

基于情境感知的个性化推荐算法^①

时念云, 李秋月

(中国石油大学(华东) 计算机与通信工程学院, 青岛 266580)

摘要: 随着互联网的快速发展, 只涉及用户和项目的传统个性化推荐已不能满足推荐要求的效率和准确率. 因此, 情景感知个性化推荐服务引起了广泛关注, 成为新的研究热点. 本文分析了情境的定义、情景感知个性化推荐模型, 并提出了一种基于情境信息降低维度的关联规则推荐模型. 最后, 以视频网站的 web 日志为数据源, 融合时间情境因素, 实现了基于时间情境划分的关联规则推荐算法, 并和传统推荐算法进行对比分析, 实验证明, 情境感知推荐算法具有更高的准确率和召回率.

关键词: 时间情境; 情境感知推荐; 关联规则推荐

引用格式: 时念云, 李秋月. 基于情境感知的个性化推荐算法. 计算机系统应用, 2017, 26(9): 135-139. <http://www.c-s-a.org.cn/1003-3254/5931.html>

Personalized Recommendation Algorithm Based on Context-Aware Technology

SHI Nian-Yun, LI Qiu-Yue

(College of Computer and Communication Engineering, China University of Petroleum, Qingdao 266580, China)

Abstract: With the rapid development of Internet, the traditional personalized recommendation involving only users and projects cannot meet demands in efficiency and accuracy of the recommendation. Therefore, context-aware recommendation has drawn wide attention and become a new research hotspot. This paper analyzes the definition of context and the model of context-aware recommendation. It also proposes an association rule recommendation model based on context information which reduces the number of dimensions. The data source of experiments is web log. Finally, this paper combines the temporal context and implements the association rule recommendation algorithm based on the temporal context partition. Compared with the traditional recommendation algorithm, the results of experiments show that the context-aware recommendation algorithm has higher accuracy and recall rate.

Key words: temporal context; context-aware recommendation; association rule recommendation

1 引言

随着电子商务的快速发展, 有越来越多的商品面向消费者, 如何快速选择出最佳商品是消费者现在面临的重要问题. 个性化推荐的出现有效地解决了这一类问题^[1]. 传统个性化推荐算法有协同过滤推荐、关联规则推荐和基于内容的推荐等^[2]. 这些传统推荐算法往往只关注“用户-项目”之间的二元关系, 只是把最相关物品推荐给用户, 却忽略了相关情境信息的影响. 而现实生活中, 用户当前所处环境对其兴趣有很大影响, 如

在考虑了温度因素后, 旅游推荐系统在冬季推荐的度假地和夏天推荐的度假地大不相同.

Panniello U^[3]等人在研究电子商务购物推荐的时候, 考虑了节日等情境因素, 在真实数据集上实现了情境感知个性化推荐和传统协同过滤个性化推荐, 并进行了对比分析. 实验结果表明情境感知个性化推荐的结果在所有评价指标上均优于传统二维协同过滤算法. 胡文^[4]等针对商场广告费用日益增长的问题, 提出了情境感知商品推荐信息的研究方案. 该方案通过分析时

^① 收稿时间: 2016-12-21; 采用时间: 2017-01-12

间情境因素,得到用户对某件商品感兴趣的程度,以便快速为用户提供个性化推荐服务,减少了商家营销成本.因此,情境感知推荐不仅能够提高推荐精度,而且能够节省商家成本,可广泛用于电子商务、旅游向导、视频和小说网站等不同领域.

基于此,本文在对情境感知个性化推荐分析的基础上,把 web 日志中时间属性作为个性化推荐中的情境因素.然后分析时间情境在个性化推荐中的影响,并和传统推荐算法结果进行对比分析.

2 情景感知个性化推荐

2.1 情境分析

情境的复杂性和多面性导致情境有 150 种不同领域的定义^[5].很难找到一个能适用于任何学科的情境定义.如在个性化电子商务中,有时把客户的购买意向作为情境的定义,不同购买意向导致不同购买行为. Baltrunas L^[6]等把用户年龄作为情境因素,按照年龄分类,得出不同年龄阶段的用户所看电影类型不同.

