Automated Essay Scoring: A Reflection on the State of the Art

Shengjie Li
Vincent Ng Human Language Technology Research Institute University of Texas at Dallas sxl180006, v

Abstract

While steady progress has been made on the task of automated essay scoring (AES) in the past decade, much of the recent work in this area has focused on developing models that beat existing models on a standard evaluation dataset. While improving performance numbers remains an important goal in the short term, such a focus is not necessarily beneficial for the long-term development of the field. We reflect on the state of the art in AES research, discussing issues that we believe can encourage researchers to think bigger than improving performance numbers, with the ultimate goal of triggering discussion among AES researchers on how we should move forward.

1 Introduction

Automated Essay Scoring (AES), the task of automatically assigning a holistic score to an essay that summarizes its overall quality, is arguably one of the most important applications in natural language processing (NLP). As an example of AES, consider the essay in Table 1, which is written in response to the prompt shown at the top of the table. Given the scoring rubric in Table 2, an AES system should assign a score of 3 to this essay for the following reasons. First, its author takes a position but fails to provide adequate support and details....

[...CONTENT MISSING: SECTIONS 2 THROUGH CONCLUSION DUE TO PDF EXTRACTION LIMITS...]

2 State of the Art

[MISSING]

3 Methodologies

[MISSING]

4 Issues and Reflections

[MISSING]

5 Conclusion

[MISSING]

References

[1] Ruosong Yang, Jiannong Cao, Zhiyuan Wen, Youzheng Wu, and Xiaodong He. 2020. Enhancing automated essay scoring performance via fine-tuning pre-trained language models with combination of regression and ranking. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1560–1569, Online. Association for Computational Linguistics.

[2] Helen Yannakoudakis and Ted Briscoe. 2012. Modeling coherence in ESOL learner texts. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 33–43, Montr'eal, Canada. Association for Computational Linguistics.

[3] Helen Yannakoudakis, Ted Briscoe, and Ben Medlock. 2011. A new dataset and method for automatically grading ESOL texts. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 180–189, Portland, Oregon, USA. Association for Computational Linguistics.

[4] Torsten Zesch, Michael Wojatzki, and Dirk Scholten-Akoun. 2015. Task-independent features for automated essay grading. In *Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages [MISSING].

Table 1: Example essay and prompt.

Prompt: [MISSING - Content from original Table 1 not visible in scan]
Essay: [MISSING - Essay content from original Table 1 not visible in scan]

Note: Content appears to be missing from extraction.

Table 2: Scoring Rubric (Example).

Score	**Criteria**
[MISSING - Content from Table 2 not visible] height	[MISSING]