

基于在线评论的课程思政 学生关注点研究

喇 磊¹, 邵立伟²

¹对外经济贸易大学信息学院, 北京

²北京理工大学珠海学院科技处, 广东 珠海

收稿日期: 2023年7月10日; 录用日期: 2023年8月10日; 发布日期: 2023年8月17日

摘 要

受众的关注点是教学设计与教学效果评估中的关键问题。教育研究领域关于思政教学中学生关注点的研究较少, 且基于问卷的研究方式受到成本、数据量与样本偏差的限制, 效果有限。文章提出一种基于深度预训练模型的课程思政评论话题识别方法。基于自主爬取的相关评论文本, 首先通过Bert算法进行特征表示和降维; 然后针对经典的分层隐树分析方法处理词共现问题的区分度不足及高时间开销等问题, 提出了基于阈值的词共现预修剪的分层隐树模型; 最后基于提出的方法, 实现了数据获取、预处理、话题挖掘及结果展示的全流程平台。实验结果表明, 提出算法具有更好的话题检测性能。研究发现, 学生对于带有思政内容的课程的关注点主要仍在于课程教学本身, 而不满主要对于思政点融入比较机械等情况。

关键词

在线评论, 话题识别, 预训练, 分层隐树模型

Research on the Concerns of Students in Curriculums with Ideological and Political Content Based on Online Reviews

Lei La¹, Liwei Shao²

¹School of Information Technology & Management, University of International Business and Economics, Beijing

²Department of Science and Technology, Beijing Institute of Technology Zhuhai, Zhuhai Guangdong

Received: Jul. 10th, 2023; accepted: Aug. 10th, 2023; published: Aug. 17th, 2023

文章引用: 喇磊, 邵立伟. 基于在线评论的课程思政学生关注点研究[J]. 计算机科学与应用, 2023, 13(8): 1547-1556.
DOI: 10.12677/csa.2023.138153

Abstract

The focus of the audience is a key issue in teaching design and evaluation of teaching effectiveness. There is a lack of research in the field of educational research on the concerns of students in ideological and political education, and the questionnaire based research method is limited by cost, data volume, and sample bias, resulting in limited effectiveness. The article proposes a topic recognition method for course ideological and political comments based on a deep pre training model. Based on self crawling of relevant comment texts, first feature representation and dimensionality reduction are performed using the Bert algorithm; then, in view of the lack of differentiation and high time cost of the classical hierarchical hidden tree analysis method to deal with the problem of word co-occurrence, a hierarchical hidden tree model based on threshold pre pruning of word co-occurrence is proposed; finally, based on the proposed method, a full process platform for data acquisition, preprocessing, topic mining, and result display was implemented. The experimental results show that the proposed algorithm has better topic detection performance. Research has found that students' focus on courses with ideological and political content mainly lies in the teaching itself, while their dissatisfaction mainly lies in the rigid integration of ideological and political points.

Keywords

Online Reviews, Topic Detection, Pre-Training, Hierarchical Latent Tree Model

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

课程思政是构建全员、全程、全课程育人格局,把“立德树人”作为教育的根本任务的关键举措。近年来,课程思政建设收到越来越多研究者的关注,涌现了一批相关成果。目前关于课程思政的研究主要集中在内容设计[1]、教学方法[2]、体系构建[3]和教学创新[4]等方面。虽然随着课程思政建设的深入,近年来相关研究者逐渐开始重视对于课程思政教学效果的评价[5][6],但以学生视角研究课程思政背景下学生的关注点与满意度等因素的研究依然稀少。

针对学生体验的课程思政教学效果研究并不多,且这些研究大多基于传统手段,例如通过学生发放问卷进行课程思政满意度研究[7],通过访谈研究学生课程思政学习体验[8]等。基于传统手段的研究,获取数据的人力和经济成本大,数据源和数据量受限易导致统计偏差[9],且获取数据的时间周期长,影响结论的时效性[10]。

