

# 基于商品引力的购物信息可视化系统<sup>①</sup>

杨笑笑<sup>1,2</sup>, 单桂华<sup>1</sup>, 田东<sup>1,2</sup>, 余敏楮<sup>1</sup>

<sup>1</sup>(中国科学院 计算机网络信息中心, 北京 100190)

<sup>2</sup>(中国科学院大学, 北京 100049)

通讯作者: 单桂华, E-mail: [sgh@sccas.cn](mailto:sgh@sccas.cn)

**摘要:** 互联网的快速发展带动了人们的购物行为, 越来越多的人群习惯于进行网购, 随之产生的是各种各样的网购交易数据, 这些数据复杂、多维, 并且具有时序性, 消费者的购买行为是众多电商及分析人员的研究热点. 本文提出了一种以个体为中心, 基于商品引力对商品层级结构数据与消费者的购物行为特征的可视分析方法, 使分析者更方便地观察每位消费者的消费行为, 从而得到他们的购物特点. 最后, 给出实验结果, 证明方法的有效性.

**关键词:** 时序数据; 层级结构; 商品引力; 购物行为; 可视化

引用格式: 杨笑笑, 单桂华, 田东, 余敏楮. 基于商品引力的购物信息可视化系统. 计算机系统应用, 2018, 27(2): 1-8. <http://www.c-s-a.org.cn/1003-3254/6209.html>

## Visualization System of Shopping Information Based on Product Gravity

YANG Xiao-Xiao<sup>1,2</sup>, SHAN Gui-Hua<sup>1</sup>, TIAN Dong<sup>1,2</sup>, YU Min-Zhu<sup>1</sup>

<sup>1</sup>(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)

<sup>2</sup>(University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** The rapid development of the Internet has led to increasing online shopping and accumulated large amount of complicated, multi-dimensional and temporal transaction data. Many e-retailers and analysts are focusing on these data to study customers' shopping behaviors. In this study, based on product gravity, a visual analytical method focusing on individuals is proposed to analyze hierarchic production data and the features of shopping behaviors, which allows analysts to observe each consumer's shopping behaviors more directly so as to discover the features of their shopping behaviors more easily. Finally, the experimental results are presented to prove the effectiveness of the method.

**Key words:** time-series data; hierarchy; product gravity; shopping behavior; visualization

随着互联网技术的迅速发展, 网络已成为日常生活中的重要组成部分, 网络购物已成为趋势. 各种购物平台每天都在进行交易, 进而不可避免地产生庞大的数据, 这些数据复杂多维, 并与时间有紧密联系. 通过对这些数据的可视化分析, 可以了解消费者的购买性格, 购物偏爱等喜好. 目前已有较多研究人员针对系统应用数据分析工作提出了多种文本分析、挖掘等研究工作<sup>[1-3]</sup>, 但对分析结果的可视化展示仍存在不足.

本文呈现一种实现多维层次数据在时序上信息可

视化的设计思想, 主要贡献如下:

(1) 基于商品引力将商品与消费者之间的联系以分层的可视化形式展现出来, 细化商品类别属性与消费者消费行为, 提出体现层级结构数据的方法.

(2) 挖掘消费者购买商品的过程与时间之间的联系, 提出体现时序数据的方法.

## 1 相关工作

目前已经有很多针对层级结构的可视化研究, 针

<sup>①</sup> 基金项目: 国家“863”高技术研究发展计划项目 (2015AA01A302); 国家自然科学基金 (91530324)

收稿时间: 2017-04-26; 修改时间: 2017-05-19; 采用时间: 2017-05-31; csa 在线出版时间: 2018-01-12

对网络数据,可采用聚集点、线,设置阈值,或者省略某些细节信息的手段来简化数据,以线图、节点图、矩阵图来展示<sup>[4]</sup>。大多研究表现的是树形结构,也就是说,他们经常将树结构表示成大量的节点以及连接父节点与子节点的边,例如节点-连线和空间填充。根树描述的是一种父节点、子节点以及连线连接多个节点之间的关系<sup>[5]</sup>。雷达树是一种根据众多节点在树中的深度,来把它们都放在同心圆中的布局。气球树是一种表达将所有节点包含在圆圈中的并连接到父节点的节点与线的表现关系。树图表达的是一种通过圈画的外壳展现大树的空间填充布局<sup>[6]</sup>。

