Exploring Extensions to Dream model: Domain Adaptation, and Model Exploration

1st Ali Ghorbanpour
Department of Control and Computer Engineering
Politecnico di Torino
Torino, Italy
s309992@studenti.polito.it

2nd Sina Hamedani
Department of Control and Computer Engineering
Politecnico di Torino
Torino, Italy
s315495@studenti.polito.it

3rd Kosar Amini
Department of Control and Computer Engineering
Politecnico di Torino
Torino, Italy
s314979@studenti.polito.it

***Abstract*—The challenge of enhancing commonsense reasoning and figurative language understanding in natural language processing (NLP) models remains a significant research problem. This project aims to address this issue by extending the capabilities of the DREAM model. We introduce the DREAM-FLUTE method, which integrates the DREAM model with T5-BASE and BERT models to improve performance on the CommonSense QA task. The methodology involves two primary systems: DREAM-FLUTE System 1, which fine-tunes BERT on the COS-E dataset to bolster baseline performance, and DREAM-FLUTE System 2, which utilizes the DREAM model to generate detailed elaborations. These elaborations are then incorporated into the T5-BASE model to enhance prediction precision. The COS-E dataset, with its comprehensive annotations and explanations for commonsense reasoning, serves as the foundation for training and evaluation. Experimental results reveal that the addition of elaborations significantly enhances the accuracy of the T5 model, while the BERT model shows minimal improvement. These findings underscore the potential of detailed context generation in advancing commonsense reasoning in NLP models. Further experimentation on various models and datasets is recommended to substantiate these results. The implementation code is available for further exploration and refinement.**

***Index Terms*—Commonsense Reasoning, Figurative Language Understanding, Natural Language Processing (NLP), Question Answering (QA), Model Fine-Tuning, Scene Elaboration**

I. Problem Statement

The challenge of natural language processing (NLP) models in understanding and analyzing linguistically contextual texts is a multifaceted issue that has garnered significant attention in recent research. One major shortcoming is the difficulty of incorporating extra-linguistic context, such as the identity of dialogue participants, into pre-trained language models like BERT. While BERT-based models can improve linguistic contextualization, they fail to integrate extra-linguistic information effectively, particularly in tasks requiring an understanding of who is speaking to whom [1]. Similarly deep neural networks have shown limitations in terms of providing context-based explanations for predictions,stemming from the discrete nature of input words and their lack of contextual verbosity [2]. Additionally,it has been shown that although contextual word representations derived from large-scale models are successful in many tasks, they struggle with tasks requiring fine-grained linguistic knowledge, such as conjunct identification and entity representations [3]. These studies collectively underscore the current models' inadequacies in fully leveraging both linguistic and extra-linguistic contexts, indicating a need for more sophisticated methods to address these gaps.

The current approaches and methodologies in NLP models for processing and comprehending linguistically contextual inputs involve several advanced techniques aimed at improving contextual understanding. Additionally, there is a focus on the incremental accumulation of linguistic context, where models generate concise summaries of previous information incrementally. This approach, which contrasts with the large-scale parallel processing of LLMs, has shown better predictions of neural activities related to narrative processing [4]. Moreover, the integration of context-specific linguistic models, as seen in patents, highlights the utility of combining context-free and context-specific models to generate accurate linguistic processing results, which are crucial for training robust NLU models [5, 6].

The T5 model, introduced as the Text-to-Text Transfer Transformer, is a versatile framework designed to convert all NLP tasks into a unified text-to-text format. This design allows for the same model, loss function, and hyperparameters to be utilized across various tasks such as translation, summarization, and question answering [7]. The T5 architecture scales up significantly, reaching up to 11 billion parameters, thus achieving state-of-the-art performance in many benchmarks. It leverages the Colossal Clean Crawled Corpus (C4), a massive dataset of English text scraped from the web, which aids in pre-training the model on a broad range of linguistic contexts.The DREAM model builds upon this robust framework. Specifically, DREAM, or Dynamically wRitten ElAborations to iMprove question-answering, uses the T5-11B model as its base to enhance situational question answering by generating detailed scene elaborations. DREAM is trained to provide additional context by elaborating on scenes described in situational questions, focusing on elements such as motivations, emotions, rules of thumb, and consequences [8]. By doing so, DREAM significantly improves the accuracy and consistency of downstream QA models like Macaw, which also relies on the T5 architecture. These scene elaborations help the QA model to form a more coherent and detailed mental picture of the scenario, thus improving its reasoning and answer accuracy beyond simple fine-tuning.

