

基于感知哈希与用户偏好的检索意图建模方法

石宏彬¹ 郭克华^{1,2}

(中南大学信息科学与工程学院 长沙 410083)¹

(南京理工大学高维信息智能感知与系统教育部重点实验室 南京 210094)²

摘 要 针对商品检索排序问题,提出结合用户查询条件与用户浏览兴趣偏好的排序方法,目的是在不增加用户输入查询条件的前提下,提高用户对商品检索结果的满意度。根据用户提交的查询条件,对数据库中的商品进行筛选和初步排序。在此基础上,以用户的浏览行为分析用户对商品的兴趣浓度,并从用户的历史浏览记录中提取出用户的兴趣偏好模型,计算商品属性信息与用户偏好模型之间的相似度大小,对返回的排序结果进行调整优化。实验表明,基于用户兴趣偏好的排序结果更加符合用户的检索意图。

关键词 用户偏好,个性化检索,排序算法,电子商务

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.3.057

Retrieval Intention Modeling Based on Perception Hash Algorithm and Browsing Preferences

SHI Hong-bin¹ GUO Ke-hua^{1,2}

(School of Information Science & Engineering, Central South University, Changsha 410083, China)¹

(Key Laboratory of Intelligent Perception and Systems for High-dimensional Information of Ministry of Education,
Nanjing University of Science and Technology, Nanjing 210094, China)²

Abstract Aiming at the problem of commodity retrieval and ranking, the ranking method based on user's query request and browsing preferences was proposed. The objective is to improve user's satisfaction to the retrieval result without increasing input requests. The commodity records from database were selected and ranked primly according to the query requests. Based on this, user's interest degree was analyzed by their browsing behavior and the user preference model was refined from browsing history. Then the similarity of the attribute information of commodities and preference model was calculated, and the ranking result was adjusted. The experimental result shows that the retrieval result taking user's preference into consideration is more fit to their retrieval intention.

Keywords User preference, Personalized retrieval, Ranking algorithm, E-commerce

1 引言

随着互联网和电子商务的迅猛发展,每天都有数以万计的商品交易在电商网站上完成。以淘宝、京东等为代表的各大电商平台的商品种类、数量呈爆炸式增长。在纷繁多样的商品信息中,如何为用户提供一种检索服务^[1],使其能够快速准确地检索到自己需要的商品,已经成为电子商务检索研究的一个重要课题。

传统的电子商务检索系统采用关键字匹配算法来实现检索功能,得到与用户输入的关键词相关的商品结果^[2]。但由于关键词不能完全反映用户的检索需求,比如存在歧义、多义现象,并且用户习惯输入简短而不完整的关键词,因此检索结果往往与用户真正的检索意图存在偏差。利用查询扩展技术虽然在一定程度上能够缓解这种矛盾^[3,4],但通常这种方式是以查询词为中心在符号匹配层次上进行的扩展,而忽略了

查询语义的扩展,因此不能从根本上消除用户检索意图和检索结果之间的语义偏差和歧义性问题。

基于商品图像内容^[5]的检索虽然能够通过图像来查找视觉上相似的商品结果,但是图像的视觉特征与它所包含的语义信息之间存在“语义鸿沟”问题^[6],这样导致得到的检索结果中可能存在外形上相似、语义内容上却与用户检索意图不相关的商品检索结果。

目前电子商务系统的检索结果通常以与用户查询条件相似度的高低作为排序依据。对于电子商务网站的各个用户,由于其具有不同的兴趣爱好,即使输入相同的检索请求,其检索意图也可能不同^[7]。因此结合用户兴趣偏好,把检索结果进行适当个性化排序调整^[8],对提高检索质量具有重要意义。

为提高用户对检索排序结果的满意度,本文设计了一个商品检索排序系统,其能够记录用户在网页上的浏览行为,并从中发现用户感兴趣的物品;从商品属性的层面上建立用户

到稿日期:2015-03-23 返修日期:2015-06-15 本文受国家自然科学基金(61202341),高维信息智能感知与系统教育部重点实验室创新基金(JYB201502),科技部国家国际科技合作专项项目(2013DFB10070),湖南省创新平台专项(2012GK4106)资助。

石宏彬(1989—),男,硕士生,主要研究方向为多媒体检索,E-mail:313768627@qq.com;郭克华(1980—),男,副教授,主要研究方向为多媒体检索、普适计算(通信作者)。

