**ExeVis Concept Based Visualization of Exercises in Online learning**

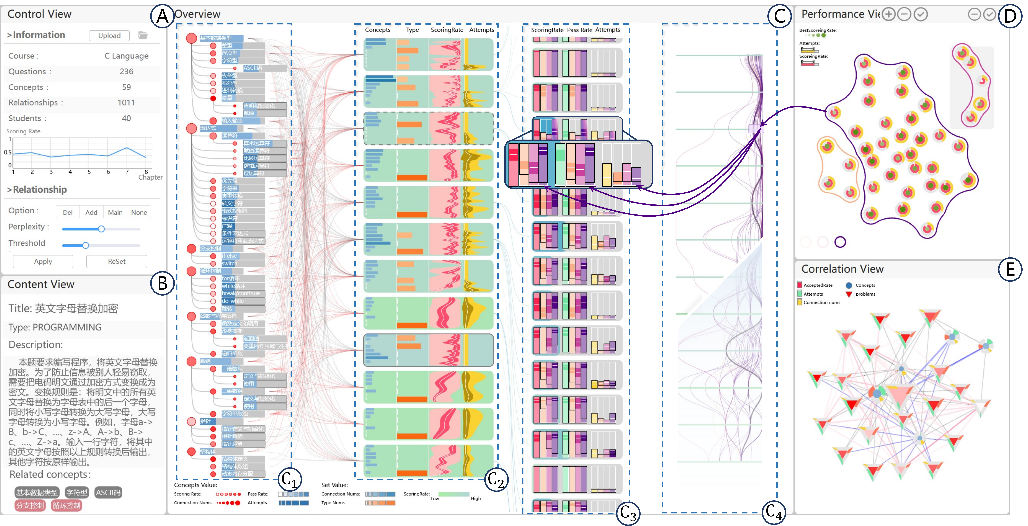


图1. ExeVis系统。 （A）提供基本信息显示和选择，（B）可视化问题内容及相关概念。概览（C）包括概念的层次结构（C1），练习和问题分布（C2和C3），以及整体/个人行为（C4）。性能视图（D）显示个人性能，而关联视图（E）揭示练习关系。

摘要 - 近年来，网络学习越来越受欢迎，并被证明是一种有效的教育方式。许多研究分析了教学材料和学习行为。然而，大多数现有研究忽略了学习概念和练习之间的关系，这些关系可以传达教学表现和学生行为。在网络学习中呈现概念和练习之间的关系不仅可以帮助教育工作者探索练习和概念的分布以巩固知识，还可以提供对在线课程中学生行为的直观反馈，从而增强教学策略。在这项工作中，我们从练习中提取学习概念，建立概念和练习之间的逻辑关系，并通过自动模型和半自动模型构建概念的层次结构。为了帮助用户有效直观地分析和评估概念和练习，我们设计并实现了一个名为ExeVis的可视化分析原型系统，集成了多个交互式可视化图形。ExeVis配备了多个交互式直观的可视化图表，包括用于选择和显示基本信息的控制视图、具有层次结构的概览以呈现概念和练习的分布和掌握情况、揭示练习之间关系的相关视图以及显示个人能力的性能视图。真实数据案例研究和专家访谈证明了ExeVis在为教育工作者提供有关练习适当性的宝贵见解并使他们能够调整教学方法方面的实用性和有效性。

**Index Terms**—Online learning, education analysis, exercise evaluation

1. 引言 近年来，网络学习已经出现并成为帮助人们获取知识和技能的最流行方式之一。随着在线学习数量和多样性的增加，许多研究通过各种元数据（如视频、幻灯片、教学大纲和练习）分析课程[22, 43, 64]。具体来说，许多当前的工作将知识转换为微观级别的单元，并将它们分类到不同的组中，以提高概念分析的效率[23, 47]。此外，许多工作也注意评估与学习材料相关的学习者行为，如论坛帖子、观看视频行为、课程笔记和退出率[17, 23, 67]。这些工作不仅帮助教育工作者了解学生的学习过程，还促进了课程的调整和定制[44, 52]。然而，大多数现有的在线学习分析工作仅将练习视为验证和认证的绩效测量手段，忽略了深入分析练习以增强学习过程的价值[6, 8, 41]。

练习源自科目教学大纲，自然向学生展示学习概念[14]。因此，练习是教育工作者不可或缺的教学工具，帮助学生巩固在课堂上获得的知识[3, 21, 41]。此外，练习在为教育者/教师提供反馈方面起着至关重要的作用[29]。目前，许多在线学习平台记录包括练习基本信息和与练习相关的学习者行为（例如，练习准确性、尝试次数、打开时间和截止时间）在内的Web日志数据，这些是许多可视化分析系统分析和可视化在线学习练习数据的基础数据[9, 20, 42]。然而，当前系统主要通过统计评估技术关注显式的学生行为，而未能深入挖掘练习以发掘潜在模式和见解。

为了弥补这一差距，从概念角度接近练习至关重要。分析练习与潜在学习概念之间的关系可以产生有价值的见解以增强教学方法。例如，在在线编程课程中，会呈现一系列编码任务。通过分析得分率和尝试次数，我们可以了解到学生的行为并确定他们可能会遇到困难的概念。

此外，为了进一步确定问题和需求，我们对教育专家进行了采访，他们强调了具有特定重点和不同格式的定制化练习[51,62]。他们强调了监控学习者行为、评估理解水平、根据解决问题能力对学生进行分组以及识别异常值的重要性。通过将这些见解与彻底的文献回顾相结合，我们对问题有了更好地理解，可以总结为三个主要挑战（CH1- CH3）。首先，由于明确和隐含概念混合在一起，从练习中全面提取和排序概念存在困难（CH1）。其次，在不同层次上外部化并可视化与练习和学生行为相关的知识结构是一个重要的挑战（CH2）[50]。最后，由于它们复杂且有时潜在的关联，建立概念和练习之间的逻辑关系具有挑战性（CH3）。

为了应对这些挑战，我们提出了一种半自动方法从练习中提取学习概念，使分析概念和练习之间的关系成为可能。此外，我们提出了ExeVis，一种全面的可视化系统，为教育工作者（例如教师、课程讲师和教育分析师）提供练习内容分析和对学生表现的直观反馈。ExeVis显示概念和练习的层次结构以分析内容和分布的适当性。同时，它提供了多尺度学习者行为的概览，包括概念、练习及其依赖关系，使用户能够识别一般学习者模式和趋势。此外，我们设计了一个网络图来揭示练习和概念之间的隐含关系。对于学生，我们以多维方式可视化个人表现，有助于识别相似之处和差异。本文的主要贡献如下：

