session 级别或者查询级别分析候选查询与原查询的相关关系,利用相应推荐方法生成查询推荐,则查询间相关关系的计算方法在很大程度上影响着最终的查询推荐效果。当前相关研究大多基于查询共现信息来衡量查询之间关系,使得最终的推荐结果是与原查询相关而并非是与用户信息需求相关的查询,如某用户输入查询“苹果手机”想获得苹果品牌的相关电子产品,其候选查询推荐“苹果沙拉”,偏离了用户初始的信息需求。

用户提交查询的主要目的是为了完成特定搜索任

务,用户任务是其从事搜索行为并且判断查询结果相关性的的重要因素,是用户信息需求的原子,理解用户任务是理解用户信息需求的主要途径。本文拟从基于用户任务级别衡量查询之间相关关系,从而实现面向用户任务的查询推荐。

2 相关研究综述

查询推荐的主流方法是基于日志方法,该方法可分为以下两类子方法:直接的查询共现挖掘与查询构图方法。其中,基于查询共现挖掘的主要思想为,若某查询与原查询共现于同一 Session 的概率越大,该查询是原查询的候选查询推荐的概率越大。相关研究有: Huang 等^[1]提出利用 Session 中查询共现信息来构建原查询的上下文,以此实现推荐,且实验结果表明该方法从多方面优于基于文档的查询推荐方法; Shi 等^[2]与 Fonseca 等^[3]提出一种基于关联规则的模型来挖掘 Session 中的相关查询; Jones 等^[4]根据 Session 中查询间的共现信息,利用互信息度量查询间相似性,以此生成候选查询。基于查询共现的推荐方法虽能对共现查询进行较好推荐,但不能对一些具有间接联系的查询实现有效推荐,为了弥补此缺陷,学者

收稿日期: 2013-12-17

收修改稿日期: 2014-01-27

*本文系武汉大学 2012 年博士生自主科研项目“网络检索用户查询意图分析与建模研究”(项目编号: 2012104010201)的研究成果之一。

们尝试利用构图方法来实现查询推荐,其相关研究有: Mei 等^[5]基于查询的点击频次来构建二元图,利用随机游走来遍历图,再以 Hitting Time 方法实现查询推荐; Boldi 等^[6]构建了 query-flow 图模型,利用一些特征来计算查询之间的转移概率,以此为候选查询排序; Song 等^[7]利用随机游走来实现查询推荐的两个主要特点: 调节随机游走中的跳转参数以及在随机游走时将用户点击和跳转的 URL 结合起来; Song 等^[8]利用用户修改查询行为如添加词、修改词、删除词来构建图,并分别从主题和词级别来实现查询推荐; 李亚楠等^[9]从 Session 中挖掘查询之间的间接联系建立查询关系图,并在图结构相似算法 SimRank 的基础上提出了加权 SimRank 用于查询推荐; Cao 等^[10]基于查询日志,先将用户查询转化为一组概念序列,并构建了这些概念序列的后缀树,以此来作为原查询上下文,最后通过在后缀树中原查询的相关概念序列来实现查询推荐; Cao 等^[11]通过在查询日志中训练可变长度隐马尔科夫模型(vIHMM)实现查询推荐,该模型主要考虑到查询以及相关点击信息,容易推荐偏离原查询任务的查询,且计算复杂度较大,适用性有一定局限。从以上两类方法的相关研究可看出,大多数研究者通过查询之间共现信息来建立查询之间关系,而很少从用户任务层级来衡量查询间关系。

其中,面向用户任务的查询研究主要体现在基于用户任务级别的查询日志切分研究中。早期的查询日志切分方法主要是基于时间间隔^[12,13]进行的,但考虑到切分后的同一 Session 中存在着多种查询任务,一些学者尝试基于用户任务级别对查询日志分析: 如在 Jansen 等^[14]时间划分 Session 的基础上,一些学者尝试通过添加其他表征用户任务的特征(如查询重构模式)来提高 Session 边界识别效果; Jones 等^[15]基于时间戳和查询词等特征从 Session 中抽取用户任务; Lucchese 等^[16]利用 Wikipedia 和 Wiki Dictionary 中的信息进一步提高用户任务识别的效果。虽当前有大量研究基于用户任务级别来探讨查询日志切分,但是几乎没有从用户任务级别来探讨查询推荐的研究。鉴于此,本文首先从用户任务级别来衡量查询之间相关关系,在此基础上,通过构图模型来实现面向用户任务的查询推荐。