的兴趣偏好模型,与用户真实的检索意图更加契合,在不增加用户输入检索条件的情况下,返回更加符合用户检索意图的个性化商品检索排序结果。

本文第2节介绍相关工作;第3节给出基于用户兴趣偏好检索的一般框架;第4节论述检索系统的实现方法;第5节为实验和检索精度的比较;最后对全文进行总结和讨论。

2 相关工作

个性化的检索排序工作主要集中在两部分,一是用户兴趣偏好的收集,二是通过收集的偏好信息建立合适的用户兴趣偏好模型,对检索排序结果进行调整。用户兴趣偏好的收集方式主要分为显式和隐式两种^[9]。现实中,由于显式收集方式存在许多不足,例如需要用户投入较多精力反馈偏好信息,以及用户更倾向于不登录就直接进行搜索或浏览,因此隐式收集方式成为了收集用户兴趣信息研究的主要方向^[10,11]。

隐式方式通过用户在页面上的鼠标点击、拖动滚动条、标记书签、停留查看等行为确定用户的兴趣浓度。对于不同类型的内容,用户的兴趣特征行为表现会不一样。针对新闻内容,文献[11]认为若页面包含用户感兴趣的内容越多,用户查看该页面花费的时间越长,因此以浏览时长作为用户的兴趣指标。文献[12]通过分析用户在地图上的鼠标点击次数以及光标停留时间来确定用户感兴趣的空间节点。文献[13]指出,一般情况下,用户的兴趣指标集中表现在页面的停留时间以及滚动条的拖动次数上。

文献[14]假定用户的具体兴趣已知,根据 Open Directory Project(ODP)中给出的类别参考信息对用户兴趣进行分类,比如把“足球、篮球”分为“球类运动”这一类,以兴趣类别建立用户的兴趣模型。这种方法需要用户花费较多精力反馈兴趣信息,限制了用户的使用体验。文献[15]依据用户使用的检索词、给出的相关评价等对用户进行分组,然后把组内的用户模型进行合并从而达到扩展用户兴趣模型的目的。文献[16]采用联合聚类的方法从用户好友的常用标签中提取信息,建立用户的标签兴趣模型。上面两种方法都存在冷启动问题,在系统启用初期个性化排序效果较差。文献[17]根据内容对用户感兴趣的网页进行归类,建立兴趣模型,再利用领域知识对模型进行扩展得到增强型的兴趣模型,此方法对某些特定领域比较有效,但不具备广泛适用性。

3 基于检索意图建模排序的一般框架

个性化的检索排序需要收集用户的浏览行为并建立兴趣模型,在以用户检索条件所得到的商品检索结果的基础上,对返回的结果进行个性化的排序调整,使得更加符合用户检索意图的商品排在前面。

本文用 G 表示一件商品, Q 表示用户输入的检索条件, P 表示用户的兴趣偏好。令 $f(G, Q, P)$ 为一个实值的检索排序度的计算函数,目的是当我们利用 $f(G, Q, P)$ 的值对商品结果进行排序时,使得与 P, Q 相关的商品结果能够排列在无关结果前面,从而实现符合用户检索意图的个性化检索。

为实现个性化的检索排序,需要考虑用户输入的检索条件以及用户的兴趣偏好,因此,将检索排序度的计算函数 $f(G, Q, P)$ 定义为:

$$f(G, Q, P) = Sim(G, Q) \times Sim(G, P) \quad (1)$$

其中, $Sim(G, Q)$ 表示商品与用户检索条件之间的相似度大小, $Sim(G, P)$ 表示商品与用户的兴趣偏好之间的相似度大小。

4 基于检索意图建模排序的实现方案

对商品与检索条件之间的相似度 $Sim(G, Q)$ 和商品与用户兴趣偏好之间的相似度 $Sim(G, P)$ 分别给出以下计算方法。尽管两种相似度的计算方法在前人的工作中有过相关使用,但目前尚未有工作在结合商品属性的层面上对商品检索结果进行排序研究。

