

# Research on Redundancy of Implicit Feedback Information in Web Browsing

Dingding Wei<sup>1</sup>, Wei Wang<sup>1,2</sup>, Xiaodan Huang<sup>1</sup>

<sup>1</sup>School of Information and Electrical Engineering, Hebei University of Engineering, Handan Hebei

<sup>2</sup>School of Internet of Things Engineering, Jiangnan University, Wuxi Jiangsu

Email: wangwei83@hebeu.edu.cn

Received: May 4<sup>th</sup>, 2017; accepted: May 21<sup>st</sup>, 2017; published: May 24<sup>th</sup>, 2017

---

## Abstract

Aiming at the problem that the amount of implicit feedback information in the user's web browsing process is large but the information is not clear, a redundancy analysis method of web-browsing implicit feedback information based on multi-dimensional aspect analysis was proposed. Taking the user's implicit feedback information on different website as time series, 6 static characteristics have been calculated. And by constructing the implicit feedback behavior feature matrix, the non-similarity matrix is calculated to reconstruct the feature matrix in the low-dimensional space. It can show the redundancy of user's different web browsing behavior, which will lay the foundation for applying the research conclusion to the recommended system. The results show that the proposed method can effectively analyze the redundancy of the implicit feedback information in web browsing, and obtain the guiding principle of implicit feedback information's selection.

## Keywords

Information Recommendation, Implicit Feedback, Redundancy, Multidimensional Scaling

---

# 网络浏览隐式反馈信息的冗余性研究

魏丁丁<sup>1</sup>, 王 巍<sup>1,2</sup>, 黄晓丹<sup>1</sup>

<sup>1</sup>河北工程大学信息与电气工程学院, 河北 邯郸

<sup>2</sup>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5月4日; 录用日期: 2017年5月21日; 发布日期: 2017年5月24日

---

## 摘 要

针对用户网络浏览过程中的隐式反馈信息数据量大但偏好信息表达不明确的问题, 提出了基于多元尺度

分析理论的网络浏览隐式反馈信息冗余性分析方法。该方法将用户对不同网站浏览的隐式反馈信息作为时间序列,分别计算6个静态特征,通过建立隐式反馈行为特征矩阵,计算不相似度矩阵,从而实现特征矩阵在低维空间的重构,以展现用户不同网络浏览行为的冗余性,为后期将研究结论应用于推荐系统奠定基础。实验结果表明,该方法可以有效地分析网络浏览隐式反馈信息的冗余性,得到具有指导性的隐式反馈信息选取原则。

## 关键词

信息推荐, 隐式反馈, 冗余性, 多元尺度分析

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

随着物联网、移动互联网等理论与技术的不断发展,将会逐步实现人与人、人与物、物与物的全面互联,从而提高对物质世界的感知能力,实现智能化决策和控制的目标[1]。在此互联的网络中,人机交互理论与技术的研究就显得尤为重要了。

在用户的网络浏览过程中,人机交互特征显著。为了使用户更高效地获取网络内容,而不迷失在纷繁的海量信息中,需要获得用户的兴趣和偏好,提供个性化推荐。同时为了使用户更专注于网络内容本身,而不被用户兴趣和偏好的收集过程所干扰,需要拓展传统的交互方式,在显式人机交互的基础上融入隐式人机交互。网络浏览中的隐式交互理论与技术的研究,可以降低用户获取个性化信息的认知负担,成为个性化信息推荐的发展方向之一。

## 2. 相关工作

隐式人机交互是交互领域中的重要研究前沿[2]。澳大利亚斯威本科技大学的 Nicole Kaiyan 在 1996 年就提出了 IHCI 的概念,但并未深入[3]。从 2005 年开始,美国、德国、中国、奥地利等国内外大学和研究所,对 IHCI 理论、技术以及应用逐步开展了深入研究。德国卡尔斯鲁厄大学的 Albrecht Schmidt 在 IHCI 的理论研究方面,进行的较早,认为隐式交互的两要素是感知和推理,并提出上下文信息对交互过程是极为重要的,同时基于 XML 语言,对交互过程进行建模[4]。美国微软研究院的 Andrew Wilson 和 Nuria Oliver 基于机器视觉,开发了四个系统,研究隐式交互技术[5]。我国自 2007 年开始,由清华大学陶霖密等人开发了自适应视觉系统,检测与理解用户行为,进行隐式交互[6]。同时,中国科学院软件研究所的田丰也从后 WIMP 用户界面的角度,研究了隐式交互的特性[7]。

