摘　要

基于 MOOC 数据的学生行为模式提取与可视分析

近年来，随着智能化技术及大数据技术的飞速发展，大规模的在线开放课程（Massive Open Online Courses, MOOCs）在世界各国迅速兴起，吸引了数百万用户在线学习和交流。MOOC 提供了涵盖各学科领域的课程内容和学习资源，帮助学习者获取更专业及系统的知识和技能。MOOC 数据不仅包含学习者信息和学习结果数据，还包含学习者与各种课程材料互动的网络日志记录，为教育相关人员深入了解在线学习行为提供了基础。然而，学生日志记录具有数据庞大、复杂异构和多元的特点，如何有效提取 MOOC 数据中蕴含的学生学习模式，并对其进行可视化是一项重要挑战。可视分析利用人类的视觉感知，通过互动探索来分析数据，让人类的认知能力参与到大规模、高维度和稀疏的数据分析中。可视分析与单纯依靠计算机自动分析的传统方法相比，可以充分利用人类认知能力优势，更有效地挖掘数据的内部模式。因此，本文提出应用先进的数据挖掘和机器学习方法，结合可视分析技术，对 MOOC 数据进行挖掘，以此帮助教师和教育专家更好地发现学生的行为模式及规律，理解各种学习行为背后的原因。

基于异构且复杂的 MOOC 数据，本文提出基于数据挖掘和可视化的方法，从不同角度提取学生行为模式，并以可交互的方式帮助教育分析者和相关研究人员挖掘和探索学生学习序列的潜在规律，为后续课程改进和针对性干预提供新见解。本文的主要工作如下：

（1）基于多属性事件序列的学生行为模式提取与可视分析：本工作将收集到的学习记录数据建模为多属性事件序列，针对属性数据格式不同且属性之间存在关联的问题，提出了一种基于最小描述长度（Minimum Description Length, MDL）的多属性事件序列模式提取方法，该方法支持对属性权重进行调整，同时考虑了属性之间的关联性。最后，设计并实现了一个交互式的可视分析系统

——SPVis，集成丰富的视图和多种交互方式，帮助用户从不同层次探索不同学习群体的学习模式。

（2）基于全课程点击流数据的学生行为与表现关系的可视分析：本工作利用高阶网络提取点击流数据中具有依赖关系的子序列，依据子序列和子序列之间的关系来总结学生学习行为，并对视频内操作过程进行建模。同时，本工作设计了四个链接视图，包括模式概览图、点击流详情图、序列视图和点击流对比图，不仅可以有效地概述大量的点击流数据，同时支持对学生个人行为进行详细的比较，有助于用户对点击流数据进行深入的理解和分析。

最后，本文基于真实的 MOOC 数据进行实验，通过案例研究和专家访谈证明可视分析系统在帮助教师分析学习行为，探索学习模式，发现学习规律和启发课程改进等方面的有效性。

关键词：

教育可视化，可视分析，事件序列，模式提取

Abstract

Pattern Extraction and Visual Analysis of Student Behavior Based on MOOC Data

In recent years, with the rapid development of intelligent technology and big data technology, Massive Open Online Courses (MOOCs) have rapidly emerged in countries around the world, attracting millions of users to learn and communicate online. MOOC data contains not only learner information and learning outcome data, but also web log records of learners' interactions with various course materials, providing a basis for education stakeholders to gain insight into online learning behavior. However, student log records are characterized by large, complex heterogeneous and multiple data, and it is an important challenge to effectively extract and visualize the student learning patterns embedded in MOOC data. Visual analytics leverages human visual perception to analyze data through interactive exploration, engaging human cognitive abilities in large-scale, high-dimensional and sparse data analysis. Visual analytics can take advantage of human cognitive abilities to explore the internal patterns of data more effectively than traditional methods that rely solely on automatic computer analysis. Therefore, this paper proposes to apply advanced data mining and machine learning methods, combined with visual analytics, to process MOOC data in order to help teachers and education experts better discover patterns and laws of student behavior and understand the reasons behind various learning behaviors.

Based on the heterogeneous and complex MOOC data, this paper proposes a approach based on data mining and visualization to extract student behavior patterns from different perspectives and help educational analysts and related researchers explore the potential patterns of student learning sequences in an interactive manner, providing new insights for subsequent curriculum improvement and targeted

interventions. The main work of this paper is as follows.

(1) Patterns extraction and visual analysis of student behavior based on multi-attribute event sequences: this work models the collected learning record data as multi-attribute event sequences and proposes a multi-attribute event sequence pattern extraction based on Minimum Description Length (MDL) for the problem of different formats of attribute data and the existence of correlations between attributes. The method supports the adjustment of attribute weights and takes into account the correlation between attributes. Finally, an interactive visual analysis system SPVis, is designed to integrate rich views and multiple interaction methods to help users explore the learning patterns of different learning groups from different levels.

(2) Visual analysis of student behavior and performance relationships based on whole-course clickstream data: This work uses higher-order networks to extract subsequences with dependencies in clickstream data, summarizes student learning behaviors based on the relationships between the subsequences, and models the operation process within videos. Meanwhile, this work designs four linked views, including pattern overview map, clickstream detail map, sequence view and clickstream comparison map, which not only can effectively overview a large amount of clickstream data, but also support detailed comparison of individual student behaviors, which helps users to deeply understand and analyze clickstream data.

Finally, this paper conducts experiments based on real MOOC data, and demonstrates the effectiveness of the visual analytics system in helping teachers analyze learning behaviors, explore learning pattern, discover learning pattern and inspire course improvement through case studies and expert interviews.

Keywords:

Educational visualization, Visual analytics, Event sequence, Pattern extraction

关于学位论文使用授权的声明

本人完全了解吉林大学有关保留、使用学位论文的规定，同意吉林大学保留或向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅；本人授权吉林大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或其他复制手段保存论文和汇编本学位论文。

（保密论文在解密后应遵守此规定）

论文级别：■硕士 □博士

学科专业：计算机技术

论文题目：基于 MOOC 数据的学生行为模式提取与可视分析

作者签名：马小宏　　指导教师签名：

2023 年 5 月 12 日

目　　录

第 1 章 绪论

1.1 研究背景和意义

近年来，大型的在线开放课程（MOOCs）发展迅速[1]，从最初在美国、法国、英国等国家开始兴起，到如今发展到世界各国，吸引了数百万的在线用户进行交流学习。MOOC 突破了传统课程中时间和空间的限制，使得用户足不出户即可学习到众多高校的课程，享受丰富的学习资源。同时 MOOC 不同于传统的课堂式学习，其采用碎片化学习，以知识点的形式呈现课堂，学生可以自由安排时间进行学习，突破了普通课堂的限制，但也要求学生有较高的自律能力和学习能力。MOOC 选课人数通常是线下教学人数的几倍到几十倍，在如此高的选课率下，教师迫切需要学生关于课程安排和讲学风格的反馈，来帮助他们提高教学质量，及时对课程做出调整。然而，与传统课程相比，MOOC 的学习者来自世界各地，具有不同的学习背景和动机，这增加了学习分析的难度。

许多 MOOC 平台记录了学习者的网络日志数据，包括学习者个人资料、视频点击流互动和论坛活动的数据，这为 MOOC 学习行为数据分析的研究提供了依据。网络日志数据包含了每个学生何时执行何种行动的时间信息，例如，观看讲座视频，向论坛发布问题，以及提交问题答案等。虽然课程视频以及问题测验的顺序是参考教学大纲精心设计的，但学生不必拘泥于此。学生可以自由的依据自身学习情况进行个性化的学习调整，这与传统的课堂教育有显著的不同。例如，一些学生可能会跳过某些视频内容，直接去参加作业和考试，以获得证书，而另一些学生可能会活跃在论坛上，但没有完成任何测验。因此，分析学习序列对于教师了解学生学习行为是至关重要的。此外，序列数据可以分析出不同学习者群体的学习习惯，区分学习者学习 MOOC 的意图，从而更好地理解学习行为与结果之间的关系。然而如何高效的利用这些数据，对其进行充分的理解和挖掘，是教育研究人员面临的挑战。

