

ABSTRACT

Feedback is a critical element of the learning evaluation process, functioning to help students recognize their mistakes and enhance their comprehension of course material. However, in higher education, the delivery of comprehensive and high-quality feedback is often hindered by time constraints and large class sizes. Consequently, many students do not receive adequate input to improve their academic performance. While automated solutions such as Automated Essay Scoring (AES) have gained popularity, they are generally limited to assigning scores without providing contextual and explanatory narrative feedback. This study aims to develop an automatic feedback generation system based on multiclass classification of student answers using the Transformer-based T5 model. The system integrates three similarity measurement techniques—BLEU, Fuzzy, and BERT—combined through a dynamic weight adjustment mechanism. The resulting similarity scores are categorized into four evaluation labels (correct, partial 2, partial 1, and incorrect), which serve as target labels during T5 model training, enabling the system to generate feedback text that aligns with students' levels of understanding. Experiments were conducted using four datasets (MPI 1, MPI 2, Siscer A1, and Siscer B1), and the system's performance was evaluated using SMAPE, Macro F1-Score, and Exact Match Accuracy. Results indicate that the proposed approach successfully produced similarity predictions that closely align with human assessments, achieving the best SMAPE score of 13.49% on the Siscer B1 dataset. The highest classification performance was observed in the MPI 1 dataset with a Macro F1-Score of 0.63. The evaluation of feedback quality using exact match accuracy showed the highest scores on MPI 1 (66.86%) and Siscer B1 (62.64%), indicating the system's ability to generate feedback that is literally identical to the reference, particularly for datasets with consistent answer patterns or multiple references. Overall, the developed system demonstrated the ability to generate contextual feedback that closely approximates human evaluation quality. The integration of the T5 model with a similarity-based dynamic weight adjustment approach has proven effective in improving the efficiency and quality of automated text evaluation. This method holds promise as a viable solution for accelerating the feedback process in educational contexts while providing meaningful guidance that supports students independent learning.

Keywords : Automatic Feedback, Multiclass Classification, T5, SMAPE, Macro F1-Score, exact match accuracy