Katarya R 等人^[7]把时间情境分为周末和工作日,观看电影的陪同人员分为家人、普通朋友和男、女朋友等.同样也得出在不同情境下,用户所看电影类型不同.因此,不同情境信息针对相同事件,可能引起不同效果.当然,相同事件根据需求不同,也可能需要不同的情境信息.如在视频网站中,如果要找到用户在不同时间喜爱访问的视频类别,需要时间情境.如果要找到不同地区用户喜爱访问的视频类别,那么需要地理情境信息.

本文以我校视频网站为例进行实验分析.我校视频网站缓冲、下载速度快,并且资源丰富、画面清晰.该视频网站包含了综艺类、科幻类和学习类等 13 类视频信息.学生和教职工经常浏览并在线观看这些视频信息,由此产生了许多用户访问日志.这些日志是 IIS 格式,日志中包含请求时间(datetime)、服务器 IP 地址(s-ip)、请求方式(method)、访问网页的 URL(url)、用户 IP 地址(u-ip)、浏览器操作系统(agent)和返回的 http 状态码(status)等数据属性.原始数据记录如表 1 所示.

表 1 原始数据记录

#	日志记录
1	2014-03-21 00:48:40 2001:da8:7007:102::244 GET /skin6/index.asp 2001:da8:7007:d21:ad2d:a586:e678:a2e7 Mozilla/5.0+(Windows+NT+6.1)+(...) 200
2	2014-03-21 01:46:21 121.251.255.244 GET /skin6/film_sort.asp id=11 211.87.191.85 Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+6.1;)+(...) 200
3	2014-03-22 02:21:16 2001:da8:7007:102::244 GET /skin6/film_view.asp id=5555 2001:da8:7007:328:3dcd:77cb:f74b:d91b Mozilla/5.0+(Windows+NT+6.1)+(...) 200

分析表中的 Web 日志,可知时间属性的记录清晰明了,以 datetime=2014-03-22 02:21:16 为例,从 date=2014-3-22 就可以查到该时间是否为周末,或者是否是上课时间,根据日历查询可知 22 日为周六.周末期间,学生看的视频类别肯定和上课时间看的略有不同.学校的上课时间为上午 8:00-12:00,下午 14:00-18:00,从 datetime 属性里截取 time=02:21:16,这样就可以判断出这不是上课时间,而是下课时间.

时间情境因素可以根据学生课上还是课下进行划分,当然课下时间包括周六和周日,还有周一到周五没课的时间.也可以根据工作日和非工作日进行划分.不同时间情境下,学生浏览的视频信息可能会有改变.如上课时间,学生肯定会浏览一些教学类视频,而在课下时间,为了放松心情,学生可能会看一些喜剧类视频.这些变化不仅包含了访问的内容和路径,还包含了喜

爱某类视频的程度.比如,传统方法只能挖掘出学生用户喜欢看学习类和电影类的视频,但基于时间情境的挖掘可以进一步挖掘出学生用户在周末更喜欢电影类视频,在课上时间更喜欢学习类视频.

2.2 情境感知推荐

传统个性化推荐模型只考虑了用户和项目之间的关系,而情境感知推荐模型是考虑了情境、用户和项目之间的关系,是一个三维数据模型.针对情境信息应用在推荐过程中的不同阶段, Adomavicius 等人^[8]将现有的情境感知推荐方法分成了三类:情境预过滤(Contextual Pre-Filtering)、情境后过滤(Contextual Post-Filtering)和情境建模(Contextual Modeling).情境预过滤就是用情境信息选择或者构建相关性最强的二维信息.情境后过滤就是在进行推荐时忽略情境信息,然后在评分排名推荐时,使用情境信息调整推荐排名.

