

MOOC 学生情感与学习效果关系分析

赵卫东 朱荣斌 李欣迪

复旦大学 软件学院 上海 200433

摘要 已有的研究证明流程挖掘和情感分析在教育数据挖掘领域中的重要性,然而两个领域的结合研究成果很少。本文提出基于约束聚类的模糊挖掘算法,在此基础上对学习流程模型进行挖掘。并通过可视化分析,将学习流程模型与情感分类进行结合,最终得到情感类别与学习效果的关系,为课程改进提供指导。

关键词 在线课程, 流程挖掘, 情感分析, 学习效果, 约束聚类

The Relationship Between MOOC Student Emotion and Learning Effects

Zhao Weidong Zhu Rongbin Li Xindi

Software School, Fudan University, 200433

Abstract—Existing studies have proved the importance of process mining and emotion analysis in the field of educational data mining, but there are few research results on the combination of the two fields. This paper proposes a fuzzy mining algorithm based on constraint clustering, and mines the learning process model on this basis. Through visual analysis, the learning process model is combined with emotion classification, and finally the relationship between emotion category and learning effects is obtained, which provides guidance for curriculum improvement.

Key words—MOOC, process mining, emotion analysis, learning effects, constraint clustering

1 引言

在互联网的快速发展下,以MOOC为代表的在线教育平台逐渐兴起,新的教育方式推动了教育行业的变革,如何有效地理解学生的学习状态和行为动机,成为一个重要问题。对在线教育平台的数据进行挖掘和分析,可以提高学生学习效率和学习体验。教育情感挖掘能够帮助理解学生的学习行为动机,学习流程挖掘能够深刻理解学生的学习行为流程,本文将教育情感挖掘和流程挖掘结合研究。

学生在学习课程中其情感扮演着重要角色。Goleman指出,合格的老师应具备分析学生各个阶段不同情感的能力,并允许学生做出适当的反应来避免学习过程中负面影响^[1]。以往的教育数据挖掘侧重点主要在于课程结束后的情感分析,或者是对系统日志进行流程挖掘改进学习流程,但是将两者结合的研究较少。

Munezero等在2013年提出了一种能够分析学生学习日志中的情感波动,并将情感变化可视化输出的系统,系统能够更深层次的分析学生的情感变化^[2]。Cairns等在2015年从课程日志中分析学生、课程章节、课程提供方三者之间的潜在关系,基于关键绩效指标提出了一种教育流程聚类分解方法^[3]。Etinger

等在2018年为了分析以及改进课程流程,对课程日志进行了流程挖掘,为了挖掘出在课程数据集和事件日志背后的隐藏信息^[4]。黄焕在2017年以学习者情感模型的构建为中心,重点研究了基于微博分析的学习者情感建模技术^[5]。刘智等在2018年采用情绪词典的特征匹配和情绪密度计算方法来研究学习者情绪特征以及群体情绪的演化趋势^[6]。然而目前这些算法大部分往往都局限于对课程结束后的反馈进行分析,缺乏对情感时间维度上变化的分析。

本文应用基于约束聚类的模糊流程挖掘算法,分析学生的在线课程学习日志数据,得到学生在课堂中的浏览行为流程模型。进一步地,通过浏览日志流程挖掘和论坛评论文本情感分类挖掘结合,将情感分类结果在流程模型上进行划分,展示学生的情感状态的时序性,并可视化不同情感状态流程图。对不同成绩的学生聚类,比较不同群体之间情感变化的差异性,研究不同情感与学习状态之间的关联性。

2 基于约束聚类的模糊挖掘算法

网上教育平台的学生在学习过程中,学习轨迹会呈现一定的时序性和流程。通常会反复地观看曾经学习过的视频,导致流程中出现复杂的环状结构。使用传统的流程挖掘技术时就会出现混乱复杂的模型。目

前的模糊流程挖掘算法在简化模型的时候,会根据节点和轨迹的重要程度以及相关性进行聚类 and 抽象,会受到重复出现的节点影响,没有考虑活动的语义和控制流一致性,所以简化后的模型和实际模型差距较大,不能正确反映出学生在学习过程中的学习流程^[7]。

