

基于对应分析与层次聚类的网购消费者行为研究

蔡钰麟^{1*}, 张佳慧¹, 余梓润², 张境洋³

1. 韩山师范学院 数学与统计学院, 广东 潮州 521041

2. 中共饶平县委党校, 广东 潮州 515700

3. 澳门科技大学 商学院, 中国澳门 氹仔 999078

摘要： 本文结合对应分析与层次聚类方法，基于抖音商城平台的大规模订单数据，深入分析了电商交易数据中多维变量之间的对应关系及其结构特征，揭露了网购消费者在支付方式、购物时间段与地域分布等维度下的行为模式差异。研究结果为平台在支付系统优化、区域运营决策及精准营销实施提供了理论依据与数据支持。

关键词： 消费者行为；对应分析；层次聚类；电子商务

Research on Consumer Behavior in Online Shopping Using Correspondence Analysis and Hierarchical Clustering

Cai Yulin^{1*}, Zhang Jiahui¹, She Zirun², Zhang Jingyang³

1. Department of Mathematics, Hanshan Normal University, Chaozhou, Guangdong 521041

2. Party School of the CPC Raoping County Committee, Chaozhou, Guangdong 515700

3. School of Business, Macau University of Science and Technology, Taipa, Macau, China 999078

Abstract： This study combines correspondence analysis and hierarchical clustering methods to conduct an in-depth analysis of the relationships and structural characteristics among multidimensional variables in e-commerce transaction data, using large-scale order data from the Douyin Mall platform. The findings reveal the behavioral pattern differences of online consumers across dimensions such as payment methods, shopping time periods, and geographical distribution. The results provide theoretical foundation and data support for platform optimization in payment systems, regional operational decisions, and the implementation of targeted marketing strategies.

Keywords： consumer behavior; correspondence analysis; hierarchical clustering; e-commerce

引言

随着我国电子商务的高速发展，消费者的在线购物行为愈发呈现出多元化与复杂化特征。以抖音、腾讯视频号为代表的短视频平台推动了“内容电商”模式的兴起，显著改变了用户的购物路径、消费时间与支付习惯。在此背景下，系统探析网购消费者的行为模式，对于优化平台运营机制、提升用户体验以及制定差异化营销策略，具有重要的理论价值与实践意义^[1]。

现阶段，已有学者围绕传统电商模式下的消费者行为与营销策略展开了较为系统的研究。崔玉宾^[2]结合问卷调查与定性定量分析方法，以淘宝与拼多多分别代表传统电商与社交电商，构建了电商情境下消费者行为的基础分析框架；张亮^[3]探讨了用户画像算法在饲料电商精准营销中的应用，提出了相应策略模型，为传统农业企业提供了数据赋能；Liu等人^[4]聚焦移动支付对消费者决策路径的影响，拓展了支付方式与消费行为之间关系的研究维度。在统计方法应用方面，凌峰和戚湧^[5]分别基于对应分析与聚类方法，分析了区域资源配置与人口行为模式；景帅等人^[6]利用对应分析对不同成因的白云岩进行结构划分，并通过点聚图进行可视化呈现；吴翔华和李旺男^[7]亦运用对应分析方法深入探讨了南京江北新区人才安居方式及其影响因素；贺畅和李可^[8]则将对应分析引入新能源汽车消费者功能偏好研究，验证了该方法在复杂消费行为识别中的适用性。

尽管当前学术界在电商消费者行为研究的理论体系与方法应用方面已取得一定进展，但针对电商订单数据中多维分类变量之间的交互关系仍缺乏系统性的量化分析框架，尤其是在“内容电商”快速发展的背景下，消费者在支付行为、时间选择及地域特征等方面所体现出的复杂模式尚未被充分揭示。为弥补这一研究空白，本文借鉴 Van de Velden 等人^[9]提出的“分类数据降维与聚类分析相结合”的方法路径，融合对应分析（Correspondence Analysis, CA）与层次聚类法中的离差平方和法（Ward 法），以抖音商城平台的大规模真实订单数据为基础，围绕支付方式、购物时间段与地域分布三个核心维度开展深入建模与实证分析。研究结果不仅揭示了消费者行为的多维特征，还为电商平台在支付系统优化、区域差异化运营及用户分群策略的制定提供了有力的理论支持和数据依据。

作者简介：

蔡钰麟（1997-），男，广东潮州人，硕士，韩山师范学院数学与统计学院教师，研究方向：数据挖掘，E-mail: caiy197@163.com；

张佳慧（2004-），女，广东梅州人，本科，韩山师范学院数学与统计学院在读本科生，研究方向：统计学；

余梓润（1991-），男，广东潮州人，硕士，中共饶平县委党校教员，研究方向：经济学；

张境洋（2004-），男，广东潮州人，本科，澳门科技大学商学院在读本科生，研究方向：商业分析。

一、数据来源及其预处理

（一）数据来源

本研究所使用的数据来源于广东省潮州市某大型日用陶瓷生产企业，由企业通过“抖音商城”商家端导出，经脱敏处理后用于分析。数据内容涵盖订单编号、地区、支付方式、支付时间、商品数量、订单金额、运费及优惠金额等多个字段，共计161326条有效记录。该数据集规模庞大、覆盖范围广泛，具备较强的统计代表性，能够精准反映消费者的实际行为特征，同时有助于降低个体差异和随机因素对分析结果的干扰，从而提升研究结论的稳健性与解释力。

（二）数据预处理

本文基于原始订单数据，选取“地区”“支付方式”及“支付时间”作为核心特征变量。针对连续型变量“支付时间”，采用离散化处理方法^[10]，将其转化为分类变量“购物时间段”，并划分为凌晨（0:00 - 6:00）、早上（6:00 - 12:00）、下午（12:00 - 18:00）和晚上（18:00 - 24:00）四个时间区间，以便刻画不同时间段下消费者的行为特征。同时，对“地区”、“支付方式”和“购物时间段”三个分类变量进行独热编码（One-Hot Encoding）^[11]，将其转换为二值变量，提升后续建模与分析的可操作性。随后，根据研究目的和数据特点，对处理后的数据进行分组聚合，生成二维列联表，为开展对应分析提供结构化的数据基础。例如，以“支付方式”作为列变量、“购物时间段”作为行变量，即可构建列联表（见表1），该列联表可呈现不同购物时间段下消费者支付方式的分布差异。

表1 支付方式和支付时间的列联表

	微信支付	抖音支付	抖音月付	支付宝	合计
凌晨	3646	4972	4498	6229	19345
上午	5277	5511	4474	7913	23175
下午	7690	7864	6499	11643	33696
晚上	16260	21011	20270	27569	85110
合计	32873	39358	35741	53354	161326

二、研究方法

（一）对应分析方法

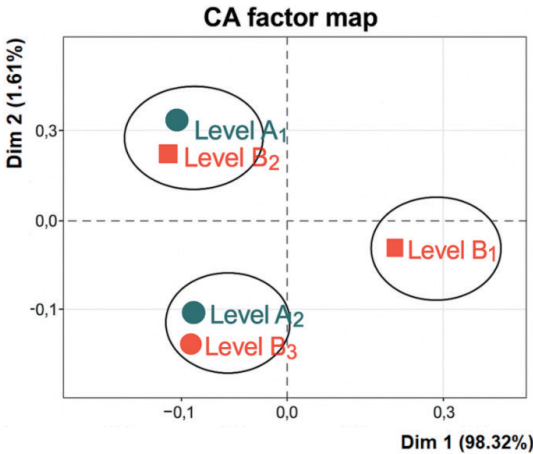
对应分析 (Correspondence Analysis, CA)^[12]是一种用于处理分类变量之间关联关系的多元统计方法，常用于可视化列联表中行列变量之间的对应模式。其基本思想是将变量各水平映射至低维空间中，以坐标点的形式展示其相互关系。其具体操作流程如下：

1. 构建列联表：基于分类变量（如支付方式与地区）生成频数矩阵。
2. 标准化处理：将频数矩阵转换为比例形式，并计算行和列的边际分布。
3. 计算残差矩阵：在边际分布的基础上，提取变量之间的“偏离”信息，反映各类别间的差异。

