

Preferences Discriminant Analysis Based on Implicit Feedback Information While Web Browsing

Tong Wang¹, Wei Wang^{1,2}, Xiaodan Huang¹

¹School of Information and Electrical Engineering, Hebei University of Engineering, Handan Hebei

²School of Internet of Things Engineering, Jiangnan University, Wuxi Jiangsu

Email: wangwei83@hebeu.edu.cn

Received: Dec. 6th, 2017; accepted: Dec. 22nd, 2017; published: Dec. 29th, 2017

Abstract

User behavior in network browsing as implicit feedback information implies the interests and preferences. This information can reduce the cognitive burden for user to get personalized information or services. It is also the basic data of personalized information recommendation. Based on regression and SVD theory, this paper proposes a method of sentiment preference discrimination based on user implicit feedback behavior. It includes coarse-grained mining and fine-grained mining methods. Based on the user behavior dimension, the category dimension of web content is added. It increases the accuracy of sentiment preference coarse-grained mining. At the same time, the user potential factor and the latent factor implicit behavior matrix are studied. The fine-grained preference discriminant model of users is obtained. Experiments show that the proposed method is effective to distinguish the user preference information.

Keywords

Information Recommendation, Implicit Feedback, Preference Discriminant, Web Browsing Behavior

基于网络浏览隐式反馈信息的偏好判别研究

王彤¹, 王巍^{1,2}, 黄晓丹¹

¹河北工程大学信息与电气工程学院, 河北 邯郸

²江南大学物联网工程学院, 江苏 无锡

Email: wangwei83@hebeu.edu.cn

收稿日期: 2017年12月6日; 录用日期: 2017年12月22日; 发布日期: 2017年12月29日

摘要

网络浏览中用户行为作为隐式反馈信息蕴涵了兴趣和偏好, 此信息可以用于个性化服务和推荐, 降低人机交互过程中的认知负担。本文基于回归和SVD理论, 提出了基于用户隐式反馈行为的情感偏好判别方法, 包括粗粒度判别和细粒度判别。在用户行为维度的基础上, 通过增加网页内容的类别维度, 提高情感偏好粗粒度判别的准确性。同时研究用户-潜在因子, 潜在因子-隐式行为矩阵的求取过程, 得到用户细粒度的兴趣偏好判别模型。实验表明本文提出的方法对判别用户兴趣偏好信息是有效的。

关键词

信息推荐, 隐式反馈, 偏好判别, 网络浏览行为

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

物联网创新应用的快速发展, 推动人与人、人与物、物与物的全面互联, 为全面感知物质世界、改变生活方式带来了契机[1]。在此互联的网络中, 人机交互成为信息交换的桥梁。

用户在网络浏览时, 获取大量信息, 经过甄别筛选, 逐步得到所需内容, 此过程中人机交互特征显著。有效的人机交互方式是提高网络内容获取效率, 提供个性化推荐的保障。为了更好地使用户关注于浏览内容, 避免显式地收集用户偏好所带来的干扰, 同时降低用户认知负担, 研究网络浏览中的隐式人机交互(Implicit Human Computer Interaction, IHCI)理论与技术是个性化信息推荐的发展方向之一。

2. 相关工作

隐式人机交互研究逐渐得到了此领域研究者的广泛关注[2] [3]。自1996年Nicole Kaiyan提出了IHCI的概念开始, 美、德、中、奥等国家的大学和研究所在此方向的理论和技术方面, 进行了诸多探索[4]。Albrecht Schmidt将感知和推理作为隐式人机交互的基础, 提出以上下文信息贯穿交互过程[5]。Andrew Wilson和Nuria Oliver则围绕隐式人机交互, 基于机器视觉技术开发了实用系统[6]。陶霖密等人基于对用户行为的感知与理解, 开发了自适应视觉系统, 完成隐式人机交互过程[7]。田丰等也从Post-WIMP用户界面的角度, 研究了隐式交互的特征[8]。

隐式人机交互研究也逐渐渗透到网络信息推荐领域, 显式和隐式人机交互相互融合, 以隐式地采集用户网络浏览行为, 部分取代传统的显式地收集用户对网页内容的评分数据, 进而提高网络信息获取效率。网络浏览中两种人机交互方式的特性如表1所示[9]。

