

模拟情境下经济决策及财富关联可视分析

刘玉华^{1,2)}, 石青¹⁾, 郑凤玲¹⁾, 谢琬滢¹⁾, 韩书焕¹⁾, 周志光^{1,3)*}

¹⁾ (浙江财经大学信息管理与人工智能学院 杭州 310018)

²⁾ (杭州电子科技大学数字媒体与艺术设计学院 杭州 310018)

³⁾ (浙江大学 CAD&CG 国家重点实验室 杭州 310058)

(zhgzhou1983@163.com)

摘要: 针对经济学领域中采用实验经济学方法产生的大规模高维动态数据, 在与实验经济学专家密切合作的基础上, 提出模拟情境下真实努力任务实验数据可视分析方法. 设计决策过程信息图以全面呈现实验过程中参与者的总体决策分布和个体行为特征; 设计自适应决策分类图、财富时序变化特征图、总体决策桑基图等方法直观地呈现实验参与者的决策过程表现与财富结果的关联关系. 集成上述可视化设计与数据挖掘算法模型, 开发面向真实努力任务实验数据的可视分析系统, 帮助领域专家探索式地发掘参与者决策行为特征以及财富变化关键因素, 为专家探索和分析贫困产生及贫困循环恶化的内在机制提供依据. 大量基于真实数据的案例分析和专家反馈信息进一步验证了该方法和系统工具的有效性和实用性.

关键词: 实验经济学; 真实努力任务; 可视分析; 人机交互

中图分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2021.18645

Visual Analysis of Economic Decision-Making and Wealth Change under Simulated Situations

Liu Yuhua^{1,2)}, Shi Qing¹⁾, Zheng Fengling¹⁾, Xie Wanying¹⁾, Han Shuhuan¹⁾, and Zhou Zhiguang^{1,3)*}

¹⁾ (School of Information Management and Artificial Intelligence, Zhejiang University of Finance and Economics, Hangzhou 310018)

²⁾ (School of Media and Design, Hangzhou Dianzi University, Hangzhou 310018)

³⁾ (State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058)

Abstract: For the large-scale high-dimensional and dynamic data generated in economic experiments, a visual analysis method is proposed to explore the experimental data of real-effort tasks under simulated situations based on our cooperation with experimental economics experts. A parallel coordinate-like information diagram is designed to give an overview of the experimental process and individual behaviors. Interactive decision tree view, time-series wealth view, and general decision Sankey view are provided to explore the correlation between participants' behaviors and wealth change. Further, a visual analysis system integrating the visual designs and data mining models is developed for domain experts to analyze the decision-making behaviors and key factors on the wealth change, and provides evidence for the internal mechanism of poverty generation and cycle deterioration. Case studies and expert feedbacks based on real data verify the effectiveness of our system.

收稿日期: 2020-09-09; 修回日期: 2021-05-01. 基金项目: 国家自然科学基金(61802339, 61872314); 教育部人文社会科学研究项目(18YJC910017); 浙江省高校重大人文社科攻关计划(2018QN021); 浙江省自然科学基金(LY18F020024); 浙江大学 CAD&CG 国家重点实验室开放课题(A2001); 浙江省统计研究重点项目(20TJZZ10). 刘玉华(1988—), 男, 博士, 讲师, 硕士生导师, CCF 会员, 主要研究方向为数据可视化与可视分析; 石青(1999—), 男, 在校学生; 郑凤玲(1998—), 女, 硕士研究生, 主要研究方向为数据可视化; 谢琬滢(2000—), 女, 在校学生; 韩书焕(2000—), 女, 在校学生; 周志光(1983—), 男, 博士, 教授, 硕士生导师, CCF 会员, 论文通讯作者, 主要研究方向为数据可视化与可视分析.

Key words: experimental economics; real-effort task; visual analysis; human-computer interaction

贫困产生的原因和如何帮助贫困人口摆脱贫困一直是经济学领域的研究热点,同时也是发展中国家在社会建设过程中亟待解决的重要任务^[1]。研究人员通常采用实验经济学方法分析贫困恶性循环及贫困陷阱的内在机制。经典的案例就是搭建受控的实验场景,让实验者以简单、抽象的方式在模拟的财富相关情境下生存和发展,记录他们的决策行为和财富变化,进而检验和探索经济理论和方法^[2]。

在搭建具体的实验场景过程中,为了提高实验的真实感和外部有效性,通常设计大量的情境模拟与真实努力任务^[3],使实验过程较为复杂,所记录的实验数据具有典型的数据维度高、决策动态变化、状态随机转移和分析结果不确定性高等特点,即便是专业的实验设定者对于实验的过程和结果都难以预期和把握,为实验数据的进一步分析和探索带来挑战。例如,马涛等^[4]在对实验经济学进行专题研讨时指出,现实的经济现象和经济行为非常复杂,而实验经济学为了模仿现实生活不得不引入更多的变量和设计更复杂的实验流程,由于缺乏有效的工具和手段,难以深入挖掘和探索。

可视化与可视分析^[5]借助人眼视觉容易感知的图形或图像,直观地呈现数据及其内部隐含的特征模式,并且支持用户根据先验知识交互式地探索和分析数据。将可视化与可视分析技术应用于经济实验数据具有2点优势:(1)有效地引入算法模型并调控参数,将原始数据、分析模型和实验结果进行无缝拼接,方便用户评估实验数据和模型,深入探索实验参与者的行为决策与财富、时间等信息之间的关联;(2)提供一个交互式的可视分析平台,允许用户对复杂多维时变实验数据进行动态过滤和展示,支持专家多角度、多粒度地观察和探索实验流程与相关细节。

本文提出面向经济行为实验数据的可视分析方法,通过与经济行为实验设计者深入交流和讨论,结合数据特点和实验设计者的分析任务,设计多尺度和多角度的可视化解决方案,支持专家进行宏观实验流程的概览与微观决策行为细节的分析;从时间、决策空间和参与者等维度出发探索实验的不同方面,从而帮助专家从新的视角解读经济行为实验数据,审视和研究以往的经济理论,并为现有扶贫工作和政策提供一定的启示。

本文基于经济行为领域的分析任务建立理论模型,并设计相关的可视化方法,包括实验流程信

息图、决策分类图、财富曲线特征图等,从多方面分析决策行为和财富收益之间的关联关系,并聚焦分析其中的时变趋势、聚类特征、行为模式等;面向经济行为实验数据开发一套可视分析系统工具,帮助专家进行数据分析、假设论证和演绎推理,提供一定的证据支持新观点的发现与启示,达到经济行为内在机制揭示与理解的目的。

