

深度学习在作文自动评分系统中的应用与进展综述

张译天

新疆师范大学计算机学院 新疆维吾尔自治区乌鲁木齐 830054

摘要: 随着科技的飞速发展,深度学习技术以其强大的特征学习和表征能力,在诸多领域展现了卓越的应用效果。在作文自动评分系统中,深度学习的应用更是日益凸显其重要性。本文旨在全面综述近年来深度学习在作文自动评分中的最新应用与显著进展。当前作文自动评分技术仍面临一些挑战,如跨语言、跨文化评分的准确性问题,以及评分系统的稳定性和可靠性等。针对这些问题,我们进行了深入剖析,并提出了可能的解决方案。展望未来,我们期待深度学习技术能在作文自动评分领域发挥更大的作用,推动教育评估的智能化和精准化。

关键词: 深度学习; 作文自动评分; 自然语言处理; 人工智能

1. 引言

作文是语言教育中的核心,反映了学生的语言能力,其评分方式随科技进步而演变。传统评分依赖教师主观判断,虽细致但标准不一、效率不高。因此,研究者探索了计算机自动化评分。深度学习技术为作文评分带来新机遇,能自动学习和提取深层特征,提供更客观、准确的评分依据。

本文综述了手工特征和深度学习的作文评分技术现状,分析各方法优劣及实际应用效果,并关注评分公正性、准确性及跨语言文化背景适应性等挑战。期望为后续研究者提供参考,推动技术完善,助力教育评估智能化,支持学生个性化学习。我们希望通过技术不断进步,实现更科学、高效的作文评估,最终提升学生的写作能力。

2. 基于手工特征的作文评分方法

基于手工特征的作文评分方法是自动化作文评分领域早期的主要研究方向。这一方法的核心在于,通过人工的方式,根据作文评分的标准和要求,精心设计和提取一系列能够反映作文质量的特征。这些特征可能包括但不限于词汇的丰富性、语法的正确性、结构的合理性以及内容的深度和广度等。

在早期的研究中,L. M. Rudner 和 T. Liang 等人(2002)^[1]便尝试利用贝叶斯分类器和 K 邻近方法对作文进行分类。他们首先根据作文的不同特点将其分为几个类别,然后在分类结果的基础上,进一步融合文本的复杂度特征,最后利用回归方法对作文进行评分。这种方法虽然简单,但在当时却为作文自动评分开辟了新的思路。

Benjamin B. Bederson 等人(2003)^[2]也沿用了类似的思路,采用贝叶斯分类器实现作文评分。他们的研究进一步验证了基于手工特征的作文评分方法的可行性。然而,这种方法也存在一定的局限性,即特征的选择和设计很大程度上依赖于研究者的经验和直觉,可能无法全面准确地反映作文的真实水平。

为了克服这一局限性,Dan Bikel 等人(2010)^[3]提出了通过对多个特征值加权求平均的方式来实现作文评分。他们综合考虑了作文的多个方面,包括词汇、语法、结构等,为每个特征分配了相应的权重,通过计算加权平均值来得出最终的评分。这种方法在一定程度上提高了评分的准确性和全面性。

随着研究的深入进行,研究者们开始探索更为精细和复杂的评分方法。Keisuke Sakaguchi 等人(2015)^[4]便引入了一种基于 pair-wise 的排序方法,他们借助支持向量机(SVM)进行作文评分。这种方法不仅考虑了作文之间的相对优劣关系,还通过机器学习算法自动学习和调整评分模型,从而进一步提高了评分的准确性。

Chen 等人(2013)^[5]则更进一步,使用了基于 list-wise 的排序方法。他们通过 LambdaMART 算法训练排序模型,该算法能够同时考虑多个作文之间的相对顺序关系,从而更加准确地评估作文的质量。这种方法在当时取得了显著的成果,为后续的作文自动评分研究提供了重要的参考。

此外,Wei Zhao 等人(2020)^[6]还尝试使用领域自适应技术来解决跨题目的作文评分问题。他们认识到,不同题目

的作文在内容和风格上可能存在较大差异,因此需要通过领域自适应技术来调整评分模型以适应不同题目的需求。这种方法为作文自动评分技术的实际应用提供了更广泛的适用性。

总的来说,基于手工特征的作文评分方法在早期为作文自动评分领域的发展奠定了坚实的基础。随着深度学习技术的兴起和发展,作文自动评分领域也迎来了新的机遇和挑战。深度学习技术能够自动学习和提取作文中的深层特征,为作文评分提供更客观、更准确的依据,有望成为未来作文自动评分领域的主流方法。