The FLUTE (Figurative Language Understanding through Textual Explanations) dataset is designed to tackle the challenge of understanding figurative language in NLP tasks, such as sarcasm, similes, metaphors, and idioms. The dataset consists of pairs of literal and figurative sentences, each annotated with entailment or contradiction labels and explanations. FLUTE leverages a model-in-the-loop approach, combining GPT-3 for initial paraphrase generation and minimal human edits for quality control, resulting in a high-quality, diverse dataset that improves the performance of language models on figurative language tasks [9]. The DREAM-FLUTE model builds upon the DREAM framework, which utilizes the T5-based sequence-to-sequence model to generate elaborations for the premise and hypothesis separately. This elaboration aids in forming a "mental model" of the situations, enhancing the model's ability to detect entailment and contradictions in figurative language [10]. DREAM-FLUTE integrates FLUTE's context-specific elaborations, such as consequence, emotion, motivation, and social norms, to improve both label prediction accuracy and explanation quality. The ensemble system of DREAM-FLUTE further enhances performance by implementing a cognitive continuum that considers various levels of intuition and analysis, demonstrating significant improvements in the figurative language understanding task [10].

The DREAM-FLUTE model, designed for understanding figurative language, utilizes the FLUTE dataset which consists of pairs of literal and figurative sentences annotated with entailment or contradiction labels and explanations. The model enhances its performance by generating detailed scene elaborations, such as consequences, emotions, motivations, and social norms, to form a "mental model" of the situations described in the text [10]. The CommonSense dataset (CoS-E) could be a compatible substitute for the FLUTE dataset in the DREAM-FLUTE model due to its comprehensive annotations and explanations focused on commonsense reasoning. Both datasets aim to enhance understanding and reasoning by providing context-rich explanations.

Integrating CoS-E into DREAM-FLUTE would involve adapting the model to utilize CoS-E's natural language explanations to form the mental models necessary for entailment and contradiction detection. The approach would leverage CoS-E's explanations to enhance the quality of scene elaborations, focusing on the implicit commonsense reasoning aspects. Given the effectiveness of DREAM-FLUTE in improving performance through detailed context generation, the integration of CoS-E could potentially yield similar or improved results in understanding figurative language by offering rich, explanatory contexts that aid in discerning the underlying meanings of figurative expressions. The reflective component added by CoS-E's explanations could further enhance the accuracy and explanation quality, as demonstrated in other commonsense reasoning tasks [10, 11]

A. Expected Input

The expected input for the DREAM model consists of questions. Additionally, for enhanced understanding, the model utilizes elaborations generated by the DREAM model to provide extra context. These elaborations include details about the social norm related to the input texts. For instance, an input could be formatted as `<question>` `<question-elaboration-from-DREAM>` , where the elaborations help in forming a more detailed mental model of the situation described by the questions.

B. Addressed Task

The primary task addressed by the study is understanding commonsense. The goal is twofold: first, to classify choices ; second, to generate a textual explanation that clarifies the answer. This task is challenging. By leveraging scene elaborations, the model aims to improve its understanding and reasoning about the underlying meaning of expressions.

C. Expected Output

The expected output includes both a classification label and a textual explanation. The classification label indicates the answer of question.The explanation provides a rationale for this classification, helping to clarify the relationship between the question and answer. This dual output helps in understanding not only the model's decision but also the reasoning behind it.

II. Methodology

A. Overview of the NLP Pipeline/Architecture

Applying the DREAM-FLUTE Method using the DREAM Model, T5-BASE, and BERT Fine-Tuning on the CommonSense Question Answering Task. The DREAM-FLUTE method leverages the capabilities of the DREAM model, T5-BASE, and BERT models to improve performance on the CommonSense Question Answering (QA) task; The application of this approach is as follows:

- DREAM-FLUTE SYSTEM 1 is designed to fine-tune the dataset specifically for the CommonSense QA task.
- DREAM-FLUTE SYSTEM 2 focuses on generating detailed elaborations to enhance the precision of the model's answers.

B. Description of each Module

In this section, we explore the application of the DREAM-FLUTE Method utilizing the DