

情景感知的移动用户行为转移模式推荐算法

张晓滨, 李园园

ZHANG Xiaobin, LI Yuanyuan

西安工程大学 计算机科学学院, 西安 710048

College of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China

ZHANG Xiaobin, LI Yuanyuan. Recommendation algorithm of behavior transfer pattern of mobile user based on context awareness. Computer Engineering and Applications, 2016, 52(20): 163-166.

Abstract: In view of the context awareness feature of the mobile user's behavior sequence, this paper proposes a behavior transfer pattern recommendation algorithm MPRC. This algorithm uses Apriori to filter the data of user's historical behavior in order to obtain a frequent pattern with length 2, and then converts these data into the decision tables for using rule extraction method of rough set to process the decision tables to mine context awareness transfer pattern. Finally, MPRC uses pattern matching method and contextual similarity calculation to rank the pattern and to make recommendation. Experimental results show that this algorithm is more effective and more accurate in user behavior pattern mining and recommendation in mobile environment.

Key words: context awareness; behavior transfer pattern; contextual similarity

摘 要: 针对移动用户行为序列的情景感知特性, 提出一种基于情景感知的行为转移模式推荐算法 MPRC。该算法首先采用 Apriori 对用户历史行为数据进行长度为 2 的频繁模式的挖掘过滤, 然后将过滤后的行为数据转换成决策表, 采用粗糙集规则提取对决策表进行处理, 挖掘情景转移模式, 最后通过模式匹配及情景相似性计算进行推荐排序。实验结果证明了该算法在移动环境下的模式挖掘及推荐方面的有效性和较高的准确性。

关键词: 情景感知; 行为转移模式; 情景相似性

文献标志码: A **中图分类号:** TP311.13 **doi:** 10.3778/j.issn.1002-8331.1603-0063

1 引言

传统的频繁模式挖掘算法是一种以用户历史记录出现频度为基础, 挖掘行为记录间关联规则的模式挖掘算法, Apriori^[1]是其经典算法。然而针对移动个性化推荐提出的情景感知新特性^[2], 传统的频繁模式挖掘算法并未考虑移动环境下用户行为的情景敏感性, 为了提高移动行为转移模式推荐准确性, 必须以“情景计算”^[3-4]为关键, 将多情景属性引入用户序列行为模式挖掘中^[5]。文献[6-9]将位置、环境等多维情景信息引入移动个性化推荐中, 考虑了移动环境下多维情景属性对用户决策的影响, 提高了推荐准确性, 但是提出的情景计算方法只是对情景数据进行简单聚合, 且对情景属性的选择比较笼统, 没有一定的选择标准, 没有深入研究各情景要素

对用户行为决策影响的不同程度。本文针对移动个性化推荐中“情景计算”仍存在的问题, 提出一种将“情景感知”贯穿于行为转移模式推荐的各个阶段的移动用户行为转移模式推荐算法 MPRC (Mobile Personalized Recommendation based on Context awareness), 此推荐算法分为两个步骤, 第一步挖掘基于情景感知的行为转移模式, 第二步对转移模式进行基于情景相似性的模式排序。

MPRC 算法的第一个步骤是: 针对文献[10]中对历史签到记录进行情景约束处理并进行模式转换生成的决策表, 采用粗糙集属性约简及规则提取方法对多情景属性进行约简并对决策表进行情景感知的行为转移模式的挖掘。由于多情景属性实时的更新, 快速增长的情

基金项目: 陕西省教育厅科学研究计划 (No.14JK1307)。

作者简介: 张晓滨 (1970—), 男, 副教授, 研究领域为数据挖掘技术、个性化技术与应用; 李园园 (1989—), 女, 在读研究生, 研究领域为个性化服务技术与应用。

收稿日期: 2016-03-07 **修回日期:** 2016-05-03 **文章编号:** 1002-8331(2016)20-0163-04

CNKI 网络优先出版: 2016-05-27, <http://www.cnki.net/kcms/detail/11.2127.TP.20160527.1530.052.html>

景数据带给多情景属性挖掘巨大的数据压力,然而各情景属性并不全都对行为转移模式有影响,且各用户行为转移模式的情景敏感性也是有差异的,以属性重要度为依据的多情景属性的约简去除了不影响用户行为转移的情景属性,减轻了数据压力。基于粗糙集规则提取的转移模式的挖掘能够发现影响用户行为转移模式的关键属性,挖掘基于不同重要度的多情景属性组合的行为转移模式。