4.1 商品与检索条件之间相似度的计算方法

检索请求分为示例图像和关键词这两种形式,用户可以选择其中一种或者两种同时上传。因此需要分别计算检索请求的示例图像与商品图像之间的相似度大小 $SimPic(G, Q)$, 以及检索请求的关键词与商品描述文本之间的相似度大小 $SimWord(G, Q)$ 。最终的商品与用户检索条件相似度的大小为:

$$Sim(G, Q) = SimPic(G, Q) \times SimWord(G, Q) \quad (2)$$

4.1.1 计算图像之间相似度的方法

为了方便图像特征的存储与快速计算,本文利用感知哈希算法来提取图像的全局灰度特征,将得到的图像感知哈希值 $Code$ 作为索引存放到数据库,通过两张图像感知哈希值之间的汉明距离 $d(Code1, Code2)$ 来计算它们的相似度 $SimPic(G, Q)$ 。

任意数字图像 I 都是由若干个像素点 Pix 所构成,按照从左至右、从上至下的顺序来统计图像的像素点,可以用像素点集合 $[Pix_1, Pix_2, \dots, Pix_N]$ 来表示 I , 即 $I = [Pix_1, Pix_2, \dots, Pix_N]$ 。对于非灰度图像,其各个像素点 Pix 由 RGB 三原色构成,即 $Pix = (R, G, B)$, 用式(3)可以计算出图像中每个像素点的灰度值;而对于灰度图像,可直接统计各个像素点的灰度值。

$$Gray = R \times 0.3 + G \times 0.59 + B \times 0.11 \quad (3)$$

通过式(3)能够将非灰度图像转换为灰度图像,因此对于所有的数字图像,都能用像素点的灰度值来表示该图像,即 $I = [Gray_1, Gray_2, \dots, Gray_N]$ 。提取图像全局灰度特征值的具体方法如算法1所示。

算法1 提取图像特征的感知哈希算法

输入: 图像 $I = [Pix_1, Pix_2, \dots, Pix_N]$

输出: 图像感知哈希值 $Code$

1. I 缩小为 $8 * 8$ 的尺寸大小的 $I' = [Pix_1, \dots, Pix_{64}]$
2. FOR $i = 1 \rightarrow 64$
3. IF I 为非灰度图像
4. 利用式(3)计算 Pix_i 对应的灰度值 $Gray_i$
5. ELSE 直接统计 Pix_i 的灰度值 $Gray_i$
6. $GrayAll += Gray_i$
7. END FOR
8. $average_Gray \leftarrow GrayAll / 64$
9. FOR $i = 1 \rightarrow 64$
10. IF $Gray_i > average_Gray$
11. $V_i = 1;$
12. ELSE
13. $V_i = 0;$

14. END IF
15. END FOR
16. RETURN $[V_1, V_2, \dots, V_{64}]$

通过算法 1 提取出商品图像和检索示例图像的感知哈希值 $CodeG$ 和 $CodeQ$, 计算它们的汉明距离 $d(CodeG, CodeQ)$, 商品图像与检索示例图像之间的相似度计算如下:

$$SimPic(G, Q) = \log_2(2 + 1/(1 + d(CodeG, CodeQ))) \quad (4)$$

4.1.2 计算关键词与商品之间相似度的方法

检索关键词相似度的计算属于文本检索领域的基本问题, 因此可以采用最常用的向量空间模型来进行处理。考虑到文章完整性, 下面简单介绍其计算方法。将商品 G 的属性描述信息作为文本的特征项 T , 采用 TF-IDF 算法计算特征权重:

$$WT = \frac{N_{T,G}}{\sum_k N_{k,G}} \times \log\left(\frac{|M|}{|M_T|} + 0.01\right) \quad (5)$$

其中, $N_{T,G}$ 表示特征项 T 在商品 G 描述信息中的出现次数, $\sum_k N_{k,G}$ 表示商品 G 中所有特征项出现的次数之和, $|M|$ 表示语料库的商品总数, $|M_T|$ 表示包含 T 的商品数, 加上 0.01 是为了防止 $\log$ 项返回为 0 的结果。

分别计算出各个特征项权重, 并且用特征权重向量 $W_{GW} = [W_{T1}, \dots, W_{TN}]$ 和 $W_{QW} = [W'_{T1}, \dots, W'_{TN}]$ 表示 G, Q , 用余弦向量夹角计算 G 与 Q 之间的关键词查询相似度:

$$SimWord(G, Q) = \cos \frac{W_{GW} \times W_{QW}}{|W_{GW}| |W_{QW}|} \quad (6)$$

