

面向在线交易日志的用户购买行为可视化分析*

贾若雨¹, 曾 昂¹, 朱 敏¹, 刘汉清¹, 李明召²

¹(四川大学 计算机学院(软件学院), 四川 成都 610065)

²(RMIT University, Melbourne 3001, Australia)

通讯作者: 朱敏, E-mail: zhumin@scu.edu.cn



摘 要: 在线交易日志,即用户通过电商平台购买商品产生的交易记录,包括用户、商品、交易及商家的相关信息,反映了用户的购买行为.现有的可视化方法未能充分结合在线交易日志的时序、层次、地理、多维等特征,实现对用户购买行为的多角度分析.对此,结合交易日志的多个特征,提出了基于径向布局的复合时序可视化方法和融合空间信息的时间轴可视化方法,设计了颜色极值映射方法和规律映射方法;并基于上述方法,设计并实现了面向在线交易日志的用户购买行为可视化分析系统 UPB-VIS,从而完成了单个用户和用户群体购买行为的全方位分析;最后,通过在京东商城在线交易日志数据集上的实验,证明了系统的易用性和相关可视化方法的有效性.

关键词: 可视化分析;用户购买行为;在线交易日志;时序可视化;多维数据可视化

中图法分类号: TP311

中文引用格式: 贾若雨,曾昂,朱敏,刘汉清,李明召.面向在线交易日志的用户购买行为可视化分析.软件学报,2017,28(9): 2450–2467. <http://www.jos.org.cn/1000-9825/5266.htm>

英文引用格式: Jia RY, Zeng A, Zhu M, Liu HQ, Li MZ. Visual analysis of consumer purchasing behavior for online transaction log. Ruan Jian Xue Bao/Journal of Software, 2017, 28(9): 2450–2467 (in Chinese). <http://www.jos.org.cn/1000-9825/5266.htm>

Visual Analysis of Consumer Purchasing Behavior for Online Transaction Log

JIA Ruo-Yu¹, ZENG Ang¹, ZHU Min¹, LIU Han-Qing¹, LI Ming-Zhao²

¹(College of Computer Science, Sichuan University, Chengdu 610065, China)

²(RMIT University, Melbourne 3001, Australia)

Abstract: Online transaction log is a set of commodity trading records generated by electronic commerce (E-commerce) platform. It incorporates information of the consumers, commodities, sellers and transactions that reflect consumer purchasing behavior. The existing visualization methods cannot fully combine the time series, hierarchical, geospatial and multi-dimensional features of online transaction log to perform multi-aspect analysis on consumer purchasing behavior. Combining with multiple features of online transaction log, this paper proposes a composite temporal visualization method based on the radial layout and a timeline visualization method incorporated with spatial information. An extreme color mapping method and an identifiable color mapping method are also designed to support the analysis. UPB-VIS is designed and implemented based on the methods above to realize the comprehensive analysis of consumer purchasing behavior. The usability of the system and the validity of the visualization methods are verified by using JD online transaction log.

Key words: visual analyze; consumer purchasing behavior; online transaction log; temporal visualization; multidimensional data visualization

* 基金项目: 四川省科技支撑计划(2013GZ0015); 四川省学术带头人培养基金(2013DTPY0010)

Foundation item: Sichuan Provincial Science and Technology Support Program (2013GZ0015); Sichuan Provincial Foundation for Leaders of Disciplines in Science (2013DTPY0010)

收稿时间: 2016-10-14; 修改时间: 2016-11-22, 2017-01-15; 采用时间: 2017-02-15; jos 在线出版时间: 2017-03-17

CNKI 网络优先出版: 2017-03-17 14:37:23, <http://kns.cnki.net/kcms/detail/11.2560.TP.20170317.1437.003.html>

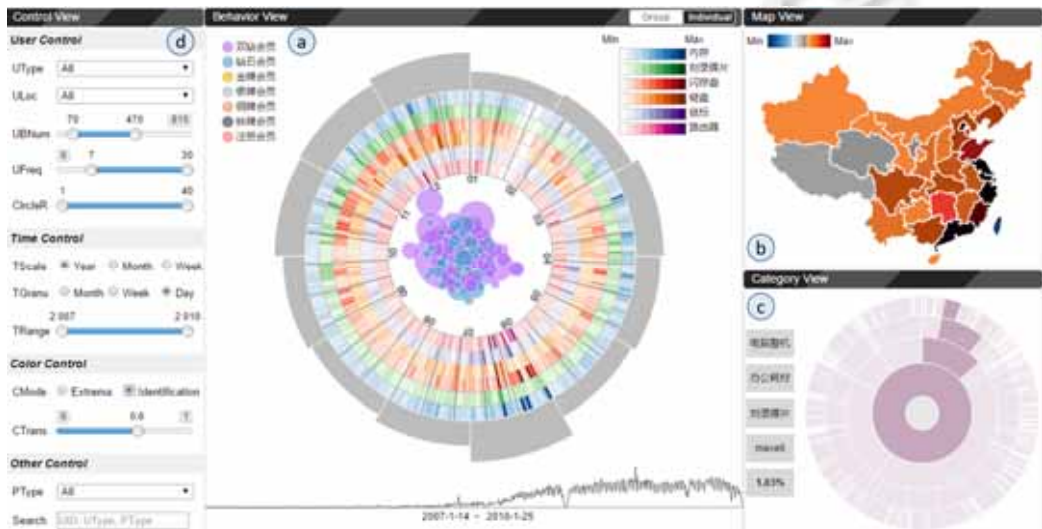
近年来,电子商务行业迅速发展,在线购物已成为国民生活的重要组成部分.用户在线购物的过程中,产生了大量的交易日志,记录了用户、商品、商家及交易的相关信息,反映了用户的在线购买行为.通过对这类数据的有效分析,可发现单个用户或用户群体的行为特征,包括购买偏好、复购行为、购买习惯,以及上述信息随时间的变化情况等.如何理解和分析这类购买行为,成为各大电商平台和电子零售商关注的焦点之一^[1-3],同时也是相关数据分析人员的研究热点^[4].

在线交易日志具有较强的时间关联性,属于时序数据,其包含的商品类别信息具有层次结构和多维属性,而用户所在地区又是地理信息.面对这种具有多重结构特征的复杂数据,如何设计合理的可视化方法来展示这类数据,并分析其蕴含的用户购买行为,是本文面临的巨大挑战.

现有的用户在线购买行为分析系统多采用统计分析方法,如淘宝的数据魔方^[5]、京东的数据罗盘^[6]、亚马逊的 Kinesis^[7]等,这类系统不便于深入分析单个用户或用户群体的购买行为特征,且系统缺乏交互性.

现有的可视化方法多基于交易日志中的时序信息,分析交易模式的时变规律或异常的购买行为,如 Keim^[4], Xie^[8]等.然而,用户的购买行为不仅与时序信息有较强的关联性,还与用户本身的特征(如等级、地域等)及购买的商品有关.现有方法未能充分结合交易日志的时序、层次、地理、多维等特征,实现对用户购买行为的多角度分析.同时,用户是否重复购买商品与电子零售商的利益息息相关,但鲜有关于用户复购行为的可视化分析方法.因此,迫切需要新的可视化分析方法和模型,来直观、多角度地研究在线交易日志和分析用户购买行为.

本文围绕上述问题展开研究工作,设计并实现了面向在线交易日志的用户购买行为可视化分析系统 UPB-VIS.UPB-VIS 从在线交易日志的多个特征出发,融合与用户购买行为相关的多种信息,完成了单个用户及用户群体购买行为的全方位分析.UPB-VIS 主要提供了 3 个可视化分析方法,分别为:(1) 基于径向布局的复合时序可视化方法(如图 1(a)所示),该方法以一种新颖的复合图形可视化布局,结合了散点图、基于像素的圆环排序方法和环状图,并提供多尺度的时间模式和时间粒度及两种颜色映射方法,从而灵活直观地分析用户群体购买行为;(2) 融合空间信息的时间轴可视化方法,该方法能够有效展示带时空信息的数据,并实现时序事件复现模式的分析,如用户的复购行为,从而完成单个用户购买行为的可视分析;(3) 辅助可视化方法,包括多层次可选的旭日图可视化方法(如图 1(c)所示)和地图可视化方法(如图 1(b)所示),主要展示商品的层次结构和用户所在的地理区域,从而实现了对购买行为相关信息的联动分析.



(a) 用户行为分析视图 (b) 地图视图 (c) 商品层次结构视图 (d) 控制视图

Fig.1 UPB-VIS system interface

图 1 面向在线交易日志的用户购买行为可视化分析系统界面示意图

本文的主要贡献如下:

- 提出了一种基于径向布局的复合时序可视化方法,解决了周期时间可视化中固定的时间模式、时间粒度与日志时间跨度不匹配的问题;
- 提出了一种融合空间信息的时间轴可视化方法,能够有效分析事件的复现模式和复现频率;
- 设计了颜色极值映射方法和规律映射方法,分别适用于展示具有幂律分布特点的数据集的整体分布特征和局部分布特征;
- 设计了丰富的交互操作,包括时间控制、颜色控制等,并最终实现了可视分析系统 UPB-VIS.