1 相关工作

1.1 经济行为实验数据分析

随着实验经济学的不断发展,研究人员陆续开展田野实验和实验室实验,研究和探索贫困恶性循环及贫困陷阱的产生原因和内在机理,其中贫困人口的风险偏好和厌恶是最受关注的焦点。风险偏好和厌恶是个体承担风险的基本态度,也是个人感知决策情景及制定风险决策的重要前导因素。风险偏好可以解释为个体面对不确定性的风险和投资时主动追求风险以获得潜在盈利;而风险厌恶指个体不愿承担风险带来的利益损失,更喜欢比较确定的投机。来自越南农村的田野实验^[6]和发薪前后行为个体风险偏好实验^[7]都发现,风险厌恶与个体的财富水平并无显著性关系。而来自赞比亚的田野实验^[8]和埃塞俄比亚的田野实验^[9]则发现,贫困人口相对于非贫困人口的风险厌恶程度要高。也有来自其他地区和国家的田野实验^[10]得到相反的结论,即贫困人口的风险厌恶程度低于非贫困人口。罗俊等^[11]在对田野实验进行综述调研时指出,目前国内经济学家尚未就田野实验进行研究,在复杂性和多样性并存的中国情境下开展田野实验仍有待进一步的分析探索。实验室实验^[12]则通过给定外生禀赋模拟不同的贫富程度,解决田野实验在考察贫困行为时面临的样本量不足和财富标准不统一等问题。但是在现实生活中,贫富状况都是由自身原因造成的,经济决策时所用的财富都是自有的,因此单纯地通过给予不同的初始禀赋代表不同的贫富状态存在外部有效性问题^[13]。所以,实验经济学家一直在改进实验室实验的环境和流程,通过引入和设计模拟情境提高决策情境的代入感、真实性和外部有效性等。但是,以上经济行为实验数据的分析仍未脱离传统分析方法的范畴。据了解,目前尚无研究工作将可视分析技术用于经济行为实验数据的探索,

而本文方法填补了可视分析方向与经济行为实验方向交叉领域的空白,通过可视分析技术解决领域问题。

1.2 经济数据可视化

经济数据可大致分为宏观经济数据(如人口普查数据、国内生产总值数据等)和微观经济数据(股票数据、商品交易数据等),本节从这 2 个方面总结可视化方法在经济数据分析中的应用。

在宏观经济数据可视化方面,周志光等^[14]实现了面向国内生产总值数据的时空多维属性可视分析系统,帮助用户快速探索经济统计数据中产业结构变化的时空特征模式。针对多维时变的金融数据, Dominik 等^[15]利用移动窗口构造相邻时间内重叠的数据集,从而设计出时序多尺度降维方法,进一步通过可视化展示帮助用户分析和探索数据的变化特征。周志光等^[16]设计面向投入产出表数据分析的经济产业结构关联特征可视化工具 VisIOT,有效地分析和探索其中复杂的关联模式和时序变化特征。

在微观经济数据可视化方面, Yue 等^[17]开发了多层级的股票可视分析工具 sPortfolio,从风险因素层面、多重投资组合层面和单一投资组合层面帮助用户理解市场发展趋势。Tanahashi 等^[18]提出的参与式的多功能可视化系统 Stock Lamp,能够持续传递市场信息,帮助用户实时跟踪股市,以应对多目标的分析任务。Wang 等^[19]提出的基于在线评论的可视分析系统 E-Comp,帮助用户在不同的细节层次上深入比较本地企业,通过在线评论、评级打分和业务之间的对比做出最佳的购买决策。胡珉等^[20]利用空间重构技术和可视分析技术表达复杂系统规律和特征的优势,提出了以股市分钟级交易信息为基础的股票市场趋势预测可视化方法。

上述方法多从结果的角度出发对历史数据进行总结和学习,进一步分析和推断发展趋势,研究角度具有一定的局限性。本文利用的经济行为实验数据采用正向模拟的方式,对经济数据进行过程式的分析和解读,为经济学中贫困因素的分析提供了新的方向和角度。

1.3 决策树模型可视分析

数据挖掘模型在经济和金融行业中应用广泛,能够帮助用户快速理解客户关系、实施精准营销以及安全与反欺诈等。其中,最常见的就是决策树模型,其利用树形结构表示决策集合,这些决策集合通过对数据集的分类产生规则,辅助用户进行决策分析。针对决策树以及基于决策树的集成模型,不少学者对其内部过程和训练结果进行可视分析,

从而直观地探索模型与数据集之间的关系以及模型数据的分布特点,进一步实现模型参数的优化设计。例如, Teoh 等^[21]设计的交互式构建决策树的可视分析系统 PaintingClass,通过改进传统的冰柱图展示用户感兴趣的局部结构,同时利用平行坐标展示多维训练数据集,辅助用户探索数据在决策树节点的分布。van den Elzen 等^[22]从用户需求和领域任务出发,开发了一个结合可视化、人机交互和算法支持的集成系统 BaobabView,使用节点链接树帮助专家理解和分析决策树的生长、剪枝和优化等过程。此外,还有一些方法使用缩进树^[23]、径向填充^[24]和树图^[25]的布局方法展示和探索决策树。为了提高提升树模型的训练效率和表现力, Liu 等^[26]结合树可视化、*t*-SNE 投影以及时序混淆矩阵,开发了可视分析工具 BOOSTVis,帮助专家有效改进提升树的训练过程和训练精度。Huang 等^[27]开发了一个渐进式的可视分析系统 GBRTVis 实时展示梯度提升树的训练过程,结合聚类视图、树图和层次树图增强用户对梯度提升树的理解。Zhao 等^[28]结合现有的机器学习模型可视化和人机交互技术开发了名为 iForest 的随机森林可视分析系统,通过展示输入特征与预测结果之间的关系以及与决策路径相关的工作机理,帮助用户更好地理解随机森林模型。还有一些工作聚焦于分析训练集和随机森林模型之间的关系,如使用多尺度降维和自组织映射方法同步展示原始多维数据的分布和基于随机森林观测的数据分布,对比和分析模型的分类效果^[29]。面向多维时变的经济行为实验数据,本文利用决策树模型对其进行处理与分析,并根据领域专家的需求来设计和定制相应的数据可视化方法,以有效地揭示关键决策行为对财富影响的重要性,并对比和总结这些影响在不同实验轮次中的关系。

2 分析任务和系统概览

2.1 模拟情境实验流程和数据介绍

本文采用的经济行为实验数据由浙江财经大学经济行为实验中心提供,实验参与者一共 304 名,全部来自在校本科生,分属于不同年级和不同专业(包括公共管理、金融、计算机等)。参与者到达实验现场后,通过实验员的讲解了解实验的全过程,并知晓实验报酬主要依赖参与者的努力程度和决策行为。每名参与者的收益由最终的实验财富排名决定,排名越高则收益越高。实验任务收益为 50~120 元,这样可以保证参与者真实有效地投入经济行为

实验中, 从而提高实验模拟的效果.

实验共 20 轮次, 如图 1 所示, 每轮的实验过程一样, 都由情境模拟和风险偏好测试组成. 情境模拟是主要流程, 包括工作、投资、保险和负面冲击等诸多环节. 例如, 在工作环节引入了真实努力任务(real-effort task), 参与者在给定时间内完成答题, 每答对一题获得 3 个财富. 此外, 为了让实验更贴合现实, 在工作环节引入失业机制, 每个参与者有 1/10 的概率失业, 这样能够对财富形成实际性的冲击, 促使财富分化; 在投资环节, 参与者可自由选择投资金额, 而系统会随机决定用户投资所得回报; 负面冲击环节模拟生活中的地震、洪水和火灾等灾害对财产造成的损失, 不同级别的灾害发生概率不同, 造成的损失程度也不同; 为进一步贴近现实, 设置保险环节削弱负面冲击的影响, 参与者可选择财富总额的 10%作为保费对财富进行投保, 若发生负面冲击, 系统将进行全额赔付. 情境模拟之后是风险测试, 系统提供多种方案供参与者选择, 方案风险级别越高则收益也越高, 这样可通过用户选择了解他们的风险厌恶程度.

