一种面向浏览式购物行为模式的LBS购书移动应用

何远舵[1] 陈之昀[2] 王亚沙[3]
(北京大学软件与微电子学院 北京 100871)[1] (北京大学物理学院 北京 100871)[2]
(北京大学北京大学软件工程国家工程研究中心 北京 100871)[3]

摘 要 为了提升用户体验,目前涌现了一批辅助购物的移动应用。这些应用为用户补充商品的信息,适用于目标明确的购物行为模式。然而,在现实生活中,还普遍存在没有明确购物目标、将购物作为休闲方式的浏览式购物行为模式。现有应用对此模式支持不足。为此,以书店为应用场景,开发了一种基于室内定位技术的浏览式购物辅助移动应用。此应用后端从书店信息系统和互联网中收集包括书籍摆放位置的各类信息,然后利用关键词提取技术,建立基于位置的关键词库;前端则利用室内定位技术,根据购物者位置向其展示附近书籍概览信息,辅助购物者对附近商品快速认知。案例评估表明,此应用可以提升浏览式购物顾客的购物体验。

关键词 浏览购物模式,LBS应用,关键词提取,信息展示
中图法分类号 TP399 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.12.007

Browse-shopping-behavior-pattern-oriented Indoor LBS Mobile Application for Book Shopping

HE Yuan-duo[1] CHEN Zhi-yun[2] WANG Ya-sha[3]
(Department of Information Science and Technology, Peking University, Beijing 100871, China)[1]
(Department of Physics, Peking University, Beijing 100871, China)[2]
(National Research Center of Software Engineering in Peking University, Peking University, Beijing 100871, China)[3]

Abstract To enhance shopping experience, a lot of shopping assistant mobile applications are emerging. They can provide supplementary information of a specific commodity, which is helpful for shoppers with clear shopping goals. However, under the browse shopping pattern, where shoppers treat shopping as a kind of relaxation, shoppers don't have clear shopping goals. Existing applications help little under this shopping pattern. This paper implemented an indoor LBS application orienting browse shopping pattern for book shopping. Its back-end collects all kinds of information of books from both the bookstore information management system and the Internet. Then keyphrase extraction method is used to build a keyphrase database according to the position. The front-end applies indoor positioning technology to display the overview information, which helps shoppers quickly understand the commodities. The evaluation indicates that the application can enhance shoppers' shopping experience.

Keywords Browse shopping pattern, LBS application, Keyphrase extraction, Information display

1 引言

随着移动互联网的发展,出现了一批辅助购物的移动应用,如MSA[1]、MP[2]、APriori[3]、CAST[4]、iCart[5]等。这些应用的共同点是:为某种给定的商品提供其他用户的评价、同类商品比较等补充信息,从而辅助购物决策,提升购物体验。对于潜在购物目标已经确定的情况,此类应用能够提供很好的支持。然而,在现实生活中,存在一种常见的购物目标不明确的购物模式:购物者没有明确的购物目标,而是将购物作为一种休闲方式,在购物场所对商品进行大范围的浏览和拣选,交易的发生具有很强的随机性。本文称这种行为模式为"浏览式购物行为模式"(简称"浏览模式"),由于在此模式下,购物者的大量时间用于无确定目标的商品浏览,现有应用难以为此过程提供支持。

针对这一问题,本文以提升浏览模式中购物体验为目标,以书店购书为具体的应用场景,设计并实现了一个具有多维信息展示能力的室内LBS购书辅助移动应用。

通过观察和访谈,购书者以浏览模式在书店中购书的流程大致分为3步。首先,购书者会根据自己的兴趣选择几个区域进行浏览,如偏爱文学的购书者会在文学图书区浏览;然后,购书者浏览周围书架上的图书直到被一本书吸引;最后,购书者通过翻阅这本书对其逐渐了解,根据了解的信息判断对其是否真正感兴趣并购买。

上述购书流程中存在着两个影响购书者购书体验的问

到稿日期:2015-02-26 返修日期:2015-05-15 本文受国家重点基础研究发展计划(973)(2011CB302604),国家自然科学基金(61121063,U1201252),国家高技术发展计划(863)(2013AA01A605)资助。
何远舵(1992—),男,博士生,主要研究方向为普适计算与移动计算等,E-mail:pkuimehyd@163.com;陈之昀(1993—),男,主要研究方向为软件工程、普适计算;王亚沙(1975—),男,教授,主要研究方向为普适计算等,E-mail:yasha.wang@163.com。