• 一个分析流程包括半自动概念提取方法和可视化分析组件，使知识探索在各种练习分析场景中成为可能。

• 具有交互式图表的可视化分析系统促进用户直观地探索练习、概念和学习者行为在多个维度上的关系。

• 来自真实在线练习数据的见解赋予教育工作者调整练习设置并增强教学方法的能力。😊

2 相关工作 2.1 在线学习分析 近年来，网络学习已成为一种流行的教育传递方式[35, 43]。为了促进学生学习和教师分析，许多研究分析了在线课程中不同的元数据，包括视频[33, 38, 54]、幻灯片[11, 60]、教学大纲[15, 32]和练习[12, 43]。例如，陈等人探讨了三种常用的视频讲座风格对学习者持续注意力、情感、认知负荷和学习表现的影响。此外，研究人员通过整合多种类型的课程内容进行全面调查。例如，Koedinger等人[34]将课程材料分类为被动/陈述性信息和主动/交互性活动，研究与每种相关的学习收益。

由于概念图在学科领域的优势[47]，大量研究集中于评估从多模型线索中提取的概念及其相互关系[64, 66]。例如，陈等人[11]提出了几种从口头课程讲座中提取关键术语的方法，而张等人[65]则使用户能够使用关键字的频率和时间戳表示来手动创建概念之间的关系。此外，刘等人[38]和Schwab等人[54]开发了概念图以帮助学习者快速将概念与视频联系起来。Cooper等人[13]结合了视频内容和从课程大纲中挖掘出来的顺序主题关系，并提供了一个交互式可视化探索推荐。

然而，对于加强学生学习并向教师提供反馈至关重要的练习评估却受到了有限的关注。而且大多数在线学习方法仅依赖练习成绩来衡量绩效并验证其他材料的有效性[6, 8, 48]，忽略了练习中包含的丰富信息。😊

2.2 学习者行为可视化

监控学生的学习行为对于有效的在线学习至关重要[30]。然而，从在线学习平台获得的数据通常复杂、庞大、嘈杂且难以理解[64]。为了解决这个问题，许多研究采用了可视化来直观地展示他们的分析[63]。

一些研究使用了基本的可视化，如散点图和折线图。例如，Mazza等人[42]使用这些图表来识别趋势和需要关注的个体，而黄等人[31]则使用堆叠条形图和河流图来可视化课程注册、参与度和表现。黄等人[33]专注于从学习者行为数据和字幕中挖掘学习概念，并使用关键字云进行表示。此外，Mubarak等人[45]将学习者与课程视频交互产生的视频点击流数据可视化为图网络。最近，出现了更全面的可视化系统。例如，陈等人[8]提出了一个专注于交互峰值的可视化系统，其中包括复杂的字形设计，如字形、流图和平行坐标来显示峰值的多个属性。陈等人[10]开发了一个可视化系统，集成了四个链接的可视化设计，帮助分析师在多个尺度上识别与退出行为相关的学习模式，并揭示学习者详细的学习活动。

总之，虽然过去的研究通常利用多面板视图和交互技术（如排序、过滤和聚类）来展示学习行为，但我们的方法提供了对练习行为的概念分析。我们超越传统的可视化，并提出一个分析系统，使用户能够全面了解学习行为。这不仅提供了对练习和概念全面而细致的理解，还提供了有关群体动态和个人行为的宝贵见解。😊

2.3 练习分析 一些研究人员以各种方式探索了练习评估。例如，Mazza等人[42]创建了一个矩阵来跟踪学生尝试与不同概念相关的问题。Grover等人[26]和Tomkins等人[56]可视化了每个练习的成绩分布。此外，Stephens等人[55]使用堆叠条形图来调查尝试分布。Emmons等人[20]在宏观和微观层面分析了练习表现，比较学生提交以识别问题区域。最近，Arruarte等人[4]开发了TEA，一种可视化分析工具，用于分析和可视化学生在基于测试的练习中的表现，当它们不符合教师的期望时提供指示器以改善练习的有效性和质量。

总之，以前的研究集中于使用练习行为来提供关于内容修订和教学方法的反馈。然而，它们忽略了练习旨在加强和巩固的潜在知识[3, 7, 14, 19, 21, 27, 41]。此外，这些研究依赖于统计评估技术，在捕捉复杂关系和识别学习者行为和表现中的模式和趋势方面存在局限性[19, 41]。没有这样的分析，就难以充分理解练习促进学习成果的有效性。据我们所知，目前还没有可以有效地在内容和行为的各个层次上可视化练习的工具。😊

3 任务和系统工作流程 3.1 数据描述 本研究中使用的练习是由实际教育工作者为C语言编程课程设计的，并由40名主修数字媒体技术的大一新生班级进行了管理。他们中的大多数人缺乏实践经验和对编程等领域的知识和技能的深入了解。根据课程大纲和学生需求，讲师创建了每周、单元和总结练习。练习及相应的学生行为记录在PTA（在线实验辅助教学平台）中。由于日志数据复杂且多维[42]，我们通过过滤掉无关字符并选择关键测量（如提交状态、尝试次数和时间）来预处理它。在审查练习后，我们从第一周到第十二周确定了236个问题，涵盖了多种问题类型，如选择题、填空题和判断题。具体来说，练习包括对编程语言教学至关重要的编码任务。

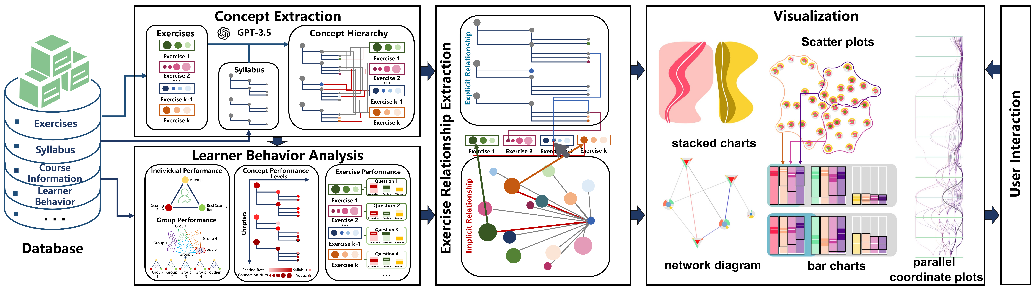
3.2 任务分析 在总结任务之前，我们与四位领域专家（E1-E4）进行了几轮采访，从各个角度收集有关练习的反馈。E1作为大学课程的管理员，负责课程开发和实施，并确保学生获得高质量的教育。E2是C语言编程课程的讲师，在在线和离线课程教育方面拥有丰富经验。此外，她还参与了教科书和其他教学材料的创建。E3是一位教学专家，在MOOCs上提供在线讲座视频已有十多年。她对教育学有深入的理解，并知道如何在在线学习环境中有效地吸引和支持学生。E4是在线学习的教育分析师，专注于理解和改善学生成果。他分析学生数据并提供关于课程或交付方法改进的反馈。此外，我们还进行了文献回顾以进一步完善任务。