本文针对网上教育平台的学习流程特征,在传统模糊流程挖掘算法的基础上,提出基于约束聚类的模糊挖掘算法对学习流程进行挖掘,通过学习活动的语义和控制流一致性来聚合流程节点,以提升最终模型的准确度。模糊挖掘在初始化时学习模型时,用流程模型中的节点表示日志中的事件(活动),通过添加有向边来表示事件之间的顺序依赖关系。在得到初始化流程模型后,通过冲突处理、边过滤、聚合与抽象等三个步骤对模型进行简化。

(1) 冲突处理

节点 A 和 B 被双向边所连接时需要进行冲突处理,有三种情况: A 和 B 之间出现了反向的异常流程,根据网上教育平台的日志特性,可知这种情况不会出现。若双向边对应的是同一用户 ID,说明 A 和 B 形成了环,否则说明 A 和 B 之间的双向边是同时出现的,不形成环。

(2) 边过滤

边过滤是指在流程模型中去除权重较低的边。在冲突处理中只是把冲突的边进行分类,并没有去除,因此在边过滤中,对一些重要性低于阈值的边需要进行去除。文献 9 提出边的效用计算方法 $util(A, B) = ur \cdot sig(A, B) + (1 - ur) \cdot cor(A, B)$, 其中 $ur \in [0, 1]$ 是可配置的效用比率, $sig(A, B)$ 表示节点 A 和 B 组成边的重要性, $cor(A, B)$ 表示节点 A 和 B 的相关性。

(3) 聚合与抽象

聚合权值较低且相互之间具有联系的节点,抽象权值较低且孤立的点将其去除。采用基于约束 K-Means 来替代传统模糊挖掘中的聚合与抽象。模型抽象时,考虑活动的语义和控制流一致性选择初始簇和约束条件。

2.1 约束 K-Means 聚类初始簇选择

Smirnov 等使用 RPST (Refined Process Structure tree) 的规范组件来优化业务流程模型 BPMN 的 K-Means 聚类,较好地保持了流程模型的控制流一致性[7, 8]。在 RPST 中,属于同一子流程的活动通常属于相同的规范组件。对于流程模型 P , 若 T 是 RPST, 按以下顺序选取 k 个初始簇组成初始簇集合 $S = (S_1, S_2, \dots, S_k)$:

如果 RPST 中规范组件 C 由多个活动组成, 则将 C

加入 S , 当 $i > k$ 或 RPST 中所有规范组件都已加入 S 时结束。如果规范组件 C 的数量大于 k , 则随机选取 k 个规范组件加入 S 作为初始簇集合。

如果 $i < k$, 则从 RPST 每层的叶结点中随机选择一个活动加入 S , 且选择的活动与 S 中其余节点不在同一层, 当 $i > k$ 或者 RPST 中每层都有叶结点加入 S 时结束。

如果 $i < k$, 则从模型中随机选择一个活动加入 S , 且选择的活动不与 S 中其余活动相邻, 当 $i > k$ 或 RPST 中没有此活动时结束。

如果 $i < k$, 随机选择一个活动加入 S , 当 $i > k$ 时结束。

2.2 语义距离和控制流一致性的约束函数

本文在计算活动之间的业务语义距离时,采用活动记录的方法来表示活动,记录通过合并活动的描述生成,可以用向量表示。

$A = \{a_1, a_2, \dots, a_n\}$ 表示流程模型 P 中的活动集合。 $D = \{d_1, d_2, \dots, d_n\}$ 为对应的活动记录集合, $S = \{S_1, S_2, \dots, S_m\}$ 表示约束 K-Means 聚类的簇集合, $\mu = \{\mu_1, \mu_2, \dots, \mu_m\}$ 为对应的簇中心。定义学习记录表示的两个向量 \vec{v}_{d_1} 和 \vec{v}_{d_2} 之间距离为

$$dist(d_1, d_2) = 1 - \cos(\vec{v}_{d_1}, \vec{v}_{d_2}) = 1 - \frac{\vec{v}_{d_1} \cdot \vec{v}_{d_2}}{|\vec{v}_{d_1}| |\vec{v}_{d_2}|} \quad (1)$$

为了对控制流顺序冲突进行约束,定义 a_k 与 S 之间的冲突值为

$$f(S, a_k) = \frac{1}{|S|(|S| - 1)} \sum_{\substack{a_i, a_j \in S \\ 1 \leq i < j \leq |S|}} W(r(a_i, a_k), r(a_j, a_k)) \quad (2)$$