题：其一，图书的信息是多维的，而被摆放到在同一个区域的图书只考虑了它们某一个维度的信息。购书者感兴趣的商品很有可能被分到了另一个区域，而购书者没有选择浏览该区域。例如《白夜行》这本书常常放置在推理小说书架上，但该书也可以被归为爱情小说、日本文学等。受制于物理空间和管理复杂度，同一本书不会被放置到不同的分区。一个偏爱爱情故事的购书者很可能错过这本书，因为其没有浏览推理小说分区。其二，购物者大多时候倾向于浏览自己视线易于到达的区域，对书架边缘尤其是底部的图书很少关注。这也会导致购书者错失自己感兴趣的图书。

为了解决上述问题，本文根据购书者在书店内的位置，为其展示临近书架上图书的概览信息——关键词。用户可以通过浏览发现感兴趣的关键词，进一步获得关键词相应的图书及其他信息。通过这种方式，本文工作解决了浏览模式下书店购物存在的问题，提升了用户的购物体验。

本文第 2 节讨论了相关工作；第 3 节描述了应用框架、整个应用的结构概况与工作流程；第 4 节以案例形式阐释了具体实现细节与解决问题的方法；第 5 节分析并总结了应用实际效果；最后总结全文并展望了未来的工作。

2 相关工作

利用普适计算技术为购物者提供商品的额外信息来提升购物体验的移动端应用有很多。MSA[1]、APriori[3]等能够为购物者提供一件商品的详细信息。MP[2]等应用能够为购物者提供一件商品的具体信息以及与其同质的其他商品信息。CAST[4]则基于预先制定的购物清单，根据对购物者所处环境的感知为其展示一件商品以及类似商品的信息。还有一些应用能够对放在购物车内的商品信息进行分析，例如 iCart[5]和 iGrocer[6]，它们能为购物者展示商品的营养信息。

上述应用都暗含一个假设：购书者的购物目标比较明确。在该假设下，这些应用通过为用户提供细节信息，帮助用户比较同质的商品。但当用户在浏览模式下购物时，用户最需要获得的信息是概览信息，而上述工作只能为购物者提供一件或几件同质商品的信息，对购物者的帮助很小。我们的工作则侧重于用户没有明确目的的浏览购物模式，通过为购物者提供粒度更广的信息（从具体一本书的信息到一个书架的信息），帮助用户寻找购物点。

3 应用框架

应用框架包括后台和前端两个模块，如图 1 所示。后台模块主要获取图书的信息并提取和整合图书的关键词；而前端模块着重于用户对手机端应用的使用，包含室内定位和多维信息的展示与交互。

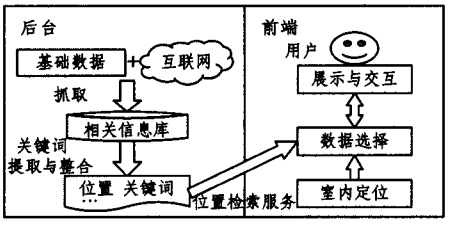

图 1 应用框架

3.1 后台模块

后台模块位于后台服务器，主要负责线下的数据获取与处理工作，其工作流程分为 3 个阶段。

（1）数据获取

数据获取基于基础数据，这些数据来自于已有的图书管理系统中的数据，包括书名、作者、ISBN 等。根据基础数据，可以在互联网上定位一本书，然后利用爬虫抓取与该书相关的信息。基础数据和互联网上的数据被存储到相关信息库中。

（2）关键词提取

对上一步从互联网中获得的图书的数据进行处理，利用关键词提取算法提取了每一本图书的关键词，同时还需要对关键词过滤。

（3）数据的整合

基于每一本图书在书架上的相对位置，整合同一个书架上所有图书的关键词，形成位置关键词数据，并对外提供基于位置信息的检索服务。

3.2 终端模块

终端模块位于用户的手机端，通过智能手机应用程序来实现。该模块主要负责对用户进行室内定位，并为用户提供多维信息的展示和交互等功能。该模块的工作流程如下。

（1）位置获取

用户的手机端应用通过室内定位技术判断用户的位置。根据不同的精度需要，可以选择不同的技术。本文根据实验场景，选择了基于 ibeacon 设备的蓝牙定位。

（2）数据选择

根据上一步获取到的用户的室内位置，手机端应用会利用后台提供的位置检索服务从后台的数据库中获取相应位置上书架的关键词信息、与关键词对应的图书的基础信息和互联网信息。

（3）展示与交互

为了能够在几次交互后为用户展示大量商品的信息，手机端应用会为用户展示有效的关键词信息，用户可以通过浏览大量的关键词信息发现自己感兴趣的内容，进而可以进一步查看具体的图书信息。