本工作的任务（T1-T4）介绍如下：

T1. 如何全面提取概念？在设计练习时，教育工作者通常会包含多种类型的问题，将相关概念结合起来以加强学习。然而，大多数日志数据仅记录与问题相关的主要概念，可能会遗漏其他重要信息。此外，一些潜在的概念可能难以被当前的实体提取模型识别。因此，解决这些挑战需要在多个领埴都有知识的模型来提取所有相关概念并确定它们与练习的相关性，确保信息以有序和易于访问的方式组织[54, 57]。

T2. 可以从学习者行为中获得什么反馈？日志数据的多样性为学习者行为提供了多维度的见解。例如，准确性是衡量学生对练习涵盖概念理解程度的最常用指标。具体来说，与传统课堂教学相比，在线学习允许多次尝试，这可以表明掌握所需的努力水平，并帮助教师识别学生最困难的领域。利用学生的多维表现可以为教育工作者提供宝贵的见解，以全面了解不同层次上的学习者行为和概念。这可以揭示个人优势和弱点，以及类似学生群体中的模式和趋势。此外，这些信息可以帮助教育工作者识别特定知识点中的薄弱环节，确定表现不佳的单元，并更好地了解课程的整体掌握情况。

T3. 练习和概念之间有哪些逻辑关系？分析练习与概念之间的关联对于有效教学和优化教学实践至关重要。这些关系可以分为两类：显式和隐式。显式关系基于从练习中提取的概念，自然地连接包含相同或相关概念的练习。通过这些关系，教育工作者可以分析概念掌握情况并找到练习所需的依赖关系和先决知识。然而，一些概念具有在提取过程中未发现的潜在联系。通过分析练习顺序和学生在多个练习中的表现，可以揭示隐含关系。这些关系可以帮助教育工作者以教育学上合理的方式安排练习。😊

T4. 如何评估练习设计的适当性？为了增强学生对课程的掌握，练习设计应与学习目标保持一致。以前的研究和专家访谈表明，内容分布和练习顺序是关键因素。因此，应当认真考虑需要加强学习目标的练习数量、范围和顺序。此外，来自学生行为的反馈可以确定练习设置是否具有适当的挑战性并为学生提供应用和扩展理解的机会。通过分析分布和反馈，教育工作者可以识别困难领域并相应地调整练习设计。😊

3.3 系统工作流程 为了解决上述任务，我们提出了ExeVis，一种可视化分析原型系统，为教育工作者提供有关练习分布的宝贵见解和对学生表现的直观反馈。如图2所示，系统流程从从练习中提取概念开始，然后通过关联规则挖掘来发现练习之间的基本隐含关系。然后计算每个练习、概念和个体的多维行为。最后，定义层次结构以呈现多层次概念和练习的分布和行为。我们的可视化系统具有四个协调视图：控制视图、概览、相关视图和性能视图，如图1所示。此外，我们还加入了丰富的交互功能，使用户能够灵活有效地探索感兴趣的区域。

4 方法 4.1 概念提取 练习中包含的概念和关系复杂且有时潜在，因此需要使用模型进行提取。为了建立清晰的概念层次结构，我们利用课程大纲作为先验知识并创建初始结构。然而，教学大纲可能无法提供我们所需的全面和详细信息，因为它们主要关注重要的教学主题[32]。为了解决这个问题，我们使用了实体提取算法，包括NLTK工具包[39]以及基于BiLSTM-CRF[68]和BERT[18]的模型。虽然这些方法在提取明显概念方面表现良好，但由于在各个课程领域的知识有限，它们难以识别隐含概念及其相互关系。此外，它们需要大量数据并且耗时较长。因此，我们需要一个既具有全面领域知识又具有高处理速度的模型。

现有研究已经验证了GPT系列模型在文本提取任务中的令人满意的性能[28, 36, 37, 61]，证明它们能够准确捕获实体边界和上下文信息以进行高质量提取。为了评估GPT-3.5在练习概念提取方面的能力，我们与课程讲师进行了比较实验。结果表明，GPT-3.5与人类专家的质量相当，并且具有速度快、输出更详细和全面的优点。鉴于现有模型由于数据可用性而在概念提取方面存在局限性，我们采用了半自动方法。首先，我们使用OpenAI API[49]访问GPT-3.5进行概念提取。为确保一致且连贯的响应与所选课程保持一致，我们赋予GPT-3.5特定课程专家的角色，并提供练习内容和课程/章节详细信息。输出被限制为与所选课程专业匹配，并存储在数据库中供将来使用。随后，我们利用

图2. ExeVis系统流水线促进了概念提取、学习者行为分析和练习关系提取的系统流动。此外，它还提供交互式可视化设计来呈现练习分布和学生表现。

GPT-3.5将提取出来的概念与教学大纲中已经记录的概念合并，以创建概念层次结构。然而，概念名称变化、错位和遗漏等挑战影响了概念层次结构的可靠性。为了缓解这些问题，我们采用了多轮基于对话的评估来评估生成结果的质量，确保更大的稳定性。此外，我们根据它们的相似性进行概念合并，并将概念之间的关系分类为主要或从属。最后，我们通过人工验证来确保结果的准确性。😊

4.2 学习者行为分析 日志记录了大量数据，提供了多种角度来分析学习者的行为。为了准确反映学生的表现，我们选择了得分率作为主要测量指标。此外，与传统的离线教育相比，在线学习平台为学生提供了多次尝试的机会。正如许多研究所证明的那样[2, 16]，多次尝试允许学生从错误中学习并获得更好的学习成果。此外，它鼓励学生积极参与练习并可以提高他们对主题的理解。因此，我们在分析中将尝试次数作为一种测量手段。

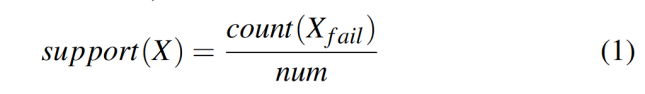
我们有一组练习E，其中包含k个练习，每个练习包含一组问题E = q11,q12,…,qi j,…,qkl。这里，qi j表示第i个练习中的第j个问题，l表示第k个练习中问题的数量。然后我们计算每个问题的平均得分率。此外，根据记录，我们确定通过率，即在截止时间之前正确回答问题的学生比例。

基于这些数据，我们进一步评估每个学生、概念和练习的行为。我们将学生对每个问题的表现表示为一个特征向量，每个问题对应一个特定维度。每个维度中的值表示学生的得分率，通过计算每次尝试的平均得分来计算。这个高维向量提供了有关学生特征的宝贵见解。为了降低数据的维度，我们在与MDS[69]、PCA[1]和UMAP[5]等其他降维算法进行比较后使用t-SNE[58]。通过将这些数据与先前提取的关系结合起来，我们可以评估概念的行为。对于概念是主要焦点的练习，我们分配0.8的权重，而对于其他练习则分配0.2的权重。这种加权方案使更准确地评估概念行为成为可能。

4.3 练习关系提取 从练习中提取概念后，我们建立概念结构和练习之间基本关系，可以揭示练习之间明确的相互关系。通过利用练习中包含的概念，更容易找到包含相同概念的其他练习。此外，可以识别出将特定概念作为关键知识以及仅包含该概念的练习分布。这些关系使我们能够获得有价值的见解来区分类似练习和识别依赖性练习。

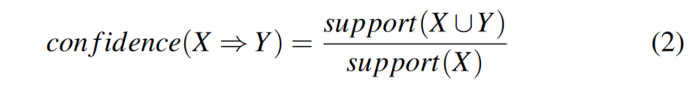
除了通过概念发现的显式关系外，一些练习具有潜在联系，可以通过分析学习者行为来揭示。例如，如果两个练习除了特殊顺序外没有其他联系，则某些学生可能表现更好。为了揭示练习之间隐含关系，我们利用关联规则挖掘技术，在改善定性教学方面广泛用于学术研究[52, 59]。

首先，我们定义关联规则X⇒Y并计算它们的支持度，其中X和Y是单独问题。函数support(X)表示在截止日期之前仍然失败但尝试过问题X的学生比例，并按如下方式计算：

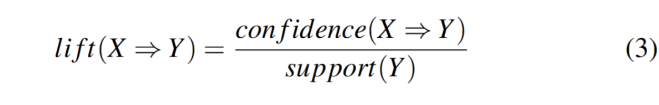


其中count(Xfail)是尝试问题X并回答错误的学生数量，num表示学生数量。

接下来，我们计算问题之间的置信度，表示在失败问题X的情况下回答问题Y错误的学生百分比。置信度可以通过以下公式计算：



其中X∪Y表示尝试问题X和Y的学生。之后，我们使用提升来筛选有价值的问题对。提升确定当学生已经掌握问题X时，回答问题Y会变得更容易还是更困难，可以使用以下公式计算：😊



最后，我们选择提升值大于1且置信度超过一定阈值的问题对。通过进行这些计算，我们可以提取有价值的问题对并帮助教育工作者更好地理解问题之间的潜在关系。

5 ExeVis 5.1 设计要求 主要设计要求（R1-R3）总结如下：

R1. 提供不同层次上内容和行为的直观表示。系统应提供练习和概念分布和行为在各个层次上的可视化表示，如个别学生表现、练习行为和概念分布。可视编码，如颜色编码方案和层次结构，帮助教育工作者识别模式和关系。清晰的可视化解释，包括数据解释和系统使用说明，支持明智的教学决策。

R2. 快速识别练习和概念之间的关联和差异。为了促进理解练习与概念之间的关系，系统应提供清晰且易于区分的主要和次要联系以及两种类型的练习之间联系的可视化表示。具体来说，相关系数需要准确测量并映射到适当的视觉提示以传达差异。此外，应提供多个层次来呈现不同层次上概念、练习和个体的多维学习行为。此外，系统应允许轻松比较不同单元内外的行为。

R3. 允许交互式过滤并立即反馈。为了支持用户干预，系统应提供熟悉的交互方式，如选择和过滤特定练习和概念、拖放关系以及套索圆圈选择用户定义组。根据这些交互的表现，系统应提供即时反馈，如淡出、高亮和工具提示，以显示可视化中的结果变化。例如，如果教育工作者在性能视图中选择特定组学生，则系统应显示相应的行为分布以指示所选学习者组的表现。

5.2 可视化设计 为了解决任务、需求和以用户为中心的设计原则[46]，我们设计了一个可视化分析系统ExeVis，帮助用户分析在线学习练习数据。ExeVis集成了许多可视化图表，包括控制视图（图1-A）用于显示基本信息并促进配置更改、内容视图（图1-B）用于呈现问题内容及其相应概念、概览（图1-C）用于展示概念与练习之间的相互联系、性能视图（图1-D）用于显示个人行为以及相关视图（图1-E）用于揭示练习之间的关系。这些视图结合了常见指标如得分率和尝试次数，并使用不同颜色和阴影来表示行为。在控制界面中，用户可以访问基本课程信息并使用滑块调整超参数[25]以微调t-SNE和布局结果。内容视图显示问题信息及相关概念。其他视图中使用的详细可视化设计将在以下章节中介绍。5.2.1 概览 概览作为用户入口点起着核心作用，提供了概念与练习的层次结构。如图1-C所示，它由四部分组成：C1显示概念结构；C2说明练习分布；C3介绍相应问题；C4展示群体行为。基于它们的层次关系(T1)，

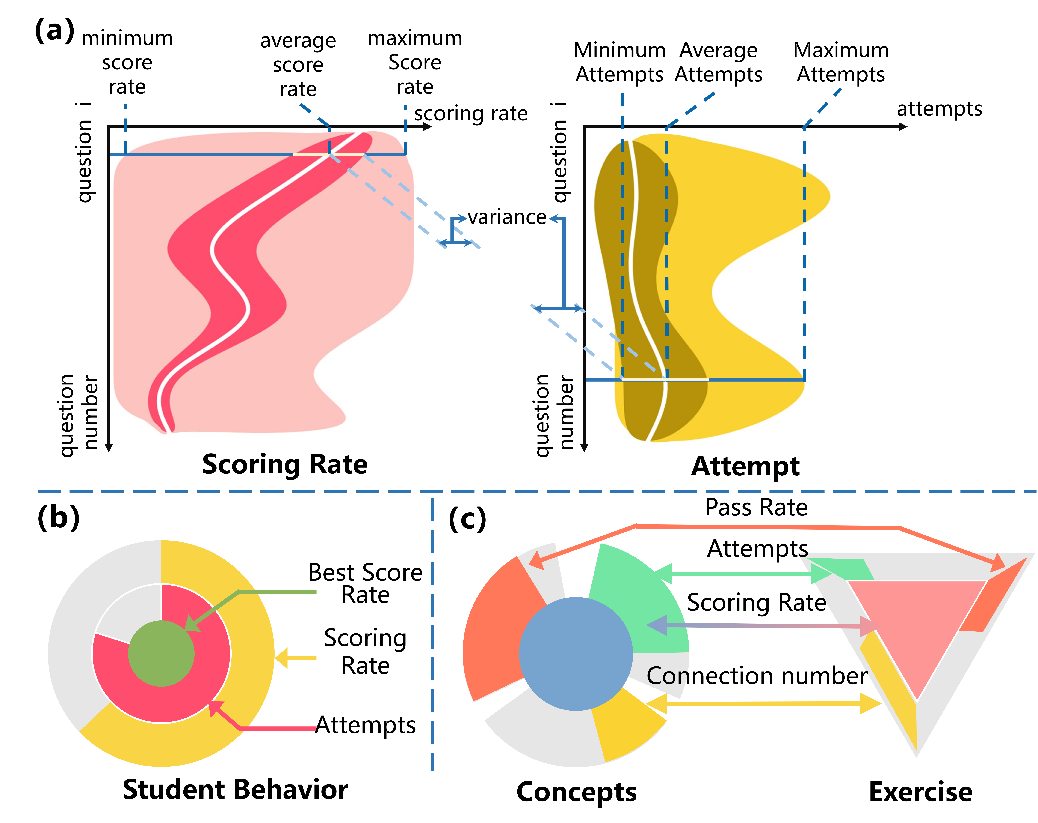


图3. (a) 堆叠图用于可视化所有练习的得分率和尝试次数的分布。 (b) 学生行为的三个关键属性，即最佳得分率、得分率和尝试次数。 © 概念和练习的属性，包括得分率、通过率、连接数和尝试次数。

