An Approach to Automate the Grading Process in IELTS Writing Task

Ankur Batta
Department of Computer Science Engineering
Baba Banda Singh Bahadur Engineering College, India

Abstract — This paper presents a novel approach to grading IELTS writing tasks using machine learning models, addressing the inherent limitations of the traditional manual grading method. The conventional approach, while widely accepted, suffers from several drawbacks including subjectivity, inconsistency, and scalability. Human graders, despite their expertise, may exhibit bias and variability in their assessments, leading to potential inconsistencies in grading. Furthermore, the manual method is time-consuming and labor-intensive, posing significant challenges in terms of scalability especially in the face of increasing numbers of IELTS test-takers globally. Our proposed machine learning model aims to mitigate these issues by providing an objective, consistent, and efficient method for grading IELTS writing tasks. Preliminary results indicate promising potential for the application of machine learning in this domain, paving the way for future research and practical applications.

Keywords: Grading Process, IELTS Writing Task

I. INTRODUCTION

The International English Language Testing System (IELTS) is a globally recognized test that assesses the English language proficiency of non-native English speakers. One of the key components of this test is the writing tasks, which require candidates to demonstrate their ability to write effectively in English. Traditionally, these tasks have been graded manually by human raters, a process that, while generally reliable, has its share of challenges.

Manual grading of IELTS writing tasks is subject to human bias and variability, leading to potential inconsistencies in scoring. It is also a time-consuming and labor-intensive process, which can be a significant bottleneck in the face of the growing number of IELTS test-takers worldwide. Furthermore, the manual method lacks the ability to provide immediate feedback, which is crucial for learners seeking to improve their writing skills in a timely manner.

In light of these challenges, there is a pressing need for a more efficient, consistent, and objective method for grading IELTS writing tasks. This paper introduces a novel approach that leverages machine learning to address these issues. Our proposed model aims to not only streamline the grading process but also enhance its reliability and objectivity, thereby improving the overall quality and effectiveness of IELTS writing assessment. The following sections will detail the development, implementation, and evaluation of this innovative machine learning model.

II. MACHINE LEARNING IN LANGUAGE ASSESSMENT

Machine Learning (ML) has revolutionized many fields, and language assessment is no exception. It offers a new paradigm for analyzing language data and making predictions or decisions based on patterns identified in these data.

In the context of language assessment, ML can be used to automate and enhance various aspects of the process. For instance, it can be used in automated essay scoring, where ML algorithms are trained on a dataset of manually graded essays and then used to predict the grades of new essays. This not only speeds up the grading process but also ensures consistency in scoring.

A. Principles of Using ML in Language Assessment

The application of ML in language assessment is guided by several key principles.

- Data-Driven Decision Making: Machine Learning models learn from data. Therefore, the decisions they make or the predictions they give are based on patterns identified in the data they were trained on.
- Consistency: One of the advantages of using Machine Learning in language assessment is that it provides consistent results. Unlike human assessors, who may have varying opinions, a Machine Learning model will always make the same prediction for the same input.
- Transparency and Interpretability: It's important that the decisions made by Machine Learning models can be understood and interpreted by humans. This is especially true in educational settings, where feedback is often necessary.
- Fairness and Bias: Care must be taken to ensure that the Machine Learning models do not perpetuate or amplify existing biases in the data. This requires careful design and testing of the models.

B. Algorithm Parameters in Machine Learning for Language Assessment

Machine Learning algorithms often have several parameters that can be tuned to optimize their performance. These parameters, also known as hyperparameters, play a crucial role in the training of the models and can significantly influence their accuracy and efficiency.

- Learning Rate: This is a crucial parameter in many Machine Learning algorithms. It determines the step size at each iteration while moving toward a minimum of a loss function. A smaller learning rate could result in a long training process, whereas a larger learning rate could lead to learning a sub-optimal set of weights too fast or an unstable training process.
- Number of Iterations: This is the number of times the learning algorithm will pass over the entire training dataset. Too few iterations can result in underfitting of the model, while too many iterations can lead to overfitting.
- Regularization Parameter: Regularization is a technique used to prevent overfitting by adding a penalty term to the loss function. The regularization parameter determines the strength of the penalty, and tuning it can help find a good bias-variance tradeoff.
- Model Complexity: The complexity of a model, such as the number of layers in a neural network or the depth of a decision tree, can significantly impact the model's

performance. A more complex model can capture more complex relationships in the data but is also more prone to overfitting and requires more data to train effectively.

C. Loss Function in ML for Language Assessment

One of the key components in training a Machine Learning model is the loss function. It measures how well the model's predictions align with the actual values. In the context of regression problems, a common choice of loss function is the Mean Squared Error (MSE).

The MSE is calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Where y_i is the actual value, $\hat{Y}_i$ is the predicted value, and n is the number of observations. The goal during training is to minimize this error, which would result in a model that makes predictions as close as possible to the actual values.