3 衡量查询间关系的模型

3.1 衡量查询间关系的一般方法

查询推荐方法的一般思路为: 通过相关模型来计算候选查询与原查询的关系权值,在此基础上,利用相关推荐方法生成候选查询推荐。衡量查询间相关关系的方法是查询推荐中的重要部分。计算原查询与候选查询关系权值的一般方法,如下所示:

$$r(q', q) = \sum_{(\lambda, \theta)} \lambda \cdot \theta(q', q) \quad (1)$$

其中, $r(q', q)$ 表示查询 q' 与原查询 q 之间的关系值。 $\theta(q', q)$ 表示利用函数 θ 计算查询 q' 与原查询 q 之间的相关关系, λ 表示函数 θ 的权值。 $\sum_{(\lambda, \theta)} \lambda \cdot \theta(q', q)$

表示若有多个函数用于衡量 q' 与 q 之间关系时,则各函数之间需进行线性加权,各权值 λ 之和为 1。在一般查询间关系模型中, Session 中共现互信息^[4]是计算 $\theta(q', q)$ 权值的常用方法。

3.2 基于用户任务的查询间关系

利用查询共现信息来衡量查询间关系的研究大多基于如下假设: 同一 Session 中的查询描述相同的用户任务。而一些研究者发现同一 Session 的查询可能表达不同的用户任务: 如 Lucchese 等^[16]通过对 AOL 查询日志分析发现,约 74% 的查询位于多任务 Session 中; Liao 等^[17]通过分析一商业搜索引擎的查询日志发现,约 30% 的 Session 中包含多任务; Jones 等^[15]通过对 Yahoo 搜索引擎日志分析发现,约 17% 的 Session 中存在着用户任务交叉现象。于是,基于共现信息生成的候选查询推荐容易偏离原查询的用户任务。本文查询推荐的目标是根据用户提交的查询来生成与原查询用户任务相同的候选查询,拟从查询间时间距离、查询表达式与查询之间语义关系来衡量查询间的用户任务相似度。

(1) 查询间时间距离

笔者假设,对于查询 q 与 q' 来说,若两查询之间时间距离越小(即在同一 Session 中,两查询之间距离越近),二者表达同一用户任务的可能性越大。设 $S = \{q_1, q_2, q_3, \dots, q_{t-1}, q, q_{t+1}, \dots\}$ 表示某一 Session 的查询集合,该集合包含原查询 q ,且该集合中查询按照提交时间先后进行排列,其中, q_1 的提交时间早于 q_2 的

提交时间, q_2 的提交时间早于 q_3 的提交时间。由以上假设可知, q_{t+1} 与 q_{t-1} 表示与原查询 q 的相邻查询, 因此, 这两个查询包含的用户任务更可能与查询 q 的用户任务相似。本文利用公式(2)中 $\text{decay}(i, j, S)$ 计算在某 Session S 中原查询 q 与查询 q' 之间的距离权值。

$$\text{decay}(i, j, S) = \beta^{j-i} \quad (2)$$

其中, j 表示查询 q' 在 S 中的排序, i 表示查询 q 在 S 中的排序, β 表示衰减系数^[6], 本文将其值设定为 0.8。

公式(3)表示当 q 与 q' 共现于多个同一 Session 中时, 两查询间的时间距离权值 $T(q, q')$ 为各 Session 中距离权值 $\text{decay}(i, j, S_k)$ 的平均值。 $T(q, q')$ 值越大, 则两查询描述同一用户任务的可能性越大。