情境建模方法就是构建多维模型,然后计算得出推荐结果.当然,为了提高推荐准确率,也可以把多种推荐方法或者推荐模型结合.杨君^[9]等就是采用的情境预过滤模式,为了进一步提高推荐准确率,对评分数据进行情境相似度计算,找到相似情境信息.因此,她们不但根据情境信息进行筛选,而且还考虑了那些相似情境信息.曾子明等^[10]主要是采用了情境后过滤的推荐模式,提高了推荐准确度.冯鹏^[11]在情境感知推荐中采用了基于矩阵分解的协同过滤推荐算法,并用真实的电影评分数据集进行了实验分析,验证了预测精度的提高.

基于此,本文提出一种基于时间情境的关联规则个性化推荐算法.该算法在关联规则推荐算法的基础上引入情境信息,建立了三维数据模型.实现多维关联规则推荐算法的基本步骤是降低维度.因此,本文采用了以时间情境信息为标准的降维策略.最后,本文以视频网站的 web 日志数据为数据源,进行了实验.实验证明,相比于传统推荐方法,本文提出的算法具有更高的准确度.

3 算法描述

情境感知关联规则推荐算法就是利用情境因素和关联规则推荐算法建模实现的,关联规则推荐算法中最典型和最常用的就是 Apriori 算法.该算法用于发现事务与事务之间的联系,不仅可以挖掘出关联程度比较强的访问路径,同时还可以得到用户对项目的喜爱程度.如在学校视频网站中,学生访问完喜剧类视频后,会更喜欢访问综艺类视频.同时,也可得到学生喜爱某个或某类视频的程度.支持度(S)和置信度(C)是 Apriori 算法的两个重要衡量指标. Apriori 算法可以很简单的通过连接步和剪枝步实现,是一种挖掘布尔关联规则的算法.但是,该算法需要不断访问数据库里大量数据集,造成时间开销较大.

基于情境感知的关联规则推荐算法模型可以形式化描述为:

$$\text{Context} * \text{User} * \text{Item} \rightarrow \text{Rating} \quad (1)$$

这是一个基于“情境、用户、项目”的三维模型.要想实现基于情境的 Apriori 算法来进行推荐,必须先降低维度.因此,可以按照相似性很大或者比较明确的情境信息进行筛选、降维.实现过程就是先基于某个情境信息找到相关项目和用户集合,然后进行关联规

则推荐算法,得到推荐结果.通过情境信息划分的方法不仅可以实现推荐算法,还可以改进 Apriori 算法中时间开销大的缺陷.

本文就是根据时间的情境信息进行划分,完成推荐算法.如,可以把周末的时间记为一个划分域标记,在周末,学生可能更喜欢看一些让人放松的电视剧或者电影之类的视频.事务数据集记为 $I(I_1, I_2, \dots, I_n)$, 把时间情境信息标记记为 $R(R_1, R_2, \dots, R_n)$. 因此,可以划分成 n 个集合,根据情境信息划分成不同集合,分别完成关联规则推荐算法,得到频繁项集集合 $F(F_1, F_2, \dots, F_n)$. 对比分析这些频繁项集以及所对应的强关联规则.

最小支持度为 $\min S = P(X \cup Y)$, 基于时间情境划分的关联规则算法部分过程如下.

算法: 基于情境划分的算法描述

输入: $I, R, \min S$

输出: 频繁项集 F

(1) $I = \{I_1, I_2, \dots, I_n\}, R = \{R_1, R_2, \dots, R_n\}$; // R 情境信息集合

(2) while $i < n$ do begin // i 为情境因素划分集合的标记

(3) $\{IR = [IR_1, IR_2, \dots, IR_n]\}$; // 根据情境 R 划分子块

(4) $L_1 = \text{FrequentItemsets}(IR_i)$; // 对每个集合内项目找到频繁一项集

(5) $m = \text{length}(IR_i)$;

