

基于长尾理论的异步在线视频学习评论主题演化分析

Analysis of Topic Evolution in Comments on Asynchronous Online Video-Based Learning from the Perspective of the Long Tail Theory

曾好¹, 许柳桐², 何涛^{3*}

深圳大学 教育学部

* tao.he2016@gmail.com

【摘要】 异步在线视频学习已成为非正式学习的重要组成部分, 然而当前研究对异步评论中蕴含的学习需求探讨仍显不足。基于长尾理论, 本研究对某流行在线平台上人文与理科领域的 22,273 条异步评论进行了主题演化分析。研究发现, 评论呈现出典型的长尾分布趋势, 且学科差异深刻塑造了参与模式。这些发现为异步学习视频设计提供了基于学科特征的学习者参与模型, 并建议平台应建立差异化的互动支持系统。

【关键词】 异步在线学习; 长尾理论; 学习者评论; 主题分析

Abstract: Asynchronous video learning is an important form of informal learning, yet the learning needs reflected in asynchronous comments remain underexplored. Drawing on long-tail theory, this study analyzes the topic evolution of 22,273 comments from humanities and science videos on a popular platform. Results show a clear long-tail distribution, with disciplinary differences significantly shaping participation patterns. The findings inform a discipline-sensitive model of learner engagement and support the development of differentiated interactive systems for asynchronous learning platforms.

Keywords: asynchronous online learning, Long-Tail Theory, learner comments, thematic analysis

1. 前言

在当今数字时代, 在线视频平台已成为关键的知识传播渠道 (Brown & Adler, 2008)。以哔哩哔哩为代表的平台, 通过异步评论系统构建了充满活力的非正式学习环境 (Cen et al., 2024), 学习者在此主动进行知识协作 (Zeng et al., 2023)。然而, 现有研究对哔哩哔哩等非结构化平台中的海量评论数据关注不足 (Wang et al., 2019)。这些评论呈现典型的“长尾分布”: 少数热门评论 (头部) 获赞无数, 而绝大多数评论 (尾部) 关注度极低 (Wang et al., 2019)。长尾理论指出尾部评论蕴含独特价值 (Siemens, 2005), 但现有研究多关注情感传播 (Qin et al., 2024), 为弥补此研究空白, 本研究基于长尾视角 (Siemens, 2005), 运用主题演化分析方法, 具体探讨以下问题: 1) 高赞 (头部) 与普通 (尾部) 评论在内容类型上存在何种本质差异? 2) 人文与理科课程的评论主题演化趋势有何不同?

2. 文献综述

2.1. 哔哩哔哩作为非正式学习空间的兴起

随着移动互联网的发展, 视频分享平台已成为非正式学习的关键场域。研究表明, 哔哩哔哩通过其独特的社区文化促进了文化内容在青年群体中的传播和内化 (Yin, H., et

al.,2012)。它不仅是内容仓库,也是一个“亲社会”的学习社区,学习者通过观察他人(社会学习)和互动反馈来维持学习动机(Cen, C., et al., 2024)。然而,随着用户量的激增,评论质量参差不齐,从海量非结构化文本中挖掘有效信息成为当前研究热点。

2.2. 在线社区中的长尾效应与行为分析

长尾理论最初由 Anderson (Siemens, G., 2005) 提出,在在线社区中,长尾效应表现为用户参与度的极端不平等。先前研究表明,传统推荐系统倾向于偏袒头部内容,从而忽视了长尾中的新颖信息(Yu, S., Androssov, A., Yan, H., & Chen, Y., 2024)。在教育情境中,研究者发现学习者在异步学习视频中的点赞行为遵循典型的长尾分布(Wang, Y., Liu, Q., Stein, D., & Xia, Q., 2019)。他们指出,头部评论常承载核心知识或强烈情感共鸣,通过信息处理的“中心路径”(Biglan, A., 1973)有效说服用户;而尾部评论尽管关注度低,却构成了社区互动的基础。在计算机视觉和机器学习领域,长尾数据分布被视为现实世界数据集的常态(Gorsky, P., et al., 2010; Hwang, S., & Foote, J. D., 2021)。