$$T(q, q') = \frac{\sum_{k=1}^n \beta^{j-i}}{n} \quad (3)$$

(2) 查询间表达式相似性

在大多数情况下, 用户常用表达式相似的查询来描述同一查询任务^[16]。本文假设, 若两查询间共有查询词越多, 二者描述同一用户任务的可能性越大。考虑到查询中某些词拼写错误以及词的前缀后缀现象会影响到这种相似性的计算, 本文综合利用文献[16]提出的计算字符串之间相似度的 Di-gram Jaccard 距离以及标准化编辑距离来计算两查询之间相似性, 如下所示:

$$u_{\text{expression}}(q, q') = e^{-\frac{(u_{\text{jac}} + u_{\text{levenstein}})}{2}} \quad (4)$$

其中, u_{jac} 表示两查询之间的 Di-gram Jaccard 距离值, $u_{\text{levenstein}}$ 表示两查询之间编辑距离值^[16]。因 u_{jac} 值与 $u_{\text{levenstein}}$ 值之和不小于 1, 为使得查询间相似性与查询间距离成反比关系, 且计算而得的查询表达式相似度值小于 1, 笔者利用 $e^{-\frac{(u_{\text{jac}} + u_{\text{levenstein}})}{2}}$ 来衡量查询之间表达式相似性。 $u_{\text{expression}}(q, q')$ 值越大, 表明两查询之间的表达式相似性越大, 描述同一用户任务的可能性越大。

(3) 查询间语义相似性

考虑到表达式不同的两查询也可能包含同一用户任务, 本文也尝试利用查询间语义相似度来衡量查询间的用户任务相似度^[3]。一般来说, 衡量查询之间语义

相似性方法大致分为从大数据集中扩展查询与直接采用语义资源两类方法, 因后一类方法相对前一类方法简单且易实现, 本文尝试借助语义资源 Wikipedia(即 ClueWeb 09 Category B^[18]中的 Wikipedia 数据集)来计算查询间语义相似性。一般情况下, Wikipedia 中的每个网页描述某一概念, 则该网页中所出现的词都与该概念相关。根据文献[3]的思想, 本文将查询词 t 的向量表示为:

$$\vec{C}(t) = (c_1, c_2, \dots, c_w)$$

其中, w 表示数据集中所包含的网页数目。 c_i 表示词 t 在第 i 个网页中的权值, 其值是由 tf-idf 原理计算得出。

公式(5)表示查询 q 的向量为各个查询词(非停用词)向量之和。

$$\vec{C}(q) = \sum_{t \in q} \vec{C}(t) \quad (5)$$

本文利用余弦值来计算两查询向量之间的相似度, 如下所示:

$$s_{\text{semantic}}(q, q') = \frac{\vec{C}(q) \cdot \vec{C}(q')}{|\vec{C}(q)| |\vec{C}(q')|} \quad (6)$$

本文在计算 $r(q', q)$ (即查询 q' 是查询 q 的候选查询概率)时, 利用三个 θ 函数来衡量查询之间相关关系, 即综合考虑查询之间时间距离、查询间表达式相似性以及查询之间语义相似性三方面, 其计算方法分别参见公式(3)、公式(4)与公式(6), 这个三函数的权值分别为 λ_1 、 λ_2 与 λ_3 , 且三者之和为 1。

4 查询推荐方法

已有实验表明^[7-9], 基于构图方法的查询推荐方法能取得较好的实验效果, 而大多数学者仅利用查询之间关系来构建图模型以实现查询推荐。考虑到点击信息也是表达用户任务(或信息需求)的重要特征, 本文通过查询间相关关系以及查询与文档间点击关系来构建图模型, 并在随机游走遍历图的基础上实现查询推荐。其中, 转移概率的设定与查询向量构建是关键部分。

4.1 转移概率设定

首先构建图 $G=(V, E)$, 节点 V 由查询词及其所点击文档构成, E 所包含的边集合有: 查询之间相关关系以及查询与文档之间的点击关系。 本文将每个节点当