(6) for ($k=2; L_{k-1} \neq \text{null}; k++$) {

(7) $C_k = \text{apriori_gen}(L_{k-1})$; // 找到候选项集

(8) While $j < \text{length}(IR_i)$; // j 标记每个子块中的每一项

(9) $\text{count}++$; // 为候选项集中每项生成一个计数器 count

(10) $L_k = \{c \in IR_i | \text{count} \geq \min S\}$; // 判断支持度是否大于最小支持度

(11) $F = \{F_i\}$; // 输出每个情境下频繁项目集的集合 F_i

4 算法实现和结果分析

本文从我校视频网站的服务器里随机截取了 3 个月的数据作为实验数据. 根据这些数据进行的实验分别为: 标准试验, 随机截取数据; 情境感知推荐算法, 对标准实验中的数据按照课上和课下情境因素进行降维, 分别得到推荐结果. 同时, 本文还把时间分为工作

日和非工作日进行实验对比分析。

标准算法实验,从得到的 Web 日志数据中随机截取一个月的数据,原始数据有 232 M,经过数据处理后得到 7204 条有意义的访问路径.进行挖掘得到频繁项集和强关联规则.本文中,最小支持度和最小置信度都设为 1%.表 2 是截取的部分频繁项集和部分关联规则.

表 2 部分频繁项集和规则表

频繁项	支持度(%)	置信度(%)
首页→动作类	10.90	11.00
首页→喜剧类	7.92	8.68
首页→电视剧类	7.54	8.27
首页→科幻类	6.78	7.44
首页→动漫类	6.76	7.41
首页→爱情类	4.84	5.32
首页→战争类	3.60	3.94
首页→学习类	3.59	3.93

由上表排序就可以把热门排行榜推荐给用户.因此,可以把动作类、喜剧类和科幻类等热门视频放在视频网站首页,这样就可以方便用户在首页直接看到自己喜欢的视频类别,不用再花费时间去寻找和浏览.

情境感知实验,根据时间情境降维后,三维模型变成课上和课下二维推荐模型,实现推荐算法,并和标准试验结果对比分析,结果对比图如图 1 所示.图 2 就是工作日和非工作日推荐结果的示意图.

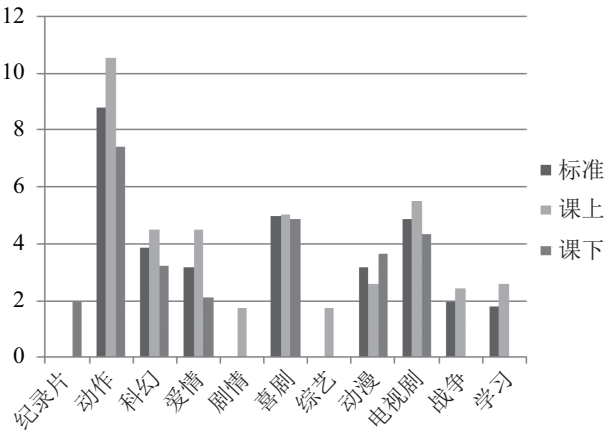


图 1 结果对比示意图

根据上图可知,引入时间情境因素也就是按照课上和课下划分原始数据,把三维模型降低成二维推荐模型,实现推荐算法.由图 1 可知,在三种情况的推荐结果中,排名趋势大体相似.然而,仅针对学习类视频,课下几乎没有用户在浏览和观看.同时,针对纪录片类

视频进行分析可知,纪录片在课下情境推荐结果中出现,表明大家在课下会看一些纪录片类视频.充分体现了不同时间情境下用户所看视频的不同,不同时间情境下推荐结果不同.因此,不同时间情境给学生推荐合理视频信息,提高视频网站的吸引力和竞争力.