在研究用户网络浏览行为隐式信息的基础上, 还在探索挖掘其蕴含的情感偏好。Katja指出利用鼠标的点击数据, 交叉比较法(interleaved comparison methods)可以挖掘出用户对不同推荐内容的情感偏好之间的微小差别, 并且能够用于在线打分[10]。同时分析了三种现有的交叉比较法, 发现他们估计情感偏好时既存在偏差, 又对推荐内容之间的一些差别不敏感。为了解决这些问题, 基于概率交叉过程, 提出了一种新的方法, 得到情感偏好的无偏估计器。Chang基于递归划分法和逻辑回归法将浏览行为按重要性进

Table 1. Features of EHCI & IHCI
表 1. 显式/隐式交互的特性

| 特性 | 隐式 | 显式 |
|--------|----|-------|
| 精确度 | 低 | 高 |
| 信息丰度 | 高 | 低 |
| 上下文相关性 | 相关 | 相关 |
| 情感偏好表现 | 正向 | 正向和负向 |
| 量测相关性 | 相对 | 绝对 |

行了区分,得到了一个预测模型,综合多种页面浏览行为和搜索结果提高了用户偏好内容的预测效果[11]。Byoungju 则对不同用户的浏览行为进行了分类,如搜索、跳过、播放列表等,构建了更精确的初始评分矩阵,以便进行后续的协同过滤[12]。

针对不同的浏览行为,研究者还基于线性回归方程 $y = \xi_0 + \xi_1 x_1 + \xi_2 x_2 + \dots + \xi_n x_n$ 建立用户兴趣度的计算方法。以用户浏览行为与兴趣度数据对进行回归拟合,从而得到用户兴趣度的线性回归模型,再根据 r^2 和 F 判断回归方程是否显著。目前的研究多集中于此种情况,其本质是对于所有商品来说,某种浏览行为的权重都是相同的。当商品的种类繁多,对于某些商品用户也许没有浏览过,或者对于被测试的用户来说,不可能将所有的商品都以显式反馈获得兴趣度,造成其时间压力,认知负担等。在此种情况下,由于从一部分商品得到的权重,去推导其余商品的权重时,会为其余商品的用户兴趣度计算精度带来影响,因此 Duke 研究了不同商品(或一类商品)具有不同的行为权重(或一类行为权重)的组合形式,首先通过研究分析了权重的可行域,发现权重的可行域很大,较难确定数值,因此转而研究对于其余商品的聚类上,将其余商品聚类到已知权重的商品类别中,进而用已知权重计算。尽管其兴趣度模型也是以线性方程表示的,但是该方法考虑了权重的更多细节,提高了情感偏好的判别效果。

一些研究工作已经在此领域展开,虽然研究还不多,但可能是未来的一个重要研究方向[13][14][15]。本文将基于线性回归和奇异值分解方法,研究用户网络浏览行为中蕴涵的偏好信息,进而得到网络浏览隐式反馈行为到情感偏好信息的映射,进而获知用户不同粒度的情感偏好信息。

3. 情感偏好判别

在用户浏览网络的过程中,会产生大量的隐式反馈行为数据,如鼠标移动时间、点击次数、页面驻留时间、滚动次数、点击上翻页和下翻页键时间等,其中蕴含了用户对于内容的偏好信息。如能通过判别分析,得到用户情感偏好,对提高信息推荐效率有较大的帮助。

用户情感偏好信息是蕴含在用户的隐式反馈行为历史中的。基于隐式反馈行为,研究用户的情感偏好判别方法,是本文的主要研究内容,具体包括粗粒度用户情感偏好判别和细粒度用户情感偏好判别两种方法。本文在用户行为维度的基础上,增加网页内容的类别维度。对不同类别的网页内容,应用回归理论确定隐式反馈行为中的不同行为集权重,以便提高用户情感偏好计算的准确度,得到用户粗粒度的情感偏好。此外,应用奇异值分解等算法,较为准确和快速地得到用户因子向量和隐式行为因子向量,进而对具体物品做出细粒度的情感偏好估计。

粗粒度用户情感偏好是指用户感兴趣的网页内容偏好,是用户兴趣的宏观展现。由于粒度较大,无法精确反映用户感兴趣的具体网页内容,快速地确定用户偏好则是其特点之一,因此,本文中选择的线性回归模型作为粗粒度用户情感偏好判别的计算模型。