作一个马尔科夫状态。查询之间转移概率,如下所示:

$$P(q|q') = (1-\epsilon) \times \frac{r_c(q',q)}{\sum r_c(q'',q)} \quad (7)$$

其中, $P(q|q')$ 表示查询与查询之间的转移概率, $r_c(q',q)$ 表示查询 q' 与查询 q 的相关关系, $\sum r_c(q'',q)$ 表示与 q 共现同一 Session 中的所有不同查询 q'' 与原查询 q 间的 $r_c(q'',q)$ 之和。 ϵ 表示从查询状态转移到文档状态的概率,其参数调节过程参见 5.2 节。

笔者在计算查询与文档之间转移概率时,首先将 URL 加以区别对待,即若一个 URL 被多个查询点击,则该 URL 具有模糊性,反之,表明该 URL 具有专指性。查询转移到专指性 URL 的概率应大于模糊类 URL。本文利用文献[19]提出的 iqf 衡量 URL 对查询的区分度(与文档的 idf 思想相似),公式如下所示:

$$iqf(d_j) = \log |Q| - \log n(d_j) = \log \frac{|Q|}{n(d_j)} \quad (8)$$

$$cfiqf(q_i, d_j) = c_{ij} \times iqf(d_j) \quad (9)$$

其中, $n(d_j)$ 表示点击了文档 d_j 的查询频次之和, $|Q|$ 表示查询日志中总查询数。 $cfiqf$ 表示点击频次与 iqf 的乘积, c_{ij} 表示查询 q_i 点击文档 d_j 的次数。

则文献[19]中 query-url 之间的转移概率公式如下所示:

$$P(d_j | q_i) = \epsilon \times \frac{cfiqf(q_i, d_j)}{\sum_{d_j \in D} cfiqf(q_i, d_j)} \quad (10)$$

其中, $\sum_{d_j \in D} cfiqf(q_i, d_j)$ 表示查询 q_i 所点击文档的 $cfiqf$ 权值之和。

公式(11)表示文档状态是一个吸收状态(即自转移状态),即转移到自身的概率为 1,而转移到查询状态的概率为 0。根据马尔科夫吸收链的性质:当状态转移到文档状态后,则会一直停留在此状态;从每一个非吸收状态出发,有限次转移后能以正的概率到达某个吸收状态。

$$p(d|d) = 1 \quad (11)$$

4.2 查询向量构建

基于如下假设:用户所点击文档能描述用户的信息需求,则点击同一文档的两查询可能具有相似用户任务^[20],本文根据每个查询所点击文档分布来对查询构建向量。以图 1 为例,图中节点由查询 q_1 、 q_2 与 q_3

以及被点击文档节点 d_1 、 d_2 与 d_3 构成。其中,查询与点击文档的实线边表示点击关系,而虚线边表示经过随机游走,查询能到达该文档状态。本文为查询构建向量的特点在于不只是考虑某查询点击的文档,而是查询到文档的整个路径,即经过随机游走该查询可能到达的文档状态,如查询 q_2 虽与文档状态 d_1 与 d_3 无直接点击关系,但经随机游走,能到达这两状态。本文采用 KL 距离来计算向量之间相似度以此来排序与原查询相关的查询,生成候选查询推荐。

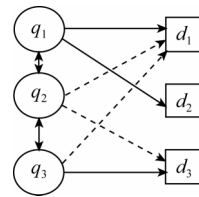


图 1 经随机游走后的查询与文档间关系

5 实验及其结果评测

5.1 实验数据集

本文采用 AOL^[21]查询日志作为数据集,其时间跨度为 2006 年 3 月 1 日到 5 月 31 日,其格式如图 2 所示,从左到右分别表示用户 ID、查询表达式、用户点击时间、被点击 URL 在结果列表中的排序、点击的 URL 地址。因原始数据集中包含许多噪音,本实验首先对其进行清理:如剔除包含色情词查询、只包含网页地址的查询等。笔者通过“15 分钟划分法”^[13]识别 Session 边界。

217	bestasiancompany.com	2006-03-20 15:15:43	1	http://www.bestasiancompany.com
217	lottery	2006-03-27 14:10:38	1	http://www.calottery.com
217	lotte ry	2006-03-27 16:34:59	1	http://www.calottery.com
217	ask.com	2006-03-31 14:31:10	1	http://www.ask.com
.....				