1 相关工作

1.1 用户在线购买行为的可视化分析

随着用户的消费习惯从线下转移到线上,在线交易数据成为新的价值源泉.一些学者提出了相关的可视化方法、模型和技术,来对其反映的用户在线购买行为进行分析.

Hao^[9]提出了基于像素的条形图(pixel bar charts),以分析交易日志不同属性(如购买时间、购买数量、购买金额总量)之间的关联;同时,通过多个条形图的对比,来分析用户购买行为的规律,并寻找最有价值的用户.但是,条形图表示的维度有限,随着用户数量及交易时间跨度的增加,条形图无法展示大量的用户信息和交易信息.Keim^[4]改进了基于像素的条形图,设计了细胞条形图(value-cell bar charts),用以可视化交易日志的整体概况和细节信息,从而发现对交易金额和交易数量贡献大的用户.然而,细胞条形图未考虑与用户购买行为相关的用户特征(如等级、地域等)和商品特征.

Xie^[8]结合基于像素的方法和时间轴(timeline),提出了 VAET 可视化模型,以发现商品购买记录中典型的交易模式,并通过交易信息的完整性来识别虚假购买行为.然而,该模型只考虑了在线交易日志的时序特征,未能充分结合日志的层次、地理、多维等特征对购买行为进行多角度分析.Liu^[10]基于树图(treemap)和时间轴,设计并实现了 SellTrend 可视化系统来分析机票交易日志,通过对交易时间、航班相关属性(如航空公司、舱位等级)的可视化,发现交易日志中的异常及用户购买机票失败的原因.但是,该系统仅适用于分析用户群体的购买行为,不适用于分析单个用户的购买行为(如复购行为特征等).Hayashi^[11]采用热图(heatmap)与时间轴可视化方法,从不同时间粒度分析用户购买行为规律,但是无法针对用户地理信息、商品层次信息等进行展示.Chang^[12]针对电汇交易日志(wire transaction)提出了 WireVis 可视化模型,该模型结合时间轴、热图和节点-链接(node-link)等可视化方法,对日志包含的账户信息、交易时间、关键词进行可视化,进而发现可疑的交易行为和潜在的诈骗人.由于 WireVis 无法展示用户的地理位置信息和等级信息,故不适用于分析商品交易日志反映的购买行为.Singh^[13]采用多视图可视化方式,协助审计人员发现交易日志中的异常和潜在的欺诈交易,但存在可视化方式简单(节点-链接、饼图、直方图等),不能展示数据当中多维信息、周期信息等问题.

总结上述可视化分析方法,发现:

- (1) 当前对在线交易日志的研究集中于时序信息的可视化,忽略了交易日志包含的商品层次结构信息、多维信息及用户的地理位置信息等,而这些信息与用户的购买行为紧密相关;
- (2) 现有的可视化方法集中于时变规律的分析 and 异常行为的发现,未见针对用户复购行为的研究.

1.2 时序数据可视化

在线交易日志具有很强的时间关联性,属于典型的时序数据(time-oriented).时序数据可视化根据展示方式的不同,分为线性时间可视化方法和周期时间可视化方法^[14,15].

- 时间轴是线性时间可视化中出现最早、应用最广泛的方法之一,其横轴表示线性时间域,纵轴表示对应时间域内的特征属性值.典型的时间轴有 LiveGantt^[16],Lifeline^[17,18],Storyline^[19,20]等.Lifeline^[17]及其改进方法 Lifeline2^[18]使用时间轴技术,对不同病人的病历记录进行可视化,包括发病、治疗、康复、复发的时间点及持续时间;Cloudlines^[21]提供了解决基于时间表示的大规模和动态事件数据集在有限空间

中的时序可视化技术;T ipoVis^[22]基于时间轴可视化方法提出了一种支持相近事件(两个同时或相继发生的事件)的可视化方法.由此可知,现有方法多用以展示事件发生的时间点、持续时间以及事件的演化.然而,事件的复现性(recurrence)在时序事件中普遍存在^[14],如用户重复购买某件商品等,但目前针对事件复现性的研究成果较少.本文提出的融合空间信息的时间轴可视化方法能够有效解决事件复现性的分析问题;

- 周期时间可视化是挖掘时序数据中隐含周期性规律的有效方法^[14],其通常采用循环的时间域,将时间序列沿螺旋轴或圆周排列. Bertini^[23]设计了 SpiralView,在该方法中,每个螺旋回路表示 24 小时,回路的数量代表天数,以此展示网络警报的时间分布情况. Zhao^[24]设计了 Ringmaps,对人类的活动进行分析,其中,一个圆环代表 24 周,圆环上不同的色段表示不同频率的人类活动情况. Visipedia^[25]设计了拥有 12 朵花瓣的花朵作为周期隐喻的事件概览视图. Episogram^[26]结合时间轴可视化方法,分析个体中心社交行为存在的模式,其中,横向时间线表示用户初始化一个事件,纵向时间线表示用户间的交互,并且采用月牙形箭头表示相对时间段. 其他的方法还有 ChronoView^[27], SpiraClock^[28], CircleView^[29], OpinionSeer^[30]等. 综上可知,现有周期时间可视化方法多采用固定的时间尺度(如,一个圆环表示一个月)来挖掘数据集中隐含的周期特征. 然而,部分时序数据集的时间跨度不确定,如,在线交易日志的时间跨度可能在 5 年以上也可能不到一年,故固定的时间尺度与这类数据的时间跨度不匹配. 对此,本文提出了一种基于径向布局的复合时序可视化方法,该方法提供了多尺度的时间模式和时间粒度,以供分析人员根据实际需求进行灵活选择.

2 数据和可视化任务描述

本节主要介绍交易日志样本数据的获取、处理和分析,并阐述系统的可视化任务.

2.1 数据描述

2.1.1 数据来源

在线交易日志包括用户、商品、商家及交易的相关信息.用户的购买行为与用户属性、购买的商品及交易时间有很强的关联性,与商家的关联不大.因此,本文基于交易日志中用户、商品及交易的相关信息来分析用户的购买行为.

本文的样本数据来源于数据堂^[31]提供的京东商城在线交易数据.原始数据包含商品信息、用户信息、交易信息和评价信息.结合系统的分析目标,本文提取了其中的商品信息、用户信息和交易信息,并对数据进行了预处理.预处理后的数据共有 165.2 万条交易记录,时间跨度 4 年,涉及到 31.5 万用户和 1.8 万件商品.

2.1.2 样本数据分析

用户的信息主要包括用户的 ID、昵称、地理位置和用户等级.其中,用户的地理位置涵盖中国大陆 31 个省级行政区及港澳台地区;用户的等级共有 7 级,包括注册会员、铁牌会员、铜牌会员、银牌会员、金牌会员、钻石会员和双钻会员.需要注意的是:用户在购买商品时地理位置和等级会发生变化,如, ID 为“215749”的用户在上海、广东、北京等 11 个地区购买过商品, ID 为“133915”的用户在 3 个月之内由铁牌会员升级为了双钻会员,升了 5 个等级.因此,需要设计有效的可视化方法来展示和分析用户地理位置、等级的变化情况.

商品的信息包括商品的 ID、名称、类别等,其中,商品的类别信息是典型的层次结构,从一级分类到四级分类共有 4 层,第 5 层(叶节点)为商品本身,如日用百货-礼品箱包-烟具-Zippo-Zippo 打火机 207.每一分类下的类别可看做不同的维度,如,一级分类有 4 个维度,四级分类有 1 223 个维度.

交易信息主要包括交易时间、商品 ID、用户 ID、购买数量等.在京东商城的在线交易日志中,每条交易记录默认商品的购买数量为 1.样本数据的时间范围为 2007 年 1 月 4 日~2010 年 1 月 25 日.表 1 为用户购买量的统计结果,通过该表可发现,用户购买量的分布近似符合幂律分布^[32],即:购买量少的用户远远多于购买量多的用户,出现了长尾现象.由于长尾用户的交易记录过少,难以反映其购买行为规律,因此,本文重点分析购买量大于 50 的用户.

Table 1 Distribution of consumer purchasing**表 1** 用户购买分布情况

购买数量(范围)	人数
815	1
300~814	3
100~299	277
50~99	2 083
10~49	41 138
1~9	270 891

2.2 可视化任务

通过与电子零售商和数据分析人员的交流,本文总结归纳了他们关注的核心问题,具体如下.

(1) 用户群体的购买行为分析

- Q.1. 不同的用户群体(如不同地区或等级的用户群体)存在购买偏好吗?或者说存在购买品类的差异吗?
 Q.2. 不同用户群体的购买行为是否存在时间规律?比如,某个地区的用户群体是否会在某个时间集中购买某种商品?
 Q.3. 能否分析某一类特定用户群体的购买特征?比如购买时间规律、购买偏好等?

