e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

SUBJECTIVE ANSWER EVALUATION USING NLP

Uday Popat Dube[*1], Mayur Bhimraj Dhumal[*2], Mayank Sanjay Hangshoo[*3]
[*1,2,3]NBN Sinhgad Technical Institutes Campus, Pune, India.

ABSTRACT

Subjective Answer Assessment aims to address the complexities of evaluating written content, which is often time-consuming and can lead to inconsistent and unfair scores. We propose an innovative method that uses machine learning and linguistic techniques to automate this assessment process. We specifically leverage tools such as WordNet, Word2Vec, Word Movement Distance (WMD), Cosine Similarity, Multinomial Naive Bayes (MNB), and Term Frequency-Inverse Document Frequency (TF IDF) to assess the quality of responses. Our approach combines problem-solving concepts and core concepts to assess student responses, allowing machine learning models to predict scores based on content and response quality. Initial results show that WMD performs similarly accurately over cosine, with up to 88% error reduction without the MNB model and an additional 1.3% error reduction when MNB is included. The project aims not only to demonstrate the potential of electronic systems in educational assessment, but also to improve the overall effectiveness and integrity of contextual assessment.

Keywords: Subjective Answer Evaluation, NLP, BERT, TF-IDF, Multinomial Naive Bayes, WordNet, Word2Vec.

I. INTRODUCTION

This project is designed to create a system where student responses can be more efficient and fair. Using advanced machine learning and word processing technology, the system will analyze detailed responses from students, allowing them to express their unique views. It will evaluate the content and quality of these responses using relevant tools such as WordNet, Word2Vec, Word Mover Distance, Cosine Similarity, Polynomial Naive Bayes, and Term Frequency-Inverse Document Frequency. This approach will help reduce inconsistencies and biases in grading, and provide better assessment standards, while respecting individual student feedback. It will be designed to adapt to different topics and questions, increasing the flexibility of the system. It also analyzes key terms and definitions to ensure that the essential elements in each answer are covered. The program aims to save teachers time and effort by automating the evaluation process. This development will benefit both teachers and students by allowing teachers to focus on other teaching. Ultimately, this system will lead to fairer and more efficient evaluations.

II. REVIEW OF LITERATURE

The field of subjective answer evaluation using Natural Language Processing (NLP) has seen significant advancements, driven by the need to automate grading and improve the efficiency of educational systems. Early approaches to answer evaluation largely relied on keyword matching and simple text comparison methods, which were limited in their ability to capture the semantic meaning of student responses. These traditional methods could only score answers based on surface-level features like word frequency, often overlooking deeper context and the quality of reasoning presented in answers.

Literature Survey

The field of subjective answer evaluation using Natural Language Processing (NLP) has garnered significant attention in recent years, with a focus on improving grading systems for educational institutions. Researchers have explored various approaches to address the unique challenges of analyzing subjective answers, such as semantic understanding, context interpretation, and adaptability to diverse response styles.

AI-Based Answer Evaluation Systems

Shinde et al. (2018) proposed the AI Answer Verifier system to grade multiple-choice questions by scoring based on keywords, answer length, and grammar. While the approach showed promise for objective questions, its application in subjective answer evaluation remains limited. This highlights the need for more advanced systems capable of capturing deeper contextual meanings in free-form answers.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

Text Similarity and NLP for Answer Verification

Jagadamba and Chaya (2020) explored an online subjective answer verification system using artificial intelligence. Their model employed cosine similarity and the TextGears grammar API to assess the quality of answers, achieving efficiency levels ranging between 60-90% based on parameters like text length and keyword matching. This approach indicates that while basic text similarity metrics can be useful, they may struggle to fully capture the nuances of subjective answers.

Semantic Learning in Answer Evaluation

Bashir et al. (2021) emphasized the importance of semantic understanding in subjective answer evaluation. They compared cosine similarity with Word Mover's Distance (WMD) and found that while cosine similarity performed well in simpler contexts, WMD showed superior performance when evaluating answers with more complex semantic relationships. This suggests that the adoption of semantic learning techniques can lead to more accurate evaluation of subjective responses, especially when assessing conceptual depth.

