Direct automated feedback for student submissions based on LLMs

MAXIMILIAN SÖLCH, Technical University of Munich, Germany

FELIX DIETRICH, Technical University of Munich, Germany

STEPHAN KRUSCHE, Technical University of Munich, Germany

A clear and well-documented LaTeX document is presented as an article formatted for publication by ACM in a conference proceedings or journal publication. Based on the "acmart" document class, this article presents and explains many of the common variations, as well as many of the formatting elements an author may use in the preparation of the documentation of their work.

1 INTRODUCTION

In the current educational landscape, providing timely and effective feedback to students remains a significant challenge. Traditionally, students must wait for course tutors or professors to review their submissions and provide feedback. This process can be time-consuming, often requiring students to arrange meetings and wait for available time slots, which are not always convenient or immediate. The inherent delays and scheduling difficulties make this approach not scalable, especially in courses with a large number of students.

These limitations hinder students' learning progress and motivation. The waiting period for feedback interrupts the learning flow, causing students to lose momentum and potentially disengage from the subject matter. Additionally, the limited availability of tutors and professors means that not all students receive the individualized attention they need to improve their understanding and skills. This situation underscores the necessity for a more efficient and scalable feedback system that can provide continuous support to students without the constraints of traditional methods.

The subsequent sections of this paper are organized to provide a comprehensive understanding of the research. Section 2 provides an overview of related work. Section 3 details the concept and methodology of Direct Automated Feedback Delivery (DAFeeD). Section 4 describes the reference implementation of DAFeeD, called Athena, including a general overview, details on the used prompts, and the system architecture. Section 5 presents the evaluation results, Finally, Section 6 concludes with a summary of findings and discusses future research directions to enhance automated feedback systems.

2 RELATED WORK

todo [1]

3 APPROACH: DIRECT AUTOMATED FEEDBACK DELIVERY (DAFEED)

DAFeeD employs large language models to deliver automated feedback on student submissions, designed to complement traditional teaching methods and provide additional support. Figure 1 illustrates the continuous feedback workflow that DAFeeD facilitates, enabling students to receive feedback at any time, thereby eliminating the need to wait for responses from human tutors or course professors.
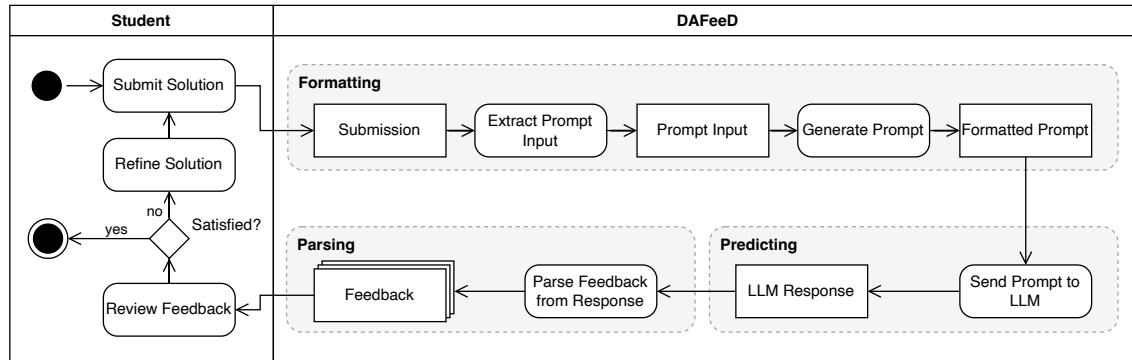

Fig. 1. Workflow of direct automated feedback delivery for students' submissions (UML Activity Diagram)

DAFeeD can provide feedback on various aspects, such as the correctness of the code, the quality of the code, and the performance of the student. Once the student submits their solution, DAFeeD initiates a three-stage process to generate natural language feedback.

The first stage, called *Formatting*, takes the student's submission and extracts the submission content, problem statement including learning objectives, and any possible grading instructions the instructor defines. This extracted information represents the prompt input. During the prompt generation step, a predefined prompt template is filled with the prompt input data, resulting in a formatted prompt.

In the second stage, called *Predicting*, the formatted prompt is sent to a large language model, which generates a response that includes detailed feedback for the student.

The final stage, *Parsing*, takes the LLM response, which comes in the JSON format, and parses feedback items from it. In addition to the feedback text, the feedback object also contains reference information indicating the part of the submission it pertains to. For programming exercises, this includes the file name and line number of the relevant code snippet to which the feedback refers. For text exercises, the reference information includes only the sentence or word range the feedback refers to.

All of the feedback is then returned to the student for review. If the student is satisfied with the feedback, the process concludes. Otherwise, the student can refine their solution and resubmit it, initiating the DAFeeD process anew.

This iterative process is designed to motivate students to continuously learn and experiment with their solutions, resulting in improved performance.

4 REFERENCE IMPLEMENTATION: ATHENA

We incorporated DAFeeD into a reference implementation named Athena, which is integrated with the learning platform Artemis through which students submit their solution and review the feedback.

4.1 Prompts

4.2 Feedback Generation

?

4.3 Architecture

Modular approach to support multiple exercise types and extend it in the future

The assessment module manager receives all requests, checks for authorization, and then forwards them to the responsible modules.

Independent of a specific learning management system, provides a REST API documented with the OpenAPI standard

Currently connected to OpenAI and Azure OpenAI Can be simply modified to use open-source models like Llama or Mistral either self-hosted or in the cloud.

As the reference implementation is used in conjuction with the Artemis learning management system for actual university courses, the system is designed to be scalable and reliable (performance, maintainability, usability) Deployed in a kubernetes cluster to ensure scalability and reliability, loadbalancing, etc. Scale each module independently, e.g., more instances of the programming module when a new programming exercise is released.

5 EVALUATION

5.1 Research Questions

5.2 Study Design

5.3 Results

5.4 Limitations

5.5 Discussion

6 CONCLUSION & FUTURE WORK

Future work includes enhancing the visualization of feedback, such as grouping and color coding the feedback items to make it easier to differentiate between critical feedback items, suggestions for improvement, and positive feedback. A high priority will be on further improving the overall quality of the feedback provided. We also aim to extend the implementation to support direct automated feedback for the remaining exercise types of Artemis. Another crucial step is to test direct automated feedback in a real-world setting by utilizing this feature in an actual course. This will allow us to collect comprehensive data to thoroughly evaluate the impact on student performance and motivation.

REFERENCES

[1] Marcelo Guerra Hahn, Silvia Margarita Baldiris Navarro, Luis De La Fuente Valentin, and Daniel Burgos. 2021. A Systematic Review of the Effects of Automatic Scoring and Automatic Feedback in Educational Settings. *IEEE Access* 9 (2021), 108190–108198. https://doi.org/10.1109/ACCESS.2021.3100890