(2) 单个用户的购买行为分析

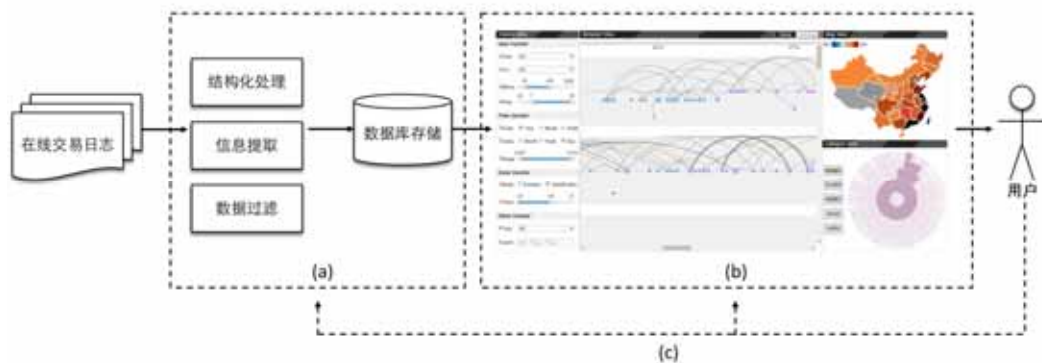
- Q.4. 用户是否会重复购买某一种或某几种商品?如果是,时间间隔是怎样的?
 Q.5. 用户有购买品类偏好吗?或者说用户是热衷于网购电子类的商品还是服饰护肤类的商品?
 Q.6. 用户的网购行为是否存在某种时间规律?比如,在每月固定的几天购买商品或者某种周期性的规律?
 Q.7. 用户网购商品时地理位置会变化吗?如果有,地理位置的变化和购买的商品之间有关联吗?
 Q.8. 用户的等级变化情况是怎样的?与其购买频率、购买的商品之间存在什么关系?
 Q.9. 哪些是有价值、需要重点关注的用户?

(3) 商品的销量分析

- Q.10. 是否能从多个方面(如,区域、时间等)分析某种商品的销量?

3 系统概览

面向在线交易日志的用户购买行为可视化分析系统 UPB-VIS 针对电商平台产生的大量在线交易日志,采用信息可视化和可视分析方法,分析日志中蕴含的用户(单个用户和用户群体)购买行为.系统共 3 个模块,分别是数据处理和存储模块、可视化模块和交互模块,如图 2 所示.



(1) 数据处理和存储模块 (2) 可视化模块 (3) 交互模块

Fig.2 Overview of UPB-VIS system

图 2 UPB-VIS 系统概览图

- 数据处理和存储模块对样本数据进行处理和解析,提取与用户在线购买行为相关的信息,转换为结构化的数据表,包括商品信息表、用户信息表和交易信息表,并存储在数据库中;
- 可视化模块首先将结构化的数据映射为图形符号、时间变量、位置变量和视网膜变量(retinal variables),然后将可视符号转换为可视化视图.系统共有 4 个可视化视图,分别是用户行为分析视图、商品层次结构视图、地图视图和控制视图,其中,用户行为分析视图包括用户群体行为分析视图和单个用户行为分析视图两个子视图;
- 交互模块用以响应用户的交互操作,系统提供了单个视图内的交互和多个视图间的交互,包括时间控制、颜色控制、数据过滤等.

4 用户在线购买行为可视化方法

基于可视化方法的设计原则^[33,34]和可视化任务,本文设计并实现了用户在线购买行为可视化分析的方法,包括基于径向布局的复合时序可视化方法、融合空间信息的时间轴可视化方法、多层次可选的旭日图可视化方法和地图可视化方法.

4.1 基于径向布局的复合时序可视化方法

设计实现基于径向布局的复合时序可视化方法,主要用以分析用户群体的购买行为,主要解决第 2.2 节中可视化任务的 Q.1~Q.3;同时,可从用户群体角度解决 Q.10.该方法提供了多尺度的时间模式和时间粒度,解决了周期时间可视化中固定的时间模式、时间粒度与日志时间跨度不匹配的问题.同时,该方法提供了两种颜色映射方法(详见第 4.2 节),以供分析人员进行灵活选择.

为提高径向布局方法的空间利用率,复合时序可视化方法的中心区域被设计为镂空,如图 3 所示:最内层(图 3,A)为散点图,中间层(图 3,B)为基于像素的圆环排序方法,最外层(图 3,C)为环状图.其中:(图 3,D)为用户等级颜色映射方案,采用京东现有会员等级颜色进行映射;(图 3,E)为商品购买量颜色映射方案.

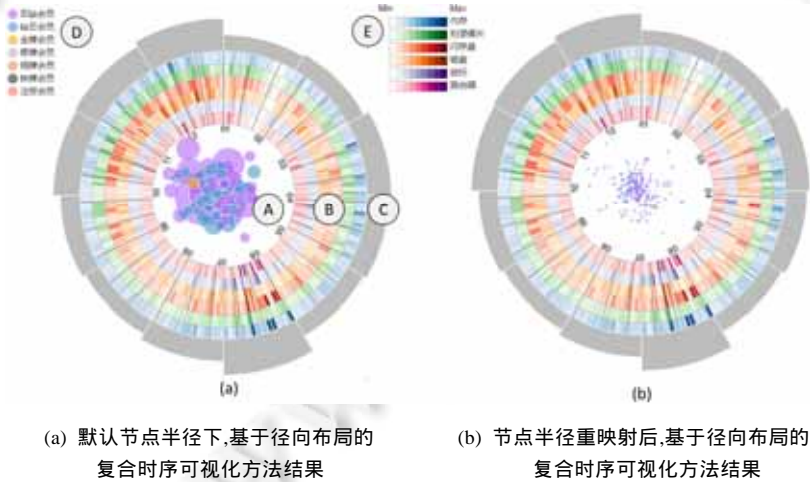


Fig.3 Composite temporal visualization method based on the radial layout

图 3 基于径向布局的复合时序可视化方法示意图

4.1.1 散点图

网购用户的数量众多,同时,每个用户具有多维属性,包括等级、地理位置等,因此,本文设计了散点图来可视化用户群体及其相关属性.

在散点图中,节点表示用户,故定义用户集合为 $U=\{u_i|i \in [1,n]\}$,其中, n 为用户总数,用户 u_i 购买时间序列为

$T_i = \{t_{ij} | j \in [1, m_i]\}$, 其中, m_i 为用户 u_i 的购买次数, 用户 u_i 在 t_{ij} 购买商品的数量为 c_{ij} , 将用户 u_i 在 t_{ij} 时刻的购买记录记为 (t_{ij}, c_{ij}) , 那么用户 u_i 在散点图中的位置 p_i 计算如下:

$$P = \left\{ p_i \mid p_i = \sum_{j=1}^{m_i} \overline{(t_{ij}, c_{ij})}, j \in [1, m_i] \right\} \quad (1)$$

p_i 的计算方法如图 4 所示, 即, p_i 由 u_i 在多个时间点购买商品的数量决定.

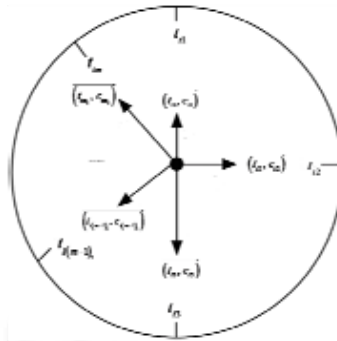


Fig.4 Node position calculation diagram in scatterplot

图 4 散点图中节点的位置计算示意图

节点的位置表示用户的购买时间倾向, 如, 节点位于 12 月和 1 月间, 且距离中心较远, 则表示该用户于 12 月、1 月进行网购的次数较多. 节点的大小表示用户的购买频次(购买商品数量), 即: 频次越高, 节点越大. 节点的颜色表示用户的等级(图 3, D). 同时, 由于屏幕空间有限, 当节点数量较多时, 节点间不可避免地会出现相互遮挡的情况. 因此, 本文采用根据节点遮挡情况调节节点半径与节点颜色的透明度的方法来解决该问题. 即: 在数据量较大情况下, 为了展示大规模节点, 可通过使用图 1(d) 中的 CircleR 调节节点半径的映射范围. 如图 3(b) 所示: 为节点半径映射范围为 2~4 的情况下散点图可视化结果, 当节点半径最大值与最小值相等时, 所有节点的大小相同; 通过对透明度的调整(如图 1(d) CTrans), 相互遮挡的节点也能被分析人员所观察和选择, 如图 3(A) 所示.

4.1.2 基于像素的圆环排序方法

用户的购买行为以商品为载体, 且与时间有很强的关联性, 因此, 本文设计了基于像素的圆环排序方法, 用以呈现用户购买商品的情况及购买行为与时间的关系. 该方法由一组半径递增的同心圆环组成, 其中, 圆环表示商品, 螺旋时间轴表示时间模式, 圆环上的像素条(pixel)表示时间粒度, 像素条的颜色表示商品的购买量(颜色编码方案详见第 4.2 节).