我们使用垂直树形结构（图1-C1）来定位练习和问题，根据它们发生的顺序进行定位(R2)。第一级显示章节名称，下面依次是顺序节点。具体来说，概念由信号点表示，捕获它们的位置和关系，并带有内容标签，提供详细信息，如概念名称。这种方法优化了可视界面中的空间，并使更多概念能够同时显示。为了增强理解，我们对概念、练习和问题使用颜色编码(R1)。概念点和练习/问题背景反映平均得分率，而内容标签表示通过率。颜色的阴影对应得分率，颜色较深表示得分率较高。此外，外径的大小表示相关概念的数量，大小较大表示连接更多(R1)。此外，内容标签使用长度表示尝试次数，而问题使用长度表示相关概念的数量，长度直接对应数量(R1)。为了区分教学大纲中的概念和新添加的概念，新添加的概念具有实心边框，而教学大纲中最初的概念没有边框。同样地，练习类型通过边框样式区分开来，每周练习没有边框，单元练习有虚线边框，总结练习有实线边框。至于关系，颜色被用来区分和突出主要和从属关系(R2)。

为了可视化练习的多样分布，我们使用了四个并行图表(R1)。前两个条形图显示章节和问题类型的分布，长度和颜色对应数量。第一个图表中的顺序表示章节顺序，而第二个图表表示真假、选择题、填空题和编程任务。为了详细了解得分率、通过率和尝试次数，我们设计了堆叠图（图3-a），展示了练习和问题的最小值和最大值/尝试次数、平均值/尝试次数以及方差。

在图1-C2中选择特定练习后，条形图显示每个问题内部得分率、通过率和尝试次数。上下边界以及区域内不同颜色阴影分别表示最大值和最小值、上下四分位数，而中间白线表示平均值。如果通过性能视图指定学生分组，则每个组的相应信息也将显示在相应条形图右侧，并且它们由与性能视图相对应的颜色区分开来。

此外，我们采用平行坐标图[24]来可视化个别学生在各个练习中的表现趋势(R1)。每个轴代表得分率，从左到右表示更高的得分率。具体来说，粗线代表所有学生的平均值。在初始界面中，每个轴对应于图1-C2中设置的特定练习集合。选择特定练习组后，在图1-C3中显示问题详细信息。然后每个轴对应于图1-C3中的每个问题。当选择学生组时，组行为的线条颜色与性能视图中显示的组颜色相对应。

5.2.2 性能视图

为了直观地表示每个学生的多维表现(R1)，我们开发了一个由填充圆圈和两个外环组成的字形（图3-b），用来展示学生参与度。中心圆圈代表最佳得分率，而内外环分别代表尝试次数和得分率。此外，圆圈的背景颜色映射到学生的最佳得分，颜色较深表示表现更好。外环被分成几个扇形，每个扇形的角度对应学生行为相对于总体表现的比例。这种顺时针编码方案使清晰地解释和比较学生表现成为可能。此外，随着字形变得更接近，这表明这些学生表现出更相似的模式和趋势。

5.2.3 相关视图 为了增强对练习关系的分析(T3)，我们利用异构网络图来可视化概念和练习之间的连接。此外，我们采用形状编码来有效地表示它们（图3-c）。练习由等边三角形表示，而概念则用圆形表示。为了结合其他行为，我们引入了一个外环，分为三个部分。每个部分都用颜色编码来表示不同的行为，每个颜色块的大小由特定行为相对于总体行为的比例决定。

为了增强对练习和概念之间不同关系的理解，我们引入颜色差异来指示练习与其主要概念之间的关系。此外，我们根据从关联规则挖掘中获得的接近度分配颜色。此外，仅显示超过置信度阈值的关系，提供有意义的见解。通过结合这些设计选择，我们为用户提供了一种直观传达关系及其重要性的可视化表示。

5.2.4 交互 ExeVis集成了一系列直观的交互方式，帮助用户使用前面讨论过的多种可视化方法分析练习。

交互1：交互式选择和自定义。用户可以使用交互式选择技术（包括点击、拖动和套索选择）无缝导航ExeVis并探索相应结果。例如，用户可以选择特定概念或练习查看相关问题，并且此功能扩展到显示章节和问题类型的条形图。相关视图调整中心字形以显示所选项目之间的关系。如果用户遇到概念-练习关系问题，则可以使用控制视图修改可视属性。通过选择所需关系类型并将相应关联从问题拖动到正确概念，用户可以进行更改并保存以供将来参考。此外，用户可以创建自定义组并对属于这些组的学生进行颜色编码，在平行坐标图中允许他们比较不同组之间每个问题的多维行为。

交互2：工具提示和高亮。为了增强有效的数据探索和提高用户参与度，ExeVis在各个组件中都加入了工具提示。当用户与概览中的概念和练习或性能视图或相关视图中的字形等元素进行交互时，工具提示会出现，提供上下文信息和解释，如内容、多维行为和相关概念/练习。在点击不适合的情况下，采用高亮来吸引注意特定数据点或感兴趣区域。例如，在相关视图中，点击字形可能会干扰识别有意义的关系。因此，我们在概览中悬停在练习或概念上时使用高亮技术。这种高亮功能也在性能视图和平行坐标图中实现，使用户能够轻松关注相关信息。

6 评估 为了证明ExeVis的有效性和实用性，在任务分析过程中与领域专家（E1-E4）合作进行了一项案例研究。专家们最初被介绍给系统以熟悉其功能。然后鼓励他们自由探索系统，参与讨论并记录他们的焦点、发现和建议。最后，我们对专家进行了深入访谈并向其他教师发放问卷。通过定性和定量分析相结合，收集并评估了他们的反馈意见，以评估我们系统的优缺点。

6.1 多维分布分析 在专家自由探索期间，我们观察到他们关注特定分布，可以总结如下：

练习中概念的分布。首先，他们检查了概览（图1-C）中概念与练习之间的关系。他们发现利用颜色差异有助于轻松区分主要和从属关系，尽管它们很复杂。然后E2和E3按顺序浏览练习，深入研究每个问题的具体关系。E2注意到教学大纲与概念层次结构之间存在一些差异。“在设计教学大纲时，我总是确保涵盖所有重要主题，但某些概念的深度有限。例如，在表达式章节中，教学大纲只提到了运算符，但边框差异指示提取源。我观察到系统提供了有关运算符的更详细信息，如算术运算符、赋值运算符、比较运算符、逻辑运算符和位运算符。它非常出色地捕获并将所有提取出来的概念与教学大纲合并(T1)。”此外，根据她的教学经验，E3将每个问题与主要和从属概念匹配。她评论道：“这种可视化很方便。当我点击一个练习或一个问题时，高亮功能使我能够轻松识别相关概念，并且特定问题包含的概念也随着其内容显示出来。它使我能够快速掌握每个练习的范围和重点，促进对它们适当性的评估，并指导我分配更合适的练习(T4)。”

