REVIEW ON AUTOMATIC FEEDBACK SYSTEMS IN EDUCATION

Rudragouda Patil[1], Prachi Channe[2], Shruti Diwate[3]

[123]School of Computer Engineering and technology, MIT Academy of Engineering, Pune

1 rgpatil@comp.mitaoe.ac.in 2 prachichanne820@gmail.com 3 shrutidiwate555@gmail.com

I. ABSTRACT

Automatic feedback generation is an emerging field of educational technology with the goal of utilising artificial intelligence (AI) and to provide students with feedback that is more effective and of higher quality. This paper provides an overview of the current state-of-the-art in automated feedback generation systems, including their capabilities, challenges and limitations, and potential applications, along with future scope. Systems using artificial intelligence (AI) for feedback production use a variety of methods, such as data analytics, machine learning, and natural language processing, to evaluate learner performance and produce individualised feedback. These tools can offer feedback on a range of learning-related skills, including language competency, material comprehension, grammar and critical thinking.

Keywords: *Feedback generation, literature review, NLP (Natural Language Processing), Semantic Similarity* **,** *Machine learning , Text Analysis*

II. INTRODUCTION

Feedback is a significant step in a student's learning process. The quality of feedback positively affects students learning. It is considered essential in speeding up the development of students, monitoring, assessing, and regulating their own learning. Students can review the feedback and work on their loopholes and perform better in the future. In the absence of feedback, students might keep repeating mistakes which hinder learning. Results are required to evaluate progress, rectify errors, and enhance performance[1,13,20]. Traditional evaluation and feedbacks were dependent on teachers and professionals. It was given as grades, written or in verbal format. It could be words of appreciation, learning guidance, comparisons, motivational messages, and so on. As we know the significant role that feedback plays in student development, it is necessary that it does not get compromised. Any kind of incorrect feedback may cause a loss to students. However, today there is a huge number of educational institutes and students, so providing feedback to each student becomes a time-consuming and complex process. Teachers are facing problems as the increasing size of classes makes it difficult to provide feedback timely and also decreases the quality of feedback eventually decreasing the quality of learning[1].

Sometimes even students don't bother to collect feedback despite the time and effort of teachers. Others might disapprove of the feedback provided by teachers manually. Students complain that feedback was too negative and not useful. Also, there are cases when teachers might be unwilling to provide honest feedback. In these cases, feedback is considered negative and is wasted. Feedback varies with the type of test or evaluation done. For example, in cases of an essay, feedback can be of regarding grammatical errors or semantics of the content. As well as graders are required to have a much deeper level of the domain. Whereas , when coming to MCQ(Multiple Choice Questions) types, feedback can be number of correct and incorrect answers out of the total attempted. It helps students to have an idea about their accuracy. The paper presents a literature review of such platforms and methodologies that are used in providing automatic feedback in learning systems.

III. APPROACHES TO AUTOMATIC FEEDBACK GENERATION

1. Natural Language Processing (NLP) Techniques

NLP methods can be used to detect errors or inaccuracies in user input. For instance, spelling correction algorithms can automatically correct misspelt words while syntactic analysis can help identify grammar mistakes. In[1], the system is focused on the English subject, and it allows learners to complete assignments in the form of multiple-choice questions and fill-in-the-blank questions with rapid feedback. To gather answers, NLP approaches like WordNet with NLTK and the Word2Vec algorithm are used. Text has been converted into audio using Python's Text-to-Speech API. The close match algorithm-containing Difflib Python package has been used as a spell checker, and TextBlob has been used to automatically suggest word corrections.NLP helps in giving accurate feedback to users by identifying and correcting misspelled words. It helps in understanding the meaning and semantics behind the corpus. It helps in addressing issues related to coherence and semantic inconsistencies in answers. Sentiment analysis is significantly used in determine the sentiment or emotional; tone reflected in corpus. It can help give the feedback on clarity or tone of text. Text summarization is another NLP sub topic, it can provide feedback by highlighting main points or keys area of improvements. Using only the text of the paper and a corpus of reviews produced by humans for other papers, Alberto Bartoli and Andrea De Lorenzo's [10] work proposed approach for automated feedback production for scientific papers creates a review suited to a particular work with a specific recommendation. A corpus-based NLG (Natural Language Generation) approach is suggested in the work. By using text examples of the desired text generator output, corpus-based approaches attempt to automatically train text generation rules.