近几年，有大量研究针对 MOOC 数据进行分析，特别是在数据挖掘和机器学习领域。研究人员已经应用了各种数据挖掘方法来研究学习行为，包括早期

发现和预测学生辍学，例如，Feng 等人[2]提出了一个上下文感知的特征交互网络（Context-aware Feature Interaction Network, CFIN）并用其来建模和预测用户的退出行为。CFIN 利用上下文平滑技术来平滑不同上下文中的特征值，并使用注意力机制将用户和课程信息结合到建模框架中。也有工作预测学生在小测验或考试上的表现，例如，Jiang 等人[3]利用学生在第一周的作业表现和 MOOC 内的社交互动来预测他们在课程中的最终表现，并发现获得外部激励的学生更有可能完成课程。推荐个性化学习路径可以提供给学生更具针对性的建议，例如，Zhou 等人[4]训练长短时记忆（Long Short-Term Memory, LSTM）模型以预测学生的学习路径和表现，并从路径预测的结果中选择个性化学习路径推荐给学习者。尽管上述工作帮助研究者发现了一些学习模式，比如识别出更容易退出课程的学习者类型，发现对学生适当的激励更有助于他们完成课程，但对于解释复杂的学习行为和学习顺序的关系仍然具有挑战。同时，这些模型大多注重预测结果的准确性，对于没有机器学习基础的分析者来说，缺乏有效的解释，并且一些对教师有意义的具体方面可能会被忽略，不利于教师在特定背景下对一些很实用的非主导特征进行调查。

为了解决上述问题，有研究引入了数据可视化和可视分析。可视化使用视觉表示和交互技术来帮助人们理解和分析数据。它利用人眼进入大脑的宽带宽，让用户可以同时看到、理解和探索大量的信息。可视分析[5][6]，作为一个高度跨学科的研究领域，结合了可视化和领域专家的知识，以及各种数据分析的算法和方法。在可视化领域方面，可视分析集成了来自各种领域的方法，如信息分析[7]、地理空间分析[8][9]、统计分析[10]、知识发现[11]、数据管理和知识表示等。可视分析使得决策者能够从庞大的数据中提取有价值的信息，并帮助他们决定如何更好地使用这些信息。

近年来，有很多研究结合可视化技术和自动分析技术，来帮助教师观察、探索、比较和理解学习记录。例如，Fu 等人[12]设计了一个名为 iForum 的可视分析系统，用于帮助教师探索复杂且异构的 MOOC 论坛数据中的动态模式。Shi 等人[13]设计多个协调视图，帮助课程讲师和教育专家从不同的方面分析 MOOC 中视频点击流的数据，Chen[14]等人设计新颖的符号来展示点击流数据的峰值，

通过大小、颜色、水平线等视觉参数将峰值的关键特征编码，有助于用户及时发现点击流峰值的异常。然而，这些工作主要集中在聚合级别上,消除了顺序信息，并且对学生在单个视频的内部使用过程缺乏研究。另外，一些研究主要集中在学习序列分析上。例如，Guo 等人[15]发现了大多数学习者表现出非线性探索的活动序列，尤其是在 MOOC 使用率不高的国家。Chen 等人[16]设计了交互式多层次视觉分析系统 ViSeq，帮助教师探索各种类型的学习序列，以检测不同的学习者群体，并了解学习序列和表现之间的潜在相关性。模式提取是学习数据分析研究中的关键任务之一，但目前工作主要集中在单属性序列模式提取，而实际数据往往是多属性的，如观看视频的行为包括视频 ID、观看时长、暂停次数、时间戳等属性。本文通过挖掘收集到的多属性事件序列，更细粒度地展现学生的学习模式。

本文主要致力于研究可视化和可视分析技术在教育领域中的应用，挖掘并解释学生学习行为记录中的潜在的学习模式和规律，帮助教育分析者和研究人员更好的理解学生行为，并为后续修改课程设计提供依据，进而提高学生的培养质量。

1.2 研究内容

本文主要的研究内容是基于异构且复杂的 MOOC 数据，从不同角度提取学生行为模式，并设计可视分析框架，以促进教师和教育专家理解、探索、分析 MOOC 数据。主要工作和贡献如下：

（1）基于多属性事件序列的学生行为模式提取与可视分析：针对教育领域的分析需求，将收集到的学习记录数据建模为多属性事件序列，并提出了一种基于最小描述长度（Minimum Description Length, MDL）的多属性事件序列模式提取方法。该方法考虑了属性之间的关联性，并且支持用户对不同属性设置权重。基于该方法设计并实现了一个交互式的可视分析系统——SPVis，系统集成隐喻的可视化视图和友好的交互机制，支持从群体层面到个体层面的学习序列探索，帮助用户从多角度分析学生学习模式。

（2）基于全课程点击流数据的学生行为与表现关系的可视分析：本工作基

于高阶网络构造方法分析、分类和总结学生的学习行为。设计四个链接视图包括模式概览图、点击流详情图、序列视图和点击流对比图，不仅可以有效地概述大量的点击流数据，而且支持对学生个人行为进行详细的比较，同时展示了学习者在视频内的非线性执行过程。最终，帮助教师了解学习行为和表现之间的潜在相关性。

基于教育领域的真实数据集进行了案例研究，以帮助教师和教育研究者获得对在线学习行为的新见解，专家反馈证明了本文方法和系统的实用性和有效性。

1.3 论文结构安排

本文主要包含以下五个部分，结构安排如下：

第 1 章：绪论。本章简单阐述了 MOOC 数据分析的研究背景和研究意义，并对本文研究内容进行概括说明。

第 2 章：相关工作。本章介绍了与本论文联系密切的相关工作，包括事件序列模式挖掘、事件序列可视化、学习分析和教育可视化的研究现状。

第 3 章：基于多属性事件序列的学生行为模式提取与可视分析。首先，介绍多属性事件序列的构建和基于 MDL 的多属性事件序列提取方法。随后，介绍了从与领域专家的访谈中总结出的分析任务。之后，对系统的可视化设计以及多个视图如何协同完成分析任务进行详细说明。最后，案例研究和专家访谈证明了方法和系统的有效性和实用性。

第 4 章：基于全课程点击流数据的学生行为与表现关系的可视分析。主要介绍基于高阶网络的学生行为模式发现，以及多个协调视图如何帮助教师了解学习行为和表现之间的潜在相关性。

第 5 章：总结与展望。本章对本文的研究工作进行总结和展望，以及分析本文存在的不足和未来工作的方向。

第 2 章 相关工作

本章将对与研究课题相关的研究方向做出总结，主要包括：1）事件序列模式挖掘，2）事件序列可视化，3）学习分析和教育可视化。

2.1 事件序列模式挖掘

模式挖掘算法是一种常用的事件序列分析方法。序列模式挖掘（Sequential Pattern Mining, SPM）即在序列集中寻找频率高于一定阈值的子序列[17]。常见的工作有基于 Apriori 的序列模式挖掘和基于频繁模式增长（Frequent Pattern Growth, FP Growth）的序列模式挖掘。GSP（Generalized Sequential Pattern）[18]算法是基于 Apriori 的序列模式挖掘方法，采用分层搜索的思想遍历整个序列集，生成每种长度下符合支持度的频繁序列模式集合，之后从中过滤包含非频繁子序列模式的序列。该算法缺点是当数据集过大时，会产生大量候选序列模式，使得运行效率低下。SPAM（Sequential Pattern Mining）[19]通过对索引进行位图操作，提高了计算效率。基于 FP 的方法引入了映射数据库的概念，将原始序列映射到一个更小的数据库中，在每一个频繁项集对应的映射数据库内进行频繁子序列搜索，这种方法允许并行索引，大大提升了搜索效率。其中常见算法的有 FreeSpan（Frequent Pattern-Projected Sequential Pattern Mining）[20]及其优化算法 PrefixSpan（Prefix-Projected Pattern Growth）[21]。许多可视化工作建立在 SPM 的基础上，来展示数据中的顺序模式，如 Fp-viz[22]、Timestitch[23]、Frequence[24]和 Peekquence[25]等。Liu 等人[26]提出了一个三阶段分析管道来探索模式和序列。该管道包括一个模式剪枝算法，可以过滤 SPM 挖掘出来的冗余模式。除了使用自动挖掘算法进行模式发现外，Vrotsou 等人[27]提出了一种交互式的数据挖掘方法，结合直观的可视化界面，帮助用户识别有意义的序列。SPM 算法往往生成大量的模式，分析人员通常需要依赖特定的阈值或合适的度量来使结果易于管理和呈现。然而，阈值设置不当，可能忽略重要但不经常出现的序列和异常值。