MPRC 算法的第二个步骤是对推荐规则的排序。推荐排序的准确性决定了用户的需求是否准确得到了满足,是否能够建立稳健的用户忠诚度。针对移动用户的实时需求,文献[11]提出一种计算当前上下文与搜索结果标签的相似性来进行推荐排序的方法,文献[12-14]针对移动环境下用户的情景感知特性,提出根据情景相似性计算结果进行推荐排序的方法,文献[14]还提出了各情景属性在推荐系统中具有不同的权重,以上排序方法根据移动用户不同的业务需求引入了不同的情景相似性计算方法,提高了推荐排序准确性。本文提出的情景感知的行为转移模式的排序是以移动环境下用户的情景相似性计算为关键,以决策信息系统的属性重要度作为属性权重计算情景相似性,这样才能得出符合移动用户当前场景的模式排序。

2 情景感知的行为转移模式的挖掘

实时更新的情景数据给情景计算带来了巨大压力,为了简化情景计算复杂性,在行为转移模式挖掘阶段提出先利用“时间”、“位置”两种情景属性将签到记录使用文献[10]的序列项约束处理算法进行数据的预处理,然后用频繁模式挖掘算法 Apriori 进行长度为 2 的频繁模式的挖掘过滤,再对挖掘出的频繁模式以文献[10]的生成决策表的模式转换方法进行决策表的模式转换,再将多情景属性中不影响用户行为决策的属性进行约简,最后以约简后的情景属性挖掘基于多情景属性的行为转移模式。

2.1 属性约简及规则提取步骤

属性约简及规则提取采用粗糙集理论的区分矩阵进行处理,其具体步骤如下所述:

步骤 1 以公式(1)计算条件属性 C 和决策属性 D 的 $w(x_i, x_j)$ 矩阵,计算公式如下^[15]:

$$w(x_i, x_j) = \begin{cases} 1, & x_i, x_j \text{ 不全属于 } Pos_c(D) \text{ 或} \\ & (x_i, x_j) \in Pos_c(D) \text{ 且 } (x_i, x_j) \notin IND(D) \\ 0, & \text{其他} \end{cases} \quad (1)$$

步骤 2 以公式(2)计算差别矩阵 M_s ^[15]。其 M_s 中各元素计算公式如下:

$$m_{(i,j)} = \begin{cases} \{a | \forall a \in C, f(x_i, a) \neq \phi, f(x_j, a) \neq \phi, \\ & f(x_i, a) \neq f(x_j, a) \text{ 且 } w(x_i, x_j) = 1\} \\ \phi, & \text{其他} \end{cases} \quad (2)$$

步骤 3 以公式(3)计算 M_s 的差别函数 $f(A_s)$, 以 $f(A_s)$ 的极小析取式的合取式集合作为约简的属性集合 C' , 根据析取式与合取式演算规则,在对差别函数 $f(A_s)$ 进行计算时可对差别矩阵中的元素进行直接化简,其元素化简规则为: $x_{i,j}$ 与 $y_{i,j}$ 是矩阵 M_s 的两个任意元素,如果 $x_{i,j} \subseteq y_{i,j}$, 则将元素 $x_{i,j}$ 置为空。 $f(A_s)$ 定义^[15]如下:

$$f(A_s) = \prod_{m(i,j) \neq \phi} \sum m(i,j) \quad (3)$$

步骤 4 对约简后的决策信息表进行规则提取,并对规则进行组合、约简。

2.2 情景感知的行为转移模式挖掘步骤

输入 历史签到记录 $I = \{I_1, I_2, \dots, I_n\}$

输出 行为转移模式集 $\{rule_i\}$

其中 $rule_i = \{X_i\{C_1, C_2, \dots, C_n\} \rightarrow X_{i+1}\}$, X 是兴趣点, $\{C_i\}$ 是情景属性集合, $X\{C_i\}$ 表示由兴趣点 X 和描述用户在该兴趣点的当前情景属性集合 $\{C_i\}$ 组成的行为项, $X_i\{C_1, C_2, \dots, C_n\}$ 是前件行为项, X_{i+1} 是决策行为项。