图 2 AOL 数据集格式

5.2 实验结果及其分析

本文随机抽取 200 个不同查询作为原查询。考虑到某些参数会影响到实验的最终效果,参数调节是本文的重要部分。图 3 表示公式(7)中随机游走参数 ϵ 取不同值时,所构建图模型中 90%的查询节点达到吸收状态时所需要的随机游走步数。可知,当 ϵ 取值为 0.6 时,所需的随机游走步数最少,则 ϵ 的最优值为 0.6。

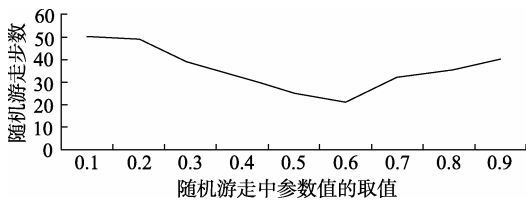


图3 随机游走参数 ϵ 取值与随机游走步数之间关系

在此基础上,本文进行了4组不同的实验:

(1) 利用共现互信息计算公式(1)中 $\theta(q',q)$ 值,并利用文献[6]中构图方法(图模型中只包含查询节点,即只考虑查询间关系)来实现查询推荐(标记为“Baseline1”);

(2) 利用共现互信息计算公式(1)中 $\theta(q',q)$ 值,并利用本文构图方法来实现查询推荐(标记为“Baseline2”);

(3) 基于用户级别计算公式(1)中 $\theta(q',q)$ 值,并利用文献[6]中构图方法实现查询推荐(标记为“Baseline3”);

(4) 基于用户任务级别计算公式(1)中 $\theta(q',q)$ 值,并利用本文中构图方法来实现查询推荐(标记为“Our Work”)。

据笔者所知,目前尚未有公认的查询推荐公共评测语料,人工评测是当前主要方法。本文选取三位专家对实验结果进行评测,对于每个原查询,三位专家分别判断每个结果的相关度,本文中相关度的标注是基于二值判断即“相关”与“不相关”两类。另外利用Cohen kappa(κ)^[22]来获取评测者两两之间同意对方评测结果的概率,平均值为0.68,说明了三位专家之间同意对方评测结果的概率较高。对于少量不一致的标注结果,标注者将根据评价标准协商后给出最终一致性结果。本文采用P@N(Precision at N)以及Recip_Rank指标来对推荐结果进行评测。其中,P@N表示前N个推荐结果中相关查询所占的比例,Recip_Rank表示返回第一个相关查询的能力。

另考虑到“Our Work”与“Baseline3”两实验从用户任务级别计算查询间相似性时,三函数(分别参见公式(3)、公式(4)与公式(6))权值 λ_1 、 λ_2 与 λ_3 的不同取值会影响到的最终效果。在这两组实验中,本文将200个原查询平均分为两份,其中一份的100个原查询作为训练集,获得三参数的最优解;而将另外一部分的100个原查询作为测试集,即在训练集中获得最优参

数解的基础上,获得实验的最终实验效果。其中,“Baseline3”最终实验结果的P@3、P@5、P@10与Recip_Rank值分别为0.47、0.54、0.67与0.68;“Our Work”最终实验结果的P@3、P@5、P@10与Recip_Rank值分别为0.50、0.60、0.72与0.73。

表1 实验结果评测表

相关方法 \ 评测指标	P@3	P@5	P@10	Recip_Rank
Baseline1	0.40	0.52	0.60	0.62
Baseline2	0.42	0.53	0.60	0.63
Baseline3	0.47	0.54	0.67	0.68
Our Work	0.50	0.60	0.72	0.73