近几年 MDL 原则逐渐被研究人员应用[28][29]，MDL 原理表明一个数据集的最佳模型会使它的描述长度最小化。有很多工作基于 MDL 原则来从序列中提取模式，进行可解释性模式挖掘。Chen 等人[30]基于 MDL 原则，通过聚合事件序列并识别每个聚合组的代表性序列模式来提供对数据的概述。AirVis[31]基于 MDL 原则挖掘空气污染的传播模式。然而，上述方法都不能很好地处理具有多个属性的事件序列数据。因此，一些学者针对多属性的事件序列数据进行了研究。Bertens 等人[32]引入了一种基于 MDL 的方法来处理多属性事件序列，该方法可以生成交叉序列模式。Wu 等人[33]提出一种多属性序列模式提取方法用于分析球拍运动中的战术模式。之后，Wu 等人[34]提出了一种基于约束的模式挖掘算法，该算法生成一组初始策略，并将专家的领域知识整合到数据挖掘算法中，以帮助发现有意义的战术模式。但是，上述方法缺乏对属性之间关联性的考虑，同时也不能满足教育领域的需求。

受上述工作启发，本文提出了一种基于 MDL 的多属性序列模式提取算法，在考虑属性关联的同时，支持用户对不同属性设置权重，以满足教育领域的分析需求。

2.2 事件序列可视化

先前已经有大量工作集中在事件序列可视化方面。最常见的可视化方法是沿着水平时间轴排列事件，例如 Wang 等人的 Lifelines2[35]和 Krstajic 等人的 Cloudlines[36]。由于顺序数据包含时间信息，上述编码方式对于探索个体序列随时间的变化是非常有效的，同时可以揭示每个事件的详细信息。然而，当处理大量的个体序列时，数据量增加，信息繁杂，增加了用户认知负荷，阻碍用户从中获取到有价值的见解。

为了有效的提取和显示序列中的信息，需要解决两方面的问题：数据规模大和模式多样性。Du 等人[63]的综述总结了解决上述挑战的一些方法，主要分为四类：提取、时间折叠、模式简化和迭代策略。Wongsuphasawat 等人提出了基于事件聚合多个序列的 LifeFlow[37]，事件序列保留公共模式聚合形成一颗树。同时，在此基础上进一步开发了 Outflow[38]作为 LifeFlow 的延伸。Liu 等人提出

了基于核心事件总结序列的 Coreflow[39]。如图 2.1 所示，Coreflow 通过对事件排序、将序列分组并根据所选事件修剪序列构造出一棵树。树的分支结构描述了序列中的一般模式，其中节点代表关键事件，链接表示关键事件之间的聚合流。Bodesinsky[40]等人设计了一个多层次可视分析系统来探索事件序列。事件视图将单个序列可视化为水平对齐的条形，其中每个条形表示一个事件。模式概述用于总结常见的事件模式，供用户在事件视图中查询和突出显示感兴趣的模式。Decisionflow[41]和谷歌 Analytics[42]等商业产品将事件序列压缩成转换图，其中节点表示不同阶段的事件，边连接序列中的连续事件。这些图通常可视化为桑基图，以显示常见的转换路径。最近有工作提出一种方法 Matrixwave[43]，它使用基于矩阵的可视化来显示图形，以避免桑基图中密集边缘引起的视觉混乱。

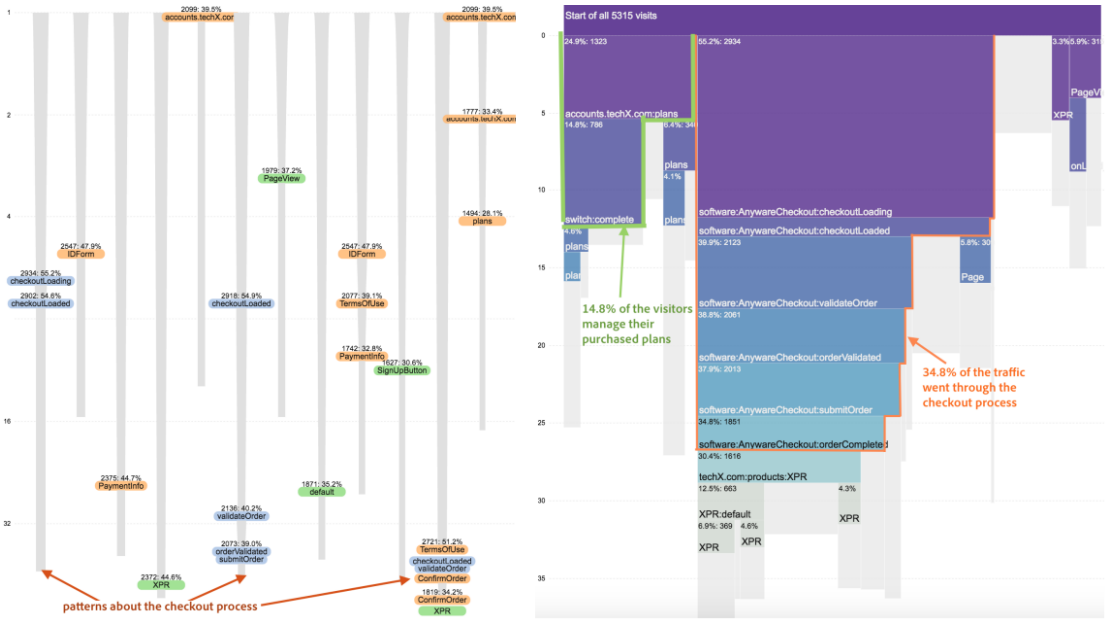

图 2.1 Coreflow：对事件序列中的分支模式进行提取并可视化[39]

交互是事件序列数据分析的重要组成部分。(s|qu)eries[44]、COQUITO[45]和 Decisionflow[41]提供可视化查询界面，帮助用户选择一个序列子集进行重点分析。Lifelines2[35]支持对选定事件的数据进行交互式对齐，用户可以轻松地发现前兆、共同发生的事件和后续事件。这些技术为序列模式分析提供了强有力的支持。然而，随着数据集的容量和复杂性的增加，这些技术的能力将受到限制。例如，随着过渡路径数量的增加，桑基图之间不同过渡路径的叠加会导致视觉混乱。Chen 等人[30]提出了一种用于事件序列汇总方法。该方法减少了序列概览中的视

觉混乱，同时最大限度地减少了概览中的信息损失。这使得分析人员即使在复杂的数据集中也能识别出显著的模式。

受上述工作启发，本文第三章提出了一种事件序列总结的可视化方法，不仅可以直观的显示顺序模式和事件的多个属性，而且利用三角形符号隐喻缺失的信息，以指导用户按需进行详细探索。第四章基于高阶网络算法来提取导致不同转移概率的关键序列。此外，高阶网络综合了这些序列之间的关系，并有效地概述大规模点击流数据。最后，通过旭日图并结合交互技术来使群体级别的学习路径可视化。