2.3. 不同学科视角下的主题演化

不同学科的知识结构深刻影响着学习者的在线互动模式。学科分类模型将学科划分为硬/软、纯/应用等维度(Biglan, 1973)。研究进一步证实,理科(硬学科)讨论更侧重于事实确认和工具应用,而人文学科(软学科)讨论则更侧重于解释和社会临场感(Gorsky et al., 2010)。先前研究表明,这些学科差异会投射到长尾分布的结构上。然而,目前尚不清楚这些理论概念如何应用于真实的教育评论数据,以及不同学科在“尾部”是否表现出不同的模式。

3. 研究方法

3.1. 数据收集与参与者

本研究选取哔哩哔哩上两门跨学科课程的评论进行分析,分别是《数据结构与算法》(理科)和《文学理论》(人文)。数据集包含 13,404 名学习者的互动记录,包括评论和点赞行为数据。从原始的 22,273 条评论中,去除了重复内容、非中文内容及系统回复,最终保留了 9,849 条有效主评论(人文 4,812 条,理科 5,037 条)。数据集包含用户信息、评论内容、点赞数(单条评论最高获赞 16,427 次)以及时间戳。数据清洗后,我们利用预训练的 Qwen-Turbo 模型(Bai, J., et al., 2023)进行评论文本分类任务。

3.2. 分析框架

我们采用两阶段分析框架:1) 应用帕累托分析识别点赞分布中的关键阈值点;2) 使用时序堆叠面积图对比人文与理科课程评论内容类型的演化特征。接着,我们应用帕累托分析构建长尾模型。针对每门课程,评论按点赞数降序排列,以计算累积点赞数及其比例贡献。累积点赞数占前 80% 的评论被定义为“头部”评论,其余则归类为“长尾”评论。最后,我们使用时序堆叠面积图对比了两门学科评论内容类型的演化特征。

4. 研究结果

4.1. 头部与长尾评论内容的内部结构差异

如图 1 所示,两门课程的累积点赞曲线在左侧急剧上升。极少数评论(理科:13 条;人文:18 条)贡献了总点赞数的 80%。其余绝大多数评论位于长尾区域,对总点赞数贡献有限,呈现出典型的长尾分布。基于大模型,我们将评论分为六种不同的内容类型,包括“资源共享”、“经验分享”、“反馈”、“情感表达”、“求助”和“其他”。在理科课程的头部评论中,“资源共享”、“反馈”和“情感表达”占显著比例,而“求助”占比较低。相反,在长尾评论