实际上,互联网已成为人们发表观点、传播经验、表达诉求和倾诉不满的主要阵地。这些海量的观点、经验、诉求和不满中蕴含着大量舆情信息,基于这些信息和机器学习手段的客户研究已成为各行业的热门课题[7][8][9],甚至催生了前沿的技术方法[10]。在教学研究领域,基于互联网用户生成内容(User Generated Content, UGC)和人工智能手段的研究近年来也成为热点,特别是在学生舆情和体验方面,相关成果颇丰[11][12][13]。然而,基于在线数据与机器学习方法的课程思政学生体验研究是缺少的。此外,传统的机器学习方法应用于互联网用户评论短文本的话题识别还存在话题区分度不足[14]、性能与时间开

销平衡[15]等问题, 制约了这一类技术在教育教学研究领域发挥效果。

基于以上原因, 针对在线学习评论数据的特征, 本文提出了一种融合预训练模型与层次聚类的话题识别方法, 基于该方法分析学生对含有课程思政内容的线上课程的评论主要包括的话题是什么, 为有针对性地提升课程思政教育教学水平提供依据, 并促进相关研究的发展。

2. 文献回顾

2.1. 在线评论话题识别与主题分析

针对在线评论的分析挖掘是人工智能与各学科交叉的热门问题, 其中对在线评论数据集运用分类、聚类或主题模型等手段进行话题发现是一个专门的方向, 在这一领域, 针对不同的行业与应用场景, 国内外学者做出了众多有特色的成果。Wang 等学者开发了一种基于边缘聚类的多维网络模型, 该模型对传统的概率模型进行改进, 不仅强调了概率分布的统计信息和核心关键词特征的共存率, 而且考虑了不同主题发布者之间的关系交互和情感属性。实验结果表明, 该方法具有较理想的话题发现能力[16]。然而, 该方法面对综合在线平台, 平台上的用户留言并不局限于某一领域, 话题差异度较大, 这与特定领域在线评论的话题识别场景不同。Pang 等学者提出了一种耦合泊松反卷积主题识别方法, 该文首先将文本特征和视觉特征编码成 k-近邻混合相似图, 然后基于泊松反卷积通过对主题的描述来细化网页之间的相似性。该方法可实现一体化的话题检测和话题描述[17]。李慧等学者提出利用词项 H 指数筛选出热点词项, 然后利用 BTM 建模和 VSM 建模的结果相融合计算文本相似度, 再利用 k-means 聚类算法发现微博热点话题的方法。该方法需要人工设置阈值以筛选对热点贡献大的词[18]。

由于教学平台中在线留言的内容大多围绕学习这一主题, 文本相似度高于综合社区, 当在线留言隶属于同一或相近领域时, 传统的基于聚类、基于主题模型的方法进行话题识别之后的结果之间缺乏差异度, 难以识别出有意义的话题[19]。在同领域主题近似文本的话题识别方面, Kim 等学者提出了一种结合 word2vec 与潜在语义分析(Latent Semantic Analysis, LSA)的主题检测模型, 通过 word2vec 进行词特征降维, 在此基础上通过 LSA 进行话题聚类可提高类簇间的差异度, 该方法用于区块链相关话题中的子话题挖掘, 取得了不错的效果[20]。Chen 等学者提出了一种基于 word2vec 的非参数模型, 该模型基于 Mahalanobis 距离度量自动决定给定的文档是否属于现有主题。实验结果表明基于词嵌入的主题模型优于直接使用主题模型进行话题识别[21]。

2.2. 研究评述

虽然从学习平台的在线评论数据中寻找教学研究或教学评价等相关问题的研究日益兴盛, 但专门研究学生对含有课程思政内容的评论中的话题检测问题并不多。话题检测是机器学习领域一个热门的方向, 这一方向有文本层面的聚类、分类和主题模型以及词特征层面的基于信号处理的方法等多种技术路线[22]。然而这些传统的方法大多应用于互联网开放主题的话题挖掘, 应用于某一领域内主题相似度较高的文本集上的效果并不理想。最近的若干研究表明, 词嵌入等特征表示模型可以提升传统聚类方法在话题检测中的性能。本文将借鉴这一思路, 研究性能更好的预训练新模型与聚类算法的融合, 并针对在线学习领域词共现高发的特点对现有聚类算法进行改进, 从而实现更有效的国外游客负面评论的话题挖掘。

3. 核心方法设计

本研究将通过对比 Bilibili、中国大学慕课和知乎等平台上关于含有课程思政内容的课程评价进行基于聚类的话题检测与识别, 发现学生对含有思政内容的课程的关注点主要集中在哪些方面, 从而提高课程思政教育教学水平提供决策依据。因此, 本研究的重点与难点是如何从学习这一窄主题领域的在线短