针对时序数据,在 ODSP 流程中,基于信息熵的流量特征来有效度量网络活动的随机特征,过滤符合条件的 IP 地址与端口,提取有意义的时序数据<sup>[7]</sup>。直方图的条状也可以很明显地展现出来,人们根据时间从左边到右边流动的图表来描述时序数据。在堆叠流图中采用堆叠基准函数和堆叠序列函数将时序数据进行分析,并综合比例、曲线方法、排序、着色和标签布局构建可视化效果<sup>[8]</sup>。还有一些其他方式,像螺旋图,基于像素的技术等。

商品数据是同时具有多层结构与时序的数据,它是复杂且多维的,与时间有很强的关联性<sup>[9]</sup>。理解并细分消费者的购买行为是众多电商及分析人员的研究热点。在目前多数的购物网站中,在消费者的角度下,所能查看的以往订单与购物车中的商品大多以表格的形式呈现,其中对商品的展示分类不甚明显,这样容易造成人的视觉疲劳。然而,针对层级结构与时序数据的结合的商品购物的研究仍不多,如 Keim 主要分析时变的规律与异常的购买行为<sup>[10]</sup>等。

## 2 数据分析

本章主要介绍的是基于商品引力的网络购物可视化系统交易日志数据的描述和分析。

消费者通过网络进行购物,在这个过程中所产生的数据复杂多变,包括消费者信息、与消费者相关联的商品属性、消费者的购物行为,以及它们与时间的关联<sup>[11]</sup>。本文重点研究商品在结构层次上的分类,并与时间、消费行为的关联可视化展示以及以消费者个体为中心的研究<sup>[12,13]</sup>。

在此之前,需要把得到的样本数据进行预处理。这些数据的主要属性特征如下:

(1) 商品: 商品数据具有显著的层次结构, 本文主要将其分为了三个层级, 它们是商品分类、分类细化、商品型号。举个例子, 某位消费者可能会购买食物, 具体的分类下 (即分类细化) 的水果, 然后该件商品是十公斤重的富士苹果。

(2) 时间: 研究时间上的三个层级划分的数据, 时间粒度分别是月、日、时。

(3) 消费者的购物行为: 根据实际经验, 在本文中消费者的购物行为细化为五种, 它们分别是浏览商品, 将商品加入到购物车中, 将商品从购物车中删除、购买商品、将商品退货。

## 3 设计思想

本章主要介绍网络购物可视分析系统的设计思想, 系统主要研究基于商品引力的商品层级分类与消费者的消费行为之间的联系, 展示一系列相互关联的可视化视图, 并且提供丰富的交互手段。

系统根据数据的特点, 联系了消费者消费行为与时间两个重要因素, 使分析者能够从多维度对购物主题进行全面的分析。

### 3.1 可视化视图

该系统的整体界面如图 1 所示, 系统主要包含 5 个视图: (1) 购物主视图; (2) 商品分类及消费行为视图; (3) 用户信息及行为细节视图; (4) 消费日历视图; (5) 消费时段视图。

本节将会重点介绍系统中各个视图的设计思想。

#### 3.1.1 购物主视图

商品的种类繁多, 例如电子类、服装类、食品类, 等。并且, 在每一类的商品下, 都有更加细化的分类, 例如, 电子类下细分电脑、手机、手表等, 而电脑下还可以细分具体的品牌和型号, 因此把有关商品的节点划为三个层级: 商品分类、分类细化、商品型号, 它们分别对应视图中的第二、三、四层节点, 而第一层级节点代表消费者节点, 即不同层级的节点代表以上不同的隐喻。

购物主视图是本系统的核心, 首先定义消费者坐标, 如式 (1), 其中  $H$  是主视图高,  $W$  是主视图宽。

$$\begin{cases} x = \text{consumer\_}x = \frac{W}{4} \\ y = \text{consumer\_}y = \frac{3*H}{4} \end{cases} \quad (1)$$

$\text{consumer\_}x$  和  $\text{consumer\_}y$  是根据主视图视口大小定位的初始化坐标。之后, 根据圆的极坐标表达式定位第二层节点, 如式 (2)。