4.2 商品与用户兴趣偏好相似度的计算方法

通过在网页上记录用户的浏览行为, 隐式判断用户对某件商品是否感兴趣^[13], 把感兴趣的记录在 cookie 中, 用户下次进行检索时, 根据历史记录建立用户的兴趣偏好模型。

假定已知从 cookie 中读取出一组用户感兴趣的记录集合 $GS_p = [G_1, G_2, \dots, G_z]$, 计算商品 G 与用户兴趣偏好的相似度过程如下。

Step1 统计 GS_p 中所有出现过的不重复的商品属性, 得到用户偏好属性集合 $VS = [V_1, V_2, \dots, V_n]$, 并且统计各属性的出现次数 $[V_1:a, V_2:b, \dots, V_n:c]$ 。

Step2 根据各个属性出现的次数计算用户对该属性的兴趣权重, 用式(7)计算:

$$Wv_i = \log_e(1 + fre)(i=1, \dots, n) \quad (7)$$

其中, Wv_i 为用户对属性 V_i 的兴趣权重, fre 为 V_i 出现次数。

Step3 以属性偏好权重向量 W_{PV} 表示用户对商品各个属性的偏好程度: $W_{PV} = [W_{V1}, W_{V2}, \dots, W_{Vn}]$ 。

Step4 用 $W_{GV} = [B_1, B_2, \dots, B_n]$ 表示商品 G 中属性信息包含 VS 集中元素的情况, 其中

$$B_i = \begin{cases} 1, & \text{if } V_i \in G \\ 0, & \text{if } V_i \notin G \end{cases} \quad (i=1, 2, \dots, n)$$

Step5 利用式(8)计算商品 G 与用户偏好 P 之间的相似度大小:

$$Sim(G, P) = \cos \frac{W_{PV} \times W_{GV}}{|W_{PV}| |W_{GV}|} \quad (8)$$

4.3 整体算法流程

用户检索意图建模与结果排序流程如图 1 所示, 主要包括以下 3 大步骤:

(1) 根据用户提交的检索条件从数据库中查找与检索示

例图像及检索关键词相关的结果。按照式(2)计算出各个结果与检索条件的相似度 $Sim(G, Q)$ 的值。

(2) 从用户兴趣记录中提取商品属性信息, 建立用户兴趣偏好模型, 对上一步的相关结果按照式(8)计算出与用户兴趣偏好之间的相似度 $Sim(G, P)$ 的值。

(3) 计算各个结果的排序度函数 $f(G, Q, P)$, 按函数值大小对结果降序排列, 并返回排序结果。

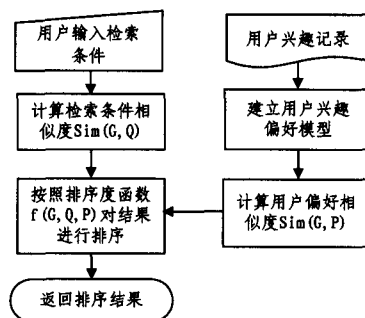


图 1 整体算法流程

对于计算商品与检索示例图像之间的相似度 $SimPic(G, Q)$, 因为需要比较两幅图像之间的哈希距离, 故算法的时间复杂度是 $O(1)$ 。假定每个商品平均有 s 个属性信息, 则计算 $SimWord(G, Q)$ 的复杂度为 $O(s)$, 对于计算商品属性与用户兴趣偏好之间的相似度 $Sim(G, P)$, 算法的计算复杂度为 $O(s)$ 。在 2.8GHz 奔腾双核处理器、4GB 内存环境运行上述算法的 Java 实现程序, 当 s 取值为 8 时, 测得的平均运行时间为 0.325ms。从以上可知本文所提出的实现方案具有较低的时间复杂度, 能达到较高的检索效率。

5 实验与结果分析

5.1 实验设置

为了获取真实的商品信息来进行实验, 通过网络爬虫获得京东商城上各类商品的信息作为测试数据, 包括衣服、家具、厨具、电器等 13 个种类的不同商品, 每个种类爬取 180 条对应的商品信息, 包括商品图片与描述信息。描述信息中包含商品的属性信息, 如产地、品牌、适用人群等。