3. 基于深度学习的作文评分方法

近年来,深度学习技术在多个领域展现出了强大的实力,作文评分领域也不例外。随着深度学习算法和模型的不断进步,越来越多的研究者开始探索其在作文评分中的应用,以期实现更准确、更客观的评分。

于华(2020)^[7]在南京大学对跨提示场景的自动作文评分技术进行了深入研究。他构建了深度学习模型,该模型能够根据不同提示场景下的作文特点进行自动评分。这一研究的亮点在于,其模型具有较强的泛化能力,可以适应不同场景下的作文评分需求。通过大量的实验验证,该模型在跨提示场景的作文评分中取得了显著的效果,为作文评分的灵活性和准确性提供了新的解决方案。

程相群(2022)^[8]在中国科学院大学则设计了面向高并发的作文自动评分系统。他针对传统作文评分系统在处理大量作文时可能出现的性能瓶颈,通过优化算法和系统架构,提高了评分系统的性能和稳定性。这一研究不仅解决了高并发场景下的作文评分问题,还为作文评分技术的实际应用提供了强有力的技术支持。

杨勇等(2021)^[9]在新疆师范大学在自动作文评分领域提出了提出了多项技术革新。他们首先开发了 MLSF 模型,通过结合 CNN 和混合神经网络,有效融合多层次语义特征,显著提升了 Kaggle ASAP 数据集上的评分性能, QWK 达到 79.17%。(2024)^[10]随后,针对长文本处理和主题相关性问题的,他们推出了 TASE 模型,利用多头注意力和 BERT 模型深入分析语义,进一步优化评分效果,并已将相关代码和数据集开源,为学术界和教育界提供了宝贵资源。

除了上述研究者外,肖平(2007)^[11]、李文娟(2023)^[12]、廖艺艺(2022)^[13]、张琳(2022)^[14]、王一凡(2022)

^[15]、杨银欣(2020)^[16]、汪美珍(2021)^[17]等也从不同角度对作文评分技术进行了深入研究。他们或从作文的语言特征入手,或关注作文的篇章结构和内容质量,或探索深度学习模型在作文评分中的最佳应用方式,为作文评分技术的发展提供了多角度、全方位的支持。

通过不断的研究和改进,我们有理由相信,这一技术将在未来的教育评估中发挥越来越重要的作用,为学生的个性化学习和发展提供更有力的支持。

4. 当前存在的问题与挑战

尽管深度学习在作文自动评分中取得了显著成果,但在实际应用过程中,仍然面临一些挑战和问题。这些问题主要集中在数据稀疏性、模型的可解释性,以及评分标准的统一与客观性等方面。

数据稀疏性是一个重要的问题。深度学习模型的训练依赖于大量的标注数据,然而在作文评分领域,标注数据的获取并不容易。一方面,标注作文评分需要大量的人力和时间成本,且标注质量受限于标注者的专业水平和主观判断。另一方面,由于作文的多样性和复杂性,即使是同一篇作文,在不同的评分标准下也可能得到不同的分数,这进一步增加了数据标注的难度。数据稀疏性不仅影响模型的训练效果,还可能导致模型过拟合,降低评分的准确性。模型的可解释性也是当前面临的一个挑战。深度学习模型通常具有复杂的网络结构和大量的参数,这使得模型的学习和决策过程变得难以解释。在作文评分中,用户往往希望了解评分结果的具体依据和理由,以便更好地理解作文的优缺点并进行改进。然而,目前的深度学习模型很难提供这样的解释性信息,这在一定程度上限制了其在作文评分中的广泛应用。评分标准的统一与客观性也是一个亟待解决的问题。作文评分涉及多个方面的评估,包括语言表达、内容质量、篇章结构等。不同的评分标准可能导致评分结果的不一致性,甚至引发争议。此外,由于作文评分具有一定的主观性,不同的评分者可能对同一篇作文给出不同的分数。因此,如何制定一个统一、客观的评分标准,以减少主观因素的影响,是当前作文自动评分技术需要解决的一个重要问题。