如第 1.2 节所述, 传统的周期时间可视化方法多采用固定的时间模式, 然而在线交易日志的时间跨度通常不确定, 因此, 本文设计了多尺度时间模式, 包括年模式、月模式和周模式. 以年模式为例, 圆环被螺旋时间轴等分为 12 份, 每个扇形表示 1 个月, 如图 3, B 所示. 此外, 本文还提供了 3 种时间粒度, 分别是日、周、月. 如图 3(B) 所示, 时间粒度为日. 不同的时间粒度用不同尺寸的像素条表示. 电子零售商或数据分析人员可根据日志的时间跨度和实际分析需求, 选择不同时间模式和时间粒度, 从而探索用户在线购买行为蕴含的时间规律.

定义圆环为 $L = \{L_i | L_i = (l_i, l_{i+1}), i \in [1, n-1]\}$, 其中, n 为圆圈数量(包含最内层圆圈), l_i 和 l_{i+1} 分别为 L_i 的内圈和外圈. 令 l_1 的半径为 r_0 , l_j 与 l_{j-1} 间的距离为 β , 则 l_j 的半径可表示为

$$R = \{r_j | r_j = r_0 + \beta(j-1), 1 \leq j \leq n\} \quad (2)$$

如图 5 所示, 定义圆环 L_i 的像素条为 $B_i = \{b_{ij} | i \in [1, n-1], j \in [1, p_i]\}$, 其中, p_i 为圆环 L_i 在给定时间粒度下的像素条总数. b_{ij} 的位置由 $(f(t_i), f(t_{i+k})) (i \in [1, n-1])$ 决定. $f(t)$ 表示为

$$f(t) = (r \cos \alpha, r \sin \alpha) \quad (3)$$

其中, 在 $(f(t_i), f(t_{i+k}))$ 中: k 为时间粒度参数, 当 $k=1$ 时, 时间粒度为日; $k=6$ 时, 时间粒度为周; $k=30$ 时, 时间粒度为月; t

为该像素条所在位置的时间. α 根据公式(4)计算得出:

$$\alpha = \frac{\pi}{2} - 2\pi \cdot \frac{t - t_0}{\gamma} \tag{4}$$

其中, t_0 为起始时间, γ 为时间模式参数.当 $\gamma=365$ 时,为年模式;当 $\gamma=31$ 时,为月模式;当 $\gamma=7$ 时,为周模式.在默认情况下, $\gamma=365, k=1$.

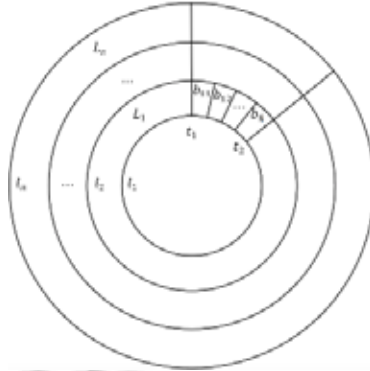


Fig.5 Donut chart combine with pixel bar chart
图5 圆环与像素条示意图

此外,考虑到基于径向布局的方法存在内圈半径小、外圈半径大的问题,本文设计了一种排序机制,即:将购买量最大的商品置于最外环,按购买量递减的顺序,购买量最小的商品置于内环,以此来平衡可视化布局并提高空间利用率.同时,由于商品种类繁多且屏幕空间有限,基于像素的圆环排序方法允许用户最多同时选择 6 种商品,即,该方法最多同时展示 6 个圆环.

4.1.3 环状图

环状图 $D=\{d_i|i \in [1,m]\}$ 由一组相邻的扇形组成, d_i 为每个扇形.扇形的尺寸包括宽度和高度,其宽度由时间模式决定,高度由商品的购买量决定.其中,高度与商品的购买量正相关.

扇形的数量和位置由时间模式决定.若为年模式,则 $m=12$,即,扇形共 12 个,每个扇形的位置即每个月的位置,如图 3,C 所示;若为月模式,则 $m=31$;若为周模式,则 $m=7$.

4.2 颜色映射

在时序可视化中,当展示多个时序变量及其之间的关联时,颜色映射显得尤为重要^[35].因此,基于像素的圆环排序方法采用颜色编码购买量的大小,故颜色编码的有效性与合理性直接影响着用户对数据的认知.

传统的颜色映射方法为线性映射,但采用该方法编码交易数据集存在如下问题:由于不同时间段商品的购买量差异较大,导致整体的颜色区分不大,规律不明显.为了更真实地反映数据的幂律分布特征,本文提出了颜色极值映射方法和规律映射方法.

4.2.1 极值映射

极值映射方法的核心思想为:将数据集分为极大值域、中间值域和极小值域这 3 段,分别采用不同的色系进行映射;同时,根据数值的分布规律,每段采用不同的颜色映射方法.

根据对数据的分析发现:极大值部分的数据稀疏,极小值部分的数据稠密.因此,这两段的数据采用分段映射(非线性映射)的方法.对于中间值部分,数据出现频次的振荡幅度较小,因此采用线性映射方法.

定义数据值为 $value_i(i \in [1,n])$,其中, n 为集合中数据值的数量,数据值出现的频次 $count_i(i \in [1,m])$,其中 m 为数据值集中出现频次的最大的数据值的频次.根据文献^[35],在数值分布不均时可采用对数或指数函数来进行分段,经过对 $value_i$ 和 $count_i$ 的分析,定义极值分段参数 φ 为

$$\varphi = \ln \left(\sum_{i=1}^m count_i \right) \tag{5}$$

本节采用 2007 年 1 月 14 日~2010 年 1 月 25 日,内存、存储卡、刻录碟片、闪存盘、键盘、鼠标这 6 种商品的交易数据作为实验数据.采用传统线性映射方法的结果如图 6(a)所示,采用本文设计的极值映射方法的结果如图 6(b)所示.

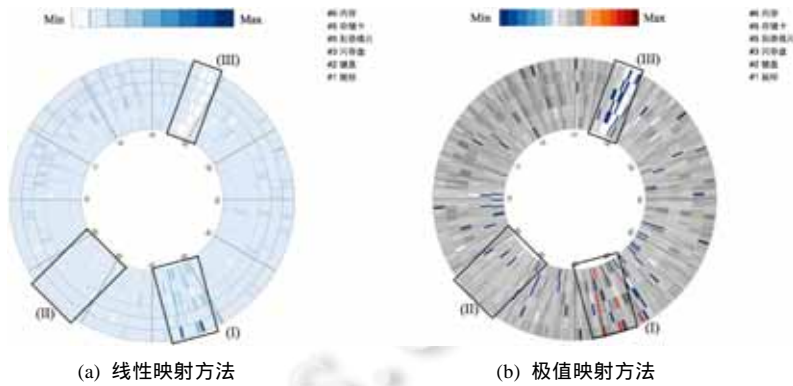


Fig.6 Traditional linear color mapping method and extreme color mapping method

图 6 传统的线性映射方法和极值映射方法的结果对比图

通过图 6 的对比可以发现:图 6(b)的商品购买量分布更加清晰,容易发现 6 种商品的整体购买规律.如:图 6(a)的(I)区域只能看出购买量大的日期;而图 6(b)的(I)区域中 6 种商品的购买量分布一目了然,能明显地看出商品购买量较大及较小的日期.此外,图 6(a)的(II)区域和(III)区域都是一种颜色,看不出规律;而图 6(b)的(II)区域和(III)区域能清晰地看出商品购买量较小的日期的购买规律.

4.2.2 规律映射

规律映射方法的核心思想为:对商品进行分类,每个分类采用不同的色系进行映射;在每个色系内,根据数值(购买量)出现频次的分布规律,采用均衡化策略进行颜色映射.

规律映射方法的相关参数说明如下:商品 L_i, L_i 在 t_i 时刻的购买量 (t_i, c_i) ;购买量 c_i 的频次 $count_{t_i}$,总频次为 $Scount$;色系 $color_i(i \in [1,6])$,每个色系的颜色最大值 $color_{imax}$,颜色最小值 $color_{imin}$.

均衡化策略,即,根据 c_i 的总频次 $Scount$ 将数值域 $[c_1, c_n]$ 分为 k 段,每段数值域的频次相等,均为 $Scount/k$ (均衡化系数).同时,将颜色域 $[color_{imin}, color_{imax}]$ 也分为 k 段,每段数值域映射到对应的颜色域.

本节采用 2007 年 1 月 14 日~2010 年 1 月 25 日,内存、存储卡、刻录碟片、闪存盘、键盘、鼠标这 6 种商品的交易数据作为实验数据.采用传统线性映射方法的结果如图 7(a)所示,采用本文设计的规律映射方法的结果如图 7(b)所示.