对于商品的图像, 事先利用算法 1 计算得出其感知哈希值, 将其与商品的属性描述信息一起存储在数据库中作为商品的索引, 以便实验检索测试。

5.2 实验结果与分析

实验 1 相似图像哈希值的汉明距离的阈值 $ThrD$ 的设定对查准率的影响

在依据用户检索条件查找商品结果时, 图像之间哈希值的汉明距离 $d(CodeG, CodeQ)$ 小于阈值 $ThrD$ 时, 认为该商品图像与用户的检索意图相近, $ThrD$ 的设定对排除不相关结果有着重要影响。分别统计阈值 $ThrD$ 设置为 1—8 时检索结果中 Top40 的检索精度, 得到图 2 所示的统计结果。

方法 1 采用本文提出的算法, 通过隐式方式收集用户的兴趣偏好记录, 并以用户最新感兴趣的商品的属性信息建立偏好模型, 从而进行个性化排序。

方法 2 以显式方式事先详细收集用户对于各种类型商品的偏好反馈信息^[14], 并以此建立用户的兴趣偏好模型, 对结果进行个性化排序。

方法3 采用相关反馈的办法^[8],在初次检索结果中,让用户选择其认为最符合检索意图的一个检索结果,采用式(6)计算此结果与其它结果之间的属性信息的相似度,按照文本属性信息的相似度高低对返回结果重排序。

方法4 仅考虑用户的检索请求对检索结果进行排序。

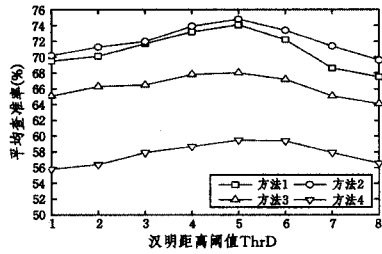


图2 ThrD变化时的平均查准率

从图2可知, ThrD 设置较小时,平均查准率较低,当 ThrD 设置为5左右时,平均查准率达到最优值。这是因为汉明距离设置越小,要求图像越相似,得到的初步检索结果总数较少,但图像相似的结果可能存在语义鸿沟而不符合用户检索意图,导致不相关的结果总能排序在 Top40。随着 ThrD 的增大,初步检索结果总数增多,会把不相关结果排在 Top40 以外,使检索精度提高。而 ThrD 超过5以后,检索结果中与用户输入检索条件不相关的结果增多,导致查准率开始下降。

实验2 建立用户兴趣偏好模型时使用的用户偏好商品信息的个数 Z 对查准率的影响

本文方法建立用户偏好模型时,以 GS_p 集合中用户最新感兴趣的 Z 个商品信息作为依据。Z 的设定对用户兴趣偏好模型的建立有重要影响。为了测试 Z 对偏好模型准确性以及检索准确度的影响,以3为步长,将 Z 分别从3取到36,统计检索结果 Top40 的平均查准率。将 ThrD 设置为5,分别采用实验1中的方法1和方法2来比较 Z 变化时的影响。

如图3所示,在开始阶段,方法1的平均查准率随着 Z 的增加而增加,在 Z 为12时,性能达到最优,但随着 Z 的继续增大,不但模型计算量增大,而且用户先前的兴趣发生迁移,影响了偏好模型的准确性,使得平均查准率有所下降。此实验说明,当从商品属性层面来分析用户的兴趣偏好时,用户的偏好更容易发生变化和迁移。从现实角度不难理解,当用户浏览了不同种类的商品时,不同类别的商品用户的属性需求也是不同的,比如浏览家具时偏好产地为台湾,而浏览电子产品时偏好为美国。因此选取合适的 Z 能够及时反映出用户对属性偏好需求的变化迁移,更为准确地表达出用户当前的兴趣偏好。

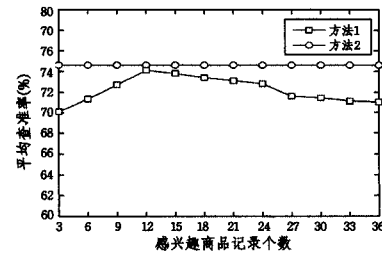


图3 使用最新 Z 个兴趣商品建模时的平均查准率

实验3 检索精度对比实验

为进一步验证算法的有效性,本实验采用的方法1—方法4与实验1中的方法一致,将 ThrD 设置为5,Z 设置为12,

得到 TopN 个返回结果的准确度如图4所示。

对于考虑了用户兴趣偏好的前3种方法,其平均检索精度均高于方法4。在考虑了用户偏好的方法中,前10的检索结果在查准率上非常接近;而之后的结果排序中,方法3的查准率开始明显低于前两种方法,因为仅根据用户反馈的某个商品检索结果信息并不能充分反映用户当前的检索偏好。而方法1的检索精度略低于方法2,它们的检索准确度随着 TopN 的增加而降低,这是由于图像检索条件并不能排除掉不相关的结果所导致的,排序时只能把不相关的结果排在后面。可以看出,通过本文隐式获取用户兴趣偏好的方法能够近似达到用户显式提供的偏好信息的效果,不需要额外增加用户的负担就能够自动为用户提供个性化的检索结果。