为了解决上述问题,研究者们正在不断探索和创新。例如,针对数据稀疏性问题,可以尝试采用无监督学习或半监督学习方法来利用未标注数据进行预训练,提高模型的泛化能力。同时,还可以考虑引入迁移学习策略,将在一个数

据集上学到的知识迁移到其他相关数据集上,从而加速模型的训练和提高性能。在模型的可解释性方面,可以研究如何设计更透明的深度学习模型,或者开发辅助工具来提供评分结果的解释性信息。此外,制定统一、客观的评分标准也需要教育领域和计算机科学领域的专家共同努力,通过深入研究和实践来不断完善和优化评分标准。

5. 未来发展方向与展望

随着深度学习技术的不断进步和教育领域的需求变化,作文自动评分技术也将迎来新的发展机遇。面向未来,深度学习在作文自动评分中的发展方向包括跨语言与跨文化的作文评分、个性化与自适应评分系统以及技术与教育融合的深入探讨等。跨语言与跨文化的作文评分是当前的一个重要发展方向。随着全球化的加速推进和教育国际化的趋势,越来越多的学生需要接受不同语言和文化的教育评估。因此,开发能够跨语言和文化进行作文评分的深度学习模型具有重要的现实意义。这将有助于学生更好地适应国际化的教育环境,提高他们的跨文化交流能力。个性化与自适应评分系统也是未来作文自动评分技术的一个重要发展方向。每个学生的写作水平和风格都有所不同,因此评分系统应该能够根据学生的实际情况进行个性化的评估和反馈。通过引入用户画像、学习路径分析等技术,可以构建更加个性化和自适应的评分系统,为学生提供更加精准和有效的写作指导。

6. 结论

在全面综述了深度学习在作文自动评分系统中的应用与最新进展后,我们不难发现深度学习在这一领域展现出的巨大潜力和广阔前景。随着技术的持续进步和科研工作的不断深入,我们有充分的理由期待深度学习在未来能够在作文自动评分中发挥更加核心的作用。其强大的数据处理和分析能力,使得作文评分更加精准、客观,从而为写作教学提供了极为智能和高效的辅助手段。然而,技术的发展总是伴随着挑战。

参考文献:

[1]Rudner, Lawrence M. and T. Liang. "Automated Essay Scoring Using Bayes' Theorem." (2002).
 [2]The Effectiveness of Automated Scoring for Complex Tasks: An Analysis of a Summarization Scoring System

[3]Feature Engineering for Automated Essay Scoring: A Multi-Dimensional Approach

[4] Effective Feature Integration for Automated Short Answer Scoring

[5]Hongbo Chen and Ben He. 2013. Automated essay scoring by maximizing human-machine agreement. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages1741 - 1752.

[6]Zhao, Wei, Hui Gao, Shuhui Chen and Nan Wang. "Generative Multi-Task Learning for Text Classification." IEEE Access 8 (2020): 86380-86387.

[7] 于华. 面向跨提示场景的自动作文评分技术研究 [D]. 南京大学,2020

[8] 程相群. 面向高并发的作文自动评分系统的设计与实现 [D]. 中国科学院大学 (中国科学院沈阳计算技术研究所),2022

[9] 周险兵. 樊小超. 任鸽. 杨勇. 基于多层次语义特征的英文作文自动评分方法 [J]. 计算机应用, 2021

[10] 陈宇航. 杨勇. 先木斯亚. 买买提明. 帕力旦. 吐尔逊. 樊小超. 任鸽. 刁宇峰. 基于主题感知和语义增强的作文自动评分方法 [J]. 计算机工程,2024

[11] 肖平. 论作文教学对学生写作个性的适应 [D]. 湖南师范大学,2007

[12] 李文娟. 管建刚“作后讲评”模式在统编本高二作文教学中的应用研究 [D]. 广东技术师范大学,2023

[13] 廖艺艺. 初中语文教材作文资源的挖掘研究 [D]. 广西师范大学,2022

[14] 张琳. 关于田上新吉生活作文教育的思想性研究 [D]. 内蒙古师范大学,2022

[15] 王一凡. 沙盘游戏在叙事类作文教学中的应用研究 [D]. 山东师范大学,2022

[16] 杨银欣. 初中作文模式化教学研究 [D]. 广西师范大学,2020

[17] 汪美珍. 胡怀琛作文教学思想探究 [D]. 华中师范大学,2021