2. Rule-Based Systems :

Rule-based systems generate feedback based on predetermined criteria using a set of specified rules. These rules are often developed by subject matter experts and used to evaluate user input. For instance, in a learning environment, rules can be developed to assess the accuracy of student responses and deliver feedback accordingly. In [4] a variety of automatic feedback methods are reviewed, including rule-based systems, machine learning algorithms, and methods for natural language processing. The effectiveness of these methods for providing feedback on various tasks, such as multiple-choice questions, short answer questions, and essays, is examined by the writers. The study offers a comprehensive assessment of the literature on automatic feedback in online learning settings and emphasizes the potential advantages of this strategy for improving student learning outcomes.

3. Machine Learning Techniques :

In order to generate feedback, machine learning techniques can be used to train models on big datasets. To learn patterns from labeled cases supervised learning techniques such as classification or regression can be utilized. Then, based on the model's predictions, feedback can be created. For instance, a machine learning model can be trained on a corpus of good texts in an automated writing assistant to recommend enhancing a user's writing. A unique framework for an automated assessment and reporting framework using a combination of supervised deep learning models and unsupervised MCMC sampling approach was presented and implemented by Chang Lu and Maria Cutumisu[9]. The performances of three models, CNN, CNN+LSTM, and CNN+Bi-LSTM, on AES tasks within the same context were specifically compared in this work. The outcomes showed that of the three algorithms, CNN+LSTM performed the best on the AES tasks. Also, on seven out of eight writing tasks, the CNN+LSTM outperformed the baseline model, demonstrating the potential of word embeddings and deep learning models for automated essay scoring. [23] used a two-layer bidirectional LSTM with score-specific word embeddings to execute AES challenges and learn essay representations. The suggested model fared better than the standard SVM model. Later, Dong and Zhang [22] developed a three-layer model architecture, outperforming Taghipour and Ng's [21] model and their two-layer CNN model, by combining CNN for character representation and LSTM for sentence representation with an additional attention-pooling layer. Later experiments using word embeddings and deep learning techniques on AES were influenced by Taghipour and Ng's attempt to combine feature engineering and deep learning models.

4. Hybrid Approaches :

Hybrid methods integrate a number of different strategies to improve the accuracy and efficiency of automatic feedback generation. A system might, for example, employ rule-based techniques to quickly deliver early input before using machine learning or NLP techniques to focus in on and modify the feedback based on user-specific patterns. [2] propose an artificial intelligence-based approach for generating essay rubrics of examination and its subsequent use of it for high- quality feedback to students. It is a NLP based annotation free approach for automatic generation of rubrics that combines TextRank and Automatic Concept Map Generation. WordNet corpus and Python's NLTK with PoS tagging are used. The formula that produced the most accurate results for determining how similar different summaries were. The automatic feedback generation systems presented in [3] analyse student essays and provide the feedback on aspects of the writing, such as Syntax, Spelling, Sentence Diversity, Supporting Ideas, and Organization, using natural language processing (NLP) and machine learning algorithms. To categorise the quality of essays in relation to each writing component, the system uses the supervised machine learning approach. Between the raters' responses, a Pearson correlation was done for each rubric evaluation.

Korbit is a large-scale, open-domain, mixed-interface, dialogue-based ITS that offers interactive, personalised learning online using ML, NLP, and reinforcement learning. According to [6] the student tries to complete the task; their attempt is evaluated using an NLP model and compared to the expectations. The inner-loop system will activate and reply with one of a dozen various pedagogical interventions if their solution is determined to be erroneous. These interventions include clues, mathematical hints, elaborations, explanations, concept tree diagrams, and multiple-choice quiz answers.

Galeena Deeva and Daria Bogdanova [7] proposed an organised review of the literature on automated feedback technology in education is presented in the study. After a four-stage search procedure, 109 papers were chosen and evaluated. This investigation led to identify the most pertinent dimensions for classification automated feedback technologies, and integrated those criteria into a general classification framework. In addition, this paper offered a summary of current trends in the area of educational systems and categorised the chosen papers in accordance with the framework.

In [5] study, a tool for the automatic analysis of spelling mistakes in texts written in free-form German is proposed. It is based on automatic annotations of spelling errors, which include both error-related attributes and linguistic properties of the target word (phonemes, syllables, and morphemes). To determine the most prevalent patterns of affective states, the data was processed using sequential pattern mining algorithms. With the aid of our own methods as well as the online service G2P of the Bavarian Archive of Speech Signals (BAS)7 [19, 20], we automatically annotate a word's attributes in accordance with the annotation scheme we outlined in the previous section.