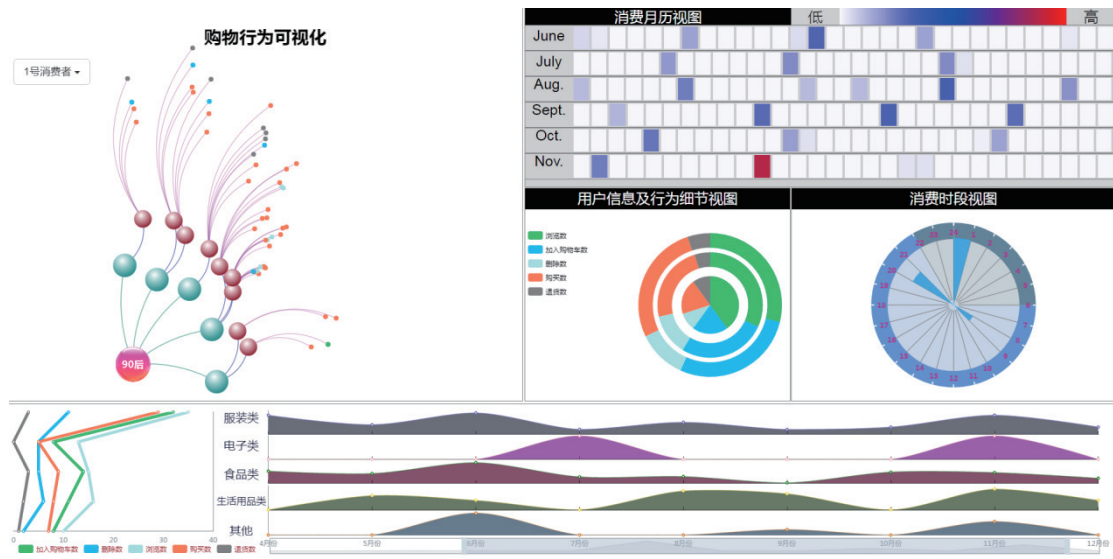


图1 系统概览

$$\begin{cases} x = consumer\_x + R * \cos(trim * angle) \\ y = consumer\_y + R * \sin(trim * angle) \end{cases} \quad (2)$$

以此类推, 后级层次节点坐标为:

$$\begin{cases} x = center\_x + R * \cos(trim * angle) \\ y = center\_y + R * \sin(trim * angle) \end{cases} \quad (3)$$

$center\_x$ 和 $center\_y$ 是上一层父节点的横纵坐标; $R$ 取主视图宽 $W$ 的 $1/6$ , 可以调整不同层级节点间的距离; $angle$ 控制后层节点围绕前层节点的角度, 取每层级节点个数平分 $15$ 度角的值; $trim$ 为同等级节点的不同下标加 $1$ , 以此判断不同的角度. 图2描述了各层级节点之间的初始化的关系, 例如节点 $1$ 处于第一层级; 节点 $2、3$ 处于第二层级; 节点 $4、5$ 是节点 $2$ 的子节点, 处于第三层级; 节点 $6、7$ 是节点 $5$ 的子节点, 处于第四层级.

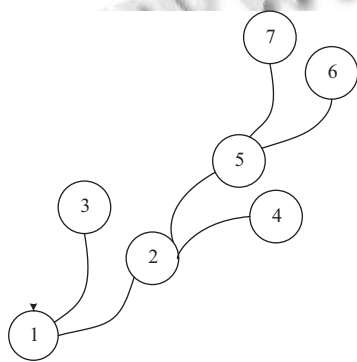


图2 各层级节点初始化定位示意图  
在物理上, 有一条众人皆知的定律——万有引力

定律, 它表述的是物体间相互作用的一条定律, 任何物体之间都有相互的吸引力, 这个力与物体间的质量成正比, 与它们之间的距离成反比.