2.3 学习分析和教育可视化

学习分析是一个横跨教育、心理和计算机等多个学科的领域，其旨在通过收集学习活动反馈的数据，并利用各种分析方法和分析技术来了解学生行为，监测学习过程，预测学习者的表现，提供课程修改的思路。学习分析吸引了教育研究人员以及教学从业人员，他们希望通过自下而上的学习行为分析和学习结果反馈来理解和改进学生的学习方式。

学习分析的一个重要研究方向是理解学习行为并对其进行建模。例如，机器学习模型可以预测辍学率[46]。同时还可以识别出学习者在课程中退出的时间或特定点[47]。Hsieh 和 Wang[48]利用基于相关性的算法推荐学习内容。然而，他们的推荐主要取决于材料的内容，缺乏对学生的行为和学习习惯的考虑。Arnold[49]等人开发了课程信号，使教师可以通过个性化的电子邮件以及信号灯上的特定颜色为每个学生提供实时反馈。

可视化技术为识别和解释行为模式提供了另一种方法。Wang 等人[50]开发了一个名为 LISSA 的学习分析仪表盘，通过可视化成绩数据来促进导师和学生之间的交流。Derick 等人开发了 AffectVis[51]来可视化学习者的情感状态，并显示它们与特定学习活动的联系。有大量工作集中在对点击流数据的分析，Shi 等人设计了一个名为 VisMOOC[13]的可视分析系统，从多个层次对 MOOC 平台的视频点击流数据进行展示，帮助分析用户的学习行为。Chen 等人[14]设计了 Peakvizor 来研究 MOOC 视频点击流数据中的“峰值”。文中设计了一种新颖

的符号用来显示峰值的多个属性，以便用户可以轻松地在概述或其它联动视图中识别特定的峰和异常点。流视图揭示关于峰值的空间和时间信息，以便详细分析学习行为。另外，MOOC 中的社会关系也受到了广泛的关注。例如，Dowell 等人探索语言和演讲风格与社会中心性[52]的潜在联系。他们发现，当学习者表现出一种叙事风格的演讲时，他们往往会在社交网络中处于中心地位。此外，Lee 等人[53]通过研究学生在不同课程之间的过渡情况，发现学生之间的社交关系与选择课程之间的潜在联系。

论坛是学生和教师互动的主要途径，评论中包含学生在学习过程中遇到的问题和对课程设计的一系列反馈。因此，对 MOOC 论坛数据进行探索可以为教师提供有价值的见解，以改进课程设计。Wu 等人[54]开发了 NetworkSeer 来了解论坛中学生之间的互动。平行坐标视图显示了学生的多个属性，用户可以根据这些属性对学生进行分组和过滤。节点链接图用来展示学生之间的互动关系。Fu 等人[12]提出了一个名为 iForum 的可视分析系统，用于帮助教师探索大量复杂且异构的 MOOC 论坛数据中的动态模式。该系统将用户发帖和相互回复的操作建模为线程，并提供多个协调视图，用于呈现活动用户和线程的概述、不同用户组随时间的详细交互以及线程的动态模式。在 MOOC 中，学习者不一定要遵循教师设计的课程资源顺序，他们可以自由地对学习材料进行探索[55]。Chen 等人[16]开发了一个名为 ViSeq 的交互式视觉分析系统，该系统侧重于对不同学习者群体的学习序列进行视觉分析。Wang 等人[56]展示了学习者访问 MOOC 资料的详细轨迹，并识别其中的学习行为模式。

在个性化学习方面，Xia 等人提出了 Peerlens[57]，一种交互式可视分析系统，可以实现同伴启发式的学习路径规划，为学习者推荐定制的、适应性强的练习题序列。该工作提出了一种基于提交类型的学习路径建模方法，并将推荐的序列采用新颖的拉链设计进行展示，有效地促进学习者对学习路径的理解和规划。问题设计者为了推断学生解题逻辑是否符合他们的设计意图，需要对学生解题步骤的过程进行分析。细粒度交互数据为分析解决问题的行为提供了机会。然而，对这类数据的总结、展示和比较仍具有挑战性。Xia 等人提出了可视分析系统 QLens[58]，如图 2.2 该系统将问题解决行为建模为一个混合状态转换图，帮

助问题设计师检查详细的解决问题轨迹，获得设计改进的见解。Han 等人设计了可视分析系统 HisVA[59]，帮助用户探索维基百科中历史事件的时间和空间信息，以发现事件之间的联系，培养学生的元认知能力。但该系统基于主题建模的方法具有黑盒特性，可能会产生用户难以理解的主题。

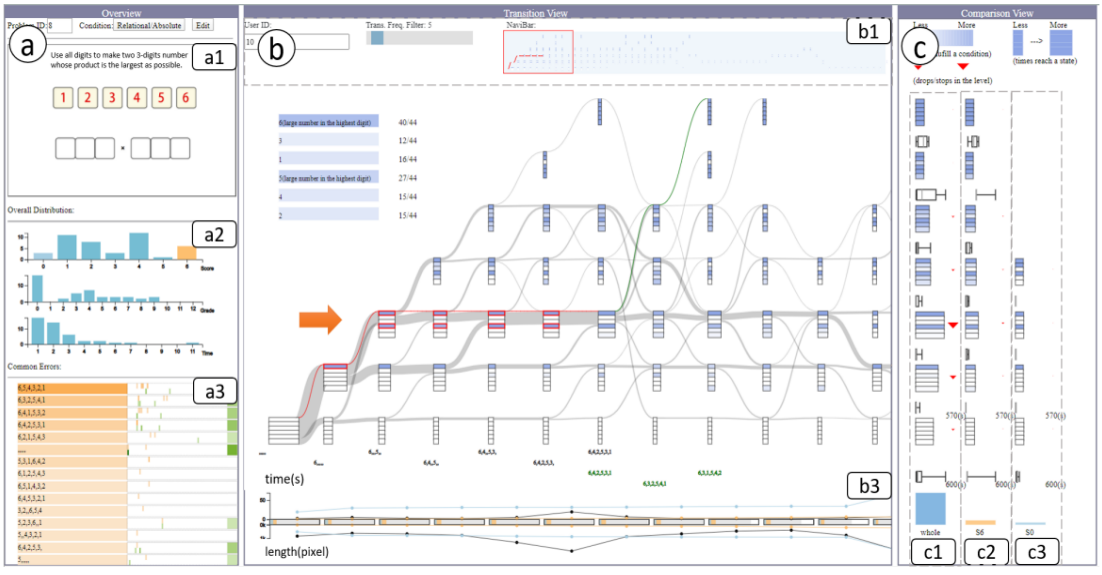

图 2.2 Qlens：用于改进问题设计的多步骤问题解决行为的可视分析[59]

与上述系统不同的是，本文第三章设计了一个交互式的可视分析系统——SPVis，支持用户在不同的数据粒度下探索学生行为模式。该工作主要针对 MOOC 中的多属性事件序列进行模式提取，并设计可视化编码来展示多个属性，更细粒度的展示学生行为。同时，系统支持用户依据不同时间、特定事件等对学生序列进行过滤。本文第四章提出新颖的视觉交互框架，该框架基于先进的机器学习方法和可视化技术，帮助用户发现点击流数据与学生表现之间的潜在联系。在数据处理模块基于高阶网络提取点击流数据中具有依赖关系的子序列，并对学生在视频中操作过程进行建模。可视化模块设计可感知的视觉表示，使得教育分析者能够在丰富的数据信息中探索潜在学习规律。

第 3 章 基于多属性事件序列的学生行为模式提取与可视分析

3.1 简介

MOOC 日志记录包含了学生的在线学习活动详情，为教育研究人员了解他们的学习行为提供了一个新的机会。然而为了充分利用这一价值，研究者必须面对数据的复杂性所带来的挑战。首先，绝大多数集成分析工具都集中在简单的统计数据上，这可能无法进行深层次的分析。例如，一个学生的活动数量经常被用来确定这个学生的活跃程度，这可能会产生误导。在本文的研究中，我们发现“访问主页→空闲→访问主页→空闲”是最常见的活动序列之一。然而，这个顺序中的活动与课程学习无关，但可以显著增加学生进行的活动的数量。其次，有效地对学生分组对于具有大量注册人数的入门课程的可扩展分析至关重要。然而，目前的学习分析工具往往无法根据学生固有的学习行为对他们进行分组。同时，了解学生在课程不同阶段的学习序列模式对于后续课程的改进至关重要。