在网络信息推荐领域,将隐式交互作为网络浏览行为的反馈通道,通过网络浏览中的隐式反馈信息提高或者替代原来的显式反馈,提高信息推荐的效果。隐式和显式反馈的特性如表 1 所示[8]。

隐式反馈可以在一定程度上缓解显式反馈信息来源不足以及可靠性等问题,但隐式反馈信息由于在实际的信息系统伴随用户的浏览过程而生,信息量巨大且连续,存在着如下的缺陷:无负反馈信息、较大的数据噪音、不明确的偏好表达、不确定的评估有效性等[9]。

上述的问题中,一类是由隐式反馈方式本身所决定的,如无负反馈信息和不确定的评估有效性,另一类是由隐式反馈信息特点决定的,如数据噪音和偏好表达不明确。针对后者,通过对网络浏览过程中

**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

的隐式反馈信息的深入研究，逐步解决。

文献[10]修改了传统的基于显式评分的矩阵分解模型，使其对于隐式反馈数据也有良好的效果。文献[11]采用基于 Hadoop 分布式平台的 MapReduce 数据处理模型，先将海量数据进行过滤、清洗等数据预处理工作，得到有效的用户隐式行为数据，再对已清洗过的用户行为数据进行进一步的统计和计算，提高对大规模数据处理的效率。文献[12]将数据利用图结构表示，利用两种不同的相似度计算方法来计算图中的推荐，一种使用基于图中边数目的方法，另一种利用一维数据模型将数据表示成特征向量，计算他们之间的相似度。文献[13]提出一个基于多源信息融合的高效综合方法，能够同时建模打分信息、文本信息和社交网络信息。

本文将基于多元尺度分析理论，研究网络浏览隐式反馈信息的相似度和冗余性，为后续对不同隐式反馈信息加权，确定信息可靠性和挖掘用户偏好奠定基础。

### 3. 行为冗余分析

通过记录用户的网络浏览行为，可以隐式地收集用户偏好信息，进而得到用户情感偏好。通过此闭环迭代，有针对性地为用户提供个性化信息，提高用户信息获取效率。用户的网络浏览行为与人机交互界面、交互设备、用户属性、数据采集方式等具有相关性，因此本文所述方法需在稳定的环境中进行。

用户网络浏览行为的统计值，如页面驻留时间、页面垂直滚动时间、页面滚动次数、鼠标移动时间、鼠标点击次数、点击上翻页键时间、点击下翻页键时间等，是具有代表性的用户隐式反馈信息。应用此类数据进行用户情感偏好挖掘前，需要分析行为信息的冗余性，即分析行为之间的差异。例如：对于两个行为  $a_1^G, a_2^G \in \mathbf{A}$  ( $\mathbf{A}$  为群体用户隐式反馈行为集)，如果他们的差异较小，则表明他们代表着类似的用户情感偏好，当同时出现时，就存在一定的信息冗余。当在线进行情感偏好挖掘时，这种冗余可以增强行为的置信度  $c_1^G, c_2^G \in [0,1]$ ，同时还可以扩展用户隐式反馈行为集  $\mathbf{A}$ ，使其由单一隐式反馈行为组成的集合，扩展为由单一隐式反馈行为和组合隐式反馈行为组成的集合，为今后隐式反馈行为序列的研究做好准备。因此，用户网络浏览行为冗余性的分析是非常重要的。本文拟采用多元尺度分析理论 (Multidimensional Scaling, MDS)，分析用户隐式反馈行为的冗余性。多元尺度分析理论是一种将多维空间的研究对象(样本或变量)简化到低维空间进行定位、分析和归类，同时又保留对象间原始关系的数据分析方法。

将用户对不同网站浏览的隐式反馈信息作为时间序列，分别计算 6 个静态特征：均值  $\mu_i$ 、标准差  $\sigma_i$ 、一阶差分绝对值均值  $\delta_i$ 、归一化一阶差分绝对值均值  $\bar{\delta}_i$ 、二阶差分绝对值均值  $\gamma_i$  和归一化二阶差分绝对值均值  $\bar{\gamma}_i$ 。

$$\mu_i = \frac{1}{N} \sum_{n=1}^N X_n^i \tag{1}$$

$$\sigma_i = \left[ \frac{1}{N-1} \sum_{n=1}^N (X_n^i - \mu_i)^2 \right]^{1/2} \quad (2)$$