引申到本文的购物系统当中来, 每相邻两个层级节点之间也有彼此之间的吸引力——商品引力 $K$ . 同一层级的节点, 当 $K$ 值越大, 则与父层级节点的距离越远; 反之, 距离越近. 在图3的示意图中, 节点 $4、5$ 都处于第三层级, 它们的父层级节点是处于第二层级的节点 $2$ , 针对第三层级的节点来说, 当该层级的 $K$ 接近于 $0$ 时, 层级中的节点越趋近于虚线 $b$ (如节点 $4$ ), 当 $K$ 接近于 $1$ 时, 该层级上的节点越趋近于虚线 $c$ (如节点 $5$ ),  $K$ 值的范围是 $0$ 到 $1$ 之间, 这样可保证节点 $4、5$ 始终处于第三层级中.

如图4, 受商品引力 $K$ 值的影响, 根据式(3)定位的处于第 $n+1$ 层级的节点 $A$ , 横纵坐标同时乘以 $K$ 值之后, 变成 $B$ 点的定位, 于是, 该节点距离处于父层级(第 $n$ 层级)的父节点的距离跟随 $K$ 值的变化而变化.

各节点基于商品引力 $K$ 的重定位公式如下:

$$\begin{cases} x = x' * K \\ y = y' * K \end{cases} \quad (4)$$

$A$ 点坐标的 $(x', y')$ , 在 $K$ 的作用下, 重定位到点 $B(x, y)$ .

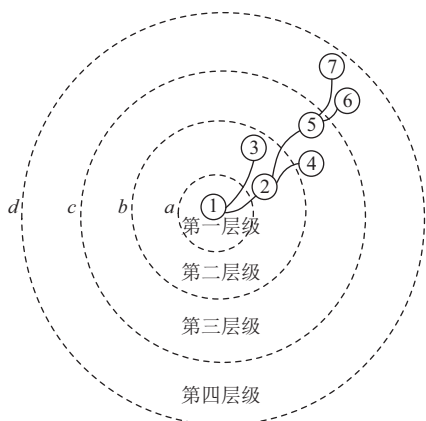


图3 各层级节点在商品引力  $K$  作用下的定位示意图

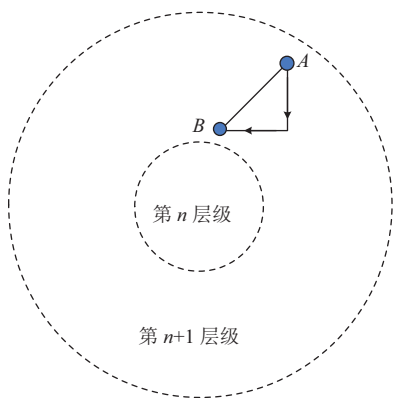


图4 商品引力  $K$  的原理图

如图5所示,第二层级上的节点代表的消费额数据是处于第三层级上的所有子节点代表消费额数据之和;同理,第三层级上的节点代表的消费额数据是处于第四层级上的所有子节点代表消费额数据之和;每个节点的商品引力  $K$  值与该节点所隐喻的消费额数据成正比.在第四层节点上,即具体的商品型号的隐喻描述,规定了5种颜色(该5种颜色与图6中的商品分类及消费行为视图、图8中的用户信息及行为细节视图的图例颜色相吻合),依次对应该件商品的最后动作:浏览、加入购物车、从购物车中删除、购

买、退货.例如,针对小米5S这件商品,消费者浏览了5次,加入购物车1件并购买了这一件,所以对这件商品的最后动作,消费者停留在了“购买”,这也意味着消费者购买了一件小米5S之后,并无退货意愿.

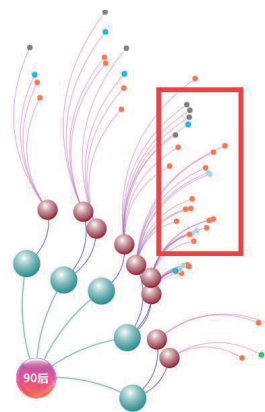


图5 第四层节点颜色隐喻商品的最后行为

### 3.1.2 商品分类及消费行为视图

本文的商品数据类别为5个大类,分别是:食品类、服装类、电子类、生活日用品类和其他.