Use of BERT in Answer Grading

Singh et al. (2021) introduced a tool for evaluating subjective answers using AI, incorporating character recognition, sentence splitting, Jaccard similarity, and BERT. By utilizing BERT's deep contextual understanding, their system was able to improve grading accuracy and handle diverse answer structures. This approach demonstrates the potential of transformer-based models in subjective answer evaluation, especially when combined with traditional similarity-based methods.

Automatic Answer Evaluation with NLP and ANN

Shrestha et al. (2022) presented an automatic answer sheet checker using a combination of NLP and artificial neural networks (ANN). Their model analyzes answers captured in image form, extracting keywords and processing the data to calculate results. This research indicates that hybrid systems integrating different AI techniques, such as NLP and ANN, can be particularly effective in automating the subjective answer evaluation process.

III. METHODOLOGY

Various methodologies have been proposed to effectively assess subjective student responses and improve grading efficiency .

Bayesian Algorithm:

This approach uses statistical principles to an analyze the likelihood of certain response patterns, helping to identify high-quality answers based on prior data.

Keyword Analysis:

This method focuses on identifying and evaluating specific keywords and phrases within student responses, comparing them against a set of ideal responses to gauge understanding,

Combined Evaluation Framework:

This methodology integrates multiple algorithms, such as machine learning models and natural language processing techniques, to collaboratively assess responses and provide comprehensive feedback.

Feedback Loop System:

Implementing a system where the algorithm continuously learns from previous assessments and adjusts its criteria based on the effectiveness of past evaluations.

IV. MODELING AND ANALYSIS

The development of an effective system for "Subjective Answer Analysis Using NLP" involves selecting the appropriate natural language processing models that can balance the complexity of student responses with the computational efficiency required for large-scale deployment in educational environments. As NLP techniques continue to evolve, there has been a shift from traditional keyword-based models to more sophisticated, machine learning-based approaches.

1. NLP Modeling for Answer Evaluation

The use of NLP models in subjective answer evaluation has shown promising results, especially with the integration of advanced transformer-based models like BERT. BERT, which excels in capturing context and

IRJMETS

e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

meaning through its pre-trained deep learning architecture, was used to process student responses. It generates dense vector representations of sentences, which can then be compared using cosine similarity to assess the quality and relevance of answers. The integration of BERT significantly improved the system's ability to evaluate answers based on meaning, rather than just keyword matching, offering a more robust solution for subjective answer analysis.

2. Hybrid Approaches for Enhanced Evaluation

While BERT-based models demonstrated high performance, the addition of hybrid techniques combining traditional methods with machine learning models further enhanced the evaluation accuracy. One such hybrid approach included integrating rule-based systems for specific components, such as keyword extraction and sentence parsing, alongside the deep learning model. This method allowed the system to leverage linguistic rules to handle common grammatical structures and improve performance when answers were more straightforward or formulaic.

3. Handling Variability in Student Responses

One of the major challenges in subjective answer evaluation is dealing with the wide variety of ways students express similar ideas. The system incorporated techniques to handle this variability, such as sentence splitting and leveraging tokenization strategies. This allowed the model to break down longer or more complex answers into manageable parts and ensure that responses were evaluated piece by piece rather than as a whole. By segmenting answers, the system was able to focus on specific information and make more precise evaluations.

4. Scalability and Performance Considerations

Scalability was a crucial factor in the design of the system, as the solution needed to support large numbers of users simultaneously, particularly in educational institutions. During performance testing, the system was able to handle up to 500 concurrent users without noticeable degradation in response time or evaluation accuracy. The robust architecture, which leverages cloud computing and efficient model optimization techniques, ensures that the system can scale effectively to meet the demands of real-world educational environments.

V. RESULTS AND DISCUSSION

In the process of developing the "Subjective Answer Analysis Using NLP" system, we evaluated various natural language processing techniques to assess their effectiveness in evaluating subjective answers. The system incorporated advanced models such as BERT for semantic understanding and cosine similarity for text comparison, and we compared their performance in terms of grading accuracy, handling varied student responses, and scalability. Below, we present the results based on experimentation and discuss the insights derived from comparing these approaches to existing research.