Ref. No	Algorithm	Environmental Setup	Pre-processing	Evaluation Parameters	Results
[1]	WordNet with NLTK and Word2Vec algorithm	MySQL database that is for storage of data which is accessed by the application layer	WordNet with NLTK and Word2Vec algorithm are used to get data from database	WordNet, Word2Vec	WordNet was found to be faster and of better performance than Word2Vec in generating feedback.
[2]	Text Rank NLP Semantic similarity	10 Essays on Artificial Intelligence course in a for-credit online MS in CS program (400-650 words)	NLTL for PoS tagging WordNet	Similarity Score for each summary	Highest similarity score is selected for rubrics
[3]	J48 Naïve Bayes SMO Baseline	1290 comments in essays of 327 English major students on 7 aspects of writing	Coh-Metrix 3.0 (feature extraction) Language Tool(spelling errors) Link Grammar	Precision, Recall, F-score	System generated ICF was useful in structure, quality writing, and conclusion Where as organisation and

			Parser(checks grammar)		coherence were difficult.
[6]	Machine Learning (Random Forest Classifier) NLP, reinforcement	796 annotated student–system interactions, collected from 183 students studying the machine learning		z-test	deep personalization model leads to the highest student learning gains at 48.53%
[9]	CNN, CNN + LSTM, and CNN + Bi-LSTM.	The corpus contains 25,000 positive reviews and 25,000 negative reviews. The training set consisted of 20,000 negative reviews and 20,000 positive reviews.	Text cleaning, tokenization, and padding. The GloVe 300-dimensional embeddings were trained on 6 billion words scraped from Wikipedia and other web texts.	Quadratic Weighted Kappa	the most accurate algorithm is CNN + LSTM. The average QWK of CNN + LSTM reaches 0.734
[21]	CNN, RNN, GRU, LSTM	Used 5-fold cross validation to evaluate our systems. In each fold, 60% of the data is used as our training set, 20% as the development set, and 20% as the test set.	Tokenized the essays using the NLTK5 tokenizer, lowercase the text, and normalize the gold-standard scores to the range of [0, 1]. D	Pearson's correlation, Spearman's correlation, Kendall's Tau, and quadratic weighted Kappa (QWK).	The attention mechanism significantly improves the results compared to LSTM without mean-overtime, but it does not perform as well as LSTM with meanover-time. The other two architectural choices do not lead to further improvements over the LSTM neural network.
[25]	Similarity analysis	4,268 reflection texts about 200 hardware concepts by 59 students	Punctuation removal, Stop words removal, Stemming,	TF-IDF score, Similarity scores	21.07% average reduction in plagiarised post ratios

Table 1 Comparison Table

IV. CHALLENGES AND LIMITATIONS OF AUTOMATIC FEEDBACK SYSTEM

Even though automatic feedback generation offers many benefits, it also comes with certain limitations and challenges. These can in terms of contextual understanding like systems might struggle with comprehending the context of user input. Language is complex, and understanding any kind of nuances, intent, and subtleties of user responses can be challenging. His may cause in inaccurate or irrelevant feedback. Another challenge is regarding preferences. Different users may require specific types of feedback that varies on the basis of skill level, learning style and type of assessment.[1] Machine learning-based approaches heavily rely on training data to generate feedback. However, obtaining large and high-quality labeled datasets for feedback generation can be challenging, particularly in specialized domains. Limited training data may affect the performance and accuracy of the feedback generation system.[27] focuses in spelling and grammatical mistakes, here error is marked out it lack in representing feedback. Hence, organization and categorization of feedback is importance. It can be difficult for automatic systems to handle some feedback forms, such giving subjective assessments or original suggestions. These kinds of feedback frequently need human judgement, creativity, and emotional understanding, which automated methods might not be able to reproduce well. Some system generated feedback are too general for correcting errors in grammar or spelling. It may even suggest too many suggestions based of features. Hence, this may impact the non-English student with less vocabulary but more semantics. [2] says that machine-learning based approach for generation of high-quality feedback to student face problems by requiring significant number of training data.