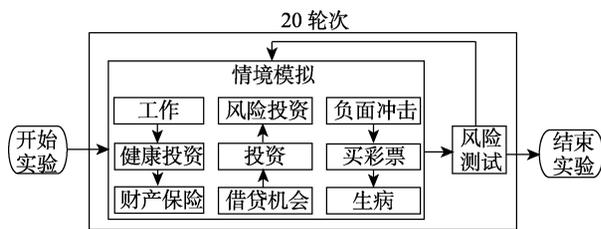


图 1 经济行为实验流程

从图 1 可以看出, 实验设置的各个环节都在模拟真实的生活情境, 尽可能地为参与者提供良好的代入感和真实感. 而参与者由于自身能力、努力程度、决策行为和运气等差异, 不同轮次不同环节的财富收益大小不同, 并且能够发生动态变化. 需要注意的是, 参与者所赚取的财富在每轮结束后并没有清零, 而是带入下一轮中用于投资和投保

等活动, 在实验进行 20 轮次结束时, 参与者根据所积累的财富进行报酬结算.

2.2 可视分析任务

本文与经济领域专家及实验设计者进行了长期深入的交流与合作, 详细地了解他们对于实验数据的分析和探索的流程. 领域专家通常利用预期性较强的统计模型进行分析, 难以获得有效的分析结果, 同时反复地存取数据和计算中间结果, 也容易导致分析过程不连续、误差累积、不确定性增加等问题. 经过与他们长期讨论, 最终确定了以下可视分析任务, 帮助他们解决以往的问题.

T1. 决策过程可视化. 提供整个实验环节的概览, 探索参与者的决策行为模式, 聚焦分析任意轮次感兴趣参与者的基本实验行为和财富收益.

T2. 财富趋势可视化. 微观层级分析参与者的财富收益随轮次的变化趋势, 从个人能力、决策行为、运气等方面呈现影响财富波动的主要因素, 进一步对比不同参与者的财富收益变化及产生原因.

T3. 决策行为和财富关联可视分析. 根据参与者的决策行为和财富收益, 探索不同决策行为对财富影响的重要性, 并深入分析其影响原因和机制.

T4. 实验经济数据可视分析系统. 支持专家和用户通过一定的交互操作, 选取和查看不同轮次、不同参与者的实验数据, 通过多视图联动协同分析决策行为和财富收益的关联关系.

2.3 系统流程

根据以上分析任务, 通过与专家的讨论合作, 迭代式地开发了一套面向经济行为实验数据的可视分析系统. 如图 2 所示, 首先设计了类似平行坐标的实验流程信息图, 展示某一轮次所有实验环节的详细流程(T1); 其次设计财富曲线图, 详细对比不同参与者在所有轮次的决策差异和财富变化趋势(T2); 再通过决策树模型探索影响财富收益的关键因素, 并用旭日图的形式展示训练过程和结果(T3), 进一步利用桑基图总结并展示所有轮次

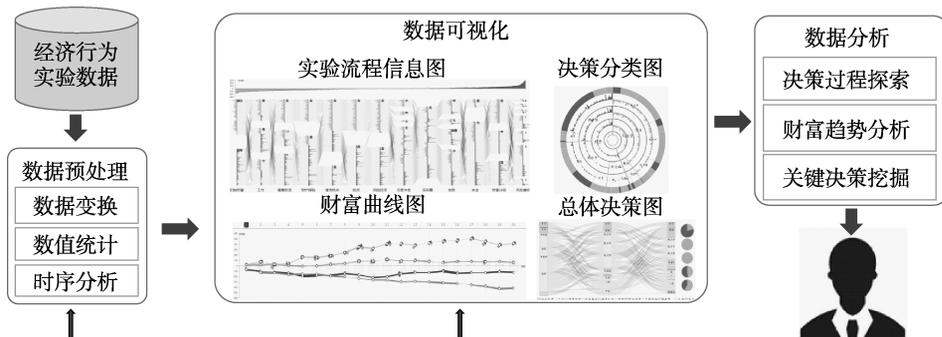


图 2 系统流程概览

的决策树综合信息；最后通过交互集成上述可视化视图和模型进行联动展示(T4)，帮助用户深入分析决策行为和财富变化关键因素。

3 可视化设计

系统整体界面如图 3 所示，本节将依次介绍不同视图的设计原则、构思过程和交互设计。

3.1 决策过程可视化

平行坐标是多维数据可视化的常用方法，而且与时间序列可视化密切相关，可以将坐标轴对应表示不同时间点的数据。考虑某一实验轮次内包含诸多先后进行的实验环节，本文设计了类似平行坐标的实验流程信息图作为系统的主视图，主要展示所有参与者在某轮实验中的详细决策过程。

如图 3 左上方的实验流程信息图所示，每列表示一个实验环节，列下方的文字对应当前实验环节的名称，如工作、健康投资和财产保险等，按照实验环节先后顺序对其进行从左到右的排序。为了更好地体现不同环节对应的决策种类和数量，将每列划分成不同段，段之间具有一定的间隔，同时在右上方标记文字，帮助用户区分相应的决策行为。例如，第 4 列表示财产保险，上下 2 段分别代表不投保和参与投保 2 种行为，而段的高度表示选择这种决策行为的参与者数量；第 8 列表示负面冲击，从上往下 4 个分段依次表示小、中、大灾害，以及无灾害，它们的高度反映了对应决策下的参与者数量。

每个参与者由一条曲线表示，曲线在不同列的不同分段中穿插游走代表了该用户的决策在不同实验环节的变化。参与者通过每环节后都会有财富的增长或减少，因此在曲线与分段相交的位置绘制水平的线条，其中，红色代表用户财富收益为富裕，橙色代表用户财富收益为中等，而绿色代表用户收益为贫穷。线条长度表示收益绝对值，线条沿轴线向左和向右分别表示收益为负和收益为正，这样可以方便用户跟踪和探索实验参与者在做出某一决策后的财富状况。为了更好地显示决策过程，借用 storyline 中的扫描优化算法^[30]最小化线条之间的交叉，使用户可以清晰地观察线条的走势和分布。

在实验流程信息图的最上方显示财富排名柱状图作为辅助视图，将每轮次所有参与者的收益情况进行汇总排序。其中，纵轴表示收益，横轴上从左到右的柱状图表示财富从低到高的参与者；绿色、橙色和红色依次表示贫穷、中等和富裕；高度代表绝对值。此外，将所有用户在上一轮的收益情况绘制成折线与柱状图同步显示，这样用户可以通过折线和柱状图的高度差异选择感兴趣的财富变化差异较大的参与者，将对应的线条高亮显示在当前视图和财富曲线图中做进一步的分析和探索。