表1是4组实验的评测结果。从评测指标值可知,“Baseline3”方法优于“Baseline1”方法,“Our Work”方法优于“Baseline2”方法,从这两组对比数据可以说明,基于用户任务级别衡量查询之间相关关系可进一步提高查询推荐的效果。另也可知,“Baseline2”方法优于“Baseline1”方法,“Our Work”方法优于“Baseline3”方法,从这两组对比数据中可知,本文构图方法优于文献[6]中构图方法(图模型中只包含查询节点,即只考虑查询间关系),即在构建图模型时,考虑查询与文档之间点击关系可进一步提高查询推荐的最终效果。

另考虑到查询的不同属性会影响到最终的查询推荐效果。表2表示将查询按照稀有性和模糊性进行分类后,利用本文推荐方法后的各类别查询的评测结果。其中,查询稀有性判断方法参见文献[23],查询的模糊性判断方法参见文献[24]。

表2 不同属性查询的实验评测结果

评价项	P@3	P@5	P@10	Recip_Rank	
稀有性	稀有查询	0.42	0.49	0.64	0.64
	非稀有查询	0.58	0.71	0.80	0.82
模糊性	模糊性查询	0.43	0.52	0.63	0.67
	非模糊性查询	0.57	0.68	0.81	0.79

可知,非稀有查询的实验效果优于稀有查询的实验效果,非模糊性查询的实验效果优于模糊性查询的实验效果,其主要原因在于:因用户提交稀有查询的频次较少,此类查询中所包含的Session信息较少,相对非稀有查询来说,此类查询较难从Session

中推荐与之相关的候选查询; 又因用户提交模糊性查询时, 自身对该查询的意图不明确, 常常构造与之无关的查询, 相对非模糊性查询来说, 同一 Session 中包含更多噪声查询, 其最终查询推荐的效果较低。所以, 如何提高稀有查询和模糊查询的推荐效果是后续一项重要研究工作。

6 结 语

查询推荐是用户与搜索引擎交互的途径之一, 是当前信息检索领域一研究热点。本文从 Session 级别, 在基于用户任务衡量查询间关系的基础上, 通过构建图模型来实现候选查询的推荐。实验结果表明, 本文的查询推荐效果优于基于查询共现衡量查询间关系的查询推荐方法。尽管如此, 本文仍存在一些不足之处, 在未来工作中还需从以下几个方面加以深入研究: 进一步采用聚类方法来衡量查询之间相似性; 在查询处理过程中, 进一步考虑查询的拼写纠错等; 如何在不借用查询日志的情况下, 基于其他推荐方法来实现面向用户任务的查询推荐; 探讨如何选取更合适的指标来对面向用户任务查询推荐进行有效评测; 为稀有查询和模糊性查询提供更有效的面向用户任务的推荐方法; 基于查询级别, 实现面向用户任务的查询推荐。

参考文献:

- [1] 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 American Society for Information Science and Technology*, 2003, 54(7): 638-649.
- [2] 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.
- [3] Fonseca B M, Golgher P B, de Moura E S, et al. Using Association Rules to Discover Search Engines Related Queries[C]. In: *Proceedings of the 1st Conference on Latin American Web Congress*, 2003.
- [4] Jones R, Rey B, Madani O, et al. Generating Query Substitutions[C]. In: *Proceedings of the 15th International Conference on World Wide Web*. 2006:387-396.
- [5] Mei Q, Zhou D, Church K. Query Suggestion Using Hitting Time[C]. In: *Proceedings of the 17th ACM Conference on Information and Knowledge Management*. 2008: 469-478.
- [6] Boldi P, Bonchi F, Castillo C, et al. The Query-flow Graph: Model and Applications[C]. In: *Proceedings of the 17th ACM Conference on Information and Knowledge Management*. 2008: 609-618.
- [7] Song Y, He L. Optimal Rare Query Suggestion with Implicit User Feedback[C]. In: *Proceedings of the 19th International Conference on World Wide Web*. 2010:901-910.
- [8] Song Y, Zhou D, He L. Query Suggestion by Constructing Term-Transition Graphs[C]. In: *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*. 2012: 353-362.
- [9] 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. (Li Ya'nan, Xu Sheng, Wang Bin. Chinese Query Recommendation by Weighted SimRank[J]. *Journal of Chinese Information Processing*, 2010, 24(3): 3-10.)
- [10] Cao H H, Jiang D X, Pei J, et al. Context-aware Query Suggestion by Mining Click-through and Session Data[C]. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*. 2008:875-883.
- [11] Cao H H, Jiang D X, Pei J, et al. Towards Context-aware Search by Learning a Very Large Variable Length Hidden Markov Model from Search Logs[C]. In: *Proceedings of the 18th International Conference on World Wide Web (WWW'09)*. 2009: 191-200.
- [12] Catledge L, Pitkow J. Characterizing Browsing Behaviors on the World Wide Web[J]. *Computer Networks and ISDN System*, 1995, 27(6): 1065-1073.
- [13] He D Q, Goker A. Detecting Session Boundaries from Web User Logs[C]. In: *Proceedings of the 22nd Annual Colloquium on Information*. 2000.
- [14] Jansen B, Spink A, Kathuria V. Define Searching Sessions on Web Search Engines[J]. *Journal of the American Society for Information Science and Technology*, 2007, 58(6):862-871.
- [15] Jones R, Klinkner K L. Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs[C]. In: *Proceedings of the 17th ACM Conference on Information and Knowledge Management*. 2008.
- [16] Lucchese C, Orlando S, Perego R, et al. Identifying Task-based Sessions in Search Engine Query Logs[C]. In: *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*. 2011: 277-286.
- [17] Liao Z, Song Y, He L W, et al. Evaluating the Effectiveness of Search Task Trails[C]. In: *Proceedings of the 21st*

- International Conference on World Wide Web(WWW'12). 2012: 489-498.
- [18] The ClueWeb09 Dataset [EB/OL]. [2013-10-14]. <http://lemur-project.org/clueweb09/>.
- [19] Deng H B, King I, Lyu M R. Entropy-biased Models for Query Representation on the Click Graph[C]. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2009: 339-346.
- [20] Chan W S, Leung W T, Lee D L. Clustering Search Engine Query Log Containing Noisy Clickthroughs[C]. In: Proceedings of the 2004 International Symposium on Applications and the Internet. 2004: 305-308.
- [21] AOL Search Data[EB/OL]. [2013-10-14]. <http://www.gregsadetsky.com/aol-data/>.
- [22] Berry K J, Mielke P W. A Generalization of Cohen's Kappa Agreement Measure to Interval Measurement and Multiple Raters[J]. Educational and Psychological, 1998, 48(4): 921-933.
- [23] Downey D, Dumais S, Horvitz E. Heads and Tails: Studies of Web Search with Common and Rare Queries[C]. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 2007: 847-848.
- [24] Clough P, Sanderson M, Abouammoh M, et al. Multiple Approaches to Analyzing Query Diversity[C]. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2009: 734-735.

作者贡献声明：

张晓娟：研究命题的提出、实验设计以及论文撰写；

唐祥彬：数据处理与分析。

(通讯作者：张晓娟 E-mail: zhangxiaojuan624@gmail.com)

Query Recommendation Based on User Task

Zhang Xiaojuan Tang Xiangbin

(Center for the Studies of Information Resources, Wuhan University, Wuhan 430072, China)

Abstract: [Objective] This paper tries to realize user task-oriented query suggestion from session level based on AOL query log dataset. [Methods] This paper firstly measures the relationship between queries based on user task, and then realizes user task-oriented query recommendation by exploiting random walk to traversal graph model. [Results] The final results show that our query recommendation method outperforms that method which measures relationship between queries by exploiting queries occurrence information. [Limitations] Misspelled candidate queries are not implemented spell correction; Query recommendation are not realized from query level; The recommendation effect of rare queries and ambiguous queries are not good. [Conclusions] Measuring the relationship between queries based on user task can improve the performance of query recommendation.

Keywords: Query recommendation User task Query log