式中 S 为 A 的子集, $r(a_1, a_2)$ 表示活动 a_1, a_2 的关系, W 表示冲突权值计算矩阵。

$conflicts(S)$ 为活动 a_k 分配到簇 S_i 导致的控制流顺序冲突。

$$conflicts(S) = \frac{1}{|A \setminus S|} \sum_{a_k \in A \setminus S} f(S, a_k) \quad (3)$$

将活动 a_k 分配到簇 S_i 时,需要使目标函数 $objective(S_i, a_k)$ 取最小值:

$$\begin{aligned} objective(S_i, a_k) &= \omega_1 dist(d_k, \mu_i) \\ &\quad + \omega_2 conflicts(S_i \cup \{a_k\}) \\ 0 &\leq \omega_1, \omega_2 \leq 1 \end{aligned} \quad (4)$$

式中 ω_1 和 ω_2 表示基于约束聚类划分子流程时,

语义距离和控制流冲突的权重，可以通过实验确定合适的值。这里选取 $\omega_1 = 0.8$ ， $\omega_2 = 0.2$ 。

模糊挖掘算法能够对模型进行冲突处理和边过滤，解决了模型的控制流顺序冲突问题。本文提出基于约束聚类的模糊挖掘算法，在传统的模糊挖掘算法上进行了改进，在聚合和抽象阶段，使用约束K-Means聚类算法进行聚合，解决了模糊挖掘中聚合时衡量标准单一的问题。使用基于约束聚类的模糊挖掘算法能够得到简洁的流程模型，且同时保留了活动的语义和控制

制流一致性。

下面通过实际数据分析实现学习流程模型挖掘。数据来源于某在线教育平台的网课数据，以2013年9月到2017年12月，总计9个学期，把收集的数据进行预处理，以符合流程挖掘算法。通过筛选，得到了966位学生的55208条有效日志数据，包含136个不同的学习资源，其中表1是事件日志片段。

表1 《公共安全管理学》学生学习的日志片段

ID	USERID	TOPICID	VIEWSTARTTIME	VIEWENDTIME
k4cl63iig	2310770104	161vabcnh59ljsekzz	2017/9/3 12:06:34	2017/9/3 13:11:12
2egbp0klq	2310770104	j6pvabcnvpikamt9ka	2017/9/6 19:21:22	2017/9/6 19:23:54
svatlalrw	2310770104	fqtvabcnlnmzx6iep	2017/9/7 21:33:51	2017/9/7 22:15:11
sh4pelpsw	2310770104	cgj6abunxbiktpaqh	2017/9/7 21:34:55	2017/9/7 22:15:12
sju4x9nea	2310770104	fqtvabcnlnmzx6iep	2017/9/12 13:57:42	2017/9/12 15:08:29
vwvbhbtq	2310770104	nca5ad6nk7tpgi-4y	2017/9/12 15:09:01	2017/9/12 15:47:50

本文首先采用传统的模糊挖掘算法得到学习流程的初始模型。挖掘结果如图1(1)所示。可以发现学生的学习流程具有规律，由于学生的反复观看和回看，形成了许多的环状结构。但初始模型节点较多，规律不是很明显，较为杂乱，通过上述本文提出的基于约束聚类的模糊挖掘算法，聚合一些重要性不高的节点，得到模型如图1(2)所示。可以明显看出课程结构，头部节点表示课程第一章节，尾部节点表示课程最后章节，部分学生对其进行多次观看，说明开始和结束的章节较为重要，一部分同学没有按照正常流程观看课程，说明部分章节前后相关性较弱，教师应对其进行调整，使章节之间更连贯。

3 学生情感与学习效果的关系

流程挖掘的模型可以直观的表现出学生的学习行为流程，但情感才是隐藏在学生学习行为背后的真正动机，学生的情感变化在整个学习过程中起着至关重要的作用。

将学习流程模型和学生情感相结合进行可视化分析，可以更好地分析学生的学习行为，帮助在线教育平台的课程进行改进及优化。还是以《公共安全管理学》课程为例，本文将可视化展示基于学习流程的学生情感变化情况，研究不同学生群体之间的学习行为情感变化。将学生在学习过程中的情感分为四种（积