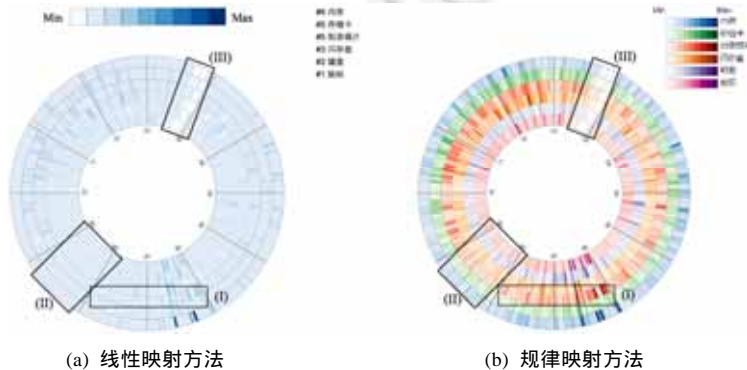


Fig.7 Traditional linear color mapping method and identifiable color mapping method

图 7 传统的线性映射方法和规律映射方法的结果对比图

通过图 7(a)线性映射方法和图 7(b)规律映射方法的对比,可以发现规律映射方法更有利于对比和发现一种商品内的购买时间规律.如:从图 7(b)的(I)区域与图 7(a)的(I)区域同为刻录碟片的销售情况,在使用规律映射的图 7(b)的(I)区域中,很明显能够发现刻录碟片 6 月~7 月中旬的销量较高,尤其是 6 月中旬的销量,且在销量较少的时间其销量规律也可以清晰地呈现;但在图 7(a)的(I)区域中只可以看到销量较大时间区域的规律,且规律不够清晰,同时,销量较少的时间区域的规律无法呈现;同时,采用规律映射方法,颜色的分布更均匀,更能真实地反映数据值的规律,如图 7(b)的(II)区域和(III)区域.

4.3 融合空间信息的时间轴可视化方法

基于 Shneiderman^[36]提出的“概览+细节”,基于径向布局的复合时序可视化方法主要用以分析用户群体的购买行为(概览).此外,本文还提出了融合空间信息的时间轴可视化方法,用以分析单个用户的购买行为特征(细节),主要解决第 2.2 节中 Q.4~Q.9;同时,可从单个用户角度解决 Q.10.

传统的时间轴可视化方法可有效展示时序事件的时间点和持续时间.然而,在线交易日志不仅具有时序属性,还有地理属性;同时,传统的时间轴可视化方法不能分析事件的复现模式.基于上述发现,本文提出了融合空间信息的时间轴可视化方法,如图 8 所示.

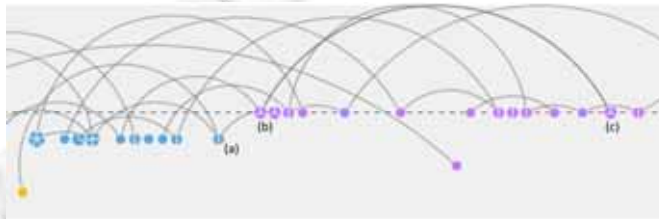


Fig.8 Timeline visualization method incorporated with spatial information

图 8 融合空间信息的时间轴可视化方法示意图

在该方法中,节点表示用户的购买事件,事件的时间属性和空间属性分别映射为节点的 X 轴坐标和 Y 轴坐标,如图 8 所示:节点(a)的 X 坐标表示 2009 年 6 月 14 日, Y 坐标表示江苏;节点(b)的 X 坐标表示 2009 年 6 月 17 日, Y 坐标表示上海.为了有效地反映地理位置的变化与购买行为之间的关联, Y 坐标的计算准则如下:

购买次数优先准则:计算用户在不同地区 Loc_i 的购买次数 $Lcount_i$,按 $Lcount_i$ 的大小对 Loc_i 进行降序排列,得到 $\{Loc_1, Loc_2, \dots, Loc_m\}$, $y(Loc_1)=0$,依此类推, $y(Loc_m)=(n-1) \cdot h$,其中: m 为用户地理区域的数量, $n \in (1, m)$; h 为相邻 Y 坐标间高度;

时间优先准则:若 Loc_x 和 Loc_z 的 $Lcount$ 相等,则按购买时间 t_i 的先后顺序对 Loc 进行排列,若 $t_x < t_z$,则 $y(Loc_x) < y(Loc_z)$.

节点的颜色表示用户的等级.为了避免分析人员产生视觉混乱,时间轴可视化中节点的颜色与散点图中节点的颜色保持一致.节点的大小表示用户的购买频次.节点的纹理表示商品的种类.当商品种类数为 1 时,节点为圆;当商品种类数大于 1 时,节点映射为饼图,各类商品的购买数量决定其在饼图中的占比.

同时,融合空间信息的时间轴可视化方法采用圆弧对包含相同商品的节点进行链接,以实现时序事件复现模式的分析,如用户的复购行为等.其中,圆弧的尺寸表示重复购买商品的数量.如图 8 所示,用户于(a),(b),(c)处均购买了相同的商品“多用途锅”,由于购买数量不同,故圆弧 ab 与圆弧 bc 的尺寸不同.

与本文设计的圆弧链接方法相比,Wattenberg^[37]设计了一种弧长链接方法,用于分析文本数据中重复出现的词语.但是,一方面,该方法针对的是文本数据,不适用于时空数据;另一方面,在该方法中,所有节点的 Y 坐标均为 0,故无法有效展示地理位置信息,且布局难度更低.

若用户 u_i 在 t_1, t_2, t_3, t_4 购买了同样的商品,则节点 a, b, c, d 用圆弧链接.为了尽量保持布局清晰并减少圆弧间的交叉,当上述 4 个节点均包含同样的商品时,仅相邻节点用圆弧链接.圆弧链接方法的示意图如图 9 所示.

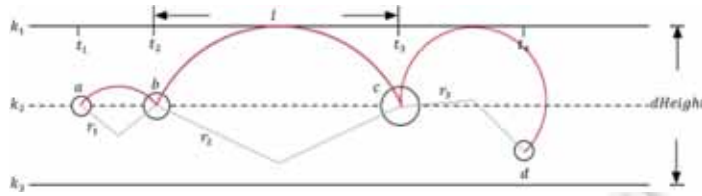


Fig.9 Arc link method diagram
图 9 圆弧链接方法示意图

设时间轴的高度为 $dHeight$, 节点 b 和 c 间的距离为 l , 则圆弧 bc 所在圆的半径为

$$r = \frac{l^2 + dHeight^2}{4dHeight} \tag{6}$$

其中, 所有圆弧均向下开口, 且若两节点在同一水平轴. 当 $l > r$ 时, 圆弧所在圆与 k_1 相切, 如由节点 b 与节点 c 所连成的圆弧; 当 $l < r$ 时, 圆弧所在圆与 k_3 相切, 如由节点 a 与节点 b 所连成的圆弧.

4.4 辅助可视化方法

用户的购买行为不仅与时间、用户属性有很强的关联性, 还与商品有关. 因此, 除了上述两个可视化方法外, 本文还设计了多层次可选的旭日图可视化方法和地图可视化方法, 用以辅助分析用户的购买行为.

4.4.1 多层次可选的旭日图可视化方法

如第 2.1.2 节所述, 在京东交易日志的样本数据中, 商品的结构共 5 层, 前 4 层分别是大类一级~四级, 第 5 层 (叶节点) 则是商品本身. 由于商品数量 (1.8 万件) 太多且屏幕空间有限, 故本文只考虑展示前 4 个层次. 除了商品的层次结构外, 各类商品的购买量也是电子零售商或数据分析人员的关注点, 故本文采用旭日图来展示商品的类别层次及其对应的购买量. 其中, 将商品分为 4 大类, 每一类及其子类均使用同一种颜色表示. 同时, 本文设计了节点的缩进 (zoom-in) 和展开 (zoom-out) 操作来查看不同层次下商品的购买分布, 解决了由于同一层次节点过多而数量关系不明显的问题. 如图 10 所示: 在图 10(a) 中选择感兴趣的节点, 随后, 该节点展开并展示其包含的层次结构和各子类商品的购买数量比, 展开后的效果如图 10(b) 与图 10(c) 所示, 其中, 图 10(b) 与图 10(c) 中, 高亮节点为同一节点, 若想回到上级层次, 则再次单击此节点.

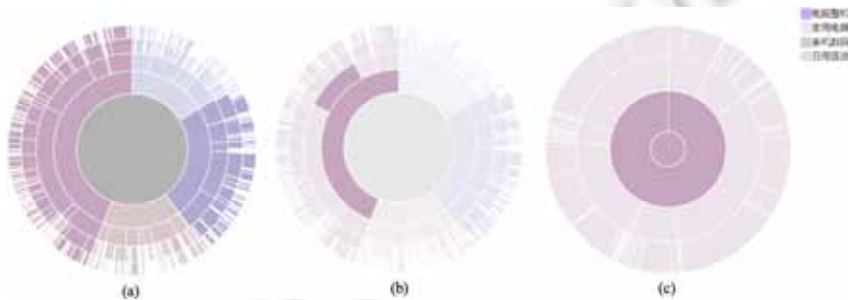


Fig.10 Multi-Level optional sunburst diagram
图 10 多层次可选的旭日图

4.4.2 地图可视化方法

地图可视化方法主要用以展示商品在不同区域的购买分布. 由于方位角投影 (azimuthal projection) 能准确表示两点之间的距离, 故地图可视化方法采用方位角投影. 另外, 本文采用颜色来映射商品的购买量. 由于不同商品的购买量差异较大, 因此, 本文采用一组对比色来区分购买量: 红色系表示购买量较大, 灰色系表示购买量适中, 蓝色系表示购买量较少. 如图 11 所示, 展示了“硬盘”这类商品在全国的购买分布. 通过该图可以发现: 东部沿海地区的购买数量明显高于全国其他地区, 而西藏、青海、宁夏和台湾地区的购买量较少.