3.2 财富趋势可视化

为了展示所有参与者的总体财富随时间的变化趋势，系统使用箱线图的形式对每轮所有参与者的财富进行统计和绘制。如图 3 左下方的财富曲

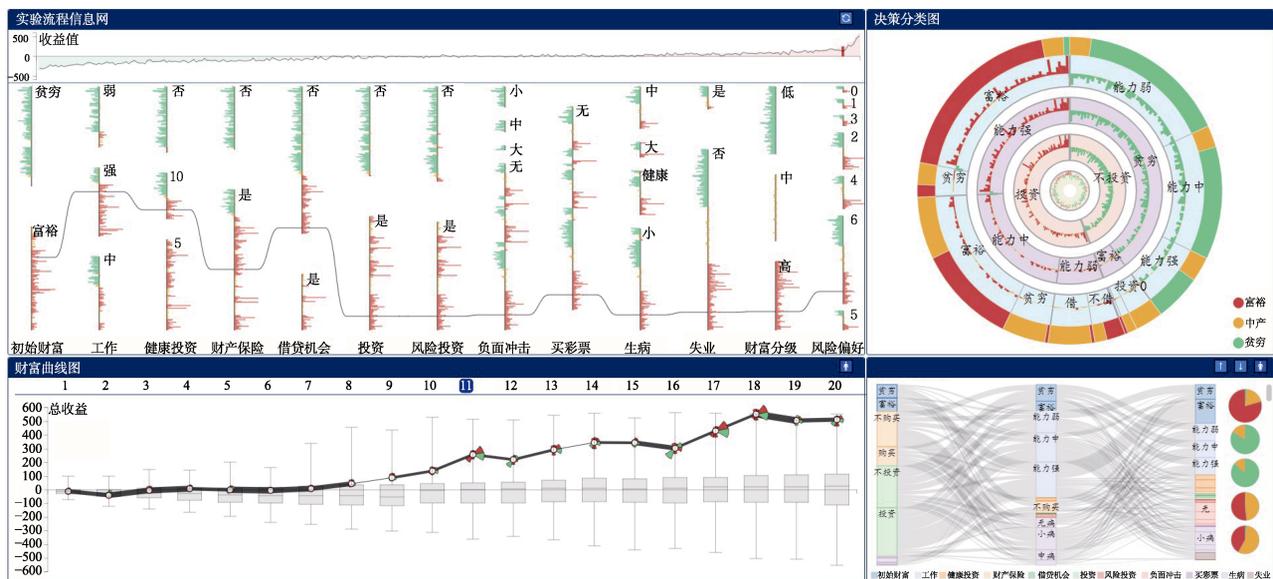


图 3 面向经济行为实验数据的可视分析系统界面

线图所示, 纵轴代表收益, 横轴代表轮次, 20 个箱线图分别对应 20 轮次实验的财富统计结果, 这样用户可以通过箱线图的差异对比财富的总体变化特征; 还可以单击上方的实验轮次编号, 以更新实验流程信息图和决策分类图对应轮次的可视化效果。

为了帮助用户聚焦分析某些参与者的个人财富变化信息, 将参与者在每轮次的收益点连成一条曲线展示他们的财富时变趋势。线条的宽度映射相邻轮次间的决策行为相似度, 线条越粗则表示差异越大, 从而帮助用户理解和判断决策行为差异对财富积累的影响。而相比于工作、投保等主动决策行为, 负面冲击、生病和失业等属于被动决策行为, 由系统随机选择, 与参与者的运气有关。为区分个人能力、决策行为和运气等方面对财富波动的影响, 本文将参与者在每轮次的具体收益用花瓣图的形式进行展示。在设计之初, 本文尝试了条形图和雷达图 2 种方案, 发现条形图在水平方向会占据较多的空间, 容易造成相邻 2 个点的图案交叉和重叠, 不利于探索和观察。而雷达图通过沿径向轴上分布的坐标点表达某一维度的数值, 使用户对收益正负和大小的感知不如花瓣图中红绿填充的扇面直观和明显。因此, 最终选择了图 4 所示的花瓣图设计, 每个扇区对应一个实验环节, 红色和绿色表示在该实验环节前后财富的增长和减少, 扇区的高度表示财富的变动程度。当用户在实验流程信息图选择感兴趣的参与者时, 财富曲线图会高亮显示他们的财富曲线帮助用户进行深入对比和探索。

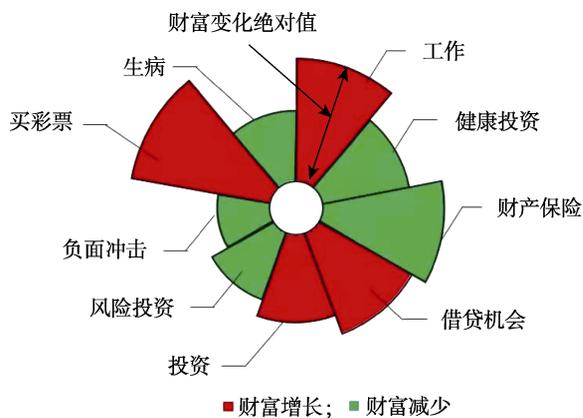


图 4 财富变化特征花瓣图

3.3 决策行为和财富关联可视化

为了分析不同的决策行为对最终财富收益的影响, 设计决策行为与财富关联可视化, 主要包括决策分类可视化和决策结果可视化 2 部分。

(1) 决策分类可视化。将每轮的实验数据作为训练集, 利用决策树模型对其进行分析和探索。因为参与者在每轮次的收益不会清零, 都会带入下一轮参与相关决策, 所以在原有的实验数据基础上添加了初始财富状态作为一个影响因子参与决策树的训练过程。而参与者在每轮开始和结束后的财富按照大小等分成富裕、中产和贫穷 3 类。因此, 初始财富和工作能力、健康、投保等实验环节都作为划分属性, 而每轮实验结束后参与者的财富状态作为决策树的分类依据。

决策树默认构造 3 层, 构建完毕后, 根据其层次数据的性质采用旭日图的可视化形式展示模型的训练过程。在实验最开始阶段, 本文分别尝试了节点链接树和冰柱图展示决策树的训练过程, 但是发现这 2 种方法都是从上到下展示层次结构, 而随着决策树往下细分, 对应的节点越多、信息量越大时, 底层空间难以清晰展示如此多的细节; 随后尝试了树图的方法, 尽管这种方法对空间利用更充分, 但是层级结构体现得不明显, 用户很难区分不同层级间的划分属性和结果差异。相比之下, 旭日图由内向外采用环形布局依次展示层级信息, 越往外可利用的展示空间越多, 更方便展示相关的细节信息。如图 3 右上方的决策分类图所示, 从内往外每层依次按照决策树生长过程中的属性选择顺序进行绘制, 每个扇环表示相应的决策种类, 附加其上的文字显示当前的具体决策行为名称。扇环的圆心角表示训练数据落在该节点的数量, 扇形内部则用一条弧线沿中心分隔开, 本轮财富最终为正的参与者表示成沿中心弧线往内的径向线条, 线条高度表示收益绝对值。类似地, 收益为负的参与者使用沿中心弧线往外的径向线条表示, 而红色、黄色和绿色依次表示富裕、中产和贫穷。最内层圆上的线条根据实验参与者的 id 号进行排列, 用户可以通过对比最内层和其他层线条的颜色分离效果, 了解训练前后数据按照最优属性进行划分的作用。最外层的圆环按照落在最后一层训练节点上的富裕、中产和贫穷用户的比例填充相应的颜色, 这样可以更好地帮助用户判断决策树训练后叶子节点的财富分类纯度。基于这种设计的决策分类图, 可以通过旭日图由内向外的可视化映射展示决策树的构建、生长和优化过程, 用户可以交互查看不同轮次的决策树训练结果, 对比分析影响最终财富收益的关键因素。

(2) 决策结果可视化。每轮的实验数据通过决策树模型进行训练和分析, 会得到 20 棵决策树。

但为了形成更具推广性的、普适性的经验和知识,有必要综合分析所有的决策以达到理解决策行为对财富收益影响机制的目的.因此,利用桑基图设计了总体决策图分析和探索决策之间的流通关系.