极、困惑、焦虑、不满），基于学习流程模型绘制情感热力图来展示学生情感流程的不同情况，探究不同学生之间的情感差异。以当前情感的最小相对强度到最大相对情感强度为绘制热力图的着色范围。

图2表示积极情感流程变化，其中蓝色的深浅表示积极情感的程度，蓝色越深表示情绪越积极，可以分析出学生在学习过程中情绪是积极向上的，说明对该课程较为满意，实际背景是学生对该课程评分较高，底下评论为“该课程很棒，学起来简单，老师很认真负责”这种类似评论。

图3表示困惑情感流程变化，绿色越深表示困惑程度越高，说明学生在该学习单元困惑情绪为主导情绪，可以看出学生在课程开始和结束时困惑情绪较高，说明学生在开始时不了解课程内容，结束时不知道课程如何考核。实际背景中课程评论大部分是“这个课程是讲什么的？”、“这个课程怎么考核？”、“老师会划重点嘛？”这种留言。

图4表示焦虑情感流程变化，黄色越深表示越焦虑，说明学生在该学习单元较为焦虑，可以分析出在课程开始时、部分中间课程单元以及课程结束时学生最为焦虑，中间黄色单元说明该课程单元难度较大或者晦涩难懂。实际背景评论区“我怎么听不懂老师在说什么”、“作业怎么这么多”、“老师有没有这节课

课的参考材料”。

图 5 表示不满情感流程变化，红色越深表示对课程越不满，图中红色节点中实际评论区“怎么卡了”、“怎么搞得，第 4 分钟突然就没声音了”、“PPT 字小的可怜”，可以知道在红色较深的课程节点可能有用户体验问题，开始和结束时不满可能是因为对教学系统的不满意。

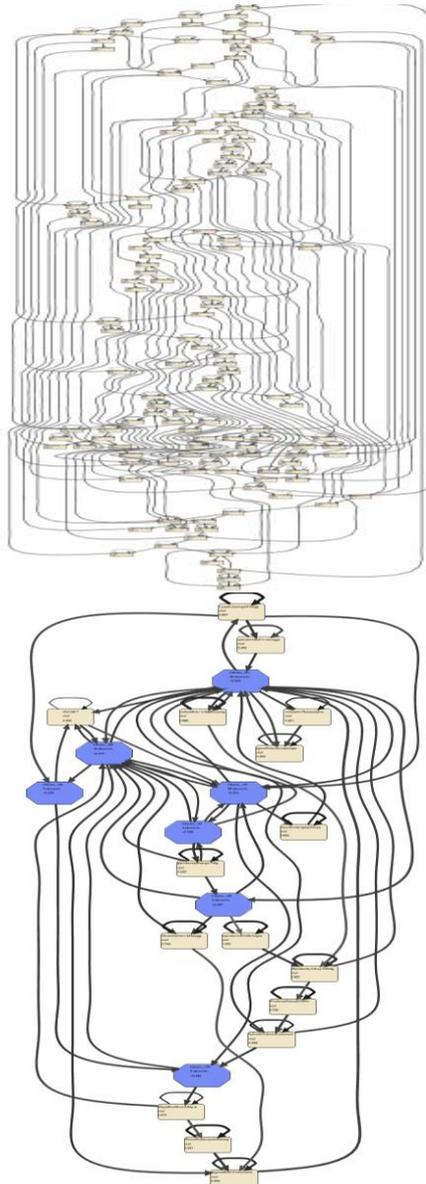


图 1 《公共安全管理学》学习流程模型比较

整体上，在该课程中学生的积极情绪为主导情绪，说明学生对课程整体较为满意。课程开始和结束时学生的负面情绪较多，说明课程开始时学生对课程相关事宜不是很了解，对新课程较为困惑，之前的问题累积起来以及期末考试导致学生在课程结束时负面情绪增多。因此，对于教师可以在课程开始之前通过一定的方法来帮助学生更全面的了解课程相关内容，在部