为了从大规模的 MOOC 学习记录数据中发现学生潜在的学习行为规律，本文提出了一种基于 MDL 的多属性事件序列模式提取方法。首先将收集到的学习记录数据建模为多属性事件序列，依据学生不同时期课程参与活跃度将其划分为几个子组，之后对不同子组提取模式，并可视化模式的细节和差异。另外，为了满足用户对特定事件前后模式的探索需求，本文使用自上而下的工作流，方便用户将初始序列逐步划分为感兴趣序列，然后对子序列提取出的模式进行细节层次的探索。

基于上述分析流程，本文设计并实现了一个交互式可视分析系统——SPVis，引导用户划分数据，探索不同学习者群体的学习序列模式，帮助用户更好地探索学习行为的潜在规律。可视化视图对同一子组所提取模式的相关信息进行展示，并支持选取该子组中的某个学生，展示其在课程周的详细学习序列，实现

了从全部学习数据到某一群体学习数据再到个体学习数据的多个层次粒度上对学习序列的可视分析，从而促进用户分析理解数据。此外，我们基于一门为期六周的真实 MOOC 课程的数据，进行了案例研究，验证了本文算法和系统的有效性，同时通过采访几位领域专家来评估该系统。专家反馈证明了本文系统在教育领域具有一定的实用性价值。

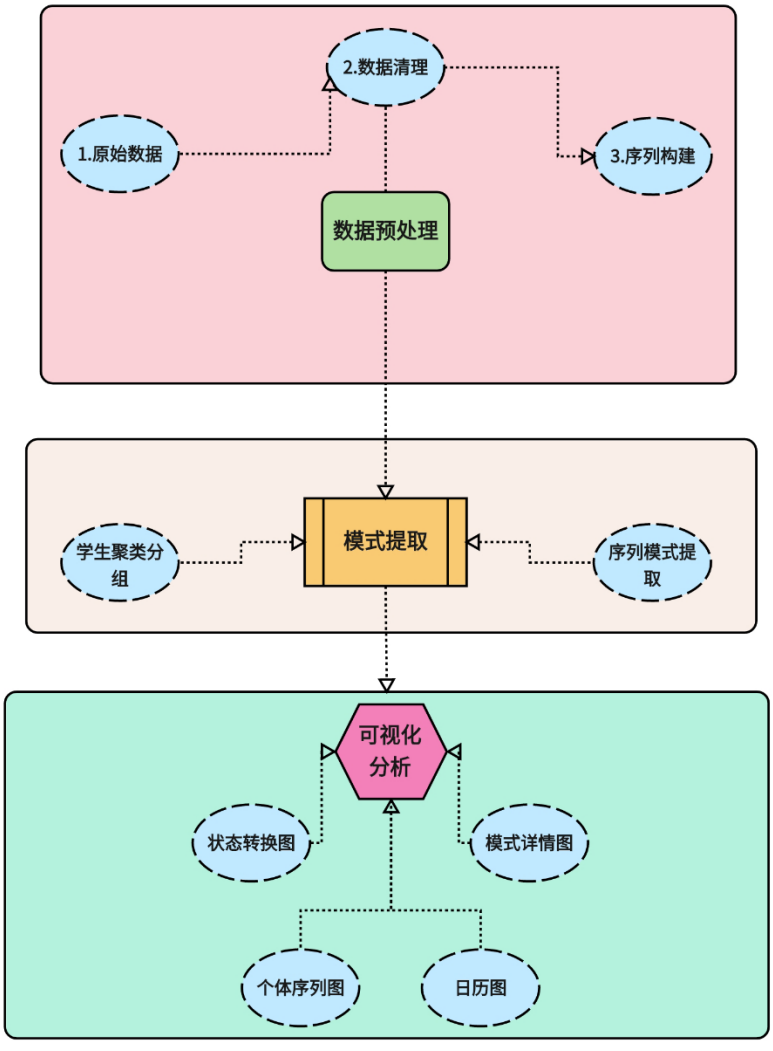

图 3.1 可视分析框架

图 3.1 展示了本文设计的可视分析框架的主要工作流程，包括以下三个模块：

（1）数据预处理模块。本文使用的数据为 XuetangX 数据集，记录多种类型的活动，包括视频观看、论坛讨论、完成作业和网页点击等。由于数据收集过程中存在一些脏数据，需对数据进行清洗，剔除其中的重复数据和无效数据。然后，将连续时间内对同一个视频的操作行为整体看作一个事件，不同操作执

行的次数作为该事件的多个属性，进而每个学习者执行的事件按发生时间顺序构成一个多属性事件序列。

（2）模式提取模块。将每个序列视为一个没有任何修正的模式，并不断地合并两个模式，以获得一个带有额外修正的新模式。在计算修正值（从模式中重建原始序列所需的修改）时使用一种新的度量方式，来满足教育领域的分析需求。

（3）可视化模块。可视化的设计要便于用户快速识别学习者群体和最常见的模式，并逐步显示足够的信息，不同视图之间要有细节上的补充。同时提供多种交互方式，支持用户自由地对感兴趣的数据进行探索分析。可视化模块总体上分为整体状态转换视图，模式探索和个体序列详情三部分。

3.2 数据分析

3.2.1 数据构建

本文的分析是基于 XuetangX 提供的数据集进行的。该平台提供了 3000 多门课程，吸引了 5000 多万名注册用户。XuetangX 的用户可以选择学习模式：教师节奏模式（IPM）和自节奏模式（SPM）。IPM 遵循与传统教室相同的课程时间表，而在 SPM，用户自己可以更灵活地安排时间进行在线学习。用户学习课程时系统会记录多种类型的活动：视频观看（观看、停止和跳转）、论坛讨论（提出问题和回答）、完成作业（正确/不正确答案和重置）和网页点击（点击并关闭课程页面）。本文选取其中一门为期六周的 IPM 模式课程进行分析研究。预处理的数据主要是学生在课程期间执行的一系列活动，以日志文件的形式记录。日志记录的详细信息如表 3.1 所示，不同类型活动的统计信息如表 3.2 所示。

学生产生的学习记录均带有对应动作发生的时间戳，执行的每个事件按时间顺序记录，本文将其整理为统一的数据格式。每个学生在课程期间的活动被转化为一个序列$S = [e_1, e_2, \cdots, e_i, \cdots, e_n]$，其中$i$为该事件的索引，$n$为序列中事件的总数。每个事件代表一个学生活动行为，包含多个描述属性，表示为$e_i =$

$\{a_1 = v_1^i, a_2 = v_2^i, \cdots, a_n = v_n^i\}$。例如，一个视频观看行为包括观看时间、观看时长、暂停次数、跳跃次数、时间戳等属性。

表 3.1 日志文件属性信息

属性	描述
enroll_id	学习者注册课程 ID
username	学习者编号
course_id	课程编号
action	操作类型
object	学习者访问的具体对象
time	时间戳

表 3.2 事件基本统计信息

类型	数量
play	52480
seek	26188
pause	42524
assignment	4038
forum	4693
click_info	45603

3.2.2 学生群体划分

由于选课人数众多，为了方便分析者探索感兴趣的学生群体，本工作依据学生日志数据中视频观看次数，回答问题个数，论坛评论次数等活动信息对学生进行聚类分组。

常用的聚类算法有很多，本文选择 k-means 算法。该算法基本思想是选取k个点作为簇的中心，计算样本中每个点与各个簇的质心的距离，将其划分给距离最近的簇，并更新簇的质心。之后，重复迭代上述过程直至收敛。k值的选取会影响聚类的效果，通常学生群体的聚类数目不会太多，一般选择 2 到 5 类，