步骤 1 数据清理。以文献[10]的序列项约束处理方法对历史记录 I 进行数据清理,生成事务数据 D 。

步骤 2 用 Apriori 算法对 D 进行长度为 2 的频繁模式的挖掘,确保将要挖掘的基于多情景属性的模式是频繁的。

步骤 3 生成决策表。以文献[11]的生成决策表的模式转换方法将长度为 2 的频繁模式集中的所有模式转换成决策表中的对象,生成决策信息系统 $IS = \{U, R, V, f\}$, 且 $U = \{X_i\}$ 。

步骤 4 使用 2.1 节中属性约简及规则提取步骤进行属性约简并挖掘基于前一多情景属性影响的行为转移模式 $rule$ 。

步骤 5 输出行为转移模式集 $\{rule_i\}$ 。

在步骤 2 的频繁模式挖掘中需注意的,由于情景感知的模式挖掘要求用户行为具有序列相关性,即在计算形如模式 $A \rightarrow B$ 与 $B \rightarrow A$ 时,由于顺序不同,要看做不同的模式计数,这也是与类 Apriori 算法的不同之处。

3 基于情景相似性的模式排序

基于情景相似性的模式排序的核心思想是:将影响用户行为转移的多情景属性的权重与模式的置信度、支持度相结合计算其与待推荐模式的情景相似性,并将与当前情景极为相似的模式排序结果推荐给用户。

多情景属性对用户行为转移模式的权重值可以用粗糙集中的依赖决策行为的属性重要度表示,用户 U_s 的一个目标行为项表示 $X_i\{C_m\}$, 其属性集合中的各属性重要度表示为 $\{S(C_1), S(C_2), \dots, S(C_m)\}$, 已有模式 $rule_i = \{X_i\{C_1, C_2, \dots, C_n\} \rightarrow X_{i+1}\}$, C_j 是 $rule_i$ 中任意一个属性,且 $1 \leq j \leq n$ 。如果 $C_j \in \{C_m\}$, 则模式 $rule_i$ 中的属性 C_j 的

权重计算公式表示如下:

$$wc_j = \frac{S(C_j)}{\sum_{i=1}^m S(C_i)} \tag{4}$$

则带属性权重的模式 $rule_i$ 可写为:

$$rule_i = \{X\{C_1(wc_1), C_2(wc_2), \dots, C_m(wc_m)\} \rightarrow X_{i+1}\} \tag{5}$$

模式集 $\{rule_i\}$ 中各模式排序按照规则各属性权重、规则置信度、支持度乘积作为排序的分数标准,其排序分数计算公式如下所示:

$$score = \sum_{j=1}^m wc_j \times cf \times support_x \tag{6}$$

基于情景相似性的模式排序算法是以带权重的模式为处理对象,利用模式排序得分公式将各属性权重、规则置信度、支持度考虑到模式排序中得出排序的计算公式。

基于情景相似性的模式排序算法的思路及主要步骤,如下所述:

输入 目标客户行为 $X_i\{C_1(S(C_1)), C_2(S(C_2)), \dots, C_m(S(C_m))\}$, 行为转移模式集 $\{rule_i\}$

输出 排序后的模式集 $rule_x$

步骤1 计算模式支持度和置信度。

步骤2 在模式集 $\{rule_i\}$ 中搜索模式前件为 $X_i\{C_n\}$ 的所有模式(搜索的模式的前件的兴趣点与目标用户行为项的兴趣点相同),并按照其模式后件的不同分组为 Gp_1, Gp_2, \dots 。

步骤3 按照公式(4)计算各组中模式的情景属性权重值,将所有模式转换成带属性权重的模式。

步骤4 根据模式排序得分公式(6)计算各组中模式的排序得分,并将各组中的所有模式分数加和,作为该转移模式的排序得分。

步骤5 根据各组排序分数,生成模式排序结果 $rule_x$ 并输出 ($x=1, 2, \dots$)。

4 实验分析

以某市区 100 名志愿者近三月来的历史签到记录作为实验数据,将志愿者三个月来的签到记录及情景数据按等差数列分为数据量大小依次减小的 10 组(1 组到 10 组数据量依次减小)进行实验,分别使用文献[7]基于多维情景影响的推荐算法(算法 1)和基于情景感知的个性化推荐算法 MPRC 对 100 名志愿者进行用户行为转移模式的挖掘,并对用户在 30 天内对两种算法推荐结果的采纳记录进行归纳总结,分析比较两种算法的准确性。