V. DISCUSSION

The utilization of machine learning algorithms has proven to be an advantageous tool in providing feedback within the educational system. Feedback is a crucial aspect for students to comprehend their strengths and weaknesses and improve their performance. Therefore, it is imperative that the models used to provide feedback generate accurate results.

In the related work literature review, it is evident that natural language processing (NLP) techniques are commonly employed, along with libraries such as NLTK, PoS tagging, and WordNet.[1,2,3,6] These methods aid the models in learning the dependency of words within the corpus, which is pivotal for generating meaningful feedback. Additionally, when generating feedback, it is necessary to consider the similarity between the model and the student's answer. Similarity methods such as cosine similarity, tf-idf, and Euclidian distance are commonly utilized to calculate similarity. Clustering is another approach that is used to group similar text and provide feedback to each cluster. Moreover, it is important to take into account grammar, syntax, and spelling while giving feedback to the student's answer, as these aspects of writing impact the overall quality of the response. Therefore, the feedback generated should also emphasize these aspects. The accuracy of the feedback generated heavily relies on the training data. If the data is biased, incomplete, or contains errors, the feedback generated will also be biased, incomplete, or contain errors.[12] Hence, it is crucial to use a diverse and representative dataset for training the models to ensure the accuracy of the feedback generated. [26] examines the influence of text mining-based automated feedback on plagiarism in online assignments. The emphasis is on how using automatic feedback affects students' propensity to plagiarise. The study employs a controlled experiment in which participants are randomised at random to receive automated feedback on their assignments or no feedback at all. Finally, the type of assessment also influences the type of feedback generated and the methods used to generate it. Different types of assessments, such as multiple-choice questions, short answer questions, and essays, require distinct approaches to generate feedback. Therefore, it is essential to select the appropriate approach based on the type of assessment used.[3]

VI. FUTURE DIRECTIONS

Future work may include the type of feedback based upon type of assessment. Increasing the training data as well can improve the accuracy when it comes to ML based feedback generation. Hence it could be replaced by artificial-intelligent based approaches. Research can be done in determining proper rubrics or rules to grade or provide feedback. Depending of this set of rules system can be developed to generate feedback. In future work, system can more focus on feedback on spelling and grammar but as well as semantics. Feedback is vital to measure growth of individual, so it cannot be compromised. By creating immersive and engaging settings, the combination of automatic feedback production and virtual and augmented reality (VR/AR) has the potential to improve the feedback experience. Feedback can also be integrated with personalized learning and self-assessment.

VII. CONCLUSION

A useful tool with enormous potential, automatic feedback creation is used in a variety of fields, including education, writing aid, language acquisition, programming instruction, and customer service. Automatic feedback systems can analyses and comprehend user input by using machine learning and natural language processing (NLP) techniques. They can then give specific suggestions for improvement and aid in learning and problem-solving. However, automatic feedback generation has a number of drawbacks and difficulties. These include concerns with subjectivity and contextual comprehension, domain specificity and restricted coverage, ambiguity and error handling, as well as scalability and customization problems. For automatic feedback generating systems to be further developed and improved, it is essential to be aware of these limits.

In conclusion, automatic feedback generation has the ability to improve learning outcomes, permit personalized instruction, and promote continual progress when properly applied and integrated. Despite challenges and limits ongoing research and technical developments are opening the way for more effective automatic feedback generation systems that will help students, teachers, and professionals from a broad range of areas.

VIII. REFERENCES

[1]. Nundlall, Janeesha, and Vidasha Ramnarain-Seetohul. "A Feedback Generation System to Enhance Learning at Primary School." *2020 3rd International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (ELECOM)*. IEEE, 2020.

[2]. Altoe, Filipe, and David Joyner. "Annotation-free automatic examination essay feedback generation." *2019 IEEE Learning With MOOCS (LWMOOCS)*. IEEE, 2019.

[3]. Liu, Ming, et al. "Automated essay feedback generation and its impact on revision." *IEEE Transactions on Learning Technologies* 10.4 (2016): 502-513.

[4]. Cavalcanti, Anderson Pinheiro, et al. "Automatic feedback in online learning environments: A systematic literature review." *Computers and Education: Artificial Intelligence* 2 (2021): 100027.

[5]. Laarmann-Quante, Ronja. "Towards a Tool for Automatic Spelling Error Analysis and Feedback Generation for Freely Written German Texts Produced by Primary School Children." *SLaTE*. 2017.