在探索之后，专家（E2和E3）将注意力转向了每个练习中的第一个统计图表，它描述了章节的分布。E2观察到每个练习中的章节分布与他们的教学计划很好地对齐。她强调了将家庭作业练习与正在进行的课程和课程进度保持一致的重要性，涵盖了课堂上教授的一系列主题，以支持学生练习和理解。E2发现在传统方法中评估对齐性和全面性具有挑战性，但承认可视化在评估练习一致性和识别异常方面的有效性。可视化使她能够轻松评估每个练习的贡献并发现潜在问题，如过早介绍、空白或课程内容覆盖重叠。她评论道：“C语言编程中的前两章为后续章节建立了基础知识。从这个视图中，我也发现了这种现象，这证明了适当性(T4)。”在进一步探索过程中，E3发现第二个练习中包含了与后面章节相关的问题，特别是第六章和第八章。通过点击第八章条形图，她迅速确定了第二个练习中与课程后期介绍的概念相关的问题。她评论道：“感谢点击交互，我可以轻松地识别与点击章节相关的问题。这非常方便。”在分析问题内容时，E3补充道：“尽管学生在学习前两章后可以完成这个问题，但在学习第八章后再做会更有益。这个系统确实具有全面提取所有问题概念的能力(T1)。”

类型和数量分布。在检查练习中概念分布之后，专家们将注意力转移到评估针对特定概念设计的练习适当性上。由于这是面向数字媒体技术专业一年级学生的编程课程，E1强调学生编程经验有限，但强调他们学习和提高编程技能的热情和渴望。因此，选择练习类型、数量和问题类型至关重要。

在分析练习边框变化之后，E2强调了点击交互以显示特定练习内所有问题的重要性。她指出：“可视化成功地区分了练习类型。显然，章节练习和总结练习都优先考虑关键概念。例如，专注于基础且简单概念的入门章节从总结练习中获益良多，以评估学生掌握情况和技能。然而，对于像指针章这样关键且具有挑战性的单元来说，必须进行单元练习来评估学生理解情况，因为结果会显著影响他们对后续主题的理解。”

此后，专家深入研究了第一个图表，在其中长度对应数量，以分析问题数量分布。他们还观察到当鼓风机时工具提示显示练习除了数量。令人惊讶的是，在第二个练习中他们发现了过多的练习，并仔细审查了内容。E1强调：“教师应包括适当数量的问题。尽管这些问题最简单，但过多的数量会带来严重的问题。它会产生高学习压力并施加沉重的认知负荷，这阻碍了有效教学。”

通过观察并点击每个练习中第二组统计图表中显示的分布，E2评论道：“在为C语言编程课程的一年级学生设计练习时，保持问题类型和数量的均衡分布至关重要。这包括选择题、填空题、判断题和编程任务。图表和交互有效地说明了基于每章特征的问题类型组合。例如，在专注于小代码片段和特定语言特征的初始章节中，前三个练习主要由针对特定概念的判断题和填空题组成。选择题可以评估更广泛的概念。此外，编程任务被用来引导学生逐步学习。值得注意的是，当观察整个练习的图表时，我还注意到编程任务逐渐成为主导问题类型，这与C语言编程课程的性质相符。”

具有显式关系的练习分布。在评估概念分布之后，专家们进一步探索了使用图1-E中描述的相关性来探索练习之间显式关系(T3)。这次检查导致了三种类型的关联的识别：预问题、类似问题和后问题。例如，如图4所示，本案例研究的中心问题涉及将明文中所有英文字母替换为字母表中的下一个字母，同时将小写字母转换为大写字母，大写字母转换为小写字母。这项任务主要涉及分支控制和循环控制，如概览和相关视图所示。此外，它还包括ASCII转换概念。从相关视图中，E3观察到，“有些问题将ASCII作为主要概念，并且应该在这个问题之前作为先决知识介绍。此外，颜色编码连接有助于识别另一个也强调分支控制和循环控制的问题，表明这些问题可以分组在一起。此外，某些问题将分支控制视为从属关系，例如涉及计算特征数量的练习。这些练习更复杂，并且可以在掌握基础练习后解决。”

6.2 多尺度行为评估 学生行为评估。专家首先检查了学生的多维表现并分析了获得的反馈(T2)。他们利用拖动、缩放和悬停在元素上等交互功能来探索可视化。在此过程中，E3轻松地识别出需要关注的学生。她指出：“颜色对比明显，我注意到学生1有一些问题。当我悬停在代表学生1的字形上时，我可以通过工具包访问全面信息，包括他们在每个练习/问题上的表现，这真的很有帮助。值得注意的是，平行坐标图清楚地描绘了准确性的波动，指出了显著时期，例如第四到第六次练习和第九到第十一次练习，在这些时期学生1的准确性为零。这些观察表明可能存在知识空缺或理解所涵盖材料的困难。”此外，E3评论道：“学生的准确性随着时间的推移逐渐改变，暗示着学生理解和表现的逐渐改善或下降。这也可能意味着学习曲线，最初挣扎后更好地掌握概念，或者可能表明整个课程中进展或参与度不一致。此外，尽管尝试次数最少，但学生1的通过率接近0.6。这表明一定程度的理解，尽管尝试次数有限可能反映出参与度或参与度不足。”除了最佳得分率较低的学生外，E2注意到学生2和学生20与其他人相比准确性较低，尽管他们获得了更高的最佳得分率。她评论道：“这些学生最初可能在理解概念方面面临挑战，但随着他们反复尝试，他们的表现稳步提高。通常情况下，这些学生以勤奋和努力著称，并且他们将受益于接受个性化指导和支持。”

在分析个人行为之后，专家利用图1-D中的套索圆圈选择用户定义组进行评估和比较。E3提到：“通过分组学生，我可以分析模式和趋势，这有助于我为每个组提供针对性支持和定制资源以满足其特定需求。例如，在查看个人表现时，我很快注意到东北部学生存在问题。当我选择它们时，我看到了它们问题行为的变化以及平行坐标图中出现的线条，显示它们之间的相似之处。然后我选择了更多组并发现来自中心地区的学生组与前一个组明显不同。这个组尝试次数较少，得分率较高，表明它们可能对材料有更深入的理解和更高水平的掌握。基于此，我们可以鼓励协作和同伴学习，在不同组之间促进互动、知识共享和支持，从而培养协作学习环境。”此外，E2强调了系统促进跨不同班级进行行为分析和比较的能力，而E1则指出其评估教师表现(T2)的有效性。