分课程中改变教学方式或者增加课时来帮助学生更好的理解较为困难的章节，在课程末期，及时解决学生的问题，帮助学生复习来减少学生的负面情绪。

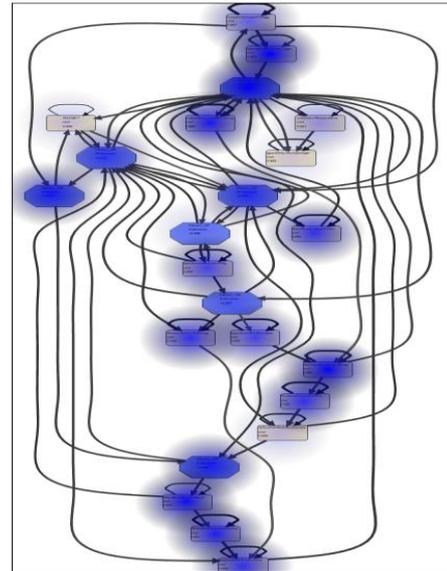


图 2 积极情感流程变化

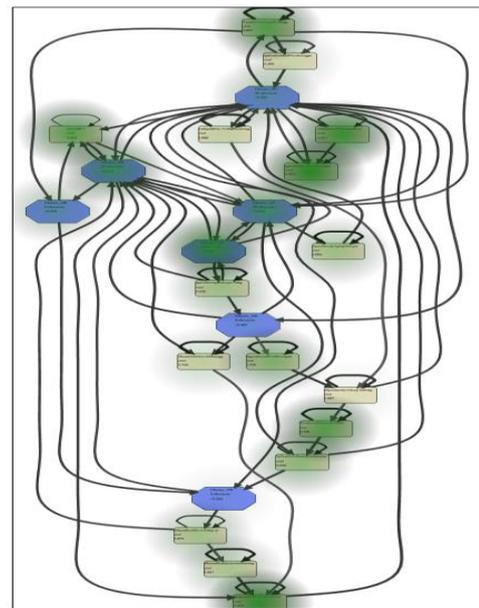


图 3 困感情感流程变化

为了发现学生的成绩和学生学习过程中情感状态关系，按照时间维度绘制出不同成绩学生之间的情感状况。横轴表示教学周，纵轴表示相对情感强度。其中图 6 表示积极情绪差异，图 7 表示困感情绪差异，图 8 表示焦虑情绪差异，图 9 表示不满情绪差异。例如图 7 可以分析出不及格的学生和其他学生之间的焦虑情绪差异较明显，一直处于较高的状态，越接近期末不及格学生的焦虑情绪越高，说明不及格学生临近期末因为害怕考核从而焦虑情绪增多，教师应对其多加关注。