文本中有效识别出有意义的话题。文本窄主题领域的挑战在于文本相似度高, 差异性小; 在线短文本的挑战在于语义信息有限。他们导致的共同结果是应用一般话题检测算法所识别的话题之间缺乏差异度, 难以具有实用价值。

3.1. 基于 Bert 的特征嵌入

词嵌入等预训练模型的初衷是为了更好地表示自然语言以便深度神经网络模型进行处理, 从而实现基于深度学习的自然语言处理。然而, 如文献回顾部分的分析, 近年来学者发现这一类特征表示方法也可以提高主题模型、聚类和分类等下游应用的效果, 原因是词嵌入等方法实现了特征降维并提高了特征的语义信息。本文将通过目前较先进的预训练模型 Bert 来进行特征嵌入, 实现降维并提升特征的语义丰度。

Bert 是一种基于深度神经网络的预训练模型。通过双向变换器的编码器, 该模型可以获得两个方向的信息。以往的预训练模型的结构受到单向语言模型的限制, 这也限制了模型的表达能力, 使其只能获得单向的上下文信息。Bert 使用掩码语言模型(Masked Language Model, MLM) [23]进行预训练, 并采用深度双向变换器组件来构建整个模型。双向转换器称为转换器编码器, 每个 token 都将附加到所有 token 上, 因此它最终生成了一个能够整合左右语境信息的深层双向语言表示。Bert 的变换器结构如图 1 所示。

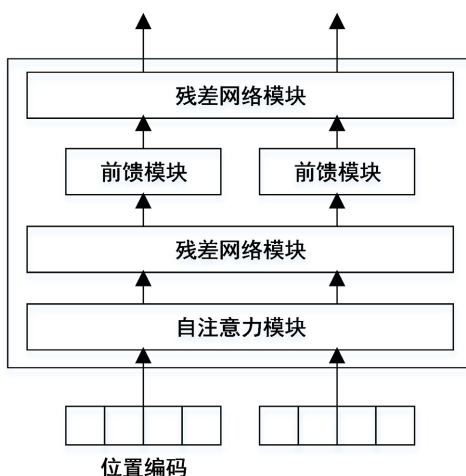


Figure 1. Transformer structure of Bert
图 1. Bert 的变换器结构

在 Bert 模型中, 多层自注意机制代替了传统的 RNN 和 CNN 神经网络, 有效地解决了自然语言处理中的长距离依存问题。在自注意机制的工作过程中, 首先利用多重自注意来连接编码器和解码器, 然后在编码器和解码器中使用自我注意来学习文本表示。在运算过程中, 当输入一个句子时, 句中的每一个单词都会受到注意, 并根据式(1)与其他单词一起计算。

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{dk}}\right)V \quad (1)$$

其中 K 、 V 和 Q 是编码器的输出和解码器中的 attention 输入。位置编码是为了解决自注意没有考虑到的语序问题。在 t 时刻输入时, 除了嵌入单词外, 还引入了位置向量。该向量与 t 相关, 采用词嵌入和位置向量相结合的方法作为模型的输入。在这种情况下, 如果两个单词出现在不同的位置, 虽然它们的词嵌入是相同的, 但是由于位置码的不同, 最终的向量是不同的。基于 Bert 的特征嵌入原理如图 2 所示。

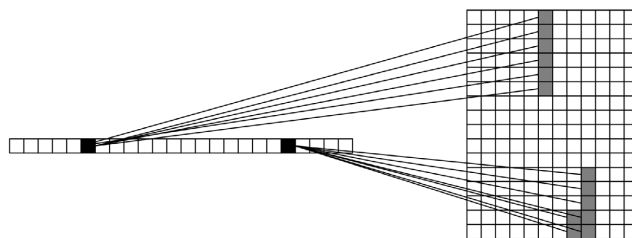


Figure 2. Feature embedding based on Bert
图 2. 基于 Bert 的特征嵌入

如图 2 所示, 通过 Bert 进行预训练能够实现类似 word2vec 的词嵌入的效果, 将语义相似的词映射到连续向量空间中的邻近点, 并可以考虑到词的顺序。相对于传统的独热编码, 经过变换, 词被映射到低维空间, 降低了特征集维度, 并提高了特征的语义丰度, 为增强下游应用中所识别出的话题之间的区分度奠定了基础。

3.2. 分层隐树模型与话题检测

隐狄利克雷分配(Latent Dirichlet Allocation, LDA)是近年来主题识别领域最经典的方法之一, LDA 及其变种版本往往通过隶属于某一主题的高频词汇来识别主题, 然而一个话题中的高频词也可能在其他话题中出现频率很高, 从而导致识别偏差。针对这一问题, Liu 等学者提出了分层隐树模型(Hierarchical Latent Tree Model, HLTM) [24] [25]。隐树模型实际上是一种无向树上的马尔可夫随机场, 叶节点代表观测变量, 内部节点代表潜在变量。从文档集中学习隐树模型的实例如图 3 所示。

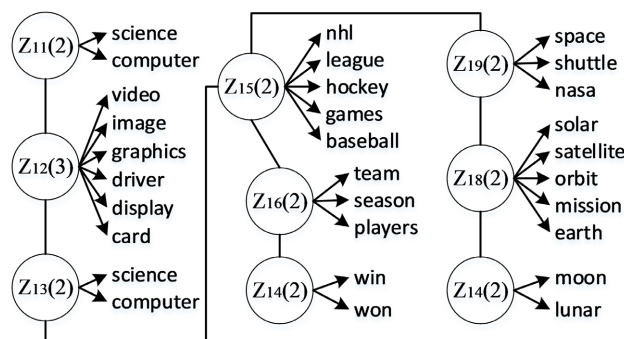


Figure 3. Example of learning hidden Tree model from documents
图 3. 从文档中学习隐树模型实例

图 3 中每列右侧的单词是二进制变量, 指示文档中是否存在这些单词。 Z_i 是潜在变量。它们是离散的, 它们的基数, 即状态数在括号中给出。为了技术上的方便, 我们通常以隐树模型的一个潜在节点为根来建立隐树结构, 并将其视为一个有向图模型, 如贝叶斯网络。然后所有的边都指向远离根的方向。模型的数值信息包括根的边缘分布和每个边的一个条件分布。与每条边相关的条件分布表征了边连接的两个节点之间的概率依赖性。所有这些分布的乘积定义了所有隐变量和观测变量的联合分布。

通常, 设有 n 个观测到的变量 $X = \{X_1, X_2, \dots, X_n\}$ 以及 m 个隐变量 $Z = \{Z_1, Z_2, \dots, Z_m\}$, 将变量 Y 的父节点记作 $pa(Y)$, 当 Y 为根节点时显然 $pa(Y)$ 为空。隐树模型中定义所有可观测和潜在变量的联合分布如式(2)所示。

$$P(X_1, \dots, X_n, Z_1, \dots, Z_m) = \prod_{Y \in X \cup Z} P(Y | pa(Y)) \quad (2)$$

基于隐树模型可以提出分层隐树模型, HLTM 的工作流程为首先分层发现词的共现模式; 其次通过递归发现模式的共现模式来构建层次结构; 然后从生成的层次结构中提取主题。在 HLTM 中, 单词被视为二进制变量, 表示文档中单词的存在与否。HLT M 的第一步是识别词的共现模式。这是通过将单词变量划分成簇来实现的, 这样每个簇中的变量是紧密相关的, 并且每个簇中变量之间的相关性可以用一个潜在变量正确地建模。与树的堂兄弟(sibling)结点类似, 图 3 中隶属于同一个隐主题 Z_i 的词簇称作堂兄弟簇, 每个堂兄弟簇中的单词在语义上是相关的, 并且倾向于共现。我们可以用一种贪婪算法实现用一个潜在变量建模簇之间的相关性, 其目标是找到具有最高贝叶斯信息准则(Bayesian Information Criterion, BIC)得分的潜在树模型, 它分为四个步骤: ① 将变量集划分为兄弟簇; ② 为每个同族簇引入一个潜在变量; ③ 将潜在变量连接起来形成树; ④ 基于全局考虑对模型进行优化。为了定义潜在的堂兄弟簇, 在此可以考虑每对变量在互信息方面的紧密程度。两个变量 X 和 Y 之间的互信息定义如式(3)。