如图6所示,右侧可视化图表的纵轴显示商品分类,各个商品的分类彼此之间是相互独立的,横轴表示该年的时间轴.每个面积图分别针对相应商品类别,来统计横坐标所选择时间段下的消费额.左侧的可视化图表和右侧的可视化图表共用一个商品分类坐标轴和时间轴,其横坐标轴对应每类商品下五种消费行为(浏览、加入购物车、从购物车中删除、购买、退货)的数据统计,横轴上的数值是由左向右依次变大的,其展示的数据随右侧可视化图表下的时间轴中选择的连续时间段的变化而变化,隐喻所选连续时间段中对应商品类别的五种消费行为的数据累积统计.比如说,在6月到11月之间,图6中的左侧可视化图表显示该消费者对服装类的购买数最多,右侧可视化图表则表明在这个时间段下,这位消费者主要集中在6月和11月购买服装,而9月和10月购买较少.

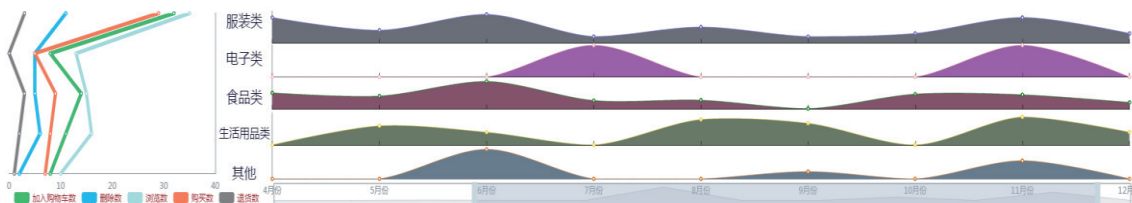


图6 商品分类及消费行为视图

### 3.1.3 用户信息及行为细节视图

该视图中可以根据操作展现四种不同的可视图形: 消费者对应的信息列表和 3 种环形图.

消费者的购物行为是时常变化的. 比如一位消费者浏览了某件商品 3 次, 然后将该件商品加入了购物车, 但是这位消费者又很快地将该件商品从购物车中删除, 在这个过程中便会产生很多数据. 为了更细化地表现消费者的消费行为, 系统为购物主视图中的各个节点加入了点击查看的功能. 例如, 点击图 5 代表的主视图中的消费者节点, 在图 7 所展示的用户信息及行为细节视图中会显示消费者相关的细节信息, 其中包含了消费者在本系统中的序号、消费者的年龄段以及年消费总额; 点击图 5 中的第四层节点, 则在用户信息及行为细节视图中显示三环图, 针对该点击节点所代表的型号商品的五种购物行为进行数据统计, 如图 8 所示; 相应的, 点击图 5 的第二、三层节点时, 用户信息及行为细节视图中显示对应的单环图和双环图, 如图 9 所示, 为单环图.

用户信息及行为细节视图

详细情况表		
序号	年龄段	消费额
1	90后	18775

图 7 点击主视图第一层节点后对应的细节可视视图

用户信息及行为细节视图



图 8 点击主视图第四层节点后对应的细节可视视图

多环图隐喻是由外向内依次展现主视图的第二、三、四层节点所代表商品类或型号的五种消费行为的数据统计. 在这里需要强调的是, 其中外层的节点对里

层相应节点的数据具有包含关系, 也就是说, 外层节点的数据包含了内层节点的数据. 例如, 某位消费者购买的第三层节点如电脑类下的所有商品的购物数据都被统计到了该电脑类节点中.

用户信息及行为细节视图

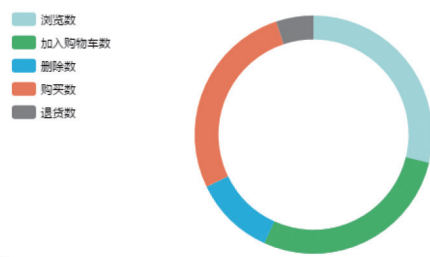


图 9 点击主视图第二层节点后对应的细节可视视图

### 3.1.4 其它视图

消费月历视图: 联系商品分类及消费行为视图 (图 6) 中的时间轴, 此视图显示该时间轴上的选择时间段内的所有日期, 并按行从上到下显示, 每月数据显示一行、每个月的数据从左到右均匀排列<sup>[14]</sup>, 如图 10 所示. 消费月历视图中的每个日期模块都渲染一定的颜色, 该颜色根据所表示日期中该消费者的消费额映射得来, 右上角配以颜色表图例, 白色代表日消费额为零, 蓝色代表日消费额较低, 红色代表日消费额较高, 这样可以使选择时间段下的日消费额情况一目了然.

