**Report: Grammar Correction Model using Generative AI**

**1. Introduction**

This project focuses on developing a grammar correction model using generative AI techniques. We used both LSTM-based and GPT-2 models to automatically correct grammatical errors in sentences. This report outlines the steps taken, analyzes the models' performance, and provides examples of sentence corrections made by both models.

**2. Data Preprocessing & Exploratory Data Analysis (EDA)**

* **Data Loading:** The dataset was imported, and missing values were checked. Each entry contained an ungrammatical sentence and its corrected version.
* **Tokenization:** Sentences were tokenized using BERT's WordPiece tokenizer, with padding and truncation to ensure uniform input lengths.
* **Dataset Splitting:** The dataset was split into training (80%), validation (10%), and test (10%) sets.
* **Error Analysis:** Various grammatical error types were analyzed, revealing that the dataset contained a mix of tense errors, subject-verb disagreements, and other common grammatical mistakes.
* **Sentence Length Distribution:** Analysis of sentence lengths provided insight into the dataset's complexity and guided the choice of model parameters, like maximum sequence length.

**3. Baseline Model: LSTM-Based Sequence-to-Sequence Model**

* **Feature Engineering:** GloVe embeddings were used to represent words in a continuous vector space.
* **Model Architecture:** The LSTM model employed a bidirectional LSTM layer followed by a TimeDistributed Dense layer for sequence prediction.
* **Training:** The model was trained on the training set with a validation split for monitoring overfitting. The model's training accuracy reached approximately 93%, while validation accuracy was slightly lower.
* **Evaluation:** The model was evaluated using BLEU score, token accuracy, and F1 score. The LSTM model showed a BLEU score close to 0.016 and low accuracy, indicating limited capability in handling grammatical corrections.

**4. Enhancements for the LSTM Model**

To improve the LSTM model's performance, the following strategies are proposed:

1. **Add More LSTM Layers:** Introduce additional LSTM layers to increase the model's capacity for learning complex patterns in the data. A potential architecture could involve stacking 2-3 bidirectional LSTM layers, with each layer containing 256 units to enhance the model's understanding of grammatical structures.
2. **Increase Hidden Units:** Increase the number of hidden units in each LSTM layer (e.g., from 128 to 256 or 512 units). This change will allow the model to capture more intricate patterns in longer sequences.
3. **Incorporate Dropout:** Apply a higher dropout rate (e.g., 0.3-0.5) between LSTM layers to prevent overfitting and improve generalization.
4. **Use Attention Mechanisms:** Integrate attention mechanisms into the model. Attention helps the model focus on relevant parts of the input sequence when generating the output, leading to improved corrections.
5. **Increase Embedding Dimensions:** Enhance the GloVe embedding dimension (e.g., from 100 to 200 or 300) to provide a more detailed word representation.
6. **Train for More Epochs:** Increase the number of training epochs and use early stopping to find the optimal point where the model's performance on the validation set plateaus.
7. **Use Regularization Techniques:** Apply L2 regularization to the LSTM layers to further prevent overfitting.
8. **Experiment with Learning Rates:** Adjust the learning rate to improve the training process. Lower learning rates often result in better convergence for complex models.

Implementing these strategies could significantly improve the LSTM model's accuracy and its ability to generalize grammatical corrections.

**5. Enhanced Model: GPT-2 Fine-Tuning**

* **Fine-Tuning:** A pre-trained GPT-2 model was fine-tuned on the dataset to improve its ability to correct sentences.
* **Experimentation:** Techniques such as beam search were used to enhance text generation quality. Beam size and sequence length were adjusted to improve the performance.
* **Evaluation:** The GPT-2 model demonstrated better performance, achieving a BLEU score of 0.0897, significantly higher than the LSTM model. This suggests that transformer-based models like GPT-2 are more effective for the task of grammar correction.

**6. Error Correction Testing**

* The GPT-2 model's corrections were compared to the baseline LSTM model. While GPT-2 showed improvements in generating coherent and grammatically correct sentences, the LSTM model struggled with longer sequences and more nuanced grammatical errors.

**7. Sample Model Outputs**

Here are examples of sentence corrections generated by both models:

|  |  |  |  |
| --- | --- | --- | --- |
| Ungrammatical Input | LSTM Output | GPT-2 Output | Corrected Reference |
| I goes to the store everyday. | the the the the the the the | I go to the store every day. | I go to the store everyday. |
| They was playing soccer last night. | the the the the the the | They were playing soccer last night. | They were playing soccer last night. |
| She have completed her homework. | the the the the the the the | She has completed her homework. | She has completed her homework. |
| He don't know the answer. | the the the the the the | He doesn't know the answer. | He doesn't know the answer. |
| The sun rise in the east. | the the the the the the | The sun rises in the east. | The sun rises in the east. |

**8. Performance Analysis**

* **LSTM Model:** The LSTM-based model failed to generalize corrections properly, often producing meaningless output ("the the the..."). Its BLEU score was very low, indicating poor grammatical correction performance. The proposed enhancements (adding more LSTM layers, increasing hidden units, etc.) aim to address these shortcomings and improve the model's accuracy.
* **GPT-2 Model:** The GPT-2 model produced more coherent and grammatically accurate sentences, indicating its superiority over the LSTM model for this task. However, its BLEU score (0.0897) suggests there is still room for improvement, particularly with complex grammatical structures.
* **Key Observations:** Transformer-based models like GPT-2 are more effective for sentence correction tasks due to their ability to capture long-term dependencies and contextual information. LSTM-based models struggle with such tasks, especially for longer or more complex sentences.

**9. Recommendations for Future Work**

* **Further Fine-Tuning:** To improve the GPT-2 model's performance, fine-tuning on a larger and more diverse dataset is recommended.
* **Enhanced LSTM Model:** Implement the proposed enhancements to the LSTM model (additional layers, attention mechanism, increased hidden units) to potentially boost its performance.
* **Human Benchmark:** Incorporating human evaluations of generated corrections will provide a more nuanced understanding of the model's accuracy and practicality.
* **Alternative Metrics:** While BLEU scores provide an indication of performance, other metrics like ROUGE or METEOR could offer a more comprehensive evaluation.

**10. Conclusion**

The GPT-2 model demonstrated superior performance in the grammar correction task compared to the LSTM model. However, with the proposed enhancements, the LSTM model's performance could be significantly improved. This work underscores the potential of generative AI in grammar correction and highlights areas for further development in both LSTM and transformer-based approaches.