图 3 右下方所示的桑基图展示了决策路径流,从左到右每列汇总了所有决策树在每层的决策信息.例如,左起第 1 列代表决策树的根节点统计信息,其中每块代表一种决策行为,块的高度反映了该决策行为在所有实验数据集被训练时作为第 1 个属性选择的占比.具有相同颜色的桑基块代表对应的行为同属于一种决策类别.而相邻 2 列桑基块间的曲线代表决策的流向模式,曲线宽度表示当前流向中包含的决策路径数量.使用饼状图编码所有决策路径形成的叶节点,饼状图的大小表示落在该叶子节点的样本数量,内部扇区的颜色和圆心角分别映射财富收益情况和所占比例.所有饼状图排列在右侧,用户可以通过拖动滑块查看任意饼状图.同时,系统提供的排序功能可以帮助用户按照人均财富收益或人数等对饼状图进行排序.当选定某一饼状图后,其轮廓会被高亮成蓝色,对应的决策路径也会同步高亮显示在桑基图中.

4 实验及结果分析

可视分析系统使用经典的 Web 框架 Flask 进行开发,原始数据经预处理后保存到 MySQL 数据库中,后端采用 Python 作为支持,前端界面采用 HTML+CSS+JavaScript,使用 Bootstrap.js, D3.js 和 jquery.js 支持可视化视图的显示和交互.在设计完成经济行为决策可视分析系统后,本文邀请实验设计者作为领域专家参与系统的使用,共同进行数据探索和案例分析.

4.1 案例分析

(1) 决策行为过程分析.加载所有经济行为实验数据后,专家首先观察到如图 5 所示的第 1 轮~第 20 轮参与者整体收入分布的箱线图.不难看出,在实验的初始阶段,参与者的财富分布多集中于 0 附近;随着实验的进行,箱线图中表示上四分位数与下四分位数距离的矩形逐渐变宽,意味着所有参与者的财富分布范围越来越大.经过讨论,专家认为情境模拟中设置的失业、负面冲击、生病等环节能够对财富形成实际性的冲击,成功地促使财富的分化,促进贫富个体的产生.

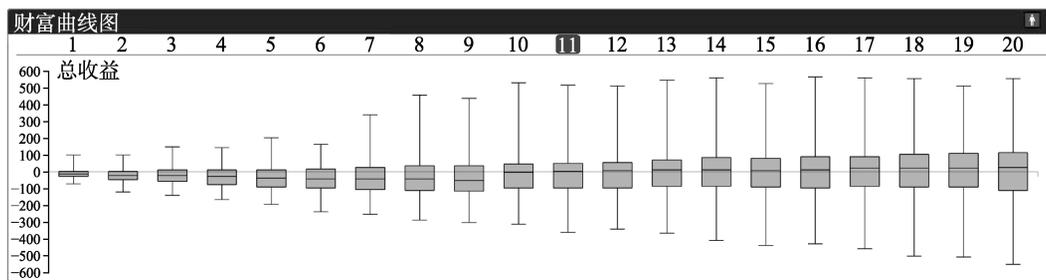


图 5 总体财富分布趋势分析

随后选择某一轮次实验时,专家通过实验流程信息图仔细观察参与者线条的走势.如图 6 所示,从健康投资、财产保险、借贷机会、投资和风险投资这 5 个主动决策环节可以看出,除借贷机会以外的 4 个环节存在很高的集群现象,即当个体财富收益为正时,参与者具有较高的行为能力,几乎都会自主决策进行健康投资、财产保险、投资和风险投资等,从而减少在生病环节的损失,规避负面冲击的影响,并且借助投资获取更多的财富.但是当个体财富收益为负时,参与者行为能力受到限制,无法进行涉及主动消耗财富的实验环节.专家特别指出,富裕人群对借贷的态度有所区别,较保守的个体不会选择借贷;而较激进的个体会选择借贷以

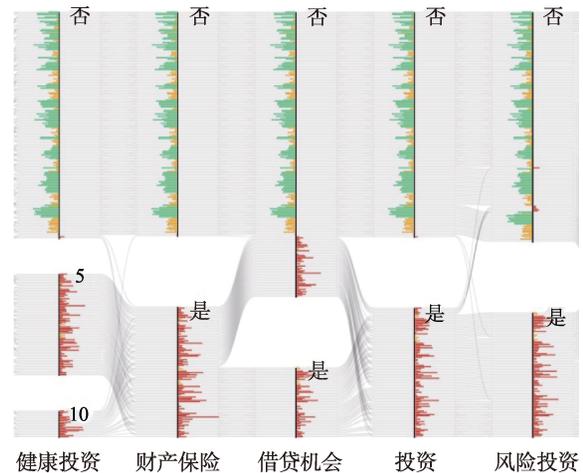


图 6 决策行为集群分析

尝试获得更多的资金进行投资等其他环节, 但同时会面临更大的风险挑战, 如果在投资中并没有获得足够的收益, 将会导致更大的损失。

最后专家分析用户财富对风险偏好的影响, 将注意力集中在实验流程信息图的最后 2 列上。可以看出, 这列中标记为 6 的分段最长, 即大部分人都会选择更高的风险选项以期获得到更高的收益。如图 7 所示, 分别高亮本轮财富为低、中、高的用户群体发现, 他们中都有最大比例的用户选择了等级为 6 的最高风险。但是在低收入人群中, 选择风险等级 0 和 1 的人数相比于中等财富和高财富要多, 而选择风险等级 5 的人数则较少。专家解释, 贫困个体的风险厌恶程度相对较高, 相当比例的贫困者不再愿意尝试在高风险的情况下博得更高的收益以改善自身的财富状况。