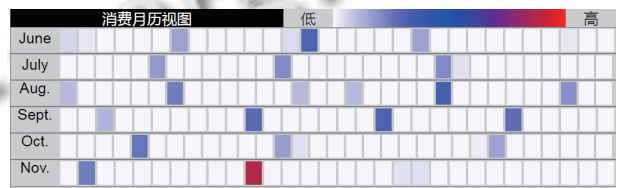


图 10 消费月历视图

消费时段视图: 联系上面的消费月历视图, 选择图 10 中某一天, 则可在消费时段视图中将该天都细化为 24 小时, 以小时为时间粒度, 将这 24 小时划分为夜间休息时间段与其他时间段, 观察该日下的消费额所发生的时间段, 数据为该小时下的消费额. 如图 11 所示, 凌晨开始到一点的时间段内, 颜色填充最多, 代表这一天中该消费者在这个小时内消费额最多, 大部分时间段下颜色填充区域全为灰白色, 代表这些时间段下没有发生消费.

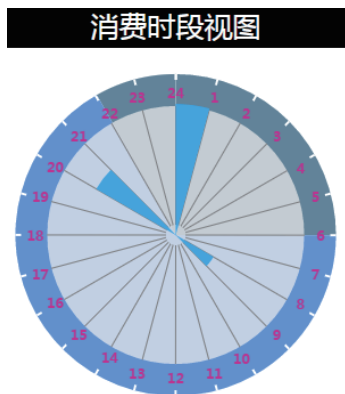


图 11 消费时段视图

### 3.2 交互设计

可视化并非仅仅是静态的形式,交互式可视化可以挖掘出更多的数据细节,便于分析者主动地去探索与分析数据,发挥人的主观能动性.因此本系统也提供了丰富的交互式操作.

#### 3.2.1 鼠标悬浮与点击

在购物主视图中,如果鼠标在任意节点上点击,即会在用户信息及行为细节视图上显示详细信息.如点击消费者节点,即第一层节点后,会显示该消费者在本系统中的序号、消费者的年龄段以及年消费总额,如图7;点击第四层节点后,会显示该节点代表的具体商品型号为中心的三环图,如图8.

如果鼠标悬浮在某节点上,即可高亮显示与该节点直接相连的节点与连接线,并在提示框中显示被悬浮节点的相关信息;在连线上悬浮,则高亮显示该连接线,且连线的提示框显示两端连接节点的名称.图12所表示的是鼠标悬浮在第四层节点上,出现提示框显示该节点代表商品魅蓝 note2,五种购物动作的数据代表该消费者针对这件商品,曾经在众多的网络购物平台中浏览过5次、加入购物车2次,之后并没有从购物车中删除,但是购买过1次,并将其退货.

#### 3.2.2 选择

本系统的数据是基于商品引力,以消费者为中心,在某位消费者的前提下挖掘其购物信息的.因此首先需要在消费者人群下拉框中(在主视图左上角)选择所要研究的消费者对象.

之后在主视图中即可显示该消费者默认时间段下的购物信息.点击不同的节点,在用户信息及行为细节视图中显示相应的表格或环图.



图 12 鼠标悬浮在某第四层节点上

在商品分类及消费行为视图中,选择不同的时间段,该视图中的众多商品分类面积图、消费行为折线图、主视图以及消费月历视图都会产生随之变化的互动展示,可以表示该消费者在此选定的时间段内的购物行为及购物量.

#### 3.2.3 选择过滤及其他

消费者的购物数据具有强烈的时间属性,包含所研究年份下的月、日、时.在商品分类及消费行为视图中,主要选择的是“月”,经过滤后,在消费月历视图中,显示的是特定的日期,通过点击消费月历视图的日期模块,在消费时段视图中显示某日下24小时下的消费额情况.

在购物主视图中,为了使分析者更加清晰地挖掘数据,系统提供了对主视图的缩放功能.将鼠标移至主视图上,鼠标滑轮上下滚动即可实现对视图的放大与缩小.