如果聚类个数过多，不利于观察各个群体之间的区别。本文根据不同的k值计算出相应的轮廓系数值，以此来确定合适的聚类个数。轮廓系数计算公式如公式 3.1 所示：

$$S(i) = \frac{b(i)-a(i)}{max(a(i),b(i))} \quad \text{.....} \quad (3.1)$$

其中，$a(i)$表示内聚度，即数据点i到同一簇内其他点不相似程度的平均值；$b(i)$表示分离度，即数据点i到其他簇的平均不相似程度的最小值。轮廓系数的值介于[-1,1]，越趋近于 1 代表内聚度和分离度都相对较优。图 3.2 给出了不同k值所对应的轮廓系数值，从图中可以看出当k取值为 3 时，轮廓系数值最大，代表相应的聚类效果最好。

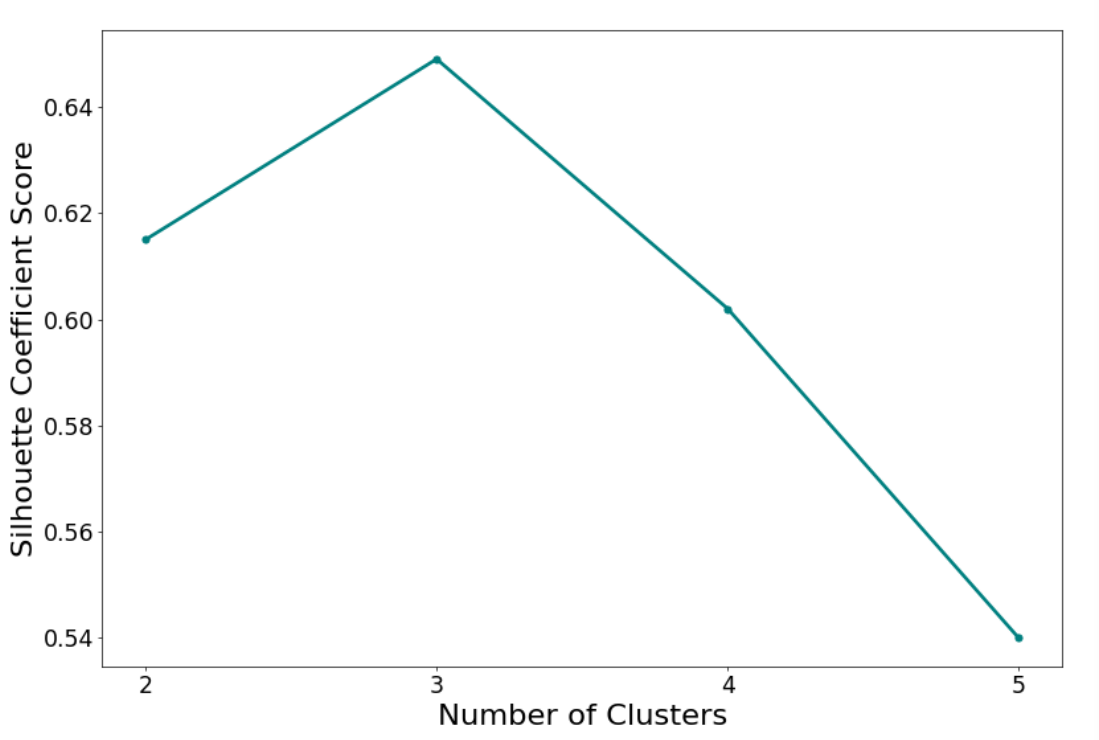

图 3.2 不同 K 值对应轮廓系数

最后，将学生分为三个类别，分别为类别 A、类别 B 和类别 C。表 3.3 展示了不同类别的活动特征的平均值的统计信息，从中可以发现各个类别的差异较明显。类别 A 参与各类活动均较少，属于不活跃的类型。类别 B 学习活跃性和投入度比较高，各类活动均频繁参与。类别 C 相较于类别 B 主要是在论坛访问和问题回答方面差别较大。因此，学生依据学习的活跃度被分为三个子组：不积极者、积极者和观看者。

- 不积极者：无论是视频观看，还是问题回答和论坛访问均很少参与。
- 积极者：在观看视频和完成问题等方面均比较活跃。
- 观看者：主要进行了访问视频等操作，但较少完成问题和访问论坛。

表 3.3 不同分组的特征统计

特征	类别 A	类别 B	类别 C
观看视频次数	14.33	152.31	90.01
回答问题次数	2.09	30.02	6.25
访问论坛次数	3.33	28.05	5.37
浏览相关信息次数	9.37	80.57	50.16

3.2.3 基于 MDL 的多属性事件序列模式提取

模式提取是学习数据分析中关键的任务之一。在以往的研究中[16]，通常使用 VMSP 或 Max SP 等算法，来提取学习序列中的频繁模式，但其仅考虑了单一的属性。本文主要研究学生行为数据中多属性时间事件序列的模式提取，并通过解决下述四个问题来支持模式提取的任务，即

（1）调整不同属性的权重

领域专家在进行分析时通常关注一个属性或多个属性的加权组合。例如，领域专家可能关注学生的视频观看时长或视频暂停次数。应该允许分析师调整不同属性的权重，以关注感兴趣的属性。

（2）关联属性的处理

时间事件序列中事件具有多个属性，通常事件的某个属性依赖于另一个属性存在，例如观看视频这一事件中包括视频 ID 和观看时长，暂停次数等属性，其中观看时长和暂停次数均依赖于视频 ID 这一属性。以往的多属性事件序列模式提取忽略了属性之间的关联，例如 Wu 等人[33]在提取序列模式时分别计算每个属性的编辑成本，然后将每个属性的转换代价相加得到序列S映射到模式P的编辑成本。各个属性的编辑成本均单独计算，这可能会造成信息损失。

（3）控制提取的模式的长度/真实性/连续性

为了使分析者更好的探索所提取的模式，应该允许其控制所提取模式的长

度、真实性、连续性等特征。

（4）基于时间信息度量

学生在课程周期中进行的所有活动均被记录，并且带有确切的时间戳。学生 A 和学生 B 均按顺序观看了{V1,V2,V3}，但学生 A 时间跨度较大，学生 B 观看较为集中，那么两个学生的学习行为有一定差异。然而，在之前的研究中，Chen 等人[16]采用 K-gram 算法提取子序列作为特征计算相似矩阵来比较学习序列的相似性，与传统的相似性度量（编辑距离）相比，虽然可以更好地反映不同长度序列的相似性，但这种度量方式丢失了时间信息。

本文改进了 MinDL[30]算法，它引入了 MDL 原则来识别一组顺序模式来概述数据，同时平衡其中的信息丢失。MDL 原则是统计模型选择中的一个著名的信息准则。MDL 原理表明一个数据集的最佳模型会使它的描述长度最小化。数据集的描述长度由两部分组成:（1）模型$L(\mathcal{M})$的编码，（2）模型$L(\mathcal{D}|\mathcal{M})$对数据的编码。最佳模型$\mathcal{M}$应最小化总描述长度，即$L(\mathcal{M})+L(\mathcal{D}|\mathcal{M})$最小。

事件序列是一个有序事件列表，表示为$S=[e_1,e_2,\cdots,e_i,\cdots,e_n]$，其中$e_i\in\Omega$，$\Omega$是一个事件字母表。对于给定一组事件序列$\mathcal{S}=\{S_1,S_2,\cdots,S_m\}$，模式提取的目标是识别一组模式$\mathcal{P}=\{P|P=[e_1,e_2,\cdots,e_l]\}$和从事件序列到模式的映射$f:\mathcal{S}\to\mathcal{P}$来最小化总描述长度：

$$L(\mathcal{P},f)=\sum_{P\in\mathcal{P}}L(P)+\sum_{S\in\mathcal{S}}L\big(S\big|f(S)\big)\quad(3.2)$$