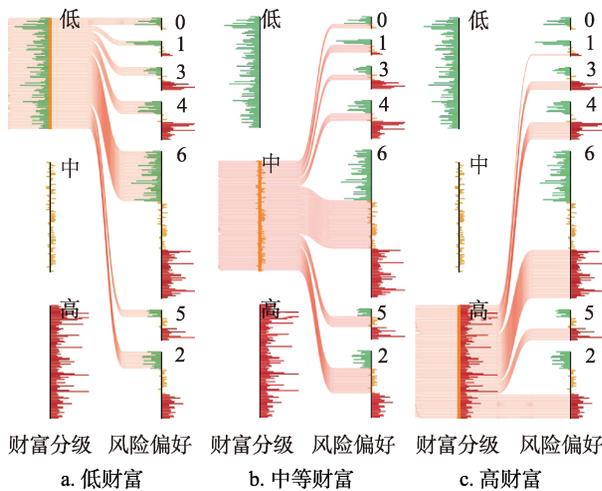
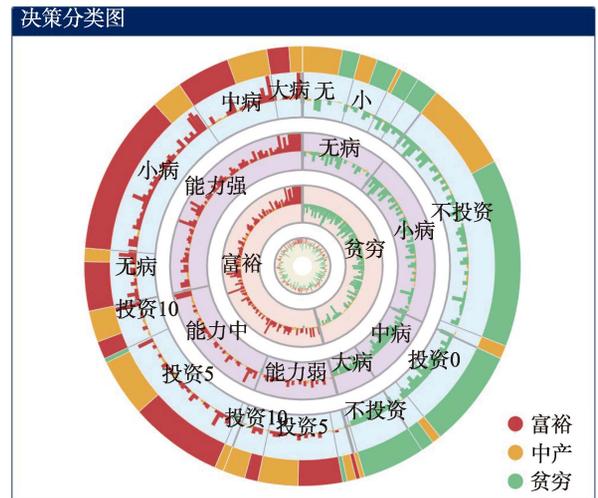
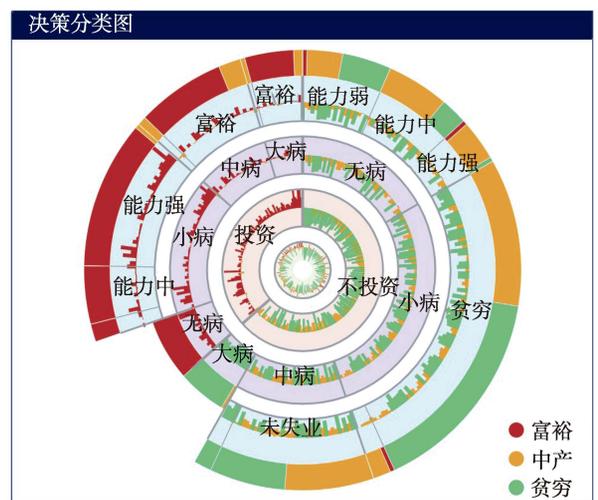


图 7 风险偏好分析

(2) 关键决策行为和财富关联分析. 专家分析影响财富收益的关键决策行为. 图 8a 所示为第 15 轮次决策树图的可视化结果, 可以看出, 最内层圆上的线条在按照参与者的 id 号进行排列时, 红色、绿色和黄色的线条相互间杂, 呈现的状态较为混乱; 而当第 1 层的属性条件是起始财富时, 起始财富为富裕的扇区几乎全是红色线条, 贫穷扇区几乎全是绿色线条, 说明该轮起始的财富状态对最终财富收益结果的影响最大. 观察下一层发现, 在初始富裕时工作能力对财富的影响最大. 值得注意的是, 再往下一层, 在工作能力强的条件之下, 影响最大的是健康, 而在工作能力中等和弱的条件下则是投资. 可以看出, 即使在初始富裕条件下, 工作能力低下并且身患疾病或投资不善也会导致赤贫. 类似地, 专家观察到在起始贫穷状态下健康对财富的影响最大, 投资次之. 经过与专家沟



a. 第 15 轮



b. 第 5 轮

图 8 决策分类图

通得知, 为改善贫困, 实验设计在 10 轮之后实施扶贫政策, 即给予贫困者一定的金额参与投资. 通过视图可以看出, 即使在初始贫穷状态下, 身体健康且投资得当的参与者就会有较大可能变为中产甚至富裕. 随后, 专家观察图 8b 所示第 5 轮的决策树图发现, 投资、初始贫富、健康和工作能力等依然是对参与者财富影响较大的决策行为. 其他轮次也是如此, 但是它们的优先次序发生了一定变化.

随后, 专家聚焦在图 3 右下方所示的总体决策桑基图上, 发现影响财富最大的初始决策行为主要包括投资、财产保险和初始财富, 而工作能力、健康多分布在第 2 层和第 3 层, 其中初始财富广泛分布在各层. 专家按照人均财富对表示叶节点的饼状图进行排序后, 分别选择了最富裕群体和最贫困群体进行决策路径查看, 如图 9a 和图 9b 所示, 发现最富裕群体往往是参与投资、工作能力强且初

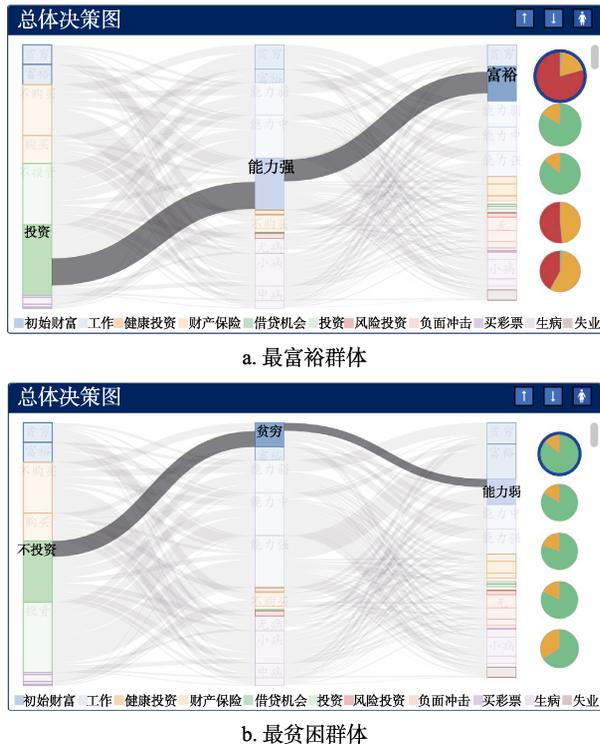


图 9 总体决策图

始富裕的个体，而最贫困群体则是由不购买财产保险、工作能力弱且初始贫困的个体组成。专家结合以上探索解读到，投资和财产保险分别是产生财富和保障财富的最佳方式，而工作能力和身体健康则是实施一系列措施的基础。初始财富分布在每层，影响力广泛且深远，客观印证了这一事实：贫困者因为缺少资金，不能参加健康投资、财产保险、风险投资等诸多环节，使得他们在遭受负面冲击、生病和失业时无法避免或削弱财富损失的影响，只能被动地接受从而变得更加贫穷。而富裕者会有更强的行为能力，通过购买保险和健康投

资等给自己的财产增加保障，还可以通过投资等来尝试获取更多财富从而变得更加富裕。专家认为这从侧面反映了贫者愈贫、富者愈富的客观现象，即贫困不仅是一种结果，也是贫困发生的原因，因为穷人无法脱离贫困最为重要的原因是他们本身就是贫困的，而贫困意味着更低的行为能力。