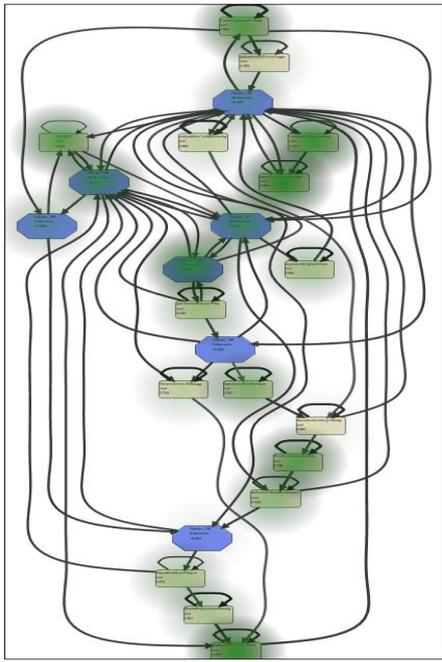


图 4 焦虑情感流程变化

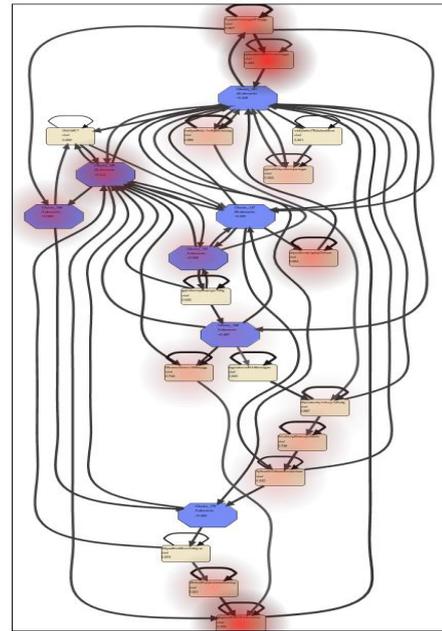


图 5 不满情感流程变化

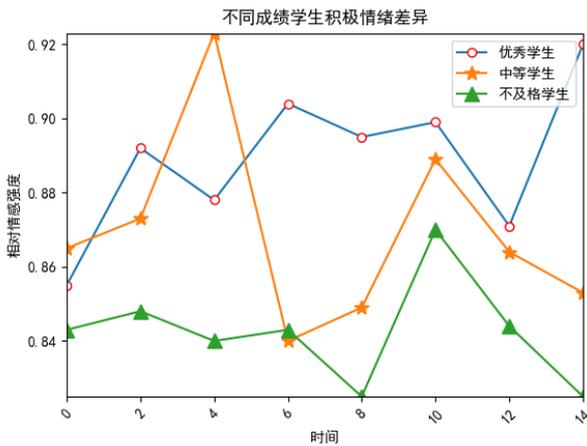


图 6 不同成绩学生积极情绪差异

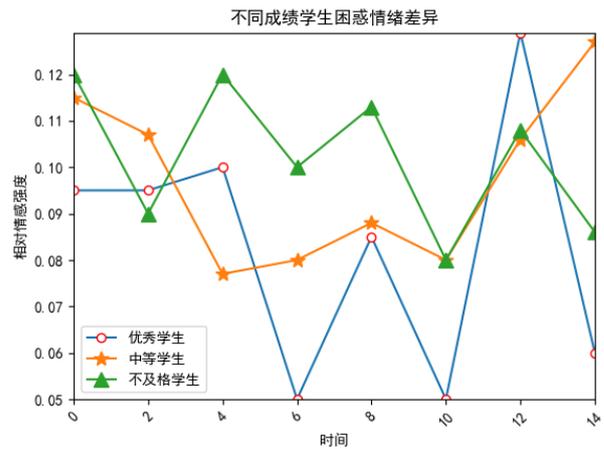


图 7 不同成绩学生困感情绪差异

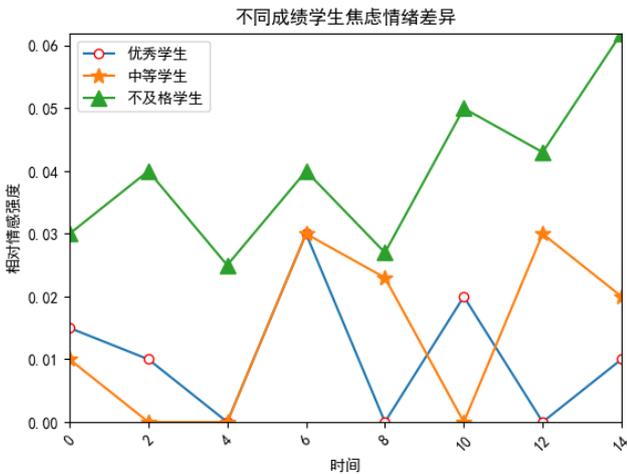


图 8 不同成绩学生焦虑情绪差异

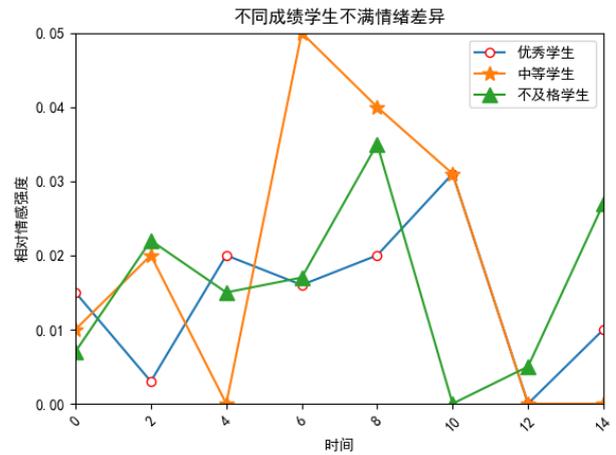


图 9 不同成绩学生不满情绪差异

可以发现积极情感与好成绩之间具有相关性,因为积极的态度可以让学生喜爱这门课程,从而想办法主动学习,主动学习后成绩明显提升积极情绪继续增高,达到良性循环。从消极情绪来看优秀学生中占比较低,而不及格学生消极情绪占比较高,因为优秀学生大部分都能自主解决学习中遇到的问题,而不及格学生遇到问题无法及时解决,导致消极情绪积累起来,期末出现高峰状态。教师在课程中应及时关注出现困惑和焦虑情绪的学生,及时解决学生提出的问题,从中找出原因,并改进课程,提高学生的积极情绪。