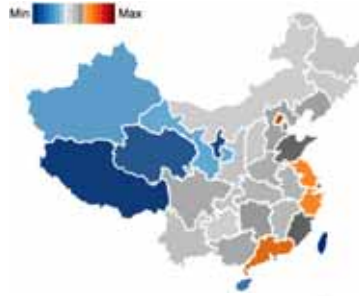


Fig.11 Map visualization method

图 11 地图可视化方法

4.5 交互设计

本文设计了丰富的交互操作,包括单个视图内的交互和多个视图间的交互.下面将介绍 UPB-VIS 中主要的交互操作.

4.5.1 单视图内交互设计

- 时间控制.

包括时间模式切换和时间粒度切换:时间模式默认为年模式,分析人员通过控制视图即可完成 3 种时间模式的切换;时间粒度切换,即日、周、月这 3 个粒度间的切换,默认粒度为日.

- 颜色控制.

包括颜色映射模式的选择(图 1(d) CMode)和颜色透明度的调整(图 1(d) CTrans).如第 4.2 节所述,系统提供了极值映射和规律映射两种颜色映射方法,系统允许分析人员根据需求进行灵活选择,默认方法为规律映射.另外,分析人员可根据实际情况调整散点图中节点的透明度,以便观察和探索用户的购买行为.

- 节点大小控制.

分析人员可以根据散点图中节点数量与节点遮挡情况,对节点的半径大小映射进行调整(图 1(d) CircleR).若节点数量较多且遮挡情况较为严重,可将节点半径映射范围调小或忽略节点半径编码(半径最大值与最小值相等情况),从而减少节点遮挡情况,以便观察和探索用户购买行为.

- 数据筛选.

数据筛选是可视化系统中重要的交互操作,该操作为用户提供了查询和过滤数据的渠道.本系统提供的数据筛选涵盖了分析用户购买行为关键指标的筛选,具体包括:(1) 用户属性筛选,包括用户等级(图 1(d) UType)、用户所在位置(图 1(d) ULoc)、购买总量(图 1(d) UBNuM)和购买频率的筛选(图 1(d) UFreq);(2) 时间范围选择(图 1(d) TimeControl);(3) 商品品类选择(图 1(d) PType).当用户进行上述 3 类的数据筛选时,系统各视图将同步更新数据并进行视图重绘.

- 用户搜索.

包括两种情况:(1) 若当前视图为单个用户行为分析视图,视图将展示被搜索用户的购买行为可视化结果;(2) 若当前视图为用户群体行为分析视图,被搜索的用户将在散点图中高亮.

- 节点缩进/展开.

当分析人员单击多层次可选的旭日图中感兴趣的节点时,该节点展开,同时以该节点为中心,展示其包含的层次关系.若想返回上级层次,则再次单击该节点,节点缩进后即可返回该节点的上级层次.

4.5.2 多视图间交互设计

- 用户行为分析视图与地图视图.

当分析人员单击用户行为分析视图中的用户节点时,用户所在区域将在地图上高亮显示;若单击地图视图中的某个区域,则用户行为分析视图将同步更新数据,并绘制和渲染该区域用户的购买行为可视化结果.

- 用户行为分析视图与商品层次结构视图.

在用户行为分析视图中,若单击其中的像素条,则该像素条表示的商品将会在商品层次结构视图中高亮显示,从而可以快速了解该商品所在的层次和购买量;当单击商品层次结构视图中的节点时,用户分析视图将同步更新数据,根据其子商品的购买量,由圆环排序算法计算其位置,并进行视图重绘.

- 商品层次结构视图与地理视图.

由于商品在全国的销量分布情况是电商平台关注的核心之一,在本系统中,当分析人员单击商品层次结构中感兴趣的节点时,地图视图将会展示该商品在全国的销量分布.

5 实验验证

本节以京东商城的在线交易日志为测试数据,对 UPB-VIS 的功能进行验证.通过用户群体购买行为的分析、单个用户购买行为的分析、商品销量的分析及特殊值的发现等多个实验,证明了系统解决实际问题的能力,尤其是可视化方法的有效性.

5.1 用户群体的购买行为分析

通过图 1(d)区域的筛选功能以及图 10 商品层次结构视图、图 11 地理视图提供的交互功能,可筛选出感兴趣的用户群体.本部分以家用电器为例,分析不同地区用户的购买行为.系统默认展示所选商品品类中购买数量排在前 6 位的子品类,通过系统可发现,全国用户购买最多的家用电器分别为插座、剃须刀、电话机、豆浆机、电吹风和电水壶;同时,购买量最大的地区为北京、上海和广东.下面分别对北京和上海的用户购买家用电器的行为进行分析.

图 12(a)和图 12(b)分别展示了北京地区和上海地区的用户购买家用电器的情况.首先,通过图例可直观地看到北京用户和上海用户购买的家用电器略有差异,除了大多数地区的用户都购买的商品外,北京用户更爱购买加湿器,上海用户更爱购买取暖电器.而这种购买偏好也符合北京和上海的地理环境及气候.

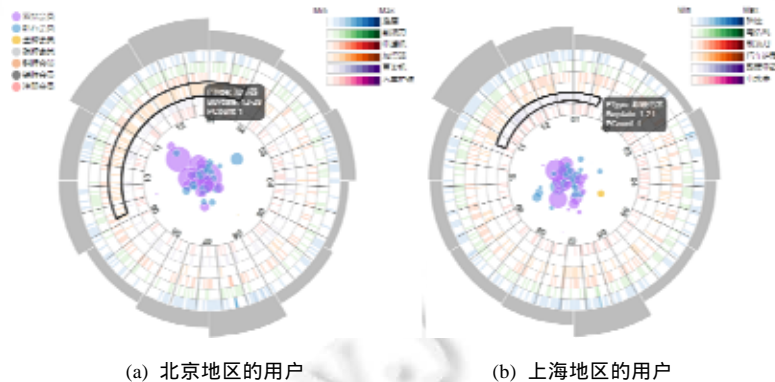


Fig.12 Visualization results of consumer from Beijing and Shanghai purchasing household electrical appliances

图 12 北京和上海的用户购买家用电器的可视化结果展示

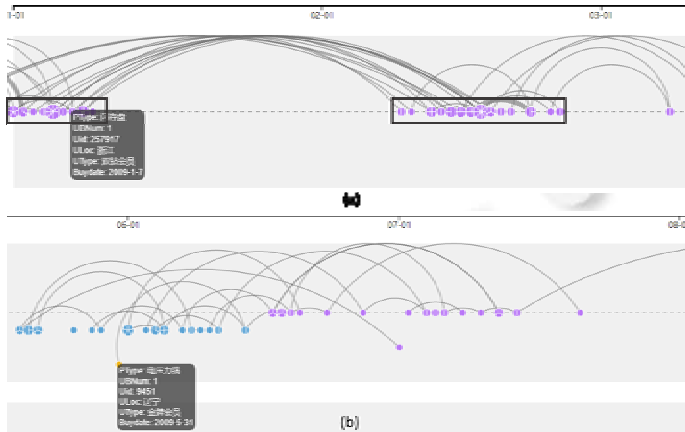
北京用户基本都在 9 月中旬~1 月中旬购买加湿器(如图 12(a)所示),而此时恰好也是北京最干燥的时候.上海用户基本都在 11 月~1 月购买取暖电器(如图 12(b)所示),这段时间也是上海寒冷的冬季.同时,通过对比图 10 中的散点图可发现:北京用户更倾向于 9 月~1 月购买家用电器,而上海用户更倾向于 6 月~8 月购买家用电器.

总结上述实验发现:系统能有效分析不同用户群体的购买行为,包括购买偏好、购买时间规律等,验证了系统能完成第 2.2 节中列出的可视化任务 Q.1~Q.3.

5.2 单个用户的购买行为分析

通过在图 3 用户行为分析视图中点击用户群体行为散点图中感兴趣的节点,或者通过本系统提供的搜索

功能,可对单个用户购买行为进行分析.如图 13 呈现了 ID 为“257917”的用户(下简称“257917”用户)和 ID 为“9451”用户(下简称“9451”用户)的购买行为可视化结果.