(3) 财富趋势分析。专家随机选择了 4 个财富收益各异的参与者，分析他们的财富时变全局趋势。如图 10a 所示，4 条折线分布在财富曲线图中，他们在第 1 轮从财富为 0 的起始状态开始进行财富的积累，直到 20 轮次实验结束后，财富收益出现较大差异。首先观察到蓝色折线的走势一直往下，黑色和橙色的折线靠近 0 且变化比较平稳；而且发现每轮次形成的花瓣图案都不太显著，原因是他们的财富收益一直为负或较少，可选择的决策行为并不多，财富基本都是消耗固定资产维持每轮实验的进行。相比之下，绿色线条财富一直为正且处于较高水平，参与者可以自主选择更多的投资和参保，财富折线起起伏伏。如图 10b 所示，黑色线条在第 11 轮~第 12 轮的财富增长明显，通过花瓣图案可以看出，在第 12 轮该参与者的风险投资和买彩票获取了大量的财富，尽管负面冲击也造成了一定损失。类似地，从第 17 轮~第 18 轮，如图 10c 所示，观察到绿色线条有明显的下降，通过花瓣图案得知该参与者在风险投资环节投入了一定量资金，但是没有获得较好的收益，反而造成了经济的损失。

之后，专家选择了第 18 轮，并通过财富柱状图选择了当前轮次与上一轮次财富差异明显的 2 个参与者。如图 11 所示，其中一个参与者财富增长较多而另一个减少较多。专家通过放大查看详

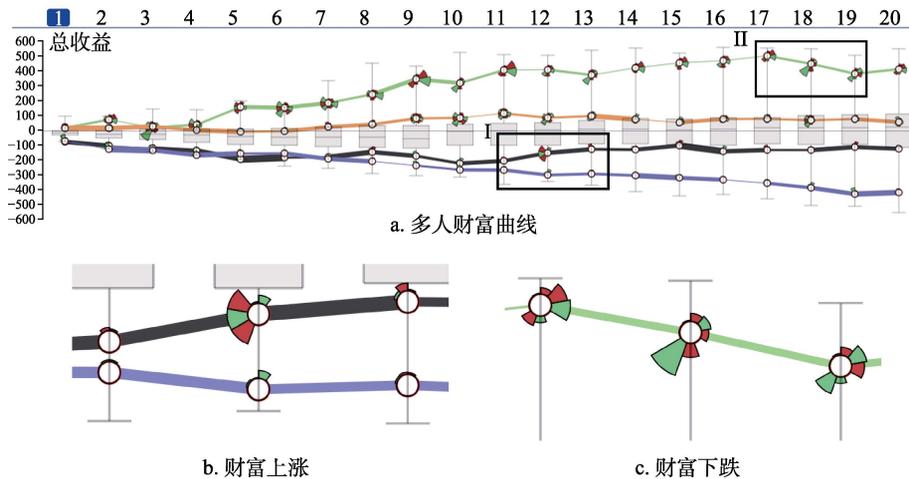


图 10 财富变化趋势分析

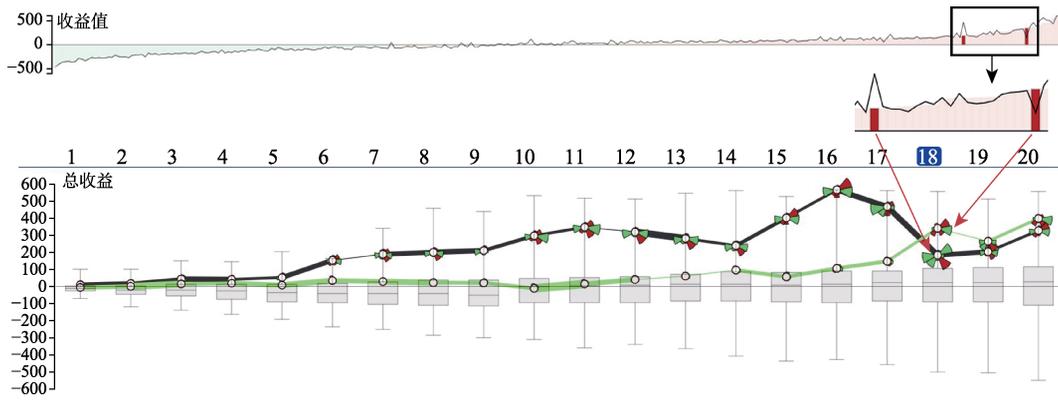


图 11 财富变化原因分析

细的第 18 轮花瓣图案后发现, 黑色线条参与者在投资环节损失惨重, 而绿色线条参与者几乎在每项决策环节都有财富进账, 积少成多最后财富收益提升明显. 在分析了这些参与者对应的实验流程信息图后, 专家认为本文设计的花瓣图案如实、客观地记录和反映了实验内容.

4.2 专家反馈

在案例分析之后, 与专家讨论了当前系统在分析实验经济数据方面的可用性和实用性. 专家首先表达了对本文工作的认可, 认为当前的经济行为实验数据可视分析系统改变了以往他们对数据分析的方式和理念, 尤其是实验流程信息图较为直观地展示了某轮次实验中所有参与者的决策过程, 可以帮助他们了解实验的具体细节, 这是以往他们在采用统计分析方法进行数据处理时达不到的; 而且系统提供的交互功能可以有效地支持他们对不同参与者、不同实验过程的筛选和对比, 方便用户进行过程式的、交互式的分析和探索, 实用性较强. 此外, 专家认为总体决策图有效地反映了所有轮次决策树结果隐藏的经验和知识, 因为每轮实验都会产生一棵决策树, 而决策树的可视分析手段面对多棵决策树时并不适用. 总体决策图简化了用户对这个复杂场景的理解, 通过综合所有的决策路径达到有效分析决策行为和财富收益关联关系的目的. 同时, 专家表示财富曲线图中的箱线图和个人财富曲线能够支持用户进行宏观层级和微观层级的财富变化趋势分析, 其中设计的花瓣图案也较为直观地解释了个体财富变化的具体原因. 最后, 专家表示掌握和理解可视化视图的映射含义需要花费一定的时间成本, 但总体而言, 整个可视分析系统的视图和交互较为直观和便捷, 用户比较容易学习和操作系统来进行经济行为实验数据的探索和分析.

4.3 讨论

专家表示, 本文的实验经济数据可视分析系统在较大程度上减轻了他们在分析数据时的压力, 提高了认知和探索效率, 并考虑将这套系统推广应用到今后的实验经济数据分析中. 但是, 专家也指出本文方法和系统需要做如下改进: (1) 财富曲线图目前只能聚焦展示少数参与者的财富时变信息, 一旦数量增多, 图中设计的连线和花瓣图会产生大量的交叉重叠, 影响用户对可视化效果的解读. 同时, 专家希望可以对参与者的财富变化模式进行分类和识别, 即根据财富曲线的变化趋势将参与者分成不同的类别. 例如, 有些参与者通过决策调整等手段使财富收益一路上涨, 这种参与者可归为学习上进型; 而有的参与者安于现状、不思进取, 使财富收益停滞不前甚至下跌, 这种参与者可归为懒惰型. 这样可以帮助专家有目的地选择和分析不同类型的参与者, 提高探索效率. (2) 尽管决策分类图可以帮助用户区分参与者的行为模式对财富收益的影响, 但并未进一步结合参与者的行为决策对其财富收益值进行预测. 本文在决策树的设计过程中将最终收益划分为 3 类, 但是专家表示财富收益是连续的数值型, 可以尝试将财富收益或净收益作为训练数据进行回归决策树的训练, 从而获得更具体的、可量化的和可预测的决策行为对财富收益大小的影响.