## 4 系统实现及案例分析

如图1所示,这是一个九零后的网上购物数据的展示.不难看出,这位消费者基本上在每月都有消费,大多数是隔一周左右进行一次购物.

研究6月到11月共半年的时间段,在月历视图下,11月份的某一天(双十一)消费额较为偏大.鼠标点击该天,在关于消费时段视图中,可以看到,该消费者主

要在这一天的凌晨一点购物量最大,上午八九点和晚上八点也有一些消费额产生. 联系实际情况,通常各大主流购物网站上,步入双十一这一天的凌晨开始,所有促销活动开放,人们习惯于双十一之前将欲购买的商品提前放入购物车,在双十一凌晨付款,所以在凌晨一点钟的消费额非常大.

在商品分类及消费行为视图中,右边的面积图与左侧的折线图综合显示,该消费者偶尔购买电子产品,对食品类和生活用品类商品的需求量也不大,对服装类商品购买量较为偏多.但是,商品引力 $K$ 值与节点隐喻的消费额成正比,第二、三层节点距离各自父层节点大体一致,可见各节点的 $K$ 值大小十分接近,由此可

推断出该消费者对服装类的购买量较大且平均单价不高,其他产品类购买量少但平均单价较高,例如购买过的电脑及配件等价格较为昂贵的电子类产品.在购买商品之前,该消费者习惯多次浏览,且退货量较少,也可体现出该消费者购物沉着谨慎.

图13展示的是一位80后消费者的网络购物情况,同样重点分析其在6月至11月的消费情况,该消费者对商品的浏览数和购买数基本保持一致,可见其购物习惯是干脆利落型,且购物偏好为电子类商品和食品类,经常连续性小额度购物.同样,在双十一的时候,重点在步入凌晨至一点的时间段中,大量下单成交商品,双十一的消费致使11月的消费额加大.

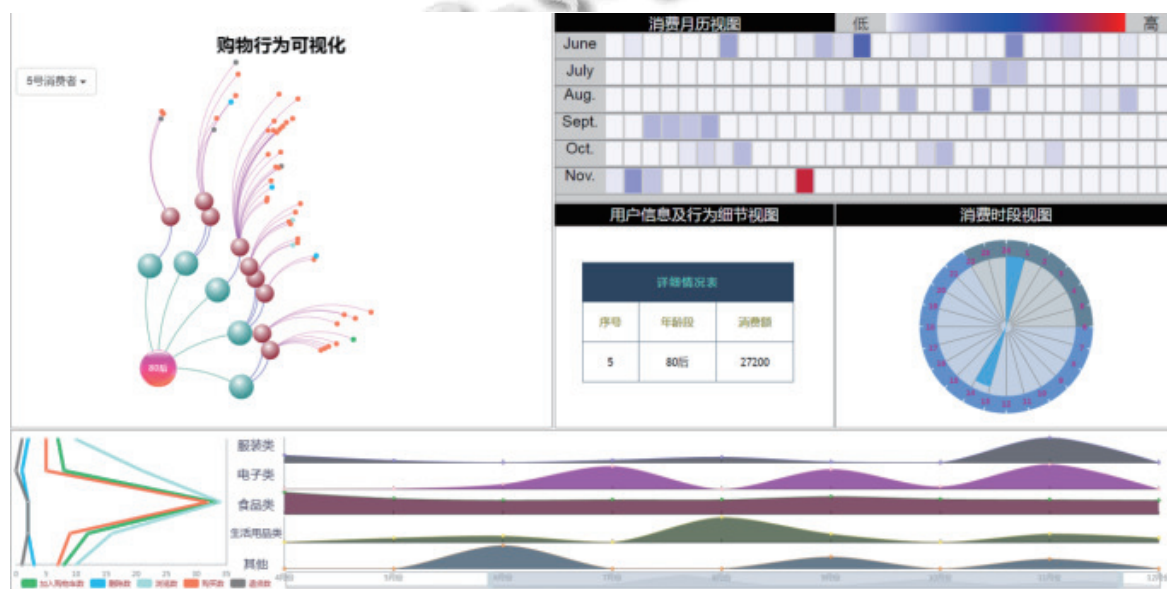


图13 一位80后消费者网络购物情况

## 5 结论

消费者的购物行为复杂多变,商品信息也具有多层次结构,由此综合的数据既多维又具时序性.