D. Advantages of Using Machine Learning in Language Assessment

– Efficiency: Machine Learning can automate the process of language assessment, making it faster and more efficient. This is particularly beneficial for large-scale assessments where manual grading would be time-consuming.
– Consistency: Unlike human assessors, who may have varying opinions, a Machine Learning model will always make the same prediction for the same input, ensuring consistency in scoring.
– Personalization: Machine Learning models can be trained on a wide range of data, allowing them to handle diverse language styles and dialects. This makes it possible to provide personalized feedback tailored to each learner's unique language use.
– Predictive Power: Machine Learning models can identify patterns in language use that humans might overlook. This can lead to more accurate predictions of language proficiency or other language-related outcomes.
– Scalability: Machine Learning models can easily be scaled to handle larger amounts of data. As more data becomes available, the models can be retrained to improve their performance.

E. Limitations of Using Machine Learning in Language Assessment

– Data Dependence: Machine Learning models are heavily dependent on the data they are trained on. If the training data is biased or unrepresentative, the model's predictions may also be biased or inaccurate.
– Interpretability: Machine Learning models, especially complex ones like deep learning models, can be difficult to interpret. This lack of transparency can make it hard to understand why the model made a certain prediction.
– Overfitting: Machine Learning models can sometimes fit the training data too closely, to the point where they perform poorly on new, unseen data. This is known as overfitting.
– Computational Requirements: Training Machine Learning models, especially on large datasets, can be computationally intensive and require significant resources.
– Privacy and Ethical Considerations: Using Machine Learning for language assessment involves processing potentially sensitive data, raising privacy and ethical considerations.

III. BUILDING A WEB INTERFACE FOR LANGUAGE ASSESSMENT WITH FLASK

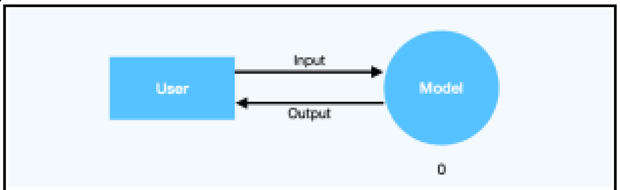
For the deployment of our Machine Learning model in a real-world application, we are using Flask, a lightweight and flexible web server gateway interface (WSGI) web application framework. Flask provides the tools and functionality required to build a web application, which can serve as the interface between the users and our Machine Learning model.

A. Overview of deployment

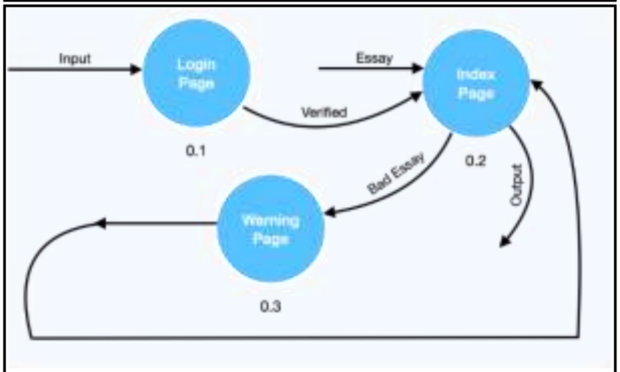
– Web Application: We're building a web application using Flask. This application serves as the front-end interface for users to interact with our Machine Learning model. Users can input their data through this application, and the application will display the model's predictions.
– Model Integration: Our Machine Learning model is integrated into the Flask application. When a user inputs data, the Flask application sends this data to the model, receives the model's predictions, and then displays these predictions to the user.
– Server Deployment: The Flask application is deployed on a server. This allows users to access the application from their web browsers, regardless of their location.
– Scalability: Flask allows for easy scalability, which is crucial when deploying Machine Learning models. As the number of users increases, the Flask application can be scaled up to handle the increased traffic.

B. Data Flow Diagrams

In this section, It is illustrated how data moves through the system.



Level 0



Level 1

IV. CONCLUSION

In this research, we have explored the application of Machine Learning in language assessment, focusing on the use of a transformer-based model. We have discussed the algorithm parameters, advantages, and limitations of this approach. We also presented a detailed description of our pipeline, including data preprocessing, model prediction, and post-processing steps.

Our findings suggest that Machine Learning, particularly deep learning models like BERT, can significantly enhance the efficiency and accuracy of language assessment. However, challenges such as data dependence, interpretability, and computational requirements need to be carefully managed.

Looking ahead, there are several potential directions for future research. One promising avenue is the exploration of other Machine Learning models and techniques. Additionally, more work could be done to improve the interpretability of these models, making them more transparent and trustworthy.

Furthermore, as our work has shown, deploying these models in a real-world application involves additional considerations. Future work could explore more efficient deployment strategies, better user interfaces, and ways to ensure privacy and security.

In conclusion, while there are challenges to using Machine Learning in language assessment, the potential benefits are significant. With further research and development, we believe that Machine Learning can play a crucial role in advancing the field of language assessment.

REFERENCES

[1] Burr Settles, Geoffrey T. LaFlair, Masato Hagiwara "Machine Learning–Driven Language Assessment," 8 247–263, April 2020.
[2] Wen-juan Hou, Jia-Hao Tsao, "AUTOMATIC ASSESSMENT OF STUDENTS' FREE-TEXT ANSWERS WITH DIFFERENT LEVELS," VOL. 20, NO. 02, April 2011
[3] Itziar Aldabe, Oier Lopez de Lacalle, Iñigo Lopez-Gazpio, Montse Maritxalar, "Supervised Hierarchical Classification for Student Answer Scoring," V1, July 2015.