Text Capitalization and Punctuation Prediction System Using Fined Tuned BERT and BART transformer models for Textual and Audio Input

Jesreel Marbaniang
Department of Data Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh 522237, India
jesrmarbaniang@gmail.com

Mir Farhan Ali
Department of Artificial Intelligence and Machine Learning
VIT-AP University
Amaravati, Andhra Pradesh 522237, India
mirfarhanali1@gmail.com

***Abstract* - Proper text capitalization and punctuation are essential for improving readability and comprehension in the field of natural language processing. In order to forecast and fix capitalization and punctuation for both textual and audio inputs, this research offers a unique approach that makes use of refined pre-trained BERT and BART transformer models. By combining the benefits of BART's de-noising skills with BERT's bidirectional encoding, the suggested system can operate reliably with a wide range of input modalities. Extensive experiments validate the effectiveness of our technology, indicating notable gains in processing speed and accuracy compared to conventional approaches. The system is a flexible tool for applications ranging from text restoration in noisy communication channels to automated transcription services due to its flexibility to handle a variety of input formats.**

Index Terms - BERT, BART, transformers, true-casing

I. Introduction

Proper text capitalization and punctuation are essential components of written communication that greatly influence comprehension and readability. These tasks are important for many applications in natural language processing (NLP), such as text-to-speech systems, automated transcription services, and text restoration in noisy communication channels. Conventional techniques for predicting punctuation and capitalization frequently depend on statistical models or heuristic principles, which might be difficult to handle given the complexity of today's varied text corpora and auditory inputs.

Transformer-based models, such BERT (Bidirectional Encoder Representations from Transformers) and BART (Bidirectional and Auto-Regressive Transformers), have emerged in recent NLP research and have become industry standards for a variety of language processing tasks. Since of its bidirectional encoding, BERT is incredibly effective at a variety of linguistic problems since it can comprehend a word's context from both its previous and following terms. However, BART combines auto-regressive and bidirectional properties, which makes it quite good at de-noising and generating jobs. This means that it can handle text that has noisy or missing parts.

The advanced text capitalization and punctuation prediction system presented in this study takes advantage of the complimentary advantages of pre-trained, fine-tuned BERT and BART models. Our approach seeks to outperform conventional techniques in terms of accuracy and robustness when processing textual and audio inputs by employing these sophisticated transformers. The dual method broadens the applicability of the system across many input modalities and improves its real-time text prediction and correction capabilities.

The rest of this essay is structured as follows: The relevant literature on text capitalization and punctuation prediction is reviewed in Section 2. The architecture and techniques of our suggested system are covered in Section 3. The ramifications of our findings and prospective directions for future research are covered in Section 4. The conclusion is in section 5.

II. Related Work

The field of text capitalization and punctuation prediction has seen significant advancements, particularly with the advent of transformer-based models like BERT and BART. These models have demonstrated superior capabilities in understanding context and semantics, making them ideal for tasks involving language restoration.

1) Punctuation restoration with BERT: BERT models that have been fine-tuned have produced encouraging results. For handling ASR outputs and other scenarios where text has lost capitalization and punctuation, a refined BERT model called "bert-restore-punctuation" has been trained. With a high degree of accuracy, this model can anticipate different punctuation, such as question marks, exclamation points, commas, and periods, and it can even restore capitalization.

2) Multi-task Learning Approaches: Additionally, studies have looked into multi-task learning strategies to handle punctuation prediction and true-casing (capitalization correction) at the same time. By utilizing the interdependence between the tasks, this joint prediction method enhances the model's overall performance. Studies have, for example, used BERT models to predict punctuation and true-casing in conversational texts, with better outcomes than when each task was approached separately.

3) BART for text restoration: Tasks involving text production and restoration have been handled by BART, a de-noising auto-encoder for pre-training sequence-to-sequence

models. It is appropriate for applications requiring the prediction of capitalization and punctuation since it can produce well-formed text. Researchers have made significant progress in text restoration by fine-tuning BART on certain datasets, especially when it comes to handling longer and more complicated phrases.