4. 奇异值分解（SVD）：对标准化残差矩阵进行分解。

5. 坐标计算与绘图：将每个变量水平投影至二维坐标系中，距离越近表示关联性越强。

假设存在两组分类变量 A（绿色，具有2个水平）与变量 B（橙色，具有3个水平），进行对应分析并可视化，可以得到结果如下图1所示。

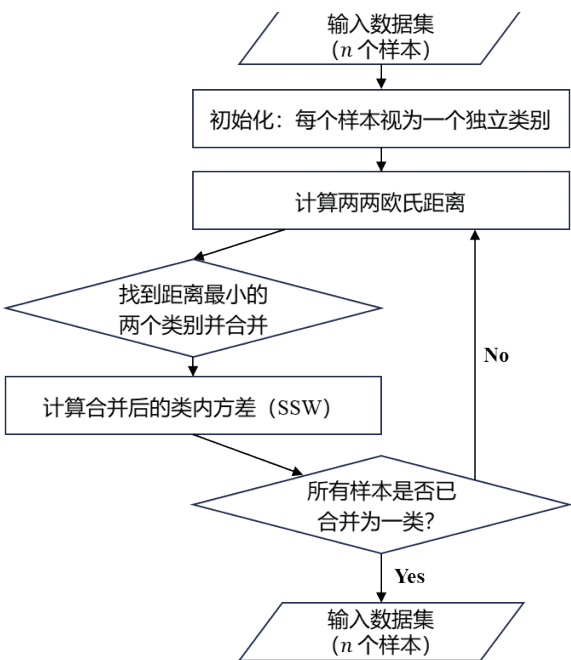


> 图1 对应分析结果示例

由图可知，变量 A 的第一个水平（Level A₁）与变量 B 的第二个水平（Level B₂）在二维空间中相互靠近，说明二者存在显著的关联关系。参考上述流程，使用对应分析将分类变量（地区、支付方式与购物时间段）的各水平映射至低维空间，以点的形式可视化呈现其对应关系和关联强度。

（二）层次聚类法

本研究基于对应分析所得的二维得分坐标，采用层次聚类法中的离差平方和法（Ward 法）对各变量水平进行聚类分析^[13]。该方法以最小化类内方差为准则，在每一步合并两个聚类时，选择使类内方差增加最小的组合，从而保证类内样本的相似性，进而获得更为合理的聚类结构^[14]。具体实施流程如图2所示。



> 图2 离差平方和聚类法的流程图

三、实验结果与分析

（一）变量独立性检验

在开展对应分析之前，需要通过卡方 χ^2 检验确认变量间是否存在统计关联^[15]。若检验结果表明变量之间存在显著的非独立性，则说明它们之间具有潜在的相关关系，适宜进一步使用对应分析探讨其各水平之间的具体对应模式。本研究对“支付方式”“购物时间段”与“地区”三组分类变量两两组合进行独立性检验，结果如表 2 所示。

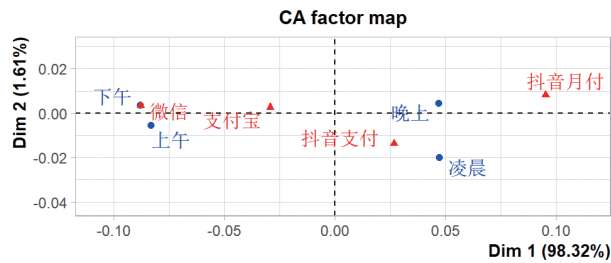
表 2 各组变量卡方检验结果

变量组合	χ^2 值	df	P 值	是否独立
支付方式、购物时间段	662.65	9	2.2e-16	否
购物时间段、地区	999.53	84	2.2e-16	否
支付方式、地区	1744.70	84	2.2e-16	否

由表可知，三组变量组合的 P 值均远小于显著性水平 0.05。因此，拒绝原假设，认为变量间均存在显著关联，有必要进一步通过对应分析方法揭示其具体的对应关系与结构特征。

（二）支付方式与购物时间段的对应关系

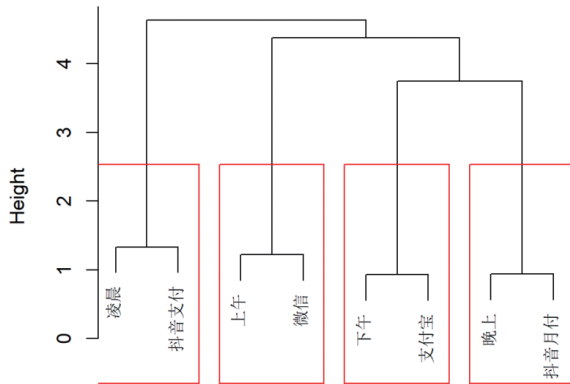
选取“支付方式”与“购物时间段”作为分析变量，构建二维列联表并进行对应分析，其结果如下图 3 所示。