4 案例研究

在应用框架下，本文研究并实现了该应用的一个案例，下面将介绍该案例的研究场景、实现细节以及在该案例中如何解决图书关键词信息的提取与整合、多维信息的展示与交互的问题。

4.1 案例研究场景

案例的实验场景是北京大学图书馆三楼自然科学英文文献区，该区域的书架分布如图 2 所示。其中过道的宽度小于 1 米，书架长度约为 1.5 米。该案例中，应用框架中涉及到的数据来自于北京大学图书馆网站以及亚马逊用户评论等。这种书架的尺寸分布及书籍摆放与一般书店相似，且书店的图书管理系统中也存有图书的基本信息。因此对该案例研究及评估的结果具有一般性。

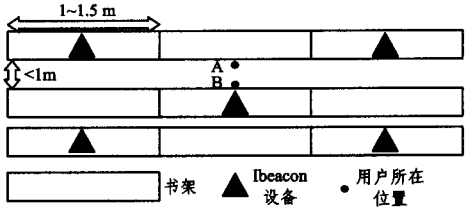

图 2 书架分布与 ibeacon 设备部署

4.2 基于 ibeacon 的室内定位

基于该实验场景下的书架尺寸与分布特点，本文采用了基于 ibeacon 设备的蓝牙室内定位，它的基本原理是根据接收到的信号强度估算手机到 ibeacon 设备的距离[7]。为了得到 ibeacon 设备工作的实际效果，本文进行了实验，并在该实验结果的基础上，结合书架的分布特点，提出了一种 ibeacon 设备室内部署方式。

4.2.1 ibeacon 设备测距实验及分析

实验使用了一台从淘宝上购买的 ibeacon 标准设备(见图 3)、一部搭载蓝牙 4.0 功能模块的三星 Galaxy S4 安卓智能手机以及该手机上的 ibeacon 测距应用。

在实验中，ibeacon 设备被固定于某处，然后将手机置于离 ibeacon 设备一定距离的位置，之后记录在该距离下 ibeacon 测距应用所判断的距离的读数；该实验分别测量了手机距 ibeacon 设备不同距离时的值并对其取平均，如图 3 所示。可以看出，0～60cm 之间，测量的平均值与理想之间的差距非常小(见虚线)。大于 1m 之后，测量均值会随实际距离的增大而增大，但是误差会增大，测量非常不准确。因此，ibeacon 室内定位能够准确地测量距离近的手机，并能大致判断离 ibeacon 设备远(>1m)的手机。

4.2.2 ibeacon 设备部署

根据上文实验结果，本文以图 2 中的方式部署了 ibeacon 设备。图中 ibeacon 放置在书架的上方。每一个过道都在一个 ibeacon 设备能够准确测距的范围内，从而能够判断用户的具体位置。图 2 中，A、B 两个位置属于同一个过道内，且过道宽度约 1m，因此能够准确地根据 A、B 的距离进行区分，进而区分出用户所面向的书架。

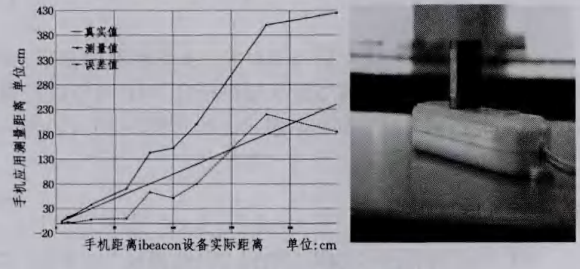

图 3 实验结果与 ibeacon 蓝牙设备

4.3 关键词提取与整合

关键词提取主要通过 KEA 算法[9]从原始数据中获取图书的关键词信息。但是研究发现该应用在关键词提取中存在着一些问题。下面将讨论关键词提取中所需的考虑与关键词提取过程。

4.3.1 问题与考虑

KEA 算法是一种通用的关键词提取算法，并没有考虑关键词的词性。因此，在进行主题词提取之前，应该根据词性过滤掉不能提供主题相关信息的动词。

KEA 提取关键词之后得到每一本图书的关键词信息，但是这些关键词之间并不是毫无关系的，常常会出现一个关键词是另一个关键词的子串，例如书《Linear integrated circuits》的关键词中包含“IC”、“linear IC devices”、“IC devices”。为了突出具体特征并减少冗余，需要过滤掉子串关键词。