本文基于商品引力将商品的信息按层次结构进行可视化分析,并对以消费者个体为中心的购物行为的时序数据进行了研究,同时挖掘出商品层次结构数据与消费者行为的时序数据之间的联系.因此,基于本系统的设计思想,可以使分析者更方便地观察每位消费者的消费行为,从而得到他们的购物特点.

不同的消费者会有不同的购物习惯与偏好,在未来的工作中,将重点研究消费者之间的消费行为差异,以消费者特征码的形式表现在购物主视图上.

## 参考文献

- 1 Chen CC, Warren J, Evans J. 'Qualities' not 'quality'—text analysis methods to classify consumer health websites. *Electronic Journal of Health Informatics*, 2009, 4(1): 1–7.
- 2 Chen CC, Warren J, Riddle P. Semantic space models for classification of consumer webpages on metadata attributes. *Journal of Biomedical Informatics*, 2010, 43(5): 725–735. [doi: 10.1016/j.jbi.2010.06.005]
- 3 Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 2010, 29(1): 24–54. [doi: 10.1177/0261927X09351676]
- 4 Becker RA, Eick SG, Wilks AR. Visualizing network data.

- IEEE Transactions on Visualization and Computer Graphics, 1995, 1(1): 16–28. [doi: [10.1109/2945.468391](https://doi.org/10.1109/2945.468391)]
- 5 Itoh T, Yamashita F. Visualization of multi-dimensional data of bioactive chemicals using a hierarchical data visualization technique “Heiankyo view”. Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation. Tokyo, Japan. 2006. 23–29.
- 6 Herman I, Melancon G, Marshall MS. Graph visualization and navigation in information visualization: A survey. IEEE Transactions on Visualization and Computer Graphics, 2000, 6(1): 24–43. [doi: [10.1109/2945.841119](https://doi.org/10.1109/2945.841119)]
- 7 赵颖, 王权, 黄叶子, 等. 多视图合作的网络流量时序数据可视分析. 软件学报, 2016, 27(5): 1188–1198. [doi: [10.13328/j.cnki.jos.004960](https://doi.org/10.13328/j.cnki.jos.004960)]
- 8 赵颖, 樊晓平, 周芳芳, 等. 多源网络安全数据时序可视分析研究方法研究. 小型微型计算机系统, 2014, 35(4): 906–910.
- 9 Shneiderman B. Tree visualization with tree-maps: 2-d space-filling approach. ACM Transactions on Graphics, 1992, 11(1): 92–99. [doi: [10.1145/102377.115768](https://doi.org/10.1145/102377.115768)]
- 10 Keim DA, Hao MC, Dayal U, *et al.* Value-cell bar charts for visualizing large transaction data sets. IEEE Transactions on Visualization and Computer Graphics, 2007, 13(4): 822–833. [doi: [10.1109/TVCG.2007.1023](https://doi.org/10.1109/TVCG.2007.1023)]
- 11 Chen SM, Chen S, Wang ZH, *et al.* D-Map: Visual analysis of ego-centric information diffusion patterns in social media. Proceedings of 2016 IEEE Conference on Visual Analytics Science and Technology. Baltimore, MD, USA. 2016. 41–50.
- 12 Liu QS, Hu YF, Shi L, *et al.* EgoNetCloud: Event-based egocentric dynamic network visualization. Proceedings of 2015 IEEE Conference on Visual Analytics Science and Technology. Chicago, IL, USA. 2015. 65–72.
- 13 Fisher D. Using egocentric networks to understand communication. IEEE Internet Computing, 2005, 9(5): 20–28. [doi: [10.1109/MIC.2005.114](https://doi.org/10.1109/MIC.2005.114)]
- 14 廖志芳, 彭燕妮, 李永, 等. AirVis: 一个基于 Web 的空气质量数据可视分析系统. 计算机工程与应用, 2015, 51(15): 174–178. [doi: [10.3778/j.issn.1002-8331.1308-0203](https://doi.org/10.3778/j.issn.1002-8331.1308-0203)]