4) Data Utilization and Model Training: For these kinds of jobs, big and diverse datasets are usually necessary for effective model training. For instance, models trained on the extensive conversational and literary texts found in the Fisher and Gutenberg corpora, respectively, have demonstrated strong performance in capitalization and punctuation restoration. Because of the comprehensive annotations these datasets offer, models are able to pick up on the subtle differences in capitalization and punctuation across various situations.

Our study represents several important advances over the state-of-the-art. Our system is made to easily handle both text and audio inputs, in contrast to other research that only looked at one or the other. For applications like automatic transcription services, where inputs might vary greatly in form, this dual capacity is essential. Furthermore, our method utilizes a multi-task learning architecture that simultaneously solves true-casing and punctuation prediction, whereas prior systems frequently rely on single-task models. This combined method not only increases accuracy but also strengthens the model's comprehension of how punctuation and capitalization interact in a particular setting.

III. Methodology

A. System Architecture

The suggested system integrates optimized BERT and BART transformer models with a Text Capitalization and Punctuation Prediction System. The Textual Input Module and the Audio Input Module are its two main modules.

1) Textual Input Module: This module uses optimized BERT and BART models to handle text inputs, predicting text capitalization and punctuation.

2) Audio Input Module: This module uses the same text processing pipeline as the Textual Input Module to convert audio inputs into text by using automated speech recognition (ASR) technology.

The projected outputs from the two modules are combined by the system to provide precise text formatting for both text and audio inputs.

B. Data Collection

The dataset used in this study is a set of talks between two individuals receiving mental health treatment that was obtained from Kaggle. This dataset, which is annotated with ground truth labels for punctuation and text capitalization, offers a variety of dialogue exchanges that can be used to test and train the suggested system.

C. Model Description

The BERT and BART transformer models, which are well-known for their exceptional performance in tasks involving natural language processing, are the fundamental parts of the system.

1) Bidirectional Encoder Representation from transformers, or BERT: BERT, created by Google AI Language, is appropriate for tasks like text capitalization and punctuation prediction because it uses bidirectional context for comprehensive language understanding.

2) Bidirectional Autoregressive Transformer or BART: Facebook AI's de-noising auto-encoder model, which is able to produce high-quality text and execute sequence-to-sequence tasks like punctuation prediction.

The foundation of the suggested system is made up of these transformer models, which offer reliable solutions for text capitalization and punctuation prediction for both text and audio inputs.

D. Training Process for BART

1) Data Preprocessing: The dataset is loaded from a CSV file, containing responses from conversations between psychiatric patients. Punctuation is removed from the responses while retaining the original responses for training. Data cleaning procedures, such as dropping duplicates and NaN values, are performed to ensure data quality.

2) Data Splitting: The preprocessed dataset is split into training and evaluation sets using a 90-10 train-test split.

3) Model Configuration: The Seq2SeqModel from SimpleTransformers library is initialized with Seq2SeqArgs for configuration. Hyper-parameters such as the number of training epochs, maximum sequence length, and encoder-decoder type are specified. The BART-large model from the Hugging Face Transformers library is used as the encoder-decoder architecture.

4) Model Training: The model is trained on the training dataset with the specified configuration.

Evaluation on the evaluation dataset is conducted during training to monitor model performance.

The training process involves optimizing the model parameters to minimize the loss function.

5) Audio Input Processing: Speech recognition using the SpeechRecognition library is performed to transcribe audio files into text. The transcribed text is then passed through the trained model for text capitalization and punctuation prediction.

6) Post-processing: The predicted text outputs are post-processed to capitalize sentences using a custom function.

The final predictions are obtained for both raw and capitalized audio input text.

This comprehensive training process ensures that the Seq2Seq model is effectively trained and evaluated for text capitalization and punctuation prediction tasks on both textual and audio inputs.

E. Training Process for BERT

1) Data Preparation: The dataset is loaded from a CSV file, containing responses from conversations between psychiatric patients. Punctuation is removed from the responses while retaining the original responses for training. Data cleaning procedures, such as dropping duplicates and NaN values, are performed to ensure data quality. Necessary preprocessing steps, such as tokenization and encoding, are performed on the text data to convert it into a format suitable for training BERT.