当整合同一个书架上所有书的关键词时，也会出现类似问题，即关键词重复或者一些关键词是另一个关键词的子串。前者由于是信息冗余因此会被滤掉，后者出现的原因是许多图书内容较为相似，但是其深度或侧重点不一样。例如“CMOS”与“CMOS VLSI design”，前一个关键词反映这本图书的内容与 CMOS 相关，后一个关键词则反映另一本图书更聚焦于 CMOS 设计。这时，这些关键词反映出了不同书的特征，不存在信息冗余，因此，这些子串关键词被保留。

书店中相似的图书被置于同一个书架上，并在书架上标注图书的类别信息，这是一种为购书者提供商品概览信息的方式。KEA 提取的部分关键词会与标注类别重复，重复关键词必须被过滤。

4.3.2 关键词的提取与整合过程

算法 1(关键词提取算法)

```
ShelfKeyhrase ←i =1…N;
Input ← Ø
OriginData← BookInfo+GetFromInternet (BookInfo)                    (1)
FOR EACH word IN OriginData
    Input. add(word) IF word. isNotVerb                            (2)
Output← KEAKeyPhraseExtraction(Input)                              (3)
KeyPhrases←FilterSubstring(Output)                                 (4)
KeyPhrases← FilterByShelfInfo(KeyPhrases)                          (5)
BuildIndex(KeyPhrases, BookInfo)                                   (6)
ShelfKeyPhrases. add(KeyPhrases)                                   (7)
```

在算法 1 中，(1)通过一本图书的 ISBN 定位这本书在互联网上的信息。(2)利用 nltk 自动词性标注功能过滤掉了其中的动词。(3)利用 KEA 算法提取了关键词。为了减少信息的冗余，(4)过滤掉了子串关键词，减少了约 1/6 的冗余词。(5)根据书架所能提供的信息对关键词进行过滤。书架提供的信息是以人工的方式来给定的。比如说，书架上的标注是“集成电路”，那么“IC”、“Integrated Circuit”等关键词将会被过滤。(6)、(7)分别建立了关键词和图书之间的索引关系并将所有关键词存到一个文件里。

4.4 多维信息的展示与交互

关键词的数量很多，而手机屏幕较小，为了高效地展示信息，本文提出了一种基于最大图书覆盖率的关键词选择算法；此外，还设计了关键词的多维展示策略以及用户与应用的交互过程，以便用户能够快速获得信息。

4.4.1 关键词选择算法

算法 2(基于最大图书覆盖率的关键词选择算法)

```
Number ← 0, Result ← Ø
WHILE Number < MaxForScreen :
  keyphrase ← rand(ShelfKeyPhrase) UNTIL keyphrase. isValid        (1)
  FOREACH book in keyphrase. books :
    book. setAllKeyphrases(invalid)                                (2)
  Result. add(keyphrase)
```

算法 2 中，(1)随机选择一个“有效”的关键词，然后(2)将包含该关键词的图书的所有关键词置为“有效”。该算法的核心思想是通过随机选择一个“有效”关键词，从一个维度描述它所包含的图书的信息，然后将这些图书中的其他关键词置为“无效”，保证被展示的关键词所对应的图书之间的交集为空，最大化图书的覆盖率，使得用户可以快速浏览到书架上大部分图书的信息。

4.4.2 多维信息展示与交互

该案例实现了通过滑动手机屏幕来刷新展示的关键词。一个书架上图书大约有 100 本，每本书约有 5 个关键词。如果不重复出现关键词，理论上 5 次滑动屏幕最多能够覆盖 90 本书的信息，在重复出现的情况下有极大概率超过一半图书。因此为用户提供的信息已经足以覆盖到大部分图书。而且，

即使在滑动前后出现了属于相同图书的不同关键词，这也是对这本书另一个维度信息的表达，即实现了对图书多维信息的展示。

图 4 中，左图展示了应用为用户呈现的关键词。点击这些关键词可以选择查看相关关键词或查看相关图书。如果选择查看相关关键词，那么应用会为用户展示包含该关键词的图书对应的其他关键词；如果选择查看相关图书，那么会展示一个包含该关键词的图书列表。用户可以点击具体的图书来获得该图书的详细信息，见图 4 中右图。例如点击了图 4 左图中的"III-V semiconductor"关键词，选择展示相关图书，最终会呈现右图中有关这本书的详细信息。

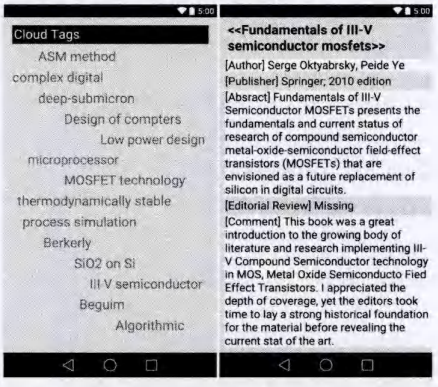

图 4　多维信息展示

5　案例评估

本文邀请了 21 名志愿者使用并评估该应用。

大多数志愿者认为该应用在判断用户面对书架中准确，但是仍有少数人认为不准确。经过仔细询问，发现这些人有时候会靠近一侧而远看另一侧的图书，此时对他们的定位就会出现偏差。

绝大部分的志愿者认为本应用为他们展示的关键词能帮助他们更好地理解该书架商品的信息。而个别的志愿者认为帮助不大，经了解，这些志愿者均为信息学院微电子系的学生，而实验区域的图书恰好包含了微电子领域的图书，因此他们认为展示的关键词没有提供额外信息。但是他们依然认为该应用能够帮助他们发现位于书架边缘容易忽视的图书；即使是非理工学院的参与者也能够根据关键词发现一些非纯技术的图书，例如《Politics and technology》。

结束语　本文设计并以案例形式研究、实现了一种面向浏览购物行为模式的室内 LBS 购书应用。该应用能够为在浏览模式下购物的用户提供图书的多维概览信息，帮助他们理解周围图书的内容，进而发现自己感兴趣的图书，产生购书的欲望。经过案例评估发现该应用能够增强用户的购物体验。

参 考 文 献

[1] Resatsch F, Karpischek S, Sandner U, et al. Mobile sales assistant: NFC for retailers[C]// Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services. ACM, 2007: 313-316

[2] Resatsch F, Sandner U, Leimeister J M, et al. Do Point of Sale RFID-Based Information Services Make a Difference? Analyzing Consumer Perceptions for Designing Smart Product Information Services in Retail Business[J]. Electronic Markets, 2008, 18(3): 216-231

[3] Von Reischach F, Guinard D, Michahelles F, et al. A mobile product recommendation system interacting with tagged products[C]// IEEE International Conference on Pervasive Computing and Communications, 2009 (PerCom 2009). IEEE, 2009: 1-6

[4] Black D, Clemmensen N J, Skov M B. Supporting the supermarket shopping experience through a context-aware shopping trolley[C]// Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open. ACM, 2009: 33-40

[5] Kallehave O, Skov M B, Tiainen N. Persuasion in situ: shopping for healthy food in supermarkets[C]// Proceedings of PINC 2011 Workshop at CHI. 2011

[6] Shekar S, Nair P, Helal A S. iGrocer: a ubiquitous and pervasive smart grocery shopping system[C]// Proceedings of the 2003 ACM Symposium on Applied Computing. ACM, 2003: 645-652

[7] Muñoz-Organero M, Muñoz-Merino P J, Delgado Kloos C. Using bluetooth to implement a pervasive indoor positioning system with minimal requirements at the application level[J]. Mobile Information Systems, 2012, 8(1): 73-82

[8] Witten I H, Paynter G W, Frank E, et al. KEA: Practical automatic keyphrase extraction[C]// Proceedings of the fourth ACM conference on Digital libraries. ACM, 1999: 254-255

(上接第 25 页)

问答信息的抽取。本文通过实验验证了方法的有效性。但是本文方法仅考虑了单一主题的邮件会话，并不适用于包含多个问题的多主题邮件会话。未来我们将针对这一不足进行深入研究。

参 考 文 献

[1] 金庸. 基于邮件列表的软件问答信息抽取工具的设计与实现[D]. 北京：北京大学，2014
Jin Yong. A design and implementation of software R&A extraction tool based on maillists[D]. Beijing: Peking University, 2014

[2] Fournier-Viger P. Spmf: A sequential pattern mining framework [OL]. http://www.philippe-fournier-viger.com/spmf, 2011

[3] 肖仁财. 序列模式挖掘算法研究与实现[D]. 南京：江苏大学，2007
Xiao Ren-cai. A research and implementation of sequential pattern mining algorithm[D]. Nanjing: Jiangsu University, 2007

[4] Belkin, Nicholas J, et al. Query length in interactive information retrieval[C]// Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2003

[5] Salton, Gerard, Wong A, et al. A vector space model for automatic indexing [J]. Communications of the ACM, 1975, 18(11): 613-620

[6] Cong G, Wang L, Lin C Y, et al. Finding question-answer pairs from online forums[C]// Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2008: 467-474

[7] Wang Kai, Chua T-S. Exploiting salient patterns for question detection and question retrieval in community-based question answering [C]// Proceedings of the 23rd International Conference on Computational Linguistics. 2010