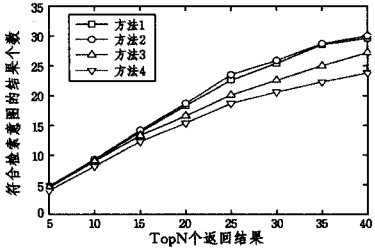


图4 TopN 检索结果中符合检索意图的个数

实验4 不同数据集下参数的设置对检索精度的影响

在实验1—实验3中,所有参数的设置是针对前文所描述的商品数据信息而得出的,将该商品数据信息集称为数据集1。为研究参数 $ThrD \times Z$ 的变化在不同商品数据集中对检索精度的影响,另采用20种商品共10000条数据作为数据集2,与数据集1进行对比实验。分别统计参数 ThrD、Z 变化时,检索结果中 Top40 的平均检索精度,所得结果如表1、表2所列。

表1 ThrD 变化时对不同数据集检索精度的影响

数据集	ThrD×Z	Top40 平均检索精度
1	1×12	0.685
1	2×12	0.702
1	3×12	0.717
1	4×12	0.732
1	5×12	0.741
1	6×12	0.720
2	1×12	0.705
2	2×12	0.714
2	3×12	0.725
2	4×12	0.714
2	5×12	0.717
2	6×12	0.698

表2 Z 变化时对不同数据集检索精度的影响

数据集	ThrD×Z	Top40 平均检索精度
1	5×5	0.712
1	5×8	0.725
1	5×12	0.741
1	5×15	0.738
1	5×17	0.734
1	5×20	0.731
2	5×5	0.704
2	5×8	0.715
2	5×12	0.717
2	5×15	0.698
2	5×17	0.702
2	5×20	0.686

(下转封三)

杨峰,王永齐,梁彦,等. 面向快速多目标跟踪的协同 PHD 滤波器[J]. 系统工程与电子技术,2014,36(11):2113-2121

- [15] Zhang Tao, Wu Ren-biao. Adaptive gating GM-CPHD for multi-target tracking [J]. Journal of Data Acquisition & Processing, 2014,29(4):523-528(in Chinese)

章涛,吴仁彪. 自适应门限 GM-CPHD 多目标跟踪算法 [J]. 数据采集与处理,2014,29(4):523-528

- [16] Jiang Tong-yang, Liu Mei-qin, Zhang Sen-lin, et al. Gating technique for the Gaussian mixture multi-Bernoulli filter [C]//IEEE American Control Conference (ACC). 2014:1096-1101

- [17] Zhang Tao, Lai Ran, Wu Ren-biao, et al. Collaborative PHD filter for fast multi-target tracking [J]. Systems Engineering and Electronics, 2013,30(12):1419-1426(in Chinese)

章涛,来燃,吴仁彪,等. 观测最优分配的 GM-PHD 多目标跟踪算法 [J]. 信号处理,2014,30(12):1419-1426

- [18] Zhang Hong-jian. Finite-set statistics based multiple target tracking[D]. Shanghai: Shanghai JiaoTong University, 2009(in Chinese)

张洪建. 基于有限集统计学的多目标跟踪算法研究[D]. 上海: 上海交通大学, 2009

(上接第 308 页)

从表 1、表 2 的结果可知,对于不同的商品数据信息集,最优检索精度的参数设置值是不完全相同的。由表 1 可知,对于数据集 1, $ThrD$ 的最优取值为 5,而对于数据集 2, $ThrD$ 取较小值时能获得良好的检索精度,其最优取值为 3;由表 2 可知,对于两种数据集,参数 Z 的最优取值均为 12。因此 Z 的最优取值在不同商品数据集上基本一致,而 $ThrD$ 的最优取值针对不同商品数据集而有所差异。