练习和概念行为评估。在分析学生行为之后，专家深入研究了平行坐标图并分析了每个练习和概念的行为。E2观察到平行坐标图与点击动作相对应，最初显示练习行为，点击后转换为问题行为。她指出：“问题类型之间的差异非常明显。大多数人只在判断题、填空题和选择题中尝试很少的次数。然而，编程任务的情况非常复杂。此外，练习、问题和概念的颜色和大小差异也表明了它们的出现。所有这些都可以帮助我快速定位需要额外关注或澄清的特定练习或概念。”此外，通过探索图1-C1中描绘的每章行为，专家们迅速获得了课程表现和变化的概览。E1评论道：“该系统不仅能够识别对学生具有挑战性的章节，还能区分教师和课程。”

此外，E4指出，在第三和第四个可视化中的准确性和尝试次数图表不仅显示了最小值、最大值和平均水平的变化，还强调了方差。他指出：“通过检查最低分和最高分，我可以确定成就水平范围并了解学生表现的极端情况。此外，平均分可以作为中心趋势度量，允许理解整体成就水平并与基准或标准进行比较。另一方面，方差表示学生成绩中分散或可变性的程度，突显分数的扩散和学生组内表现的多样性。此外，练习和问题的垂直对齐使每个问题的表现一览无余，并促进跨问题比较。”

此外，E1指出，在某些练习中，他们能够通过相关视图识别与其他练习之间的潜在关系。这可以在图4中描绘的练习中看到，其中彩色关联表示这些关系。进一步探索发现，这些练习通过

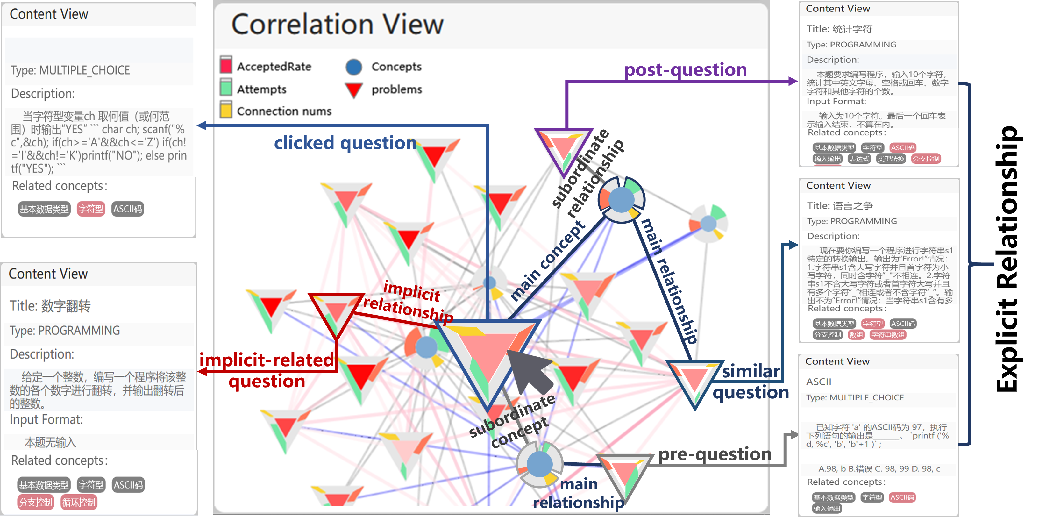


图4. 练习之间从概念中推导出的显式关系和从学生表现中推断出的隐式关系。

ASCII概念。E2补充说：“通过检查平行坐标图，我观察到大多数人都在前后问题上都有困难。值得注意的是，尽管后面的问题可以使用各种解决方案来解决，但也可以使用ASCII来解决。尽管这两个问题可能没有明显的关系，但它们具有进一步探索和调查的潜力。隐式关系确实揭示了一些有趣的事实(T3)。”

6.3 专家访谈 基于前面的探索，我们与领域专家进行了深入访谈，收集他们的见解，并邀请6位具有相关背景的教师参加问卷调查，旨在验证我们系统的有效性和实用性。

领域专家确认我们的系统设计得很好，考虑到了练习的特点，并赞扬了它的用户友好界面。E1提到：“该系统在多个层次上可视化内容和学生行为。通过可视化设计，我可以直观地观察整体行为并识别显著事件。我想将这个系统应用于其他科目，以评估不同课程或教师的行为，这可能会揭示有趣的现象。”E2强调了我们设计在展示差异和关联方面的有效性，使分析练习适当性和识别潜在问题变得更加容易。她强调：“具体来说，该系统为练习类型、问题类型和数量提供了有价值的建议。此外，它允许我调整练习分布并提示教师更多地关注需要注意的概念。然而，当前数据处理是半自动的，可以改进以获得更大便利。自动化应在系统中进一步开发。”作为教学专家，E3强调该系统的交互（如选择、工具提示和高亮）满足了她快速识别值得注意学生并访问相关信息以提供进一步帮助的需求。她还建议：“该系统作为监控学生基于练习表现的行为的便捷工具。如果它能够整合其他行为，如论坛帖子或视频观看，并分析更大样本量，我就可以探索更多可能原因并考虑适当措施。”与前面的专家不同，E4（分析师）注意到系统内部统计数据种类繁多且组织良好。他评论道：“在在线学习练习中，尝试次数提供了有关学生学习进度和毅力的宝贵信息。然而，它可能无法完全展现学生的学习历程或对材料的理解。因此，考虑其他因素至关重要。”

总的来说，专家反馈突出了系统的优势，包括其视觉设计、分析练习适当性的有效性、提供有价值的建议以及支持监控和帮助学生。此外，专家们对扩展系统能力以包括额外的行为和主题表示兴趣，从而使对学生行为的评估更加全面，并告知适当的干预措施。

在访谈之后，我们调查了来自不同背景的10位教师，以验证ExeVis的有效性和实用性。参与者具有正常或矫正到正常视力，并事先了解研究目标。问卷遵循推荐指南[40, 53]设计，使用7点Likert量表，范围从“完全不同意”（1）到“完全同意”（7）。调查分为三个步骤，持续约30分钟。参与者被介绍给ExeVis并接受10分钟的培训课程。然后他们有15分钟自由探索系统，最后5分钟完成问卷。

图5中的结果表明，参与者对系统普遍持积极看法，所有七个问题的平均评分在4.99至5.77之间。具体来说，Q1-Q3获得了5.65的平均评分，表明一致的积极评价和对系统的总体满意度。在视觉组件方面（Q4-Q7），系统获得了7分中5.20分的平均评分，表明参与者认为视觉设计引人入胜且有效地传达信息。然而，由于它们复杂性和抽象性，参与者在识别练习之间的关系（Q6）时遇到了一些挑战。总体而言，问卷回答反映了系统的积极表现。