> 图 3 支付方式与购物时间段的对应分析图

由上图可知，支付方式与购物时间段之间存在显著的关联性。第一主维度（Dim1）方差解释率高达 98.32%，表明该维度主要反映了支付方式与购物时间段之间的整体变异趋势。而第二维度（Dim2）方差解释率仅为 1.61%，说明其主要捕捉了支付方式在不同时段上的局部差异。

为了更直观地揭示支付方式与购物时间段的关系，采用 Ward 聚类法将上述二维平面图中的点进行聚类分析，并将其划分为四个主要类群。（见图 4）。



> 图 4 支付方式与购物时间段的聚类图

初步分析可以得到：

1. 抖音支付与凌晨购物归为同一类，表明选择该种支付方式的用户更倾向于在深夜进行消费。这可能与短视频平台的即时性、娱乐属性及其引发的冲动消费特征密切相关。

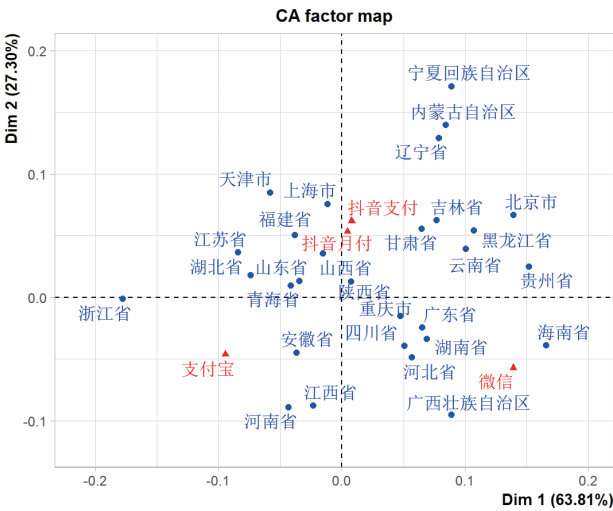
2. 微信支付与上午购物聚类在同一类别，反映出微信支付用户的消费高峰主要集中在日间，可能受工作日通勤、生活服务及商务支出等因素的影响。

3. 支付宝用户主要在下午时段活跃，与该支付方式在理财、分期付款及综合金融服务方面的特性相关，其用户可能具有更为理性的消费决策行为。

4. 抖音月付则与晚上购物群体高度契合，表明该信用支付工具更受夜间消费人群青睐，可能与消费者的超前消费习惯以及短视频电商平台的购物氛围相辅相成。

（三）支付方式与地区的对应分析

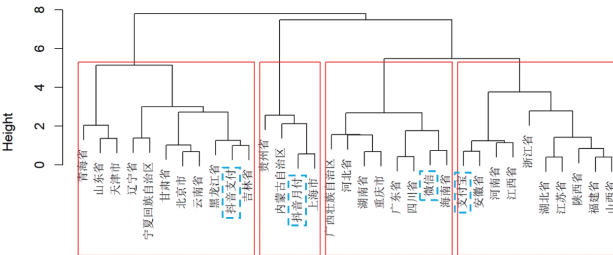
选取“支付方式”和“地区”作为分析变量，构建二维列联表并进行对应分析，其结果如下图 5 所示。



> 图 5 支付方式与地区的对应分析图

对应分析结果表明，消费者所在地与支付方式之间存在显著关联，第一维度方差解释率为 63.81%，累计解释率达 91.11%。整体来看，不同地区的支付方式选择呈现区域化特征。