$$\delta_i = \frac{1}{N-1} \sum_{n=1}^{N-1} |X_{n+1}^i - X_n^i| \quad (3)$$

$$\tilde{\delta}_i = \frac{1}{N-1} \sum_{n=1}^{N-1} |\tilde{X}_{n+1}^i - \tilde{X}_n^i| = \frac{\delta_i}{\sigma_i} \quad (4)$$

$$\gamma_i = \frac{1}{N-2} \sum_{n=1}^{N-2} |X_{n+2}^i - X_n^i| \quad (5)$$

$$\tilde{\gamma}_i = \frac{1}{N-2} \sum_{n=1}^{N-2} |\tilde{X}_{n+2}^i - \tilde{X}_n^i| = \frac{\gamma_i}{\sigma_i} \quad (6)$$

不同用户网络浏览隐式反馈行为的统计数据差异可以用向量  $\mathbf{Att}_i = [\mu_i \ \sigma_i \ \delta_i \ \tilde{\delta}_i \ \gamma_i \ \tilde{\gamma}_i]$ ,  $i \in [1, L]$  描述, 即隐式反馈行为  $i$  的特征向量为  $\mathbf{Att}_i$ , 进而建立相应的特征矩阵  $\mathbf{Att} = [\mathbf{Att}_1 \ \mathbf{Att}_2 \ \cdots \ \mathbf{Att}_L]^T \in \mathcal{R}^{L \times 6}$ . 根据度量多元尺度分析理论, 在隐式反馈行为特征矩阵的基础上, 构建不相似度矩阵  $\mathbf{\Delta}_{L \times L}$ ,  $\mathbf{\Delta}$  中的元素  $\delta_{ij}$  为:

$$\begin{aligned} \delta_{ij} = d_{ij} &= \left[ (\mathbf{Att}_i - \mathbf{Att}_j)(\mathbf{Att}_i - \mathbf{Att}_j)^T \right]^{1/2} \\ &= \left[ \sum_k (\mathbf{Att}_{ik} - \mathbf{Att}_{jk})^2 \right]^{1/2} \quad \forall i, j \in \{1, 2, \dots, L\}, k \in \{1, 2, \dots, 6\} \end{aligned} \quad (7)$$

再计算内积矩阵  $\mathbf{\Gamma}$ , 其元素  $\gamma_{ij}$  为:

$$\gamma_{ij} = -0.5 \times (\delta_{ij}^2 - \delta_{i\cdot}^2 - \delta_{\cdot j}^2 + \delta_{\cdot\cdot}^2) \quad (8)$$

$$\text{其中, } \begin{cases} \delta_{i\cdot}^2 = \frac{1}{L} \sum_j \delta_{ij}^2 \\ \delta_{\cdot j}^2 = \frac{1}{L} \sum_i \delta_{ij}^2 \\ \delta_{\cdot\cdot}^2 = \frac{1}{L^2} \sum_i \sum_j \delta_{ij}^2 \end{cases} .$$

把隐式反馈行为特征矩阵在低维空间上的重构矩阵记为  $\hat{\mathbf{\Omega}}_{L \times M}$  ( $M < 3$ ). 其表示的网络浏览行为不相似度矩阵记为  $\mathbf{D}$ , 则根据度量多元尺度分析理论,  $\mathbf{\Delta}$  与  $\mathbf{D}$  在某种意义上近似. 且有:

$$\mathbf{\Gamma} = \hat{\mathbf{\Omega}} \hat{\mathbf{\Omega}}^T \quad (9)$$

求解式(9)可得到在低维空间上的隐式反馈行为特征重构矩阵  $\hat{\mathbf{\Omega}}$ , 即网络浏览过程中的隐式反馈行为差异可以在低维空间被表示出.

同时, 还可求出隐式反馈行为特征重构矩阵第  $j$  维对应的特征值  $\lambda_j$ :

$$\lambda_j = \sum_i \hat{\omega}_{ij}^2 \quad (10)$$

其中,  $\hat{\omega}_{ij}$  是  $\hat{\mathbf{\Omega}}$  的元素值.

隐式反馈行为特征在低维空间得到重构, 进而更好揭示不同用户隐式反馈行为的冗余性. 在此基础上, 还可综合更多的因素, 扩展隐式反馈行为  $i$  的属性向量  $\mathbf{Att}_i$ , 进行更精确的用户网络浏览行为的冗余性分析.