对其他几门在线课程的学习日志进行分析,发现了类似《公共安全管理学》课程的分析结果。通过上述分析,能够快速定位学生在学习过程中遇到问题的具体位置。任课教师可以找到学习流程中的问题,进一步确认是否由于教学方法或者教学内容不合理而引起学生的情感异常,及时对学生的进行学习进行监控和干预。

4 学习干预与教学指导

根据实验结果,通过挖掘出的流程图,我们可以快速发现并定位课程中使学生产生困扰的章节,发现问题章节后,教师 and 平台管理人员可以依据课程流程模型的相对情感热力图,来进一步分析课程章节可能出现的问题,研究是否因为课程内容或者教学流程设置不当而导致学生的情绪异常。对于相关教育研究人员,可以通过分析课程的情感热力图,来改进课程内容和课程章节顺序,来进一步提高学生的学习体验。

通过对不同成绩的学生情感差异分析,积极情感对最终取得好成绩具有明显的正向作用,而困惑、焦虑以及不满情绪明显与最终的学习成果负相关。所以教育相关人员应在课堂中,需要对那些产生大量负面情绪的学生加以关注,及时发现问题并分析原因对相关学生进行干预和疏导,帮助学生建立积极情绪,提高学生在学习体验。

教师和管理者可以通过基于学习流程模型的相对情感强度热力图,找到学习流程中的问题。另外,对于教师来说,能够基于相对情感热力图,优化教学内容和流程的设置,提升学生整体的学习体验和学习效果。

5 结束语

对于不同学习表现的学生群体情感进行差异性分析得出,积极情绪对于最终的学习表现有着明显的正向促进作用,而困惑和焦虑情绪与最终的学习表现负相关。因此教师在教学过程中,需要特别关注学生在学习时产生的困惑情绪和焦虑情绪,及时定位负面情绪产生的原因,及早对出现学习情感异常的学生进行干预和指导,帮助学生摆脱负面情绪,提升学习表现。

本文对在线教育平台中课程的事件日志进行收集与处理,使用约束聚类的模糊流程挖掘算法,挖掘学习流程模型,结合情感分类绘制不同情感流程热力图,以及利用可视化分析不同情感和成绩之间的关系,发现不同情感和学生学习成绩之间的关系,研究结果有利于在线教育平台对课程进行优化。

参考文献

- [1] Goleman D. Emotional intelligence[M]. Bantam, 2006.
- [2] Munezero M, Montero C S, Mozgovoy M, et al. Exploiting sentiment analysis to track emotions in students' learning diaries[C]. Proceedings of the 13th Koli Calling International Conference on Computing Education Research. ACM, 2013: 145-152.
- [3] Cairns A H, Gueni B, Fhima M, et al. Process mining in the education domain[J]. International Journal on Advances in Intelligent Systems, 2015, 8(1): 219-232
- [4] Etinger D, Orehovački T, Babić S. Applying Process Mining Techniques to Learning Management Systems for Educational Process Model Discovery and Analysis[C]. International Conference on Intelligent Human Systems Integration. Springer, Cham, 2018: 420-425.
- [5] 黄焕. 面向 e-Learning 的学习者情感建模及应用研究[D]. 武汉: 华中师范大学硕士学位论文, 2014.
- [6] 刘智, 杨重阳, 彭晔, 等. SPOC 论坛互动中学习者情绪特征及其与学习效果的关系研究[J]. 中国电化教育, 2018(4): 102-110
- [7] Günther C W, Van Der Aalst W M P. Fuzzy mining-adaptive process simplification based on multi-perspective metrics [C]. International conference on business process management. Springer, Berlin, Heidelberg, 2007: 328-343
- [8] Li J, Liu D, Yang B. Process mining: Extending α -algorithm to mine duplicate tasks in process logs[M]. Advances in Web and Network Technologies, and Information Management. Springer, Berlin, Heidelberg, 2007: 396-407.