(a) “257917”用户的购买行为可视化结果 (b) “9451”用户的购买行为可视化结果

Fig.13 Purchasing behavior visualization result of consumer “257917”and “9451”

图 13 “257917”用户和“9451”用户的购买行为可视化结果

通过图 13(a)和图 13(b)的对比可发现:

- (1) 地理位置的变化:“257917”用户(如图 13(a)所示)一直都在浙江,没有地理位置的变化;而“9451”用户(如图 13(b)所示)从 5 月下旬到 6 月中旬基本都在江苏,只有 5 月 31 日去了辽宁,同时,该用户从 6 月下旬到 7 月中旬几乎都在上海,只有 7 月 1 日时在浙江.因此,我们可以推测,“257917”用户的工作比较稳定,而“9451”用户常出差,并且出现了一次工作地点的变动;
- (2) 等级的变化:从图 13(a)中可发现:“257917”用户一直都为双钻会员;而“9451”用户经历了钻石会员-金牌会员-钻石会员-双钻会员的等级变化过程,如图 13(b)所示,2009 年 5 月 31 日时,该用户从钻石会员降级为金牌会员,但 6 月 1 日又恢复为钻石会员;
- (3) 重复购买行为的分析:如图 13(a)所示,“257917”用户在 1 月初和 2 月中旬的重复购买率非常高,而且基本都买相同的几种商品;“9451”用户虽然在一定的周期内也会重复购买商品,如图 14 所示,但复购的模式却与“257917”用户不同;
- (4) 购买偏好的分析:通过交互操作可进一步发现:“257917”用户常购买电子产品,如硬盘、路由器、键盘、笔记本配件等;“9451”用户频繁购买家用电器,包括空调(如图 14 所示)、洗衣机、电压力锅、消毒柜等.因此我们推测:“257917”用户可能是 IT 工作者;而“9451”用户因工作地点的变动,故常购买家用电器.



Fig.14 Consumer “9451” repeat purchase air conditioning in a time period

图 14 “9451”用户在一定周期内重复购买空调

通过上述分析可发现:“257917”用户和“9451”用户网购频率高且重复购买率也很高;同时,两个用户都是双钻会员,并都有各自的购买品类偏好.像这类用户都是有价值的用户,电子零售商可以重点关注.

总结上述实验发现:系统能有效分析单个用户的购买行为,包括复购行为特征、购买偏好、购买频率、等级变化、地理位置变化等,验证了系统能完成第 2.2 节中的可视化任务 Q.4~Q.9.

5.3 商品的销量分析

本文实现的可视化系统 UPB-VIS 不仅能用以分析单个用户和用户群体的购买行为,还能用以分析商品的销量.针对商品销量,本系统提供了 4 个分析角度,包括不同时间的商品销量分析、不同种类商品的销量分析、不同地区的商品销量分析和多视图联动的综合分析.下面以不同时间的商品销量分析为例进行说明.

通过单击图 10 商品层次结构视图中感兴趣的节点,即可查看该类商品的销量分布.图 15 展示了在年模式/月粒度下,时尚影音类商品的销量分布.图 15(a)为采用规律映射方法的结果,通过该图可观察和对比某类商品各月的销量,如:音箱的销量在 10 月~12 月较高,而高清播放器的销量在 6 月和 12 月较高.图 15(b)采用的是极值映射方法,其呈现的结果更容易对比各类商品每月的销量,例如:耳机/耳麦在 10 月~12 月的销量相对较高,而高清播放器 1 月、2 月的销量相对较低,从而验证了系统能完成第 2.2 节中的可视化任务 Q.10.

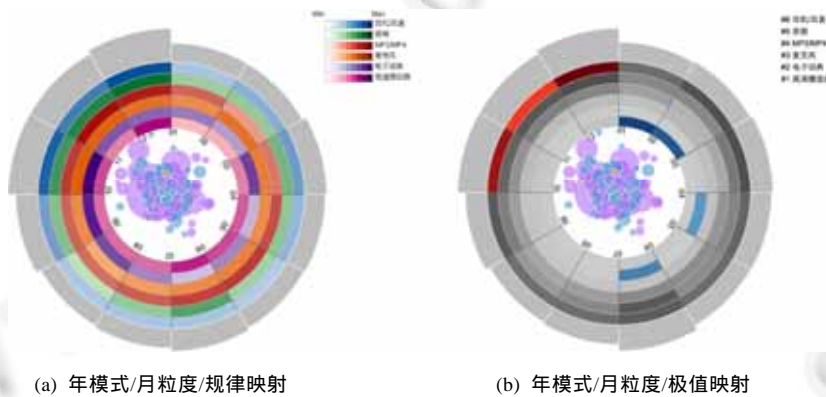


Fig.15 Sales of audio-visual goods distribution

图 15 时尚影音类商品的销量分布

5.4 其他实验结果分析

在进行上述实验的过程中,发现了一些有趣的特殊值.图 16 展示了 ID 为“4766”的用户(下简称“4766”用户)的购买行为可视化结果.注意到,“4766”用户的购买行为与第 5.2 节中两个用户的购买行为差异较大.“4766”用户在 2009 年 3 月 27 日购买了 40 种商品,包括睡衣家居服、烟具、电热毯、钱包、手机等,其中,购买睡衣家居服 15 件,烟具和电热毯各 5 件,所有商品共 78 件.但是,该用户在其他时间很少网购,如图 16 所示,除了 3 月 27 日这天的网购,在 3 个月的时间内仅网购 3 次,且每次仅购买一件或两件商品.

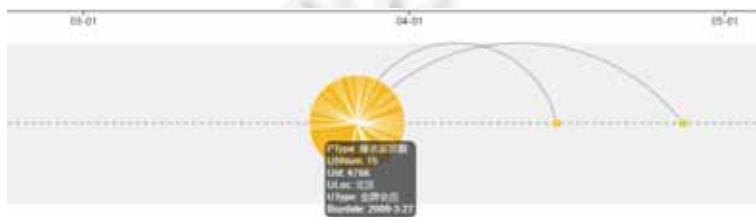


Fig.16 Purchasing behavior visualization result of consumer “4766”

图 16 “4766”用户的购买行为可视化结果

图 17 展示了电脑整机类和手机数码类商品的购买情况可视化结果,从图中可发现:两类商品在 6 月 18 日当天的购买量都很高,而两类商品均属于电子类商品.通过查询得知,6 月 18 日是京东商城的店庆,而京东商城又

是一家以 3C(计算机、通讯和消费电子产品的统称)类商品为核心的电商,其电子类商品最受欢迎。

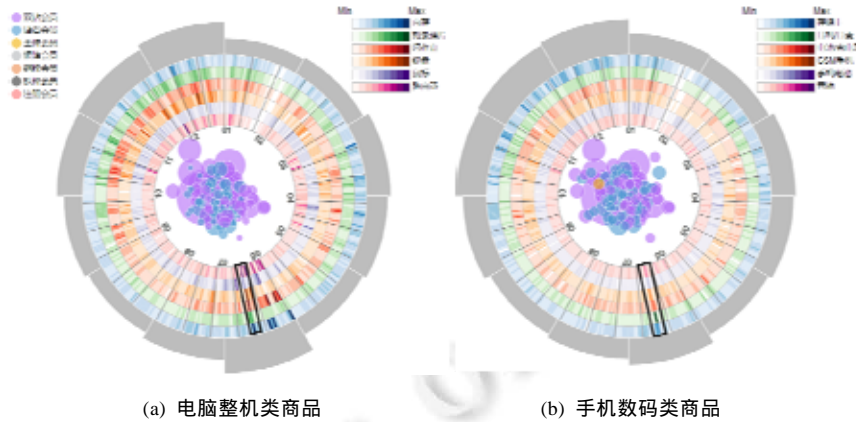


Fig.17

图 17

6 总 结

本文针对用户购买行为的可视化分析展开研究,面向在线交易日志数据,结合时序可视化、层次可视化、地图可视化等信息可视化与可视分析方法,设计并实现了用户购买行为可视化分析系统 UPB-VIS.通过京东在线交易数据集上的实验,验证了本文提出的基于径向布局的复合时序可视化方法、融合空间信息的时间轴可视化方法和两种颜色映射方法的有效性,并完成了用户(包括单个用户及用户群体)复购行为、购买偏好、购买时间规律等行为特征的分析。

对于其他具有时序、层次、地理、多维特征的数据集,或具备其中两到三个特性的数据集,可直接或经过一定的调整后,将本文提出的可视化模型和方法应用到该类数据集上。

致谢 在此,我们向对本文的工作给予支持和建议的同行表示感谢。

References:

- [1] Chiu CM, Wang ETG, Fang YH, Huang HY. Understanding customers' repeat purchase intentions in B2C e-commerce: The roles of utilitarian value, hedonic value and perceived risk. *Information Systems Journal*, 2014,24(1):85–114. [doi: 10.1111/j.1365-2575.2012.00407.x]
- [2] Yun CH, Chen MS. *Mining Web Transaction Patterns in an Electronic Commerce Environment*. Berlin, Heidelberg: Springer-Verlag, 2000. 216–219. [doi: 10.1007/3-540-45571-x_28]
- [3] King RC, Schilhavy RAM, Chowa C, Chin WW. Do customers identify with our website? The effects of website identification on repeat purchase intention. *Int'l Journal of Electronic Commerce*, 2016,20(3):319–354. [doi: 10.1080/10864415.2016.1121762]
- [4] Keim DA, Hao MC, Dayal U, Lyons M. Value-Cell bar charts for visualizing large transaction data sets. *IEEE Trans. on Visualization and Computer Graphics*, 2007,13(4):822–833. [doi: 10.1109/TVCG.2007.1023]
- [5] Taobao data cube. 2016 (in Chinese). <http://mofang.taobao.com/>
- [6] Jingdong data compass. 2016 (in Chinese). <http://luopan.jd.com/>
- [7] Kinesis. 2016. <http://aws.amazon.com/cn/kinesis/>
- [8] Xie C, Chen W, Huang X, Hu Y, Barlowe S, Yang J. VAET: A visual analytics approach for e-transactions time-series. *IEEE Trans. on Visualization and Computer Graphics*, 2014,20(12):1743–1752. [doi: 10.1109/TVCG.2014.2346913]
- [9] Hao MC, Ladisch J, Dayal U, Hsu M, Krug A. Visual mining of e-customer behavior using pixel bar charts. In: *Proc. of the ACM KDD/2001*. 2001.