结束语 检索排序对于电子商务系统提供便捷有效的服务具有重要意义。尽管通过计算商品与用户检索条件的相关度进行排序能够取得不错的效果,但是其忽略了用户个体之间的差异,不能提供个性化的检索结果。将用户的浏览历史记录与用户的检索条件相结合,考虑用户对商品属性的偏好,避免了仅仅使用检索条件对检索结果进行排序的局限性,进而提出了基于浏览偏好的商品图像检索排序算法。

实验结果表明,隐式地获取用户对商品属性的兴趣偏好能够达到与显式获取兴趣信息相近的效果。从获取的偏好信息中提取用户的偏好模型,能够提供更加符合用户检索意图和偏好的检索排序结果。

参考文献

- [1] Liu Q, Tan C C, Wu J, et al. Towards Differential Query Services in Cost-Efficient Clouds [J]. IEEE Transactions on Parallel & Distributed Systems, 2014,25(6):1648-1658
- [2] Chen Ming-jing, Yao Jian-rong, Tang Zhi-hao. Researching on Algorithm of Searching Engine for E-Commerce[J]. Computer Engineering and Application, 2006,42(3):209-211(in Chinese)
- 陈明晶,姚建荣,唐志豪. 电子商务系统的商品搜索算法研究[J]. 计算机工程与应用,2006,42(3):209-211
- [3] Chen L, L'Abbate M, Thiel U, et al. Increasing the customer's choice: query expansion based on the layer-seeds method and its application in e-commerce[C]//IEEE International Conference on E-Technology, E-Commerce and E-Service. 2004:317-324
- [4] Zhou D, Lawless S, Wade V. Improving search via personalized query expansion using social media[J]. Information Retrieval, 2012,15(3/4):218-242
- [5] Eakins J, Graham M, Franklin T. Content-based image retrieval [R]. JISC Technology Applications Programme Report, 1999
- [6] Chen Zhi-gang, Wang Jiang-tao, Deng Xiao-heng. A trust model based on semantic distance for pervasive environments[J]. Secu-

rity & Communication Networks, 2014,7(9):1322-1330

- [7] Liu Q, Tan C C, Wu J, et al. Cooperative private searching in clouds[J]. Journal of Parallel & Distributed Computing, 2012, 72(8):1019-1031

- [8] Wang Y, Li H X, Yen G, et al. MOMMOP: Multiobjective Optimization for Locating Multiple Optimal Solutions of Multimodal Optimization Problems[J]. IEEE Transactions on Cybernetics, 2015,45(4):830-843

- [9] Speretta M, Gauch S. Personalized search based on user search histories[C]//The 2005 IEEE/WIC/ACM International Conference on Web Intelligence, 2005. IEEE, 2005:622-628

- [10] Tang Xiao-ling, He Tian-yun. A Research on Topic Preference-Based Personalized Search Model[J]. Journal of Intelligence, 2011,30(4):133-136(in Chinese)

唐晓玲,何天云. 基于主题偏好的个性化检索模型研究[J]. 情报杂志,2011,30(4):133-136

- [11] Liang T P, Lai H J. Discovering User Interests from Web Browsing Behavior: An Application to Internet News Services[C]//Proceedings of Annual Hawaii International Conference on System Sciences. Hicss, 2002:2718-2727

- [12] Mac Aoidh E, Bertolotto M, Wilson D C. Analysis of implicit interest indicators for spatial data[C]//Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems. ACM, 2007:1-4

- [13] Claypool M, Le P, Wased M. Implicit interest indicators [C]//Proceedings of the 6th International Conference on Intelligent User Interfaces. ACM, 2001:33-40

- [14] Ma Z, Pant G, Sheng O R L. Interest-based personalized search [J]. ACM Transactions on Information Systems (TOIS), 2007, 25(1):5

- [15] Teevan J, Morris M R, Bush S. Discovering and using groups to improve personalized search [C]//Proceedings of the Second ACM International Conference on Web Search and Data Mining. ACM, 2009:15-24

- [16] Lu D, Li Q. Personalized search on Flickr based on searcher's preference prediction[C]//Proceedings of the 20th International Conference Companion on World Wide Web. ACM, 2011:81-82

- [17] Kumar R, Sharan A. Personalized Web search using browsing history and domain knowledge[C]//2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT). IEEE, 2014:493-497