$$I(X;Y) = \sum_{X,Y} P(X,Y) \log \frac{P(X,Y)}{P(X)P(Y)} \quad (3)$$

为了确定第一个同级堂兄弟簇, 需要维护一组变量构成工作集 S , 这些变量最初由具有最高互信息的一对变量组成, 其他变量逐个添加到集合中。在每个步骤中, 选择添加使 $\max_{Z \in S} I(X;Z)$ 最大化的变量 X 。在每一步扩展后, 进行贝叶斯统计检验, 以确定 S 中变量之间的相关性是否可以用一个单独的潜在变量进行适当的建模, 一旦测试为否, 膨胀停止。对于测试, 首先将原始数据集 D 投影到工作集 S 上, 以获得更小的数据集 D' 。然后从 D' 得到了仅包含 1 个潜在变量或不超过 2 个潜在变量的最优隐树模型 m_1 和 m_2 , 从而满足式(4)。

$$BIC(m_2|D') - BIC(m_1|D') \leq \delta \quad (4)$$

其中 δ 为阈值参数, 在确定第一个同级堂兄弟簇之后, 从数据集中删除集群中的变量, 并重复该过程以查找其他同级集群, 重复这一过程直到所有变量被分组到同级集群中。在确定了同级堂兄弟簇之后, 为每个同级簇引入一个潜在变量, 在学习过程中自动确定潜在变量的基数。此后将所有的潜在变量进一步连接起来形成树状结构。最后, 基于全局考虑对结构进行调整。如图 3 所示, 这一策略可以生成扁平的潜在树模型, 其中潜在变量捕捉单词共现的模式。模式本身可能同时发生, 这种更高层次的共现模式可以通过递归应用上述策略来发现, 这也恰是 HLT M 算法的设计目的。HLT M 模型的伪代码如算法 1 所示。

算法 1. 分层隐树模型分析算法(HLT M)

输入: 初始数据集 D_m , 贝叶斯阈值 δ , 隐变量最大层数 k

输出: HLT M 模型及话题

```

1. Begin
2.  $D \leftarrow D_m, Level \leftarrow 0, M_{whole} \leftarrow \phi$ 
3. while  $Level < k$  or  $D$  含多于两个变量 do
4.  $M \leftarrow \text{Bridge-Islands}(D, \delta)^{[25]}$ 
5. If  $M_{whole} = \phi$  then
6.  $M_{whole} \leftarrow M$ 
7. else
8.  $M_{whole} \leftarrow \text{MergeModel}(M_{whole}, M)$ 
9. end if
10.  $D' \leftarrow$  项目数据集( $M_{whole}, D_m$ )
11.  $D \leftarrow D', Level \leftarrow Level + 1$ 
12. endwhile
13. 输出不同层次的话题并返回 HLT M 模型
14. end

```

Algorithm 1. Analysis process of hierarchical latent tree model

算法 1. 分层隐树模型分析算法

如算法 1 所示, 在迭代计算词共现和模式共现的末尾, HLTM 构建了一个包含多个潜在变量的模型。该模型顶层由单个潜在变量或多个潜在变量组成, 它们以树状结构连接起来; 其他层次由多个潜在变量组成, 每一个潜在变量分别与上一层次的一个潜在变量和下一个层次的几个变量相连; 底层由单词变量组成。模型中的每个潜在变量代表文档的软分区, 其状态可以解释为主题。通过以上流程, 可实现完整的基于词共现与模式共现的文档主题识别。

3.3. 基于领域停用词的冗余共现预消除

由于本文所使用的文档集来源于在线学习社区, 文档主题都隶属于学习这一窄领域, 话题的相关度较大, 有些教学中的通用词对于领域内的话题识别实际上是无效的, 反而可能干扰话题检测的结果。