首先根据文献[10]的情景约束的预处理算法进行数据预处理,根据分析签到历史记录特征,设置其参数值 $C_{\max - next} = 80 \text{ min}$, $C_{\min - next} = 5 \text{ min}$, $C_{\max - whole} = 240 \text{ min}$,

$C_{\min - whole} = 10 \text{ min}$, $\delta = 0.38$ 分别采用算法 1 和 MPRC 算法对实例数据进行行为转移模式的挖掘并生成各自的行为转移模式知识库,再以 30 天用户的需求进行模式推荐,其算法在不同数据集上的执行时间比较图如图 1 所示。

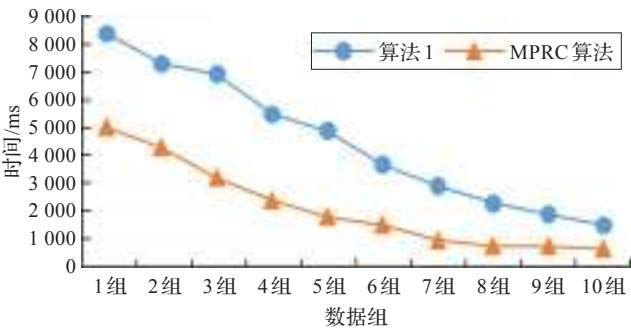


图1 算法执行时间

由图 1 可得:无论是在较大数据集还是较小数据集中,基于情景感知的行为转移模式挖掘算法 MPRC 相比于多维情景影响的推荐算法的执行时间短很多,更高效。

用户对推荐排序模式的采纳情况可以反映推荐算法的准确性,采纳的规则越是靠近推荐排序模式的前面,则推荐越准确。如果一次推荐结果共有排序模式 k 项,则排序为 $p(1 \leq p \leq k)$ 的规则的用户采纳权重公式如下:

$$s_p = \frac{k - p + 1}{k} \tag{7}$$

如果共有 N 次推荐,则 N 次推荐的平均准确率 $AverageP$ 的计算公式如下:

$$AverageP = \frac{1}{N} \sum_{i=1}^N s_p \tag{8}$$

其中 s_p 是每一次推荐中用户采纳模式的权重值。

针对算法 1 和 MPRC 算法对 30 天用户的需求进行模式推荐,并根据用户在 30 天内的模式采纳记录计算其推荐的平均准确率,其算法在不同数据集上的准确性比较图如图 2 所示。

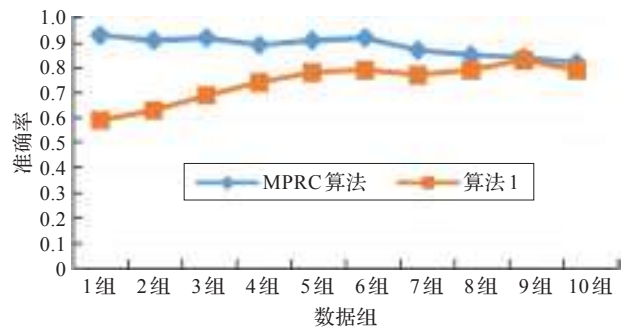


图2 准确率比较图

由图 2 可得:在较大数据集上,MPRC 算法的准确性明显高于算法 1,然而在较小数据集上,MPRC 算法与算法 1 的准确性几乎相等,且都维持在较高水准。

5 结束语

针对移动环境下的情景感知特性,MPRC算法挖掘出基于情景感知的转移模式,利用情景相似性进行了模式的排序,并且在模式挖掘过程中对多情景属性进行了约简,提高了情景感知的模式挖掘的效率,通过研究多情景属性对转移行为的不同影响度,提高了情景感知的移动个性化推荐的准确性。

参考文献:

- [1] Agrwal R, Srikan R. Fast algorithms for mining association rules in large databases[C]//Proceedings of the 20th International Conference on Very Large Databases. San Francisco: Morgan Kaufmann Publishers, 1994: 487-499.
- [2] Zhu H, Chen E, Xiong H, et al. Mining mobile user preferences for personalized context-aware recommendation[J]. ACM Transactions on Intelligent Systems & Technology, 2014, 5(4): 1-27.
- [3] Mehra P. Context-aware computing: beyond search and location-based services[J]. IEEE Internet Computing, 2012, 16(2): 12-16.
- [4] Perera C, Zaslavsky A, Christen P, et al. Context aware computing for the Internet of things: a survey[J]. IEEE Communications Surveys & Tutorials, 2014, 16(1): 414-454.
- [5] Chen T S, Chou Y S, Chen T C. Mining user movement behavior patterns in a mobile service environment[J]. IEEE Transactions on Systems Man & Cybernetics Part A Systems & Humans, 2012, 42(1): 87-101.
- [6] 曹怀虎, 朱建明, 潘耘, 等. 情景感知的P2P移动社交网络构造及发现算法[J]. 计算机学报, 2012, 35(6): 1223-1233.
- [7] Hosseini-Pozveh M, Nematbakhsh M, Movahhedinia N. A multidimensional approach for context-aware recommendation in mobile commerce[J]. International Journal of Computer Science and Information Security, 2009, 3(1): 86-91.
- [8] Baltrunas L, Ludwig B, Peer S, et al. Context relevance assessment and exploitation in mobile recommender systems[J]. Personal & Ubiquitous Computing, 2012, 16(5): 1-20.
- [9] Weng S S, Lin B, Chen W T. Using contextual information and multidimensional approach for recommendation[J]. Expert Systems with Applications, 2009, 36(2): 1268-1279.
- [10] 张晓滨, 李园园, 郭斌. 基于情景感知与约束的移动用户序列行为研究[J]. 计算机工程与应用, 2015, 51(19): 138-140.
- [11] Han J, Woo W, Schmidtke H R, et al. Adaptive content recommendation for mobile users: ordering recommendations using a hierarchical context model with granularity[J]. Pervasive & Mobile Computing, 2014, 13(4): 85-98.
- [12] Liu Liwei, Mehandjiev N, Xu D L. Context similarity metric for multidimensional service recommendation[J]. International Journal of Electronic Commerce, 2013, 18(1): 73-104.
- [13] Zheng Y, Mobasher B, Burke R. Integrating context similarity with sparse linear recommendation model[J]. Lecture Notes in Computer Science, 2015, 9146: 370-376.
- [14] Zheng Y, Burke R, Mobasher B. Recommendation with differential context weighting[J]. Lecture Notes in Computer Science, 2013, 7899: 152-164.
- [15] Yao Y, Zhao Y. Discernibility matrix simplification for constructing attribute reducts[J]. Information Sciences, 2009, 79(7): 867-882.

(上接153页)

- [6] Li Qi, Zheng Jinsong, Tsai A, et al. Robust endpoint detection and energy normalization for real-time speech and speaker recognition[J]. IEEE Trans on Speech and Audio Processing, 2002, 10(3): 146-157.
- [7] Wu Yadong, Li Yan. Robust speech/non-speech detection in adverse conditions using the fuzzy polarity correlation method[C]//Proc of IEEE International Conference on Systems, Man, and Cybernetics, 2000: 2935-2939.
- [8] Chen Shihuang, Wang Jhingfa. A wavelet-based voice activity detection algorithm in noisy environments[C]//Proc of the 9th IEEE International Conference on Electronics, Circuits and Systems, 2002: 995-998.
- [9] Pan Nenghuang, Yu Mingshing, Wu Mingjer. A Mandarin intonation prediction model that can output real pitch patterns[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, 2003, 1: 496-499.
- [10] Huang N E, Shen Zheng, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-station time series analysis[C]//Proceedings of the Royal Society of London, 1998: 903-995.
- [11] Couvreur L, Couvreur C. Wavelet-based non-parametric HMM's: theory and applications[C]//IEEE International Conference on Acoustics, Speech and Signal Processing-Proceeding, 2000, 1: 604-607.
- [12] Li Manman, Yang Hongwu, Hong Ni, et al. Endpoint detection based on EMD in noisy environment[C]//Proceedings of 6th International Conference on Computer Sciences and Convergence Information Technology, 2011: 783-787.
- [13] 李晔, 崔慧娟, 唐昆. 基于能量和鉴别信息的语音端点检测算法[J]. 清华大学学报: 自然科学版, 2006, 46(7): 1271-1277.
- [14] Kopsinis Y, McLaughlin S. Development of EMD-based denoising methods inspired by wavelet thresholding[J]. IEEE Transactions on Signal Processing, 2009, 57(4): 1351-1362.
- [15] Cover T, Thomas J. Elements of information theory[M]. New York: John Wiley & Sons, Inc, 1991.