现有的类似方法,对于不同的网页内容,行为权重取值相同,无法体现网页内容之间的差别。本文

在用户行为维度的基础上，增加网页内容的类别维度。对不同类别的网页内容，隐式反馈行为中的行为集权重也有所不同，以便提高用户情感偏好计算的准确度，实现方法如图 1 所示。

基于前述的隐式反馈行为样本和用户评分信息，应用线性回归得到行为权重矩阵 $WEIGHT^{con} \in \mathfrak{R}^{M^G \times N^l}$ ，其中， $M^G = |\mathbf{A}|$ 为群体用户隐式反馈行为集的基数， N^l 为情感推荐系统中推荐网页内容的类别数。则粗粒度的用户情感偏好为：

$$\overline{INT}^{con} = \overline{ACT}^T \times WEIGHT^{con} \tag{1}$$

细粒度用户情感偏好是指用户感兴趣的具体网页内容的偏好信息，是用户兴趣的微观展现。综合考虑情感偏好判别算法的精确性和复杂性，本文采用潜在因素模型(Latent Factor Models, LFM)中的奇异值分解(Singular Value Decomposition, SVD)算法。

在用户与情感推荐系统交互过程中，用户的隐式反馈行为序列被记录。随后分析此序列中有关群体用户隐式反馈行为集中包含的行为，并对被记录行为的相关属性进行量化，将量化值作为因子，则用户 u 的因子向量为 $\mathbf{x}_u \in \mathfrak{R}^{M^G \times 1}$ ，物品 i 的因子向量为 $\mathbf{y}_i \in \mathfrak{R}^{M^G \times 1}$ ，基于隐式反馈行为，用户 u 对物品 i 的情感偏好值为 $\hat{p}_{ui} = \mathbf{x}_u^T \mathbf{y}_i$ 。根据前述隐式反馈行为效用的分析，可计算得到此情感偏好值对应的置信度为 c_{ui} 。

$$c_{ui} = \sum_1^{M^G} w_i c_i^G \tag{2}$$

其中， c_i^G 为用户各隐式反馈行为的置信度， w_i 为其权重。

根据 SVD 算法，基于用户 u 对一些物品的已有情感偏好 p_{ui} ，通过最小化代价函数，如式(3)所示，便可求得此用户 u 对其余未浏览物品 i 的用户因子向量 \mathbf{x}_u 和物品因子向量 \mathbf{y}_i ，进而通过 $\hat{p}_{ui} = \mathbf{x}_u^T \mathbf{y}_i$ 计算出未浏览物品的情感偏好估计。当计算出用户 u 对所有物品的情感偏好值时，取最大的 Top N 个物品进行后续推荐。

$$\min_{\mathbf{x}, \mathbf{y}} \sum_{u,i} c_{ui} (p_{ui} - \mathbf{x}_u^T \mathbf{y}_i)^2 + \kappa \left(\sum_u \|\mathbf{x}_u\|^2 + \sum_i \|\mathbf{y}_i\|^2 \right) \tag{3}$$

此外，还可将上述两种不同粒度的用户情感偏好判别方法相结合，先计算得到粗粒度的用户情感偏好，即用户感兴趣的物品类别，再在该类别中，运用 SVD 方法，进而获得用户感兴趣的具体物品的偏好

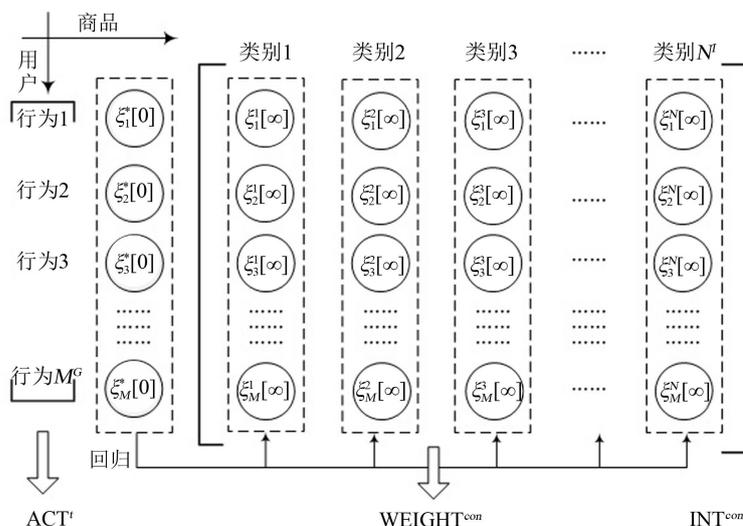


Figure 1. Implicit feedback sentiment preference computing model
图 1. 隐式反馈情感偏好计算模型