进一步地，对上图中的点使用 Ward 法进行聚类，将其划分为四个主要类群，得到聚类树状图如下图 6 所示。



> 图 6 支付方式与地区的聚类图

初步分析可以得到：

1. 抖音支付与黑龙江省、吉林省、云南省等地区聚为一类，表明该类群中多数为北方及西部地区，消费者更倾向于通过短视

频平台完成购物，尤其在娱乐化、社交化的使用场景中，抖音支付的接受度较高。

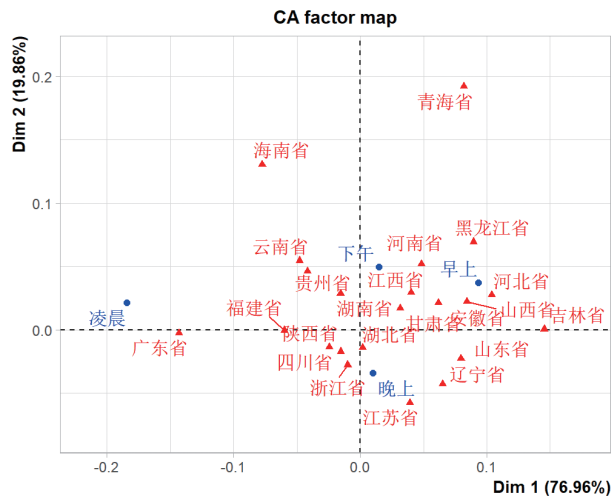
2. 抖音月付的使用群体主要集中在内蒙古自治区、上海市等地，用户更易受到短视频平台内容驱动的消费影响，偏好“先享后付”式的信用支付方式，该现象与平台的内容生态与分期推广策略密切相关。

3. 微信支付在广东省、海南省等地区的使用偏好明显，可能得益于其在社交支付、日常转账及本地生活场景中的高度融合，形成稳固的使用习惯。

4. 支付宝则在浙江省、江苏省、安徽省等东部沿海及中部省份占据主导地位，显示出其在上述地区具备较强的用户基础和金融服务渗透力，这与其早期推广策略和本地化运营优势密切相关。

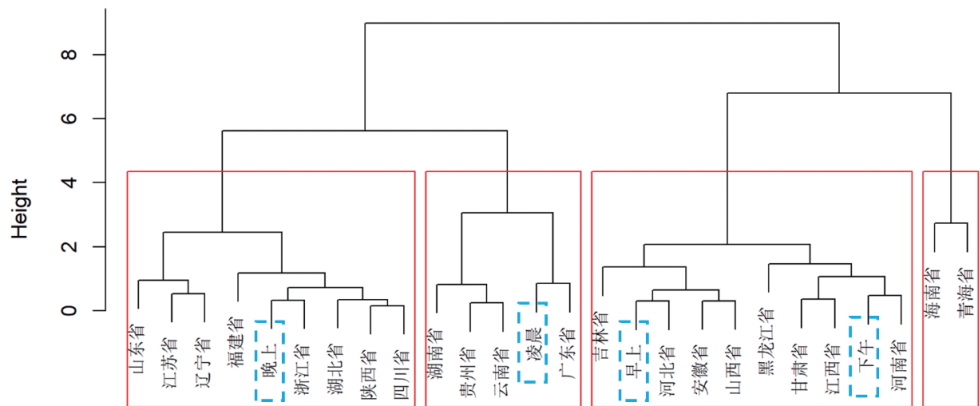
（四）地区与购物时间段的对应分析

选取“地区”和“购物时间段”作为分析变量，构建二维列联表并进行对应分析，其结果如下图 7 所示。



> 图 7 地区与购物时间段的对应分析图

由对应分析结果可知，消费者所在地与购物时间段之间存在显著关联性。第一主维度（Dim1）方差解释率为 76.96%，累计解释率达 96.82%，表明二维空间能较好地揭示两者关系。针对上图进行聚类，得到系统聚类图如下图 8 所示。



> 图 8 地区与购物时间段的聚类图

由聚类结果初步分析可以得到：

1. 江苏省、浙江省、湖北省、辽宁省等地区与“晚上”购物时间段聚合，说明这些地区的消费者更倾向于夜间购物，这与当地的夜间经济的活跃度以及年轻消费者群体的偏好有关。

2. 广东省、贵州省、云南省等地区形成一个独立类群，该类群中的消费者主要集中在“凌晨”消费，其购物时间分布也更加均衡，不易被归入单一时间段的聚类中。

3. “早上”和“下午”被归为同一类，表明这两个时间段在消费者行为模式上较为相似。在大多数地区，工作日的上午和下午都是消费者进行网购的主要时段，网购行为可能集中于工作间隙或休息时间，从而使其聚类结果趋同。

4. 海南省与青海省单独成类，这一现象可能是由于这两个省份的消费模式在数据中较为独特，有较强的地方特色且数据样本量相对较少，因此未能与其他地区或购物时间形成强关联。