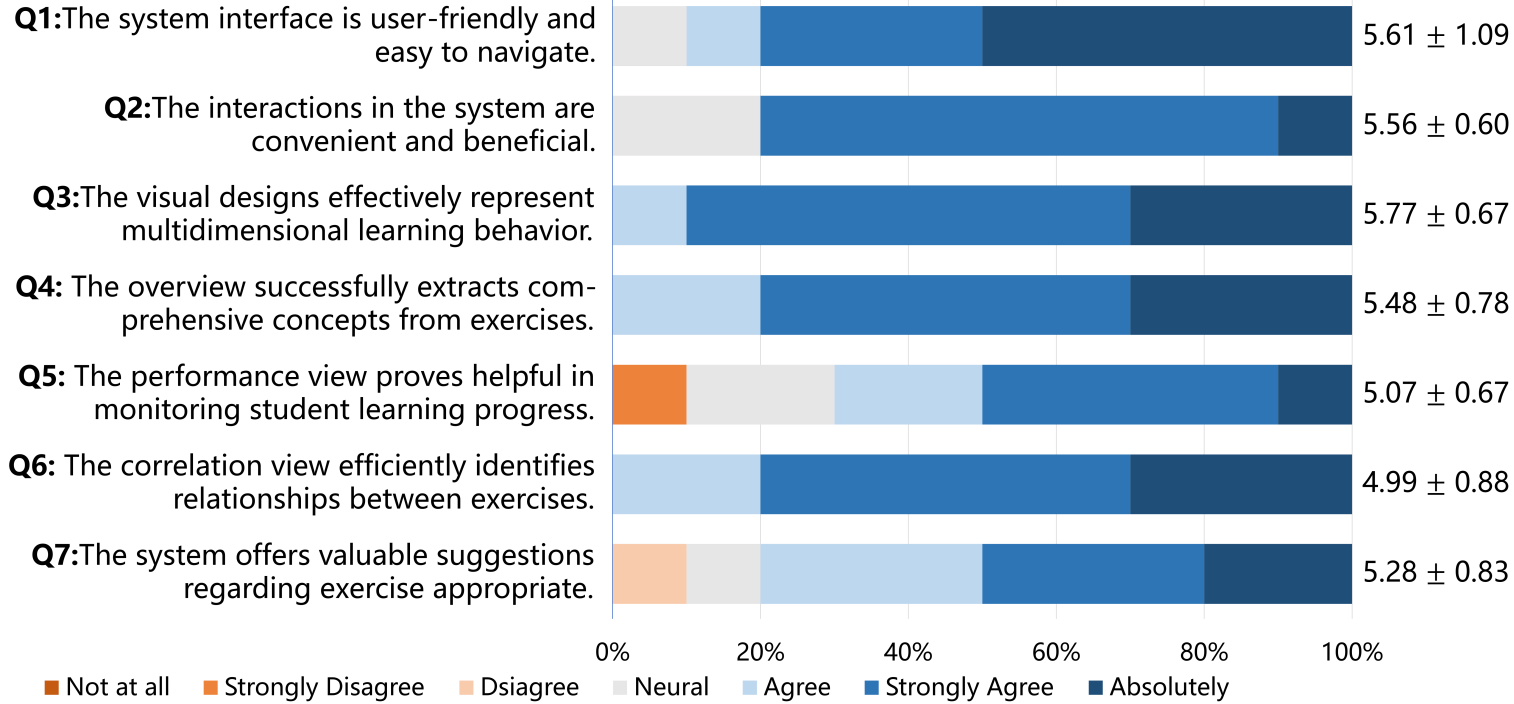


图5. 问卷的描述和行为，最右边一列代表平均值±标准差。

6.4 讨论 根据案例研究和专家访谈，我们确定了系统的局限性和相应的反思：

当前实现依赖于自动和半自动方法的组合，可能导致潜在的不便。此外，用户定义的组选择方法可能引入偏见或主观性。为了克服这些局限性，需要进一步研究以增强自动化能力并开发更稳健、数据驱动的组选择方法。这需要开发先进的算法和技术来自动化数据处理、可视化生成和交互机制。一个完全自动化的系统将为用户提供无缝、流畅的体验，使他们能够专注于洞察力和分析。

其次，为了克服这一局限性，我们建议扩大分析范围以包括更多变量和维度。这可以通过整合额外数据源（如在线论坛和视频观看行为）来实现，以获得对学生参与度和学习模式的更全面理解。

此外，我们研究中使用单一数据集限制了我们捕捉不同教育背景下变化的能力。为了获得对可视化系统有效性和适用性的更全面理解，我们必须将分析扩展到包含多种教育背景的多个数据集。

通过解决这些局限性并结合自动化技术、全面分析和多个数据集，我们可以提高ExeVis提供的能力和洞察力。这将有助于更有效、更用户友好地支持教育决策并改善在线学习成果。

7 结论 在本文中，我们介绍了ExeVis，一个用于分析在线学习练习的综合可视化系统。它引入了几种创新的视觉编码和多个交互式可视化视图，帮助教育者从练习分布和学习者行为中获得宝贵的洞察力。此外，设计了丰富的交互集来关注感兴趣的局部区域，并测量并直观呈现一组比较指标。基于真实世界数据集的案例研究和领域专家访谈进一步证明了我们系统在评估不同维度下练习适当性并使教育者相应调整教学方法方面的有效性。

致谢 本工作部分得到国家自然科学基金（62277013, 62177040, 62132017）、国家统计科学研究项目（2022LY099）、中国浙江省科技计划项目（2021C03137）、浙江省科技厅公益计划研究项目（LTGG23H260003）和浙江省统计科学研究项目的支持。

下面是每篇论文的来源列表：

1. H. Abdi 和 L. J. Williams 的 “Principal component analysis” 发表在 Wiley Interdisciplinary Reviews: Computational Statistics，第2卷，第4期，第433-459页，2010年。
2. A. Ahadi, R. Lister 和 A. Vihavainen 的 “On the number of attempts students made on some online programming exercises during semester and their subsequent performance on final exam questions” 发表在 2016 年 ACM 创新与技术计算机科学教育会议论文集，第218-223页，2016年。
3. J.R. Anderson 的 “The architecture of cognition” 由 Psychology Press 在 1996 年出版，作为 Cognitive Science 系列的第 5 卷。
4. J. Arruarte, M. Larrañaga, A. Arruarte 和 J.A. Elorriaga 的 “Measuring the quality of test-based exercises based on the performance of students” 发表在 International Journal of Artificial Intelligence in Education，第31卷，第585-602页，2021年。
5. E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I.W. Kwok, L.G. Ng, F. Ginhoux 和 E.W. Newell 的 “Dimensionality reduction for visualizing single-cell data using umap” 发表在 Nature Biotechnology，第37卷，第1期，第38-44页，2019年。
6. L. Breslow, D.E. Pritchard, J. DeBoer, G.S. Stump, A.D. Ho 和 D.T.Seaton 的 “Studying learning in the worldwide classroom research into edx’s first mooc” 发表在 Research & Practice in Assessment，第8卷，第13-25页，2013年。
7. P.C Brown、H.L Roediger III 和 M.A McDaniel 的 “Make it stick: The science of successful learning” 由 Harvard University Press 在 2014 年出版。