信息。由于是在一类物品中搜索 x_u 和 y_i ，用户情感偏好的判别速度会有所提高。

4. 实验

在基于用户网络浏览行为的情感偏好判别过程中，本小节首先应用线性回归进行粗粒度用户情感偏好判别。随后，应用 SVD 方法进行细粒度用户情感偏好判别。

(1) 粗粒度用户情感偏好

基于前述的隐式反馈行为样本和用户评分信息，应用线性回归得到行为权重矩阵 $WEIGHT^{con} \in \mathfrak{R}^{M^G \times N^I}$ ，其中， $M^G = 7$ ，进而可通过(1)式判别用户对不同类别网页内容的偏好程度。图 2 反映了对于某一类网页，综合 48 名用户的网络浏览隐式行为数据，应用线性回归建立模型，当置信区间为 95%时的拟合数据残差，其中有 2 个离群点。

在得到网络浏览隐式行为的线性回归模型后，将用户的网络浏览隐式行为数据作为测试样本，分别针对 2 类网页内容进行粗粒度用户情感偏好判别，结果如图 3 所示。

由图 3 可以看出，用户针对不同类别的网页内容，通过其网络浏览过程中的隐式行为，可以以较粗的粒度区分不同用户对网页内容的偏好。例如，基于第一类网页内容所建立的网络浏览隐式行为回归模型，通过用户的网络浏览隐式行为预测得到的评分，都在 0~5 分的合理区间内，而基于第二类网页内容所建立的网络浏览隐式行为回归模型，则有预测评分超出合理区间，进而可粗略判断用户对某类网页的兴趣偏好。

(2) 细粒度用户情感偏好

本文采用潜在因素模型(Latent Factor Models, LFM)中的奇异值分解(Singular Value Decomposition, SVD)算法。

潜在因素模型通过假设一个潜在因子空间，分别得到用户-潜在因子，潜在因子-隐式行为矩阵，然后通过矩阵相乘得到用户兴趣偏好细粒度判别模型。

本文建立用户-隐式行为矩阵，矩阵中的元素表示用户在进行某网页内容的浏览时，所做出隐式浏览行为的归一化信息。一般来说，此矩阵是一个非常稀疏的矩阵，因为我们不可能要求所有用户对于各网页内容采用所有的隐式行为。于是利用稀疏的矩阵，填充得到一个满矩阵就是我们的目的。本文应用奇异值分解 SVD 算法，通过稀疏矩阵中的已知值，分解得到两个矩阵，按需要进行约简后再相乘，填充稀