领域性是停用词的一大特征, 同一个词在某个领域是停用词, 在另一个领域未必是停用词。由于缺乏开放的学习领域停词表, 因此我们通过自行爬取 34 万篇相关文档构建学习领域停词表。传统的停止词模型只根据累积频率提取停止词, 而未考虑词在文档中的分布情况。作为改进, 我们根据词的总体分布来提取停止词, 由于均值和方差是分布的两个重要度量, 因此我们基于这两个标准来提取停止词。对于每个单词 w_j , 计算其在文档 D_i 中的频率 $f_{i,j}$ 。为了规范文档长度, 计算文档 D_i 中单词 w_j 的概率 $P_{i,j}$, 因此可得每个词在一篇文章中的平均概率如式(5)。

$$P_{\text{mean}}(w_j) = \frac{\sum_{1 \leq i \leq N} P_{i,j}}{N} \quad (5)$$

停止词不仅需要具有较高的出现概率 P_{mean} , 还需要具有稳定的分布, 因此计算词的平均方差 V_{mean} 如式(6)。

$$V_{\text{mean}}(w_j) = \frac{\sum_{1 \leq i \leq N} (P_{i,j} - \bar{P}_{i,j})^2}{N} \quad (6)$$

一个词成为停止词的概率与概率均值成正比, 但与概率方差成反比。因此可得式(7)。

$$\theta(w_j) = \frac{P_{\text{mean}}(w_j)}{V_{\text{mean}}(w_j)} \quad (7)$$

根据 θ 值可构建合适的停词表。将自建的旅游领域停词表引入话题检测模型, 先进行双重去停词(通用停词与领域停词), 再通过 Bert 实现特征表示, 将降维后的数据集供 HLTM 进行话题检测, 从而构成了本项目核心方法的完整流程。

4. 数据获取与平台构建

为了验证本文提出方法在教学话题挖掘中的有效性, 我们构建了一个包含数据爬取、预处理、分析和展示的全流程平台, 将本文开发的核心方法作为数据分析模块嵌入平台。

在数据爬取方面, 针对目标网站, 我们应用 selenium + ChromeDriver 方法和 request 解析方法开发了定制化的爬虫, 并通过分布式和虚拟 IP 技术应对网站的封禁措施。目标数据源为 Bilibili、中国大学慕课和知乎平台上包含思政内容的课程评论、弹幕和话题。

在数据预处理方面, 我们通过正则表达式去除了文档中的噪声数据, 并通过设置阈值剔除了长度过短的、缺少语义信息的文本, 此外, 通过开源关键字过滤工具 keywordspy, 我们初步过滤了文档集中的广告文本。