四、结论和建议

本研究围绕电商交易数据，系统探讨了支付方式、购物时间段与地域特征三者之间的关联关系，揭示了网购消费者在多维属性下的消费行为差异。通过卡方独立性检验与对应分析方法，验证了变量间的显著相关性，前两个主维度的累计方差解释率均超过 90%，并以散点图形式对变量关系进行了可视化呈现，从而直观反映了各变量水平之间的对应结构。进一步，采用层次聚类法对高度相关的因素进行了聚类分析。研究结果显示，消费者在支付方式的选择上存在显著的时间偏好与地域差异，不同地区用户的购物时间分布亦表现出明显特征。基于上述发现，本文提出以下针对平台界面优化、区域精细化运营及用户精准营销的策略建议。

（一）支付工具与产品设计优化

针对不同地区用户的支付偏好，平台可在支付界面中优化支付方式次序的呈现逻辑。例如，在江苏、浙江等地优先显示支付宝，在广东、湖南则默认推荐微信支付。此举有助于简化用户的支付选择流程，提升支付便捷性与交易转化率。

（二）用户分群与精细化运营

依据地域维度构建用户分群模型，并联合支付平台推出区域

专属优惠方案，如“江浙支付宝红包专区”或“广东微信支付福利包”。同时，结合消费者在不同平台的使用偏好，开展定向广告投放与差异化营销策略。针对地区间的购物时间偏好差异，可在用户活跃高峰时段实施精准促销活动（如直播带货、限时折扣等），进一步提升用户参与度与购买转化率。

（三）系统资源调度与服务体验优化

平台应结合地域与时间特征，动态优化资源配置。例如，在夜间消费活跃地区（如江苏、浙江），可增强夜间物流配送能力，提升晚间购物履约效率；同时，根据各地不同时段的用户活跃度

动态调配服务器资源，从而提升平台的响应速度与系统稳定性，降低整体运维压力与技术成本。

（四）未来展望

本研究基于静态订单数据开展实证分析，未来可结合动态行为数据（如用户点击路径、停留时间等）进一步挖掘消费者支付偏好演变趋势。同时，引入人口统计学特征（如年龄、性别、收入等）进行多维用户画像建模，有助于实现更高粒度的市场细分与个性化营销策略优化。

参考文献

[1]陈红. 电子商务环境下消费者购买行为研究 [J]. 商场现代化, 2025, (02): 10-12.
[2]崔玉宾. 电子商务对消费者行为的影响分析——基于两种电商模式的比较 [J]. 商业经济研究, 2020, (10): 89-93.
[3]张亮. 基于用户画像算法的饲料电商精准营销策略研究 [J]. 中国饲料, 2024, (20): 100-103.
[4]Liu Y, Luo J, Zhang L. The effects of mobile payment on consumer behavior [J]. Journal of Consumer Behaviour, 2021, 20(3): 512-520.
[5]凌峰, 戚湧. 基于聚类与对应分析的江苏科技资源配置现状研究 [J]. 科技进步与对策, 2019, 36 (10): 41-48.
[6]景帅, 王起琮, 程浩, 等. R-Q 因子分析在塔里木盆地寒武系白云岩岩相分析中的应用 [J]. 桂林理工大学学报, 2020, 40 (04): 695-702.
[7]吴翔华, 李旺男. 基于对应分析的人才安居方式及影响因素研究——以南京江北新区为例 [J]. 现代城市研究, 2023, (09): 8-12.
[8]贺畅, 李可. 基于对应分析的新能源汽车智能化功能偏好研究 [J]. 汽车文摘, 2024, (12): 8-13.
[9]Van de Velden M, D'Enza A I, Palumbo F. Cluster correspondence analysis [J]. Psychometrika, 2017, 82(1): 158-185.
[10]Pivkin V Y, Pivkina I V. Processing of experimental data using discretization of their domains [J]. Optoelectronics, Instrumentation and Data Processing, 2010, 46: 507-509.
[11]Qiu J, Lin Z, Li Y. Predicting customer purchase behavior in the e-commerce context [J]. Electronic commerce research, 2015, 15: 427-452.
[12]Greenacre M J. Correspondence analysis [J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2010, 2(5): 613-619.
[13]Nielsen F, Nielsen F. Hierarchical clustering [J]. Introduction to HPC with MPI for Data Science, 2016: 195-211.
[14]Eszergár-Kiss D, Caesar B. Definition of user groups applying Ward's method [J]. Transportation Research Procedia, 2017, 22: 25-34.
[15]McHugh M L. The chi-square test of independence [J]. Biochemia medica, 2013, 23(2): 143-149.