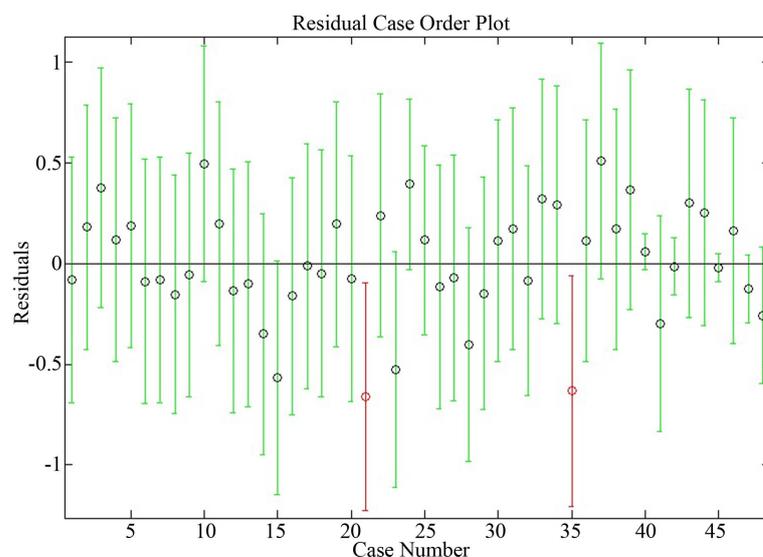


Figure 2. Residuals of linear regression fitting data
图 2. 线性回归拟合数据残差

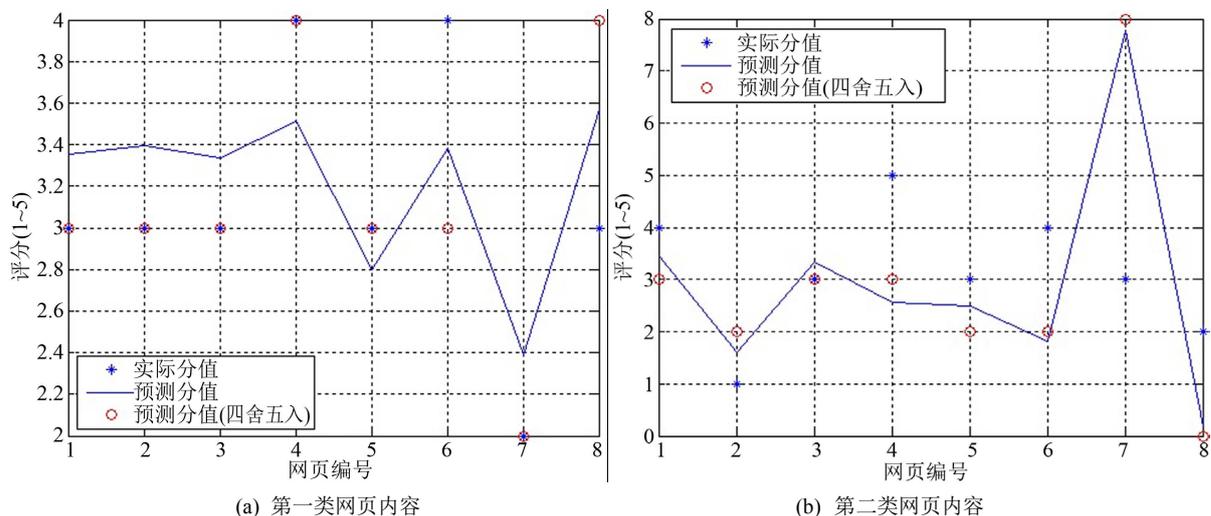


Figure 3. Discriminant distinguish ability of coarse-grained preference for different types of web contents

图 3. 不同类别网页内容的粗粒度偏好判别可区分性

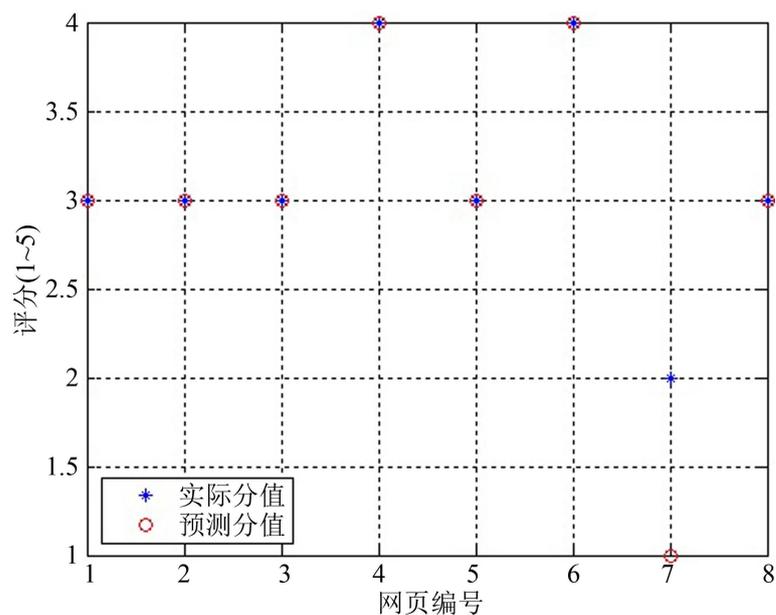


Figure 4. Fine grained user preference discrimination

图 4. 不同类别网页内容的粗粒度偏好判别可区分性

疏矩阵后可得到一个满的用户-隐式行为矩阵，即用户兴趣偏好细粒度判别模型。

随后，当新的用户浏览此类网页时，记录下其隐式浏览行为，并进行量化和归一化，通过与上述矩阵模型中的行向量进行相似性比较，本文中采用余弦相似度计算后，即可找出与新用户最相似的已知用户，可认为两者具有相似的情感偏好。

基于细粒度用户情感偏好模型，新用户的兴趣偏好判别效果如图 4 所示。

图 4 中“*”是新用户按照自己的兴趣偏好对不同网页进行显式打分的分值，“○”表示通过新用户的网页浏览隐式行为预测的兴趣偏好分值，即与新用户相似性较高的已知用户的显式打分分值。从图 4 中可以看出，用户的实际兴趣偏好量化分值与预测分值较为接近，预测成功率较高，基本反映了用户对此类网页内容中各网页的偏好程度。

5. 结论

本文基于回归和 SVD 理论, 提出了基于用户隐式反馈行为的情感偏好判别方法, 包括粗粒度判别和细粒度判别。通过实验, 验证了在用户行为维度的基础上, 通过增加网页内容的类别维度, 可以提高情感偏好粗粒度判别的准确性。同时基于所提出的用户细粒度兴趣偏好判别模型, 对偏好判别也具有较高的成功率, 可以反映用户对此类网页内容中各网页的偏好程度。

基金项目

河北省自然科学基金(F2015402108); 邯郸市科学技术研究与发展计划(1625202042-1); 江苏省博士后科研资助计划(1601085C)资助课题。

参考文献 (References)

- [1] 徐光祐, 陶霖密, 等. 普适计算模式下的人机交互[J]. 计算机学报, 2007, 30(7): 1041-1053.
- [2] Schmidt, A., Spiessl, W., et al. (2010) Driving Automotive User Interface Research. *IEEE Pervasive Computing*, **9**, 85-88. <https://doi.org/10.1109/MPRV.2010.3>
- [3] 贾堃阳. 基于隐式反馈的分布式推荐算法研究[D]: [硕士学位论文]. 浙江: 浙江大学, 2015.
- [4] Kaiyan, N. (1996) Exploratory Study of Implicit Theories in Human Computer Interaction. *Proceedings of the 6th Australian Conference on Computer-Human Interaction*, Computer Society Press, USA, 338-339. <https://doi.org/10.1109/OZCHI.1996.560158>
- [5] Schmidt, A. (2000) Implicit Human Computer Interaction through Context. *Personal Technologies*, **4**, 191-199. <https://doi.org/10.1007/BF01324126>
- [6] Wilson, A. and Oliver, N. (2005) Multimodal Sensing for Explicit and Implicit Interaction. *Proceedings of the 11th International Conference on Human-Computer Interaction*, Lawrence Erlbaum Associates, USA, 1-10.
- [7] 王国建, 陶霖密. 支持隐式人机交互的分布式视觉系统[J]. 中国图像图形学报, 2010, 15(8): 1133-1138.
- [8] 田丰, 邓昌智, 等. Post-WIMP 界面隐式交互特征研究[J]. 计算机科学与探索, 2007, 1(2): 160-169.
- [9] Jawaheer, G., Szomszor, M. and Kostkova, P. (2010) Comparison of Implicit and Explicit Feedback from an Online Music Recommendation Service. *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, ACM, 47-51.
- [10] Katja, H., Shimon, W. and Maarten, D.R. (2011) A Probabilistic Method for Inferring Preferences from Clicks. *Proceedings of CIKM'11*, Glasgow, 249-258.
- [11] Chang, L., Nicholas, B. and Michael, C. (2011) Personalization of Information Retrieval in Different Types of Tasks. *Proceedings of SIGIR 2011 Workshop on Enriching Information Retrieval (ENIR 2011)*, Beijing, 1-4.
- [12] Byoungju, Y., Sangkeun, L., Sungchan, P., et al. (2012) Exploiting Various Implicit Feedback for Collaborative Filtering. *Proceedings of WWW 2012 Companion*, Lyon, 639-640.
- [13] 龚佼蓉. 基于大规模用户隐式行为反馈的书籍推荐方法研究[D]: [硕士学位论文]. 浙江: 浙江工业大学, 2015.
- [14] 王贺玉. 利用关联数据中隐式反馈的 Top-N 推荐系统研究[D]: [硕士学位论文]. 黑龙江: 哈尔滨工业大学, 2015.
- [15] 胡光能. 推荐系统中多源信息融合和隐式反馈挖掘的研究[D]: [硕士学位论文]. 江苏: 南京大学, 2016.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2325-2286，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：sea@hanspub.org