3.2 式中$L(P)$为模式P的描述长度，$L(S|f(S))$是S在模式$f(S)$下的描述长度。由于模式P可以通过列出其中的事件来描述，编辑距离可以由事件和事件所涉及的位置来指定，其描述长度可以被视为一个常数，因此公式 3.2 可以改写为：

$$L(\mathcal{P},f)=\sum_{P\in\mathcal{P}}len(P)+\alpha\sum_{S\in\mathcal{S}}\|edits(S,f(S))\|+\lambda\|\mathcal{P}\|\quad(3.3)$$

其中，$len(P)$是模式中事件的数量，$edits(S,f(S))$是一系列可以将$f(S)$转换为S的编辑操作。公式 3.3 中引入参数α，以控制最小化信息损失对减少视觉混乱的影响。第三项加上λ参数用于控制生成模式的总数。λ越大生成的模式总数越少。因此，通过适当设置λ可以提高模式概述的可扩展性。

该算法采用了一种自底向上的启发式方法来确定序列的分组，以及每个组的代表模式，从而使总描述长度最小化。在初始化阶段，每个序列被视为一个

单独的集群，集群的模式为该序列本身。然后进行迭代合并集群对，并计算新集群的代表性序列模式。合并时采用贪婪算法的思想，算法总是选择可以最大的减少总描述长度的两个集群进行合并。当算法无法再找到一对可合并的集群来进一步减少总描述长度时，该算法将停止。但是，MinDL 无法解决上述提到的四个问题，不能直接应用于学习数据分析。因此，本文引入了一种新的度量方法（算法 3.1）来计算描述成本，从而满足领域需求。

在介绍算法前，首先介绍与其相关的附加信息，这四条信息可以帮助解决上面提出的四个问题。

（1）属性的权重。分析师可以通过输入不同的权重来调整分析结果。属性a_i的权重表示为w_i。

（2）关联属性。分析师可以自定义属性之间的关联，将属性分组。分组后每组由一个主属性和多个副属性构成，副属性的计算依赖于主属性。不同主属性之间相互独立。

（3）控制所提取模式的特征的三个参数。以下参数控制了所提取的模式的三个特征。

- $cost_{in}$定义为控制插入事件。一个事件在原始序列中出现，但是模式序列中未出现，该事件在模式提取过程中可能会影响模式的长度。
- $cost_{mis}$定义为控制丢失事件。一个事件在模式序列中出现，但在原始序列中未出现，该事件在模式提取过程中可能会影响模式的真实性。
- $cost_{con}$定义为控制不连续的事件（即在模式中相邻但在原始序列中不相邻的两个事件），这会影响模式的连续性。

（4）哨兵事件。分析师可以选择某种类型事件作为哨兵事件。当选择哨兵事件后，时间尺度从绝对时间变为相对时间。哨兵事件发生时间变为 0。

接下来，利用上述信息来进行度量距离的计算，以提取满足领域专家需求的学习模式。算法 3.1 是相似性度量算法的核心部分，下面使用图 3.3 中的例子来对其进行说明。算法以序列S和模式P作为输入（图 3.3（a）），输出成本ΔL_{cost}。其核心思想是遍历所有独立属性，分别计算每个属性的成本，并按权重相加，其中关联属性的成本依赖主属性进行计算。以独立属性a_1为例来演示

这个过程, 其中属性a_2是它的副属性。

算法 3.1：相似性度量算法

Input：序列S、模式P

Output：ΔL_{cost}（序列S转换到模式P所需开销）

1: $\Delta L_{cost} = 0$

2: **for** $i \leftarrow 1$ *to* $len(attributes)$ ***do***

3: $P_i \coloneqq the\ sequence\ of\ v_i\ in\ P$

4: $S_i \coloneqq the\ sequence\ of\ v_i\ in\ S$

5: $P\_lcs_i, S\_lcs_i, lcs = LCS(P_i, S_i)$

6: $mismatch\_cost = transform(S_i, lcs, P_i)$

8: $match\_cost = distance(P\_lcs_i, S\_lcs_i)$

9: $\Delta L_{cost} += (mismatch\_cost + match\_cost) \times w_i / w_{total}$

10: **end for**

首先，分别提取模式P和序列S的属性a_1对应的值（图 3.3（b））, 形成单变量模式P_1和单变量序列S_1。然后，计算P_1和S_1的公共子序列，记作lcs_1。另外，为了方便后续计算副属性a_2的成本，需分别记录模式P_1和序列S_1中公共序列对应的部分，记作P_lcs_1，S_lcs_1。接下来总成本由两部分组成：匹配和不匹配。匹配部分，在主属性a_1的值相同时（图 3.3（c））, 计算副属性a_2的欧式距离，记作$match_cost_1$。如果，主属性对应多个副属性，则匹配部分的成本为各个副属性的欧式距离和，且不同属性被赋以不同的权重。不匹配部分，依据P_1和S_1缺失和额外的值的个数计算（图 3.3（d））, 记作$mismatch_cost_1$。

类似 Levenshtein distance 方法，在计算不匹配部分时分别计算插入和删除的次数。但是本文的度量方法进一步使用插入/删除成本来控制模式的长度/真实性。此外，定位S_1和P_1中lcs_1的第一个事件和最后一个事件，来确定lcs_1在S_1和P_1中的跨度，即$span_1^S$和$span_1^P$，在示例中分别为 2 和 3。之后，分别计算三个部分的成本，公式 3.4 计算插入事件的成本，公式 3.5 计算删除事件的成本，公式 3.6 计算连续性的成本。最后，$mismatch_cost_1$的计算公式 3.7 如下：

$$\Delta L_i^{in} = (len(S_i) - len(lcs_i)) \times cost_{in} \quad \ldots\ldots\ldots\ldots (3.4)$$

$$\Delta L_i^{mis} = (len(P_i) - len(lcs_i)) \times cost_{mis} \quad \text{(3.5)}$$

$$\Delta L_i^{con} = \left(span_i^P - len(lcs_i)\right) + \left(span_i^S - len(lcs_i)\right) \times cost_{con} \quad \text{(3.6)}$$

$$mismatch\_cost_i = \Delta L_i^{in} + \Delta L_i^{mis} + \Delta L_i^{con} \quad \text{(3.7)}$$

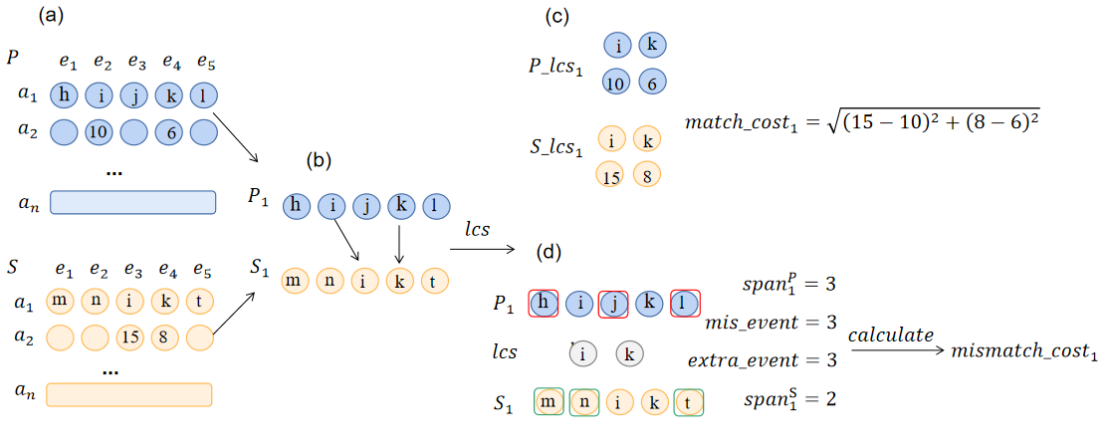

图 3.3　举例说明算法 3.1 的执行过程

3.3 分析任务

基于对数据的理解，通过和教育分析者的多次讨论后，确定了以下分析任务，来帮助分析者分析学习者的不同行为模式。

T1：根据学生行为特征来确定常见学生分组有哪些，以及不同子组随时间如何过渡。

通常，一门 MOOC 有上千人观看，且不同于线下课，学生可能在课程中途进入或退出。教育分析者无法用单纯的成绩高低和课程表现来对参与人员分组。通过可视化不同子组随时间的状态转换，可以帮助分析者快速挖掘到感兴趣的组，以此来进行后续分析研究。