### 4. 实验

根据用户的网络浏览行为记录，如页面驻留时间、页面垂直滚动时间、页面滚动次数、鼠标移动时间、鼠标点击次数、点击上翻页键时间、点击下翻页键时间分别计算 6 个静态特征形成隐式反馈行为特征矩阵  $\mathbf{Att} \in \mathcal{R}^{7 \times 6}$ ，并基于上节所述的度量多元尺度分析理论，在低维空间上对隐式反馈行为特征进行重构，分析样本的类间距离。根据克鲁斯卡压力值的要求，本文中，用户的网络浏览隐式反馈行为特征向量在二维空间的重构如图 1 所示。

图 1 中的 ToP、TVS、NoSE、TMM、NoMC、THPUK 和 THPDK 分别表示上述 7 种隐式反馈行为。以图 1(a)为例，对于第 1 个用户，左图中的点 THPUK、THPDK 和 TMM 距离较近，说明当采用 6 个静态特征描述网络浏览的隐式反馈行为时，点击上翻页键时间、点击下翻页键时间和鼠标移动时间三种行为之间仅存在少许差异，隐式反馈行为模式存在较强的相关性，冗余性较大。点 ToP、NoSE、和 NoMC

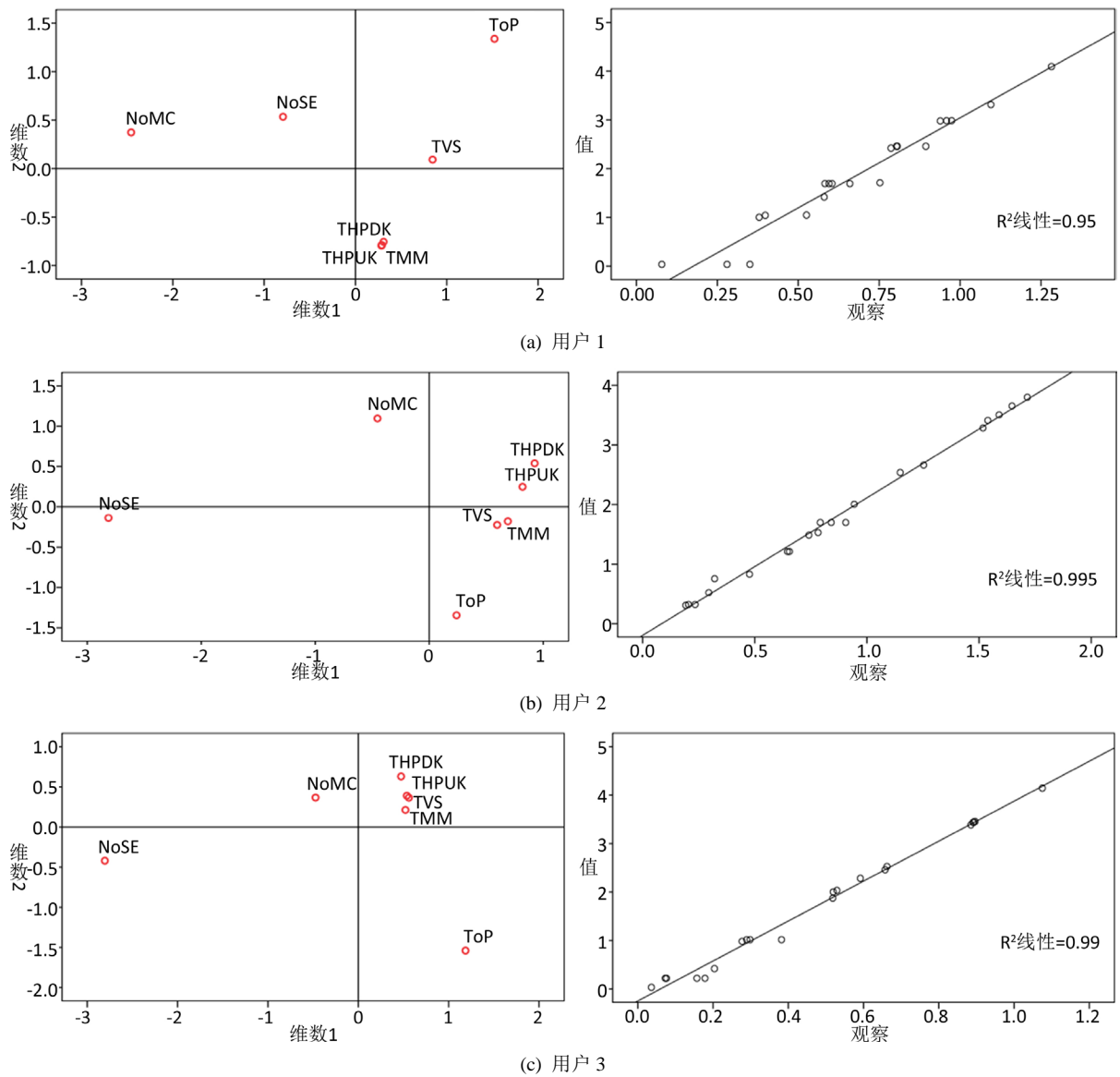


Figure 1. Reconstruction of feature vector of web browsing implicit feedback behavior  
图 1. 网络浏览隐式反馈行为特征向量重构