[6]. Kochmar, Ekaterina, et al. "Automated personalized feedback improves learning gains in an intelligent tutoring system." *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II 21*. Springer International Publishing, 2020.

[7]. Deeva, Galina, et al. "A review of automated feedback systems for learners: Classification framework, challenges and opportunities." *Computers & Education* 162 (2021): 104094

[8]. Bartoli, Alberto, et al. "Your paper has been accepted, rejected, or whatever: Automatic generation of scientific paper reviews." *Availability, Reliability, and Security in Information Systems: IFIP WG 8.4, 8.9, TC 5 International Cross-Domain Conference, CD-ARES 2016, and Workshop on Privacy Aware Machine Learning for Health Data Science, PAML 2016, Salzburg, Austria, August 31-September 2, 2016, Proceedings*. Springer International Publishing, 2016.

[9]. Lu, Chang, and Maria Cutumisu. "Integrating Deep Learning into an Automated Feedback Generation System for Automated Essay Scoring." *International Educational Data Mining Society* (2021).

[10]. Fossati, Davide, et al. "Data driven automatic feedback generation in the iList intelligent tutoring system." *Technology, Instruction, Cognition and Learning* 10.1 (2015): 5-26.

[11]. Dzikovska, Myroslava, et al. "BEETLE II: Deep natural language understanding and automatic feedback generation for intelligent tutoring in basic electricity and electronics." *International Journal of Artificial Intelligence in Education* 24 (2014): 284-332.

[12]. Hatziapostolou, Thanos, and Iraklis Paraskakis. "Enhancing the impact of formative feedback on student learning through an online feedback system." *Electronic Journal of E-learning* 8.2 (2010): 111-122.

[13]. Carvalho, Justin. "The Design of an Educationally Beneficial Immediate Feedback System." PhD diss., University of Guelph, 2017.

[14]. Keuning, Hieke, Johan Jeuring, and Bastiaan Heeren. "Towards a systematic review of automated feedback generation for programming exercises." *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*. 2016.

[15]. Xiong, Ye, and Yi-Fang Brook Wu. "An Automated Feedback System to Support Student Learning in Writing-to-Learn Activities." *Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale*. 2019.

[16]. Zhang, Jialu, et al. "Automated Feedback Generation for Competition-Level Code." *37th IEEE/ACM International Conference on Automated Software Engineering*. 2022.

[17]. Delmonte, Rodolfo. "Feedback generation and linguistic knowledge in ‘SLIM’automatic tutor." *ReCALL* 14.2 (2002): 209-234.

[18]. Hattie, John, and Helen Timperley. "The power of feedback." *Review of educational research* 77.1 (2007): 81-112.

[19]. Barnes, Tiffany, and John Stamper. "Automatic hint generation for logic proof tutoring using historical data." *Journal of Educational Technology & Society* 13.1 (2010): 3-12.

[20]. Ahea, Md, et al. "The Value and Effectiveness of Feedback in Improving Students' Learning and Professionalizing Teaching in Higher Education." *Journal of Education and Practice* 7.16 (2016): 38-41.

[21]. Taghipour, Kaveh, and Hwee Tou Ng. "A neural approach to automated essay scoring." *Proceedings of the 2016 conference on empirical methods in natural language processing*. 2016..

[22]. Dong, Fei, and Yue Zhang. "Automatic features for essay scoring–an empirical study." *Proceedings of the 2016 conference on empirical methods in natural language processing*. 2016.

[23]. Alikaniotis, Dimitrios, Helen Yannakoudakis, and Marek Rei. "Automatic text scoring using neural networks." *arXiv preprint arXiv:1606.04289* (2016).

[24]. Xu, Weiqi, and Fan Ouyang. "The application of AI technologies in STEM education: a systematic review from 2011 to 2021." *International Journal of STEM Education* 9.1 (2022): 1-20.

[25]. Akçapınar, Gökhan. "How automated feedback through text mining changes plagiaristic behavior in online assignments." *Computers & Education* 87 (2015): 123-130.

[26]. Akçapınar, Gökhan. "How automated feedback through text mining changes plagiaristic behavior in online assignments." *Computers & Education* 87 (2015): 123-130.

[27]. Laarmann-Quante, Ronja. "Towards a Tool for Automatic Spelling Error Analysis and Feedback Generation for Freely Written German Texts Produced by Primary School Children." *SLaTE*. 2017.