经过预处理后的文本通第 2 节描述的核心方法进行话题挖掘, 分析结果通过平台的展示模块呈现给

用户。在展示模块,我们对检测到的话题按文本数量进行倒序排列,并通过词云展示该话题下的关键词。平台的工作流程如图 4 所示。

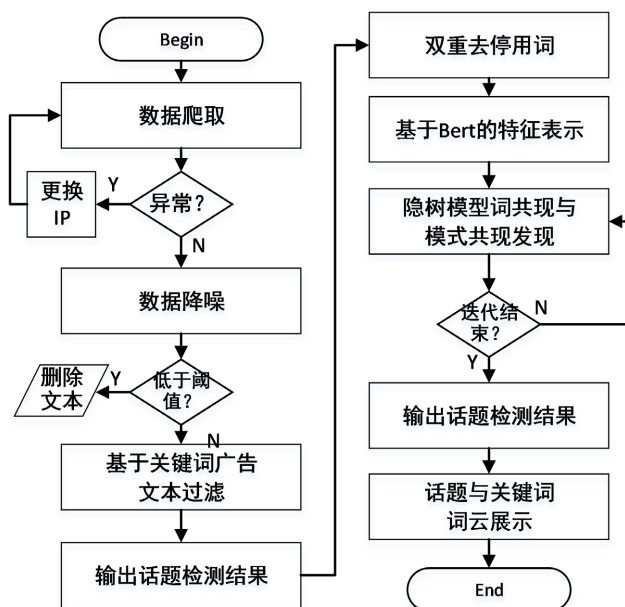


Figure 4. Flow chart of the platform
图 4. 平台工作流程图

5. 话题挖掘结果与分析

经过模型的自动分析,我们获取了带有课程思政内容课程评论中的十大话题,经过人工经验的筛选,我们最终选择出 7 大有关话题,如表 1 所示。

Table 1. Results of student evaluation topic mining

表 1. 学生评教话题挖掘结果

负面话题	示例文本
课程内容	课程从大数据的概念、产生背景、技术挑战、应用场景等方面进行了全面而系统的介绍。此外,课程还介绍了数据收集、处理和分析的方法。
方式与态度	老师平易近人,能循循善诱,很注意启发和调动学生的积极性,能够鼓励学生踊跃发言,课堂气氛较为活跃。
语音与字母	就不太懂吐字都不清楚是怎么当上老师的,感觉好多细节都没 get 到,好多都是课后翻书才搞明白。
时间安排	感觉前半部分讲的有点慢,后半部分一直在赶进度,是不是可以把前半部分压缩一下,特别是概述那一章,其实后面都有展开……
收获效果	收获非常大,通过课程学习一下子弄明白了计算机视觉、NLP 和推荐等不同方向的特点和主要问题、主要方法。
融入生硬	典型的为了思政而思政,感觉打断了内容的学习,也没有很大的关联,不知道是不是硬加进去的……
成绩证书	为什么一门慕课给分压得这么狠,这样我还不如选本校的课程呢,强烈建议看到这条的都别选。

此外,我们对学生评价话题中 TOP10 中有意义的七个话题所包含的特征(词)进行了词云分析,结果如图 5 所示。



Figure 5. Word cloud distribution of main topics in student evaluation
图 5. 学生评价主要话题的词云分布

如表 1 及图 5 所示, 学生对于含有思政内容的课程的关注点, 与一般课程的关注点大体相同, 主要集中在教学内容、教学方法、教学进度等安排上。从词云图可以看出, 学生对在线课程的评价以正面评价为主, 多数词语为褒义词, 然而, 当思政内容融入生硬时, 会导致负面评价。

6. 总结

学生对带有思政内容的课程的关注点是思政教学建设中的重要问题, 针对当前这一领域的研究空白, 以及传统实证研究方法再获取数据的成本与效果方面的局限, 本文提出了基于在线数据与机器学习方法的课程思政学生关注点研究模型。实验结果表明, 虽然增加了课程思政的内容, 学生对课程的多数关注点与思政无关, 教师还是要做好内容设计、进度设计和教学方法设计, 注重教姿教态、注重配套资料与答疑。增加课程思政, 并不会提高学生的容忍度。此外, 对于课程思政, 学生的关注点主要在于思政点的融入是否平滑, 是否打破了原有的教学逻辑, 是干扰还是促进对知识点的学习, 这也是课程思政教学应当着力思考的问题。

由于各大在线学习平台没有对含有思政内容的课程添加专门的分类或标签, 致使数据筛选困难, 不得已补充了短视频平台和问答平台有关课程思政的数据, 即便如此, 总的数据量还是偏小, 以至于分析结果中与课程思政相关的主题偏少。如何解决这一问题, 将是这一领域未来有意义的研究方向。

基金项目

本研究受对外经济贸易大学 2022 年度课程思政研究课题资助。

参考文献

- [1] 何茂刚, 刘向阳, 张颖. 课程思政及热力学的思政元素[J]. 高等工程教育研究, 2023, 51(1): 213-216.
- [2] 李健, 杨晓晨, 门桐宇. “课前 + 课中 + 课后”全过程高中数学课程思政[J]. 中国教育学报, 2023(52): 76-78.
- [3] 葛晨光, 朱清慧, 张水潮. 高校课程思政与思政课程协同育人的基本要求及实现路径[J]. 学校党建与思想教育, 2023(10): 25-28.
- [4] 翟乐, 杜茹. 高校思政教育网络课程开发创新研究[J]. 中国高等教育, 2023(6): 27-30.