- [10] Liu Z, Stasko J, Sullivan T. SellTrend: Inter-Attribute visual analysis of temporal transaction data. *IEEE Trans. on Visualization and Computer Graphics*, 2009,15(6):1025–1032. [doi: 10.1109/TVCG.2009.180]
- [11] Hayashi A, Kohjima M, Matsubayashi T, Sawada H. Regularity measure and influence weight for analysis and visualization of consumer's attitude. In: *Proc. of the 2015 19th Int'l Conf. on Information Visualisation*. 2015. 290–299. [doi: 10.1109/IV.2015.59]
- [12] Chang R, Ghoniem M, Kosara R, Ribarsky W, Yang J, Suma E, Ziemkiewicz C, Kern D, Sudjianto A. WireVis: Visualization of categorical, time-varying data from financial transactions. In: *Proc. of the 2007 IEEE Symp. on Visual Analytics Science and Technology*. 2007. 155–162. [doi: 10.1109/VAST.2007.4389009]
- [13] Singh K, Best P. Interactive visual analysis of anomalous accounts payable transactions in SAP enterprise systems. *Managerial Auditing Journal*, 2016,31(1):35–63. [doi: doi:10.1108/MAJ-10-2014-1117]
- [14] Aigner W, Miksch S, Schumann H, Tominski C. *Time & Time-Oriented Data*. London: Springer-Verlag, 2011. 45–68. [doi: 10.1007/978-0-85729-079-3_3]
- [15] Chen W, Shen ZQ, Tao YB. *Data Visualization*. Beijing: Publishing House of Electronics Industry, 2013 (in Chinese).
- [16] Jo J, Huh J, Park J, Kim B, Seo J. LiveGantt: Interactively visualizing a large manufacturing schedule. *IEEE Trans. on Visualization and Computer Graphics*, 2014,20(12):2329–2338. [doi: 10.1109/TVCG.2014.2346454]
- [17] Plaisant C, Mushlin R, Snyder A, Li J, Heller D, Shneiderman B. LifeLines: Using visualization to enhance navigation and analysis of patient records. In: *Proc. of the AMIA Symp.* 1998. 76–80. [doi: 10.1016/B978-155860915-0/50038-X]
- [18] Wang TD, Plaisant C, Shneiderman B, Spring N, Roseman D, Marchand G, Mukherjee V, Smith M. Temporal summaries: Supporting temporal categorical searching, aggregation and comparison. *IEEE Trans. on Visualization and Computer Graphics*, 2009,15(6):1049–1056. [doi: 10.1109/TVCG.2009.187]
- [19] Liu S, Wu Y, Wei E, Liu M, Liu Y. StoryFlow: Tracking the evolution of stories. *IEEE Trans. on Visualization and Computer Graphics*, 2013,19(12):2436–2445. [doi: 10.1109/TVCG.2013.196]
- [20] Tanahashi Y, Ma KL. Design considerations for optimizing storyline visualizations. *IEEE Trans. on Visualization and Computer Graphics*, 2012,18(12):2679–2688. [doi: 10.1109/TVCG.2012.212]
- [21] Krstajic M, Bertini E, Keim D. CloudLines: Compact display of event episodes in multiple time-series. *IEEE Trans. on Visualization and Computer Graphics*, 2011,17(12):2432–2439. [doi: 10.1109/TVCG.2011.179]
- [22] Han Y, Rozga A, Dimitrova N, Abowd GD, Stasko J. Visual analysis of proximal temporal relationships of social and communicative behaviors. *Comput Graph Forum*, 2015,34(3):51–60. [doi: 10.1111/cgf.12617]
- [23] Bertini E, Hertzog P, Lalanne D. SpiralView: Towards security policies assessment through visual correlation of network resources with evolution of alarms. In: *Proc. of the 2007 IEEE Symp.m on Visual Analytics Science and Technology*. 2007. 139–146. [doi: 10.1109/VAST.2007.4389007]
- [24] Zhao J, Forer P, Harvey AS. Activities, ringmaps and geovisualization of large human movement fields. *Information Visualization*, 2008,7(3-4):198–209. [doi: 10.1057/PALGRAVE.IVS.9500184]
- [25] Sun Y, Tao Y, Yang G, Lin H. Visitpedia: Wiki article visit log visualization for event exploration. In: *Proc. of the 2013 Int'l Conf. on Computer-Aided Design and Computer Graphics*. 2013. 282–289. [doi: 10.1109/CADGraphics.2013.44]
- [26] Cao N, Lin YR, Du F, Wang D. Episogram: Visual summarization of egocentric social interactions. *IEEE Computer Graphics and Applications*, 2016,36(5):72–81. [doi: 10.1109/MCG.2015.73]
- [27] Shiroy S, Misue K, Tanaka J. ChronoView: Visualization technique for many temporal data. In: *Proc. of the 16th Int'l Conf. on Information Visualisation*. 2012. 112–117. [doi: 10.1109/IV.2012.29]
- [28] Dragicevic P. SpiraClock: A continuous and non-intrusive display for upcoming events. In: *Proc. of the CHI 2002 Extended Abstracts on Human Factors in Computing Systems*. Minneapolis: ACM Press, 2002. 604–605. [doi: 10.1145/506443.506505]
- [29] Keim DA, Schneidewind R, Sips M. CircleView: A new approach for visualizing time-related multidimensional data sets. In: *Proc. of the Working Conf. on Advanced Visual Interfaces*. Gallipoli: ACM Press, 2004. 179–182. [doi: 10.1145/989863.989891]
- [30] Wu Y, Wei F, Liu S, Au N, Cui W, Zhou H, Qu H. OpinionSeer: Interactive visualization of hotel customer feedback. *IEEE Trans. on Visualization and Computer Graphics*, 2010,16(6):1109–1118. [doi: 10.1109/TVCG.2010.183]
- [31] DataTang. 2016 (in Chinese). <http://www.datatang.com/data/15516>

[32] Clauset A, Shalizi CR, Newman MEJ. Power-Law distributions in empirical data. *SIAM Review*, 2009,51(4):661–703. [doi: 10.1137/070710111]

[33] Moere AV, Purchase H. On the role of design in information visualization. *Information Visualization*, 2011,10(4):356–371. [doi: 10.1177/1473871611415996]

[34] Plaisant C. The challenge of information visualization evaluation. In: *Proc. of the Working Conf. on Advanced Visual Interfaces*. Gallipoli: ACM Press, 2004. 109–116. [doi: 10.1145/989863.989880]

[35] Aigner W, Miksch S, Schumann H, Tominski C. *Visualization Aspects*. London: Springer-Verlag, 2011. 69–103. [doi: 10.1007/978-0-85729-079-3_4]

[36] Shneiderman B. The eyes have it: A task by data type taxonomy for information visualizations. In: *Proc. of the '96 IEEE Symp. on Visual Languages*. 1996. 336–343. [doi: 10.1109/VL.1996.545307]

[37] Wattenberg M. Arc diagrams: Visualizing structure in strings. In: *Proc. of the IEEE Symp. on Information Visualization (INFOVIS 2002)*. 2002. 110–116. [doi: 10.1109/INFVIS.2002.1173155]

附中文参考文献:

[5] 淘宝数据魔方.2016. <http://mofang.taobao.com/>

[6] 京东数据罗盘.2016. <http://luopan.jd.com/>

[15] 陈为,沈则潜,陶煜波.数据可视化.北京:电子工业出版社,2013.

[31] 数据堂.2016. <http://www.datatang.com/data/15516>



贾若雨(1992 -),女,河北宁晋人,硕士,CCF 学生会员,主要研究领域为信息可视化,可视分析.



刘汉清(1989 -),男,硕士,主要研究领域为信息可视化,可视分析.



曾昂(1991 -),女,硕士,主要研究领域为信息可视化,可视分析.



李明召(1990 -),男,博士生,主要研究领域为信息可视化,可视分析.



朱敏(1971 -),女,博士,教授,CCF 高级会员,主要研究领域为信息可视化,可视分析,生物信息处理.