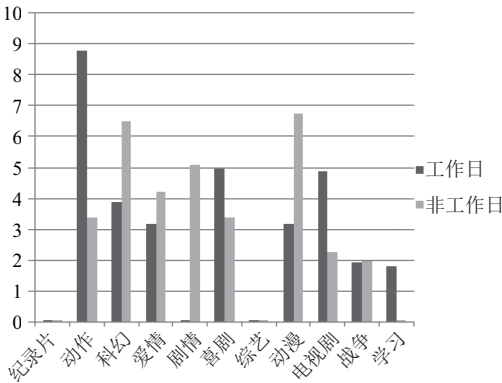


图 2 结果对比示意图

从图 2 中可知,工作日和非工作日情境下推荐的结果有很大不同,更明显的表明了情境因素对推荐结果的重要性.在工作日情境下排名较前的有动作片和喜剧等,而在非工作日情境里动漫和科幻等类的视频排名较前.

根据得到的推荐结果,可以分别计算出标准时推荐结果的准确率和召回率以及情境感知推荐算法的准确率和召回率.前者的准确率和召回率分别为 0.727 和 0.615,而后者分别为 1 和 0.846.因此得出的结论是:引入时间情境后,推荐精度有相对的提高,情境感知推荐算法不仅提高了推荐精度,为用户提供更人性化、个性化服务,同时也为视频网站排列视频提供合理指导.

5 小结

随着电子商务的快速发展,用户购买时要求越来越高,基于用户情境的个性化推荐算法研究也越来越受到重视.不同情境下,用户需求不同.基于情境的推荐系统不仅考虑了用户偏好,而且还把用户情境因素作为推荐的一部分.情境因素用以提高推荐的精度和准确度.本文结合本校视频网站,实现基于时间情境降低维度的关联规则推荐算法.根据时间情境因素,划分为多个情境子集进行关联规则算法推荐.用实际数据进行实验验证,得到理想结果,进行分析,使得推荐结

果得到进一步优化和精确化。但是,随着原始数据集数量的增加,单机环境已经不能满足推荐要求的速度,因此,下一步工作是基于分布式的情境感知个性化推荐算法研究。

参考文献

- 1 马瑞敏,卞艺杰,陈超,等.基于Hadoop的电子商务个性化推荐算法——以电影推荐为例.计算机系统应用,2015,24(5): 111-117.
- 2 王国霞,刘贺平.个性化推荐系统综述.计算机工程与应用,2012,48(7): 66-76.
- 3 Panniello U, Tuzhilin A, Gorgoglione M, *et al.* Experimental comparison of pre- vs. post-filtering approaches in context-aware recommender systems. Proc. of the 3rd ACM Conference on Recommender Systems. New York, NY, USA. 2009. 265-268.
- 4 胡文,李良学.基于情景感知的商场导购推荐商品信息的研究.哈尔滨商业大学学报(自然科学版),2012,28(1): 58-61.
- 5 Bazire M, Brézillon P. Understanding context before using it. Proc. of the 5th International Conference on Modeling and Using Context. Paris, France. 2005. 29-40.
- 6 Baltrunas L, Ricci F. Context-based splitting of item ratings in collaborative filtering. Proc. of the 3rd ACM Conference on Recommender Systems. New York, NY, USA. 2009. 245-248.
- 7 Katarya R, Verma OP, Jain I. User behaviour analysis in context-aware recommender system using hybrid filtering approach. Proc. of the 4th International Conference on Computer and Communication Technology (ICCCCT). Allahabad, India. 2013. 222-227.
- 8 Adomavicius G, Tuzhilin A. Context-aware recommender systems. Ricci F, Rokach L, Shapira B, *et al.* Recommender Systems Handbook. Berlin Heidelberg, Germany: Springer, 2011. 217-253.
- 9 杨君,吴菊华,艾丹祥.一种基于情景相似度的多维信息推荐新方法研究.情报学报,2013,32(3): 262-269.
- 10 曾子明,陈贝贝.移动环境下基于情境感知的个性化阅读推荐研究.情报理论与实践,2015,38(12): 31-36.
- 11 冯鹏程.基于情境感知的个性化推荐算法的研究[硕士学位论文].上海:东华大学,2014.