- [5] 崔正贤, 马万利. 新时代课程思政建设的功能效用、问题症结与着力方向研究[J]. 中国电化教育, 2022(11): 82-89.
- [6] 张欣茹, 姜泽毅, 蒋滨繁, 等. 面向新工科建设“五维并举”的“能量转换与利用”课程教学设计、创新与实践[J]. 高等工程教育研究, 2023(S1): 57-61.
- [7] 吴姗姗, 吴晋峰, 王云弟, 等. 中国国家旅游形象研究——以北京和西安外国游客为例[J]. 浙江大学学报(理学版), 2015, 42(4): 466-473.
- [8] 樊姗姗. 系统科学视野下“虚拟仿真 + 思政课”亲和力提升路径探析[J]. 系统科学学报, 2024, 32(2): 97-102.
- [9] Jia, S.S. (2020) Motivation and Satisfaction of Chinese and U.S. Tourists in Restaurants: A Cross-Cultural Text Mining of Online Reviews. *Tourism Management*, **78**, Article ID: 104071. <https://doi.org/10.1016/j.tourman.2019.104071>
- [10] Malbon, J. (2013) Taking Fake Online Consumer Reviews Seriously. *Journal of Consumer Policy*, **36**, 139-157. <https://doi.org/10.1007/s10603-012-9216-7>
- [11] Lui, T.W., Bartosiak, M., Piccoli, G. and Sadhya, V. (2018) Online Review Response Strategy and Its Effects on Competitive Performance. *Tourism Management*, **67**, 180-190. <https://doi.org/10.1016/j.tourman.2018.01.014>
- [12] Cheng, M.M. and Jin, X. (2019) What Do Airbnb Users Care About? An Analysis of Online Review Comments. *International Journal of Hospitality Management*, **76**, 58-70. <https://doi.org/10.1016/j.ijhm.2018.04.004>
- [13] 史达, 王志敏. 绿色饭店用户体验——基于在线评论的深度学习研究[J]. 旅游科学, 2019, 33(6): 62-76.
- [14] 周楠, 杜攀, 靳小龙, 等. 面向舆情事件的子话题标签生成模型 ET-TAG[J]. 计算机学报, 2018, 41(7): 1490-1503.
- [15] Choi, H.J. and Park, C.H. (2019) Emerging Topic Detection in Twitter Stream Based on High Utility Pattern Mining. *Expert Systems with Applications*, 115, 27-36. <https://doi.org/10.1016/j.eswa.2018.07.051>
- [16] Wang, G.H., Chi, Y.X. and Liu, Y.J. (2019) Studies on a Multidimensional Public Opinion Network Model and Its Topic Detection Algorithm. *Information Processing & Management*, **56**, 584-608. <https://doi.org/10.1016/j.ipm.2018.11.010>
- [17] Pang, J.B., Tao, F., Huang, Q.M., Tian, Q. and Yin, B.C. (2019) Two Birds with One Stone: A Coupled Poisson Deconvolution for Detecting and Describing Topics from Multimodal Web Data. *IEEE Transactions on Neural Networks and Learning Systems*, **30**, 2397-2409. <https://doi.org/10.1109/TNNLS.2018.2872997>
- [18] 李慧, 王丽婷. 基于词项热度的微博热点话题发现研究[J]. 情报科学, 2018, 36(4): 45-50.
- [19] 张仰森, 段宇翔, 黄改娟, 等. 社交媒体话题检测与追踪技术研究综述[J]. 中文信息学报, 2019, 33(7): 1-10, 30.
- [20] Kim, S., Park, H. and Lee, J. (2020) Word2vec-Based Latent Semantic Analysis (W2V-LSA) for Topic Modeling: A Study on Blockchain Technology Trend Analysis. *Expert System with Applications*, **152**, Article 113401. <https://doi.org/10.1016/j.eswa.2020.113401>
- [21] Chen, J.Y., Gong, Z.G. and Liu, W.W. (2019) A Nonparametric Model for Online Topic Discovery with Word Embeddings. *Information Science*, **504**, 32-47. <https://doi.org/10.1016/j.ins.2019.07.048>
- [22] Jelodar, H., Wang, Y.L., Yuan, C., et al. (2019) Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey. *Multimedia Tools and Applications*, **78**, 15169-15211. <https://doi.org/10.1007/s11042-018-6894-4>
- [23] Conneau, A. and Lample, G. (2019) Cross-Lingual Language Model Pretraining. *33rd Conference on Neural Information Processing Systems*, Vancouver, 8-14 December 2019, 7059-7069.
- [24] Liu, T.F., Zhang, N.L., Chen, P.X., et al. (2017) Latent Tree Models for Hierarchical Topic Detection. *Artificial Intelligence*, **250**, 105-124. <https://doi.org/10.1016/j.artint.2017.06.004>
- [25] Liu, T.F., Zhang, N., Chen, P.X., et al. (2015) Greedy Learning of Latent Tree Models for Multidimensional Clustering. *Machine Learning*, **98**, 301-330. <https://doi.org/10.1007/s10994-013-5393-0>