5 结语

本文在与实验经济学专家密切合作的基础上, 提出模拟情境下真实努力任务实验数据可视分析方法, 开发了一套经济行为决策可视分析系统, 分别设计实验流程信息图, 以全面地呈现实验过程中参与者的总体决策分布和个体行为特征; 设计

自适应决策分类图、财富时序变化曲线图等直观地呈现实验参与者的决策过程表现与财富结果的关联关系。分析以实际数据为基础的实验经济决策案例,帮助领域专家探索式地分析和理解参与者决策行为特征及财富变化关键因素,为专家探索和分析贫困产生及贫困循环恶化的内在机制提供了客观依据。

通过专家的反馈和讨论,本文将在下一步的工作中不断改进和优化当前的可视化设计,降低用户的学习时间和成本,支持参与者的财富时变模式识别与分类、行为决策对财富收益影响的预测等,并尝试引入更高级的数据挖掘模型和机器学习算法分析和处理数据,以期获得更多的隐藏规律。

参考文献(References):

- [1] Ye Puwan. Poverty economics: a survey[J]. *The Journal of World Economy*, 2005, 28(9): 70-79(in Chinese)
(叶普万. 贫困经济学研究: 一个文献综述[J]. *世界经济*, 2005, 28(9): 70-79)
- [2] Caplin A. Experimental economics: rethinking the rules, by Nicholas Bardsley, Robin Cubitt, Graham Loomes, Peter Moffat, Chris Starmer, and Robert Sugden[J]. *Economics and Philosophy*, 2011, 27(2): 179-182
- [3] Brügggen A, Strobel M. Real effort versus chosen effort in experiments[J]. *Economics Letters*, 2007, 96(2): 232-236
- [4] Ma Tao, Li Shaodong. Studies on experimental economics (symposium)[J]. *Academic Monthly*, 2006, 38(2): 83-86(in Chinese)
(马涛, 李绍东. 实验经济学研究(专题讨论)——实验经济学方法论面临的难题[J]. *学术月刊*, 2006, 38(2): 83-86)
- [5] Ren Lei, Du Yi, Ma Shuai, *et al.* Visual analytics towards big data[J]. *Journal of Software*, 2014, 25(9): 1909-1936(in Chinese)
(任磊, 杜一, 马帅, 等. 大数据可视分析综述[J]. *软件学报*, 2014, 25(9): 1909-1936)
- [6] Tanaka T, Camerer C F, Nguyen Q. Risk and time preferences: linking experimental and household survey data from Vietnam[J]. *American Economic Review*, 2010, 100(1): 557-571
- [7] Carvalho L S, Meier S, Wang S W. Poverty and economic decision-making: evidence from changes in financial resources at payday[J]. *American Economic Review*, 2016, 106(2): 260-284
- [8] Wik M, Kebede T A, Bergland O, *et al.* On the measurement of risk aversion from experimental data[J]. *Applied Economics*, 2004, 36(21): 2443-2451
- [9] Yesuf M. Risk, time and land management under market imperfections: applications to Ethiopia[D]. Gothenburg: University of Gothenburg, 2004
- [10] Henrich J, McElreath R. Are peasants risk-averse decision makers?[J]. *Current Anthropology*, 2002, 43(1): 172-181
- [11] Luo Jun, Wang Dingding, Ye Hang, *et al.* Experimental economics bridging to real world: the literature review of field experiments[J]. *China Economic Quarterly*, 2015, 14(3): 853-884(in Chinese)
(罗俊, 汪丁丁, 叶航, 等. 走向真实世界的实验经济学——田野实验研究综述[J]. *经济学: 季刊*, 2015, 14(3): 853-884)
- [12] Shah A K, Mullainathan S, Shafir E. Some consequences of having too little[J]. *Science*, 2012, 338(6107): 682-685
- [13] Falk A, Fehr E. Why labour market experiments?[J]. *Labour Economics*, 2003, 10(4): 399-406
- [14] Zhou Zhiguang, Xu Yangbing, Liu Fang, *et al.* Visual analysis of GDP with spatiotemporal multidimensional features[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2016, 28(5): 712-722(in Chinese)
(周志光, 徐杨炳, 刘芳, 等. 面向 GDP 数据时空多维属性的可视分析方法[J]. *计算机辅助设计与图形学学报*, 2016, 28(5): 712-722)
- [15] Dominik J, Fabian F, Tobias S, *et al.* Temporal MDS plots for analysis of multivariate data[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2016, 22(1): 141-150
- [16] Zhou Zhiguang, Shi Chen, Hu Miaoxin, *et al.* VisIOT: visual analysis of economic structure correlation[J]. *Statistical Research*, 2019, 36(11): 3-13(in Chinese)
(周志光, 石晨, 胡淼鑫, 等. VisIOT: 经济产业结构关联可视分析[J]. *统计研究*, 2019, 36(11): 3-13)
- [17] Yue X W, Bai J X, Liu Q H, *et al.* sPortfolio: stratified visual analysis of stock portfolios[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2020, 26(1): 601-610
- [18] Tanahashi Y, Ma K L. Stock lamp: an engagement-versatile visualization design[C] // *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. New York: ACM Press, 2015: 595-604
- [19] Wang Y, Haleem H, Shi C L, *et al.* Towards easy comparison of local businesses using online reviews[J]. *Computer Graphics Forum*, 2018, 37(3): 63-74
- [20] Hu Min, Sun Yufeng. A visual stock market trend forecasting method based on state evolution[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2014, 26(2): 302-313(in Chinese)
(胡珉, 孙瑜峰. 基于状态演化的股市变化趋势可视化预测方法[J]. *计算机辅助设计与图形学学报*, 2014, 26(2): 302-313)
- [21] Teoh S T, Ma K L. PaintingClass: interactive construction, visualization and exploration of decision trees[C] // *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2003: 667-672
- [22] van den Elzen S, van Wijk J J. BaobabView: interactive construction and analysis of decision trees[C] // *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Los Alamitos: IEEE Computer Society Press, 2011: 151-160

- [23] Poulet F. Towards effective visual data mining with cooperative approaches[M] //Visual Data Mining. Heidelberg: Springer, 2008: 389-406
- [24] Ankerst M, Elsen C, Ester M, *et al.* Visual classification: an interactive approach to decision tree construction[C] //Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 1999: 392-396
- [25] Thomas M, Lorenz L, Torsten M, *et al.* TreePOD: sensitivity-aware selection of pareto-optimal decision trees[J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 24(1): 174-183
- [26] Liu S X, Xiao J N, Liu J L, *et al.* Visual diagnosis of tree boosting methods[J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 24(1): 163-173
- [27] Huang Y F, Liu Y H, Li C H, *et al.* GBRTVis: online analysis of gradient boosting regression tree[J]. Journal of Visualization, 2019, 22(1): 125-140
- [28] Zhao X, Wu Y H, Lee D L, *et al.* iForest: interpreting random forests via visual analytics[J]. IEEE Transactions on Visualization and Computer Graphics, 2019, 25(1): 407-416
- [29] Płoński P, Zaremba K. Visualizing random forest with self-organising map[C] //Proceedings of International Conference on Artificial Intelligence and Soft Computing. Heidelberg: Springer, 2014: 63-71
- [30] Liu S X, Wu Y C, Wei E X, *et al.* StoryFlow: tracking the evolution of stories[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12): 2436-2445