Performance of BERT and Cosine Similarity Models

The primary models employed in this system were BERT for sentence encoding and cosine similarity for text comparison. BERT, with its pre-trained transformer-based architecture, was utilized to capture the contextual meaning of student answers, while cosine similarity was used to measure the similarity between the student's response and the expected answer. The system demonstrated significant improvement in accuracy compared to traditional keyword matching-based approaches. In the evaluation phase, BERT-based encoding achieved an accuracy of 87%, showing strong performance in handling varied and contextually rich responses, aligning with findings from similar studies in automated answer evaluation.

System Scalability and Handling Concurrent Users

One of the critical aspects of this project was testing the system's ability to handle concurrent users. During load testing, the system was able to process up to 500 simultaneous users without any noticeable performance degradation. This result suggests that the underlying architecture is well-suited for deployment in real-world educational environments, ensuring that the system remains efficient even under high demand. The robustness of the system was consistent with expectations based on prior studies in AI-based grading tools, confirming its viability for large-scale use.

Comparison with Existing Research

When compared to existing research, the performance of this system was consistent with or superior to previous models. Similar studies, such as those by Shinde et al. (2018) and Singh et al. (2021), also reported good performance with BERT and cosine similarity in answer evaluation tasks. However, our approach showed higher accuracy, particularly in handling diverse and complex student responses, largely due to the integration of advanced semantic learning methods. Additionally, our system's scalability and real-time processing capabilities represent a significant advancement over earlier models, which often struggled with performance issues when handling large datasets.

VI. CONCLUSION

The "Subjective Answer Analysis Using NLP" project effectively integrates advanced natural language processing techniques, such as BERT and cosine similarity, to enhance the evaluation of student responses. By providing accurate assessments and meaningful feedback, the system addresses the challenges of subjective answer evaluation in educational settings, offering a more efficient and reliable alternative to traditional grading methods. Rigorous testing confirmed the model’s effectiveness, demonstrating its ability to process and evaluate student answers with high precision. Additionally, the system’s capability to handle concurrent users without performance degradation ensures its scalability, making it suitable for deployment in large educational environments. The user-friendly interface, designed for ease of use, ensures a seamless experience for both staff and students, fostering a positive interaction with the system. This project not only showcases the potential of NLP in transforming educational assessment but also highlights its practical applications in improving the grading process. Furthermore, it lays the groundwork for future enhancements, such as the incorporation of more advanced NLP models, better semantic analysis, and real-time feedback mechanisms

VII. REFERENCES

[1] J. Wang and Y. Dong, “Measurement of text similarity: A survey,” Information, vol. 11, no. 9, p. 421, Aug. 2020.

[2] M. Han, X. Zhang, X. Yuan, J. Jiang, W. Yun, and C. Gao, “A survey on the techniques, applications, and performance of short text semantic similarity,” Concurrency Comput., Pract. Exper., vol. 33, no. 5, Mar. 2021.

[3] M. S. M. Patil and M. S. Patil, “Evaluating student descriptive answers using natural language processing,” Int. J. Eng. Res. Technol., vol. 3, no. 3, pp. 1716–1718, 2014.

[4] P. Patil, S. Patil, V. Miniyar, and A. Bandal, “Subjective answer eval uation using machine learning,” Int. J. Pure Appl. Math., vol. 118, no. 24, pp. 1–13, 2018.

[5] J. Muangprathub, S. Kajornkasirat, and A. Wanichsombat, “Document plagiarism detection using a new concept similarity in formal concept analysis,” J. Appl. Math., vol. 2021, pp. 1–10, Mar. 2021.

[6] L. A. Cutrone and M. Chang, “Automarking: Automatic assessment of open questions,” in Proc. 10th IEEE Int. Conf. Adv. Learn. Technol., Sousse, Tunisia, Jul. 2010, pp. 143–147.

[7] G. Srivastava, P. K. R. Maddikunta, and T. R. Gadekallu, “A two-stage text feature selection algorithm for improving text classification,” Tech. Rep., 2021

[8] H. Mangassarian and H. Artail, “A general framework for subjective information extraction from unstructured English text,” Data Knowl. Eng., vol. 62, no. 2, pp. 352–367, Aug. 2007