T2：探索不同学生分组的典型学习序列模式。

学习平台收集到海量的日志数据记录，但教育分析者很难查看所有学生的学习序列。在根据行为特征对学习者进行分组后，用户需要更详细的了解关于感兴趣学习组在某周中如何学习的信息。例如：他们以某种顺序观看了哪些视频，访问了某个论坛，回答问题等。因此，需要在系统中提供模式查询功能。

T3：可视化不同分组观看视频的差异。

不同于线下课程，教师可以直接从课堂上获得反馈，MOOC 主要是通过上传提前录制的教学视频进行教学，教师需要通过分析这些教学视频的使用情况来了解学生的学习情况。分析者可能关注哪些视频得到了较高的关注量，哪些视频被反复观看，以及哪个视频被暂停次数较多，这些活动是否符合最初教学预期，后续应该对哪些视频进行改进。

T4：可视化给定学习者的详细学习路径。

在探索完一个学习分组的行为模式以后，用户可能对组内某个个体学习者感兴趣，想要了解他在课程某一时间段所做的事情。例如，在完成作业前，观看了哪些视频做准备。可视化应提供个体层面的详细信息，方便用户进一步分析。

T5：探索不同学生分组的时间管理。

学生的时间管理与花在学习上的时间有关，这可能会直接影响学生的参与度。研究它们之间的关系有助于理解学生参与的变化，并为教学干预提供基础。

3.4 可视分析设计

基于上述分析任务和可视化设计基本准则，本文设计四个视图用于辅助分析，包括：状态转换图，模式详情图，个体序列详情视图和日历视图。状态转换图向用户展示了不同时间段学生状态的流向分布。模式详情图允许用户查看不同学生分组的典型序列模式。个体序列视图展示了学生在某一时间段的详细学习路径。日历视图提供不同群体或个人的学习参与度对比。本节将详细介绍每个视图的可视化和交互设计。

图 3.4 SPVis 系统用户界面

3.4.1 状态转换视图

本文根据历史行为数据(即学生与 MOOC 互动的方式，如访问视频、提交评论、访问论坛等)，将所有学生聚类分组。如图 3.4（a）所示，系统使用桑基图来可视化三个子组之间的过渡模式（T1），其中黄色代表积极者、蓝色代表视频观看者、绿色代表不积极者，辍学的学生被标记到一个不同的子组，用红色部分表示。从左至右按周展示了属于每个子组的学生的比例，以及不同子组在周之间的过渡模式。教育分析者可以从中发现一些感兴趣模式，例如，除了第一周，在接下来的几周内，没有学生从不积极者子组（NA）转变为积极者子

组（Positive），这表明，如果没有任何干预，一个高度不活跃的学生不太可能在短时间内变得非常活跃。因此，一旦某个学生被检测到不活跃，教师应该尽早对其进行干预。例如，通过发送邮件提醒学生加入课程学习。

设计选择：该视图经过迭代设计。最初系统设计使用堆叠柱状图，堆叠柱状图可以很好的展示不同时期人数的占比。然而，它很难显示学生在每周之间的状态过渡。桑基图没有这样的限制，最后选定使用该方案。

交互：当光标移动到某一子组上时，该子组的过渡情况被高亮显示，便于使用者查看感兴趣子组。

3.4.2 模式详情视图

如图 3.4（b）所示，该视图向用户展示了不同学习群体的典型学习序列模式。该视图中垂直列出算法识别的所有顺序模式（T2 和 T3），每个模式代表一组原始学习序列。如图 3.5 所示，对于每个模式，我们从左到右布局事件，并将它们显示为矩形。矩形顶部的颜色条编码了事件的所在的周，内部的条形图从左到右依次代表该模式所代表的一组学习序列的多个属性，包括视频观看总时长，视频暂停次数，跳转次数和访问视频相关信息次数。用户可以直观的对比不同事件在各个属性的区别。

由于每个模式代表了一组相似的序列，因此模式中除了显示按顺序排列的事件外，还需显示更正部分中的编辑次数。我们利用放置在相邻事件之间或模式的开始/结束处的倒三角符号表示序列转换为模式的编辑次数，它们的大小与在相应位置的插入数成正比，相应位置的插入数由该模式中所有相似序列累积得出。

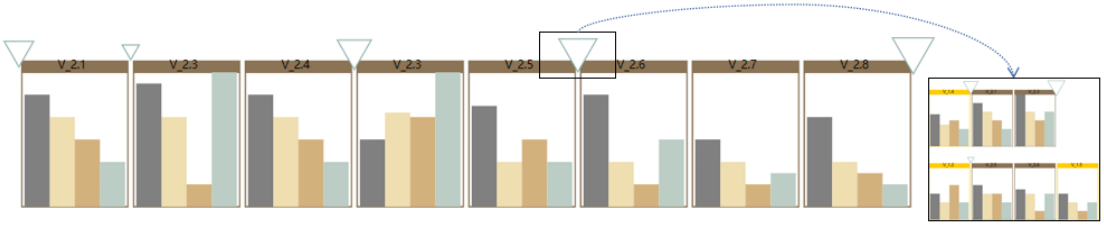

图 3.5 模式详情视图

考虑到用户教师通常不具有计算机领域知识，本文使用简单的可视编码设

计该视图，从而易于用户快速理解主要模式。该视图利用三角形的大小直观地表明了信息丢失的数量，这种设计有助于使用者识别具有高/低相似性的簇，从而引导他们更详细地探索数据。例如，用户点击相应三角形，将弹出缺失部分序列的常见模式。

3.4.3 个体学习序列详情视图

用户从整个学生群体深入探索到一个特定群体之后，下一个探索层次是个体视图(T4)。由于之前的分析都集中在群体层面，为了减轻用户在上下文之间转换的认知负担，本文将个体序列视图设计在下方一个单独的弹出窗口中。在该窗口中，针对个体水平的探索提供了两种不同数据粒度的展示。宏观层面上为了让用户一次看到更长的序列，系统提供聚合模式，将课程安排上为同一周的事件压缩为带颜色的条形图，不同颜色代表了事件所在的周，用户可以快速定位感兴趣的片段。微观层面中展示了学生学习的顺序，如图 3.6 不同形状代表了不同类型的事件，圆圈代表观看视频，三角形代表回答问题，矩形表示访问论坛。同时展示了更细粒度的信息，用一个饼状图来展示事件的不同属性，如视频观看行为即包含学生观看视频时长，视频暂停次数，跳转次数等，颜色编码与模式详情图中设计一致。用户通过该视图可以逐级地探索学生的详细活动。另外，在序列下方添加折线图方便用户比较相应操作在不同视频的次数。通过点击右上角不同颜色图标，来展示相应的操作情况。

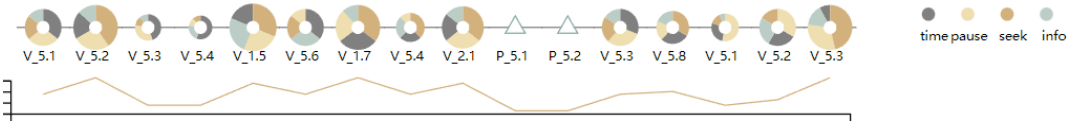

图 3.6 个体序列编码图

3.4.4 日历视图

为了方便用户比较不同学生的学习参与度(T5)，如上图 3.4（d）所示，本工作设计日历视图来展示学生在整个课程周的学习情况。日历被设计为一个 7*W 的矩阵，其中 W 代表周数，日历中每一格代表一天，颜色编码为学生参与