中，“求助”的比例显著上升，成为尾部最大的类别。“反馈”和“情感表达”是主要驱动力，而“资源共享”在尾部的存在感远低于头部。

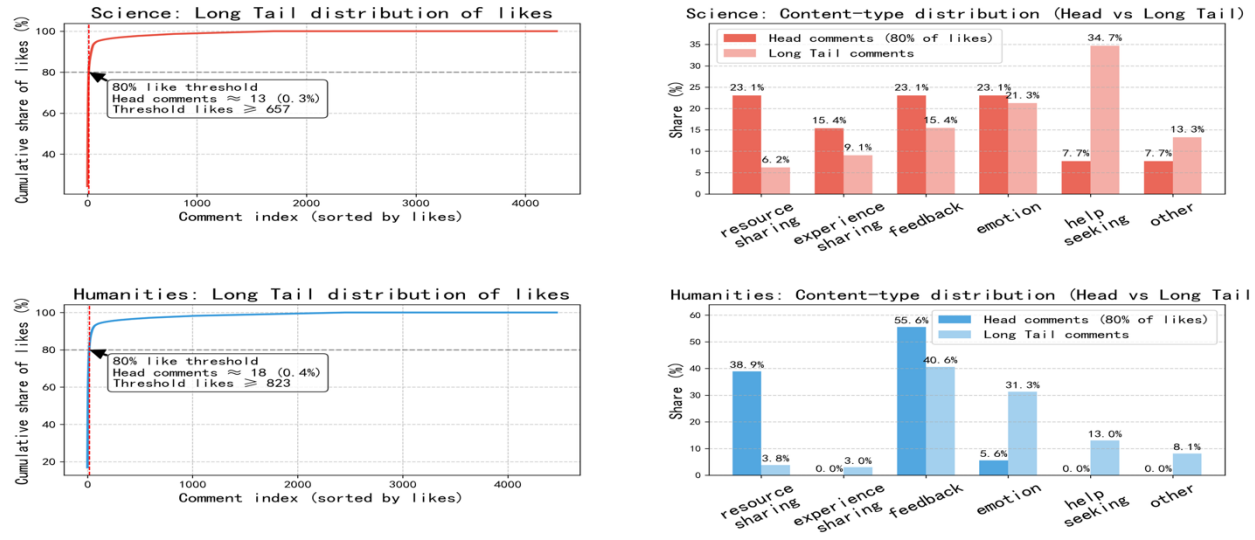


图 1 不同学科点赞评论的长尾分布及内容构成

4.2. 不同学科间评论的主题演化

图 2 显示，理科课程的评论在所有阶段主要由“求助”驱动，其次是“情感表达”和“反馈”。相比之下，人文课程的评论在所有阶段均以“反馈”和“情感表达”为主导。具体而言，在视频发布的初期阶段，人文课程中“求助”的比例显著低于理科。随着时间的推移，理科，人文课程中“求助”的占比也呈现出上升趋势。

Temporal Evolution of Comment Content Types (Science vs Humanities)

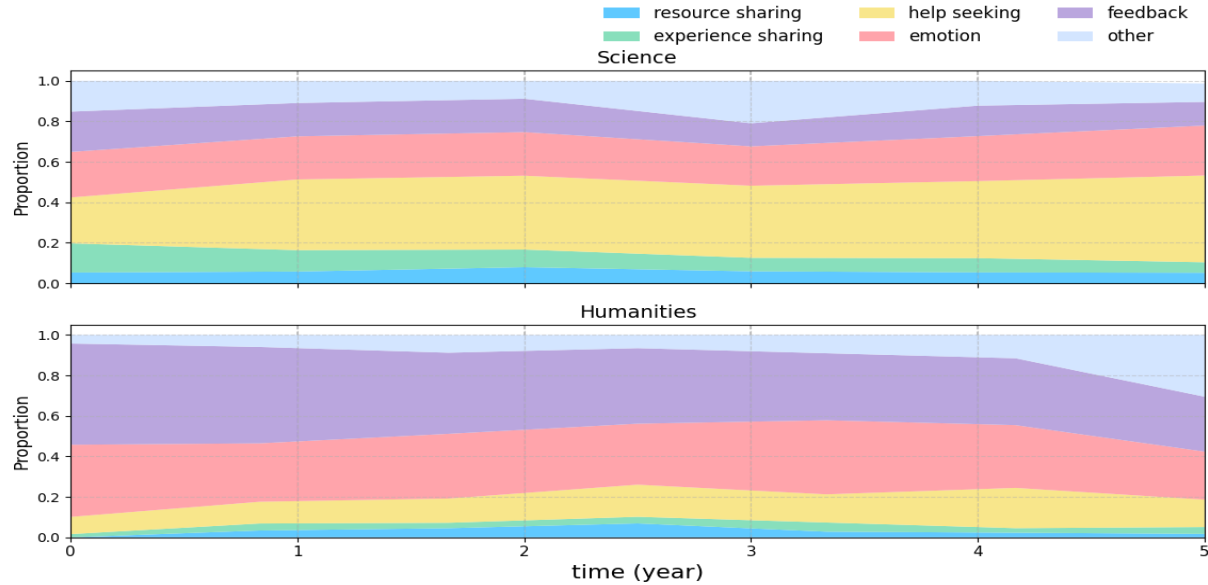


图 2 理科与人文课程评论内容类型的时间演化

5. 讨论

针对研究问题一，研究结果揭示了头部与尾部评论之间的根本区别。先前研究证实，分享高质量资源有助于其他学习者解决问题，同时促进高阶思维技能的发展（Rodriguez, M. A., 2014）。与此一致，本研究发现头部评论侧重于传播此类资源并提供深入反馈，相比之下，尾部评论主要由即时求助和简短情感表达构成。这种低门槛行为使学习者能够获得及时支持，并被认为是促进学习社区形成的关键机制（Wenger, E., 1998）。总体而言，这两种评

论类型展现出互补的动态关系：头部评论促进信息集中，而尾部评论支持点对点交流。二者共同维系了异步学习环境中的知识流动和社区凝聚力。针对研究问题二，时序分析进一步阐明了维持学习社区活力的动态力量。在理科课程中，技术性求助在所有学习阶段都普遍存在甚至有所增加，而经验分享则主要集中于早期阶段。这种模式表明了一种社区动态：投入的学习者在初期积极贡献，之后更广泛的参与者群体进行持续的求助。

6. 结论

本研究分析了异步在线视频平台上两个不同学科的评论数据，揭示了异步视频学习环境中学习者评论类型的结构特征与跨学科动态演化。这与先前研究一致，硬学科更关注工具使用，而软学科更关注情感共鸣（Gorsky, P., et al., 2010）。本研究为学习平台设计提供了启示：应根据学科属性实施差异化的功能设计，理科课程加强求助系统和知识库建设，而人文课程优化支持情感表达。未来研究可以扩展到更广泛的学科领域，以进一步验证和完善这些发现。

参考文献

- Bai, J., et al. (2023). Qwen Technical Report. ArXiv, abs/2309.16609.
- Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology*, 57(3), 195–203.
- Brown, J. S., & Adler, R. P. (2008). Minds on fire: Open education, the long tail, and learning 2.0. *EDUCAUSE Review*, 43(1), 16–32.
- Cen, C., et al. (2024). Enhancing the dissemination of Cantonese Opera among youth via Bilibili: a study on intangible cultural heritage transmission. *Humanities and Social Sciences Communications*, 11, 1038.
- Gorsky, P., et al. (2010). The relationship between academic discipline and dialogic behavior in Open University course forums. *International Review of Research in Open and Distributed Learning*, 11(4).
- Qin, W., et al. (2024). Uncovering the Dynamics of Epistemic Emotions in Asynchronous Online Video Learning: The Long-tail Effects of Like Behavior. In *Proceedings of ICET*.
- Rodriguez, M. A. (2014). Content Analysis as a Method to Assess Online Discussions for Learning. *SAGE Open*, 4, 215824401455901.
- Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2(1).
- Wang, Y., Liu, Q., Stein, D., & Xia, Q. (2019). Measuring students' affective states through online learning logs: an application of learning analytics. *International Journal of Information and Education Technology*, 9(5), 356–361.
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.
- Yin, H., et al. (2012). Challenging the Long Tail Recommendation. *arXiv preprint arXiv:1205.6700*.

- Yu, S., Androsov, A., Yan, H., & Chen, Y. (2024). Bridging computer and education sciences: A systematic review of automated emotion recognition in online learning environments. *Computers & Education*, 220, 105111.
- Zeng, L., et al. (2023). Behavior Analytics of Danmaku from Online Electronics Courses. *International Journal of Information and Education Technology*.
- Zhang, C., et al. (2024). A Systematic Review on Long-Tailed Learning. arXiv preprint arXiv:2408.00483.