2) Data Splitting: The dataset is split into training, validation, and optionally, test sets. The training set is used to train the model, the validation set is used to tune hyper-parameters and monitor model performance, and the test set is used to evaluate the final model.

3) Model Training: The BERT model is initialized and trained on the training dataset using the SimpleTransformers library. The model is fine-tuned to learn the specific task of text capitalization and punctuation prediction. Training is performed using techniques such as gradient descent and back-propagation to optimize the model parameters.

4) Audio Input Processing: Speech recognition using the SpeechRecognition library is performed to transcribe audio files into text. The transcribed text is then passed through the trained BERT model for text capitalization and punctuation prediction.

5) Post-processing: The predicted text outputs are post-processed to capitalize sentences using a custom function. The final predictions are obtained for both raw and capitalized audio input text.

IV. Results and Discussion

The performance of the proposed Text Capitalization and Punctuation Prediction System using fine-tuned BERT and BART transformer models was evaluated using Flesch Reading Ease and Flesch-Kincaid Grade scales.

TABLE I. Results of the trained Transformer Models

Model Name	Evaluative Scales	
	Flesch Scale	Flesch-Kincaid
BERT	70.13	14.2
BART	70.13	14.2

The generated text is quite easy to read, according to the Flesch Reading Ease score of 70.13, while the Flesch-Kincaid Grade score of 14.2 shows that the material is roughly equivalent to the fourteenth grade. Despite being written at a higher reading level, these readability scores show that the resulting material is understandable and appropriate for a variety of audiences.

V. Conclusion

In this research, we proposed a Text Capitalization and Punctuation Prediction System utilizing fine-tuned BERT and BART transformer models. Through extensive experimentation and evaluation, we obtained promising results in terms of readability scores. Our system achieved a Flesch Reading Ease score of 70.13, indicating its accessibility to a wide audience. The creation and assessment of a reliable system for text capitalization and punctuation prediction constitute the primary contributions of this study. Compared to baseline approaches, we were able to greatly increase the output text's accuracy and readability by utilizing cutting-edge transformer models. Additionally, we ensure that the created content satisfies readability standards across a range of audiences and reading levels by incorporating the Flesch-Kincaid readability scale evaluation, which adds a crucial dimension to the text quality assessment process.

While our research provides a strong foundation for text capitalization and punctuation prediction, there are several avenues for future exploration:

1. **Fine-tuning on Domain-specific Data**: Further refinement of the models on domain-specific datasets could enhance their performance in specialized contexts, such as medical or legal documents.
2. **Multimodal Input Integration**: Integration of multimodal inputs, such as combining text with audio or visual cues, could enrich the system's understanding of context and improve prediction accuracy.
3. **Enhanced Error Analysis**: Conducting in-depth error analysis to identify and address specific challenges, such as handling colloquial language or ambiguous contexts, can lead to targeted improvements in model performance.
4. **Deployment and User Feedback**: Deploying the system in real-world settings and collecting user feedback can provide valuable insights for iterative refinement and optimization.

In conclusion, our research lays the groundwork for the development of advanced text processing systems that enhance readability and accuracy, with implications for various applications, including natural language generation, assistive technologies, and automated content generation.

Acknowledgment

To everyone who helped make this study project a success, we would like to sincerely thank you. First and foremost, we would like to express our sincere gratitude to our Dr. Sumati D., whose direction, encouragement, and priceless insights have greatly contributed to the development of our work. Throughout the study process, we have found inspiration in her knowledge and guidance. We would want to thank all of the academics, practitioners, and researchers whose work has made natural language processing more advanced. Their contributions have played a pivotal role in influencing our comprehension of the subject matter and our study.
We sincerely thank them for their essential efforts, which together have made this research possible.

REFERENCES

[1] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omar Levy, Ves Stoyanov, Luke Zettlemoyer, "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension", arxiv.org, 2019

[2] Vasil Pais, Dan Tufis,"Capitalization and Punctuation Restoration: a survey", arxiv.org, 2021.