**Table 2.** Summary of the feature reconstructing process of the web browsing implicit feedback**表 2.** 网络浏览的隐式反馈行为特征重构过程数据总结

User	Young's S-stress	决定系数 RSQ	冗余性较强的行为
1	0.00383	0.9992	THPUK、THPDK 和 TMM
2	0.00637	0.99768	THPUK、THPDK、TMM 和 TVS
3	0.00188	0.99970	THPUK、THPDK、TMM 和 TVS

间的距离较远，其表示的页面驻留时间、页面滚动次数和鼠标点击次数三种隐式反馈行为之间存在较大的差异，行为模式存在较弱的相关性，冗余性较小。点 TVS 与其他点的距离不定，页面垂直滚动时间的冗余性不好确定。

图 1(a)的右图为重构过程中的转换散点图。从图中可以看出拟合系数  $R^2 = 0.95$ ，线性程度较好。其 Young's S-stress 和克鲁斯卡压力值分别达到了 0.00383 和 0.00443，决定系数 RSQ 为 0.9992，重构可信度较高，总变异中能够被相对空间距离解释的比例较大。上述 3 个用户网络浏览的隐式反馈行为特征重构过程数据总结如表 2 所示。

通过对 30 个用户网络浏览过程中的上述 7 种隐式反馈行为进行分析，具有相同的结论。点击上翻页键时间、点击下翻页键时间和鼠标移动时间的隐式反馈信息冗余性较大，尤其是上翻页键时间、点击下翻页键时间。而页面驻留时间、页面滚动次数和鼠标点击次数的隐式反馈信息冗余性较小。因此，通过用户的网络浏览隐式反馈信息推断其对网络信息的兴趣时，可保留页面驻留时间、页面滚动次数、鼠标点击次数、鼠标移动时间四个维度。

## 5. 结论

伴随着用户的网络浏览过程，存在海量的隐式反馈信息，如能合理地在推荐系统中应用此类信息，将提高用户获取信息的效率。本文基于多元尺度分析理论，研究了部分网络浏览隐式反馈信息的相似度和冗余性。页面驻留时间、页面滚动次数、鼠标点击次数、鼠标移动时间存在较弱的相关性，作为重要隐式反馈信息应得到保留，且在推荐方法应占较大的权重，以提高推荐系统的准确性和可靠性。

## 基金项目

河北省自然科学基金(F2015402108)；河北省教育厅科学研究计划(QN20131152)；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.
- [3] Kaiyan, N. (1996) Exploratory Study of Implicit Theories in Human Computer Interaction. *Proceedings of the 6th Australian Conference on Computer-Human Interaction*, Hamilton, 24-27 November 1996, 338-339.
- [4] Schmidt, A. (2000) Implicit Human Computer Interaction through Context. *Personal Technologies*, **4**, 191-199.
- [5] Wilson, A. and Oliver, N. (2005) Multimodal Sensing for Explicit and Implicit Interaction. *Proceedings of the 11th International Conference on Human-Computer Interaction*, Las Vegas, 22-27 July 2005, 1-10.
- [6] 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.
- [7] 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.
- [8] 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*, Barcelona, 26-30 September 2010, 47-51.

- [9] Hu, Y., Koren, Y. and Volinsky, C. (2008) Collaborative Filtering for Implicit Feedback Datasets. *Eighth IEEE International Conference on Data Mining*, Pisa, 15-19 December 2008, 263-272.
- [10] 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]. 杭州: 浙江大学, 2015.
- [11] 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]. 杭州: 浙江工业大学, 2015.
- [12] 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]. 哈尔滨: 哈尔滨工业大学, 2015.
- [13] 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]. 南京: 南京大学, 2016.

**期刊投稿者将享受如下服务:**

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

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

期刊邮箱: [csa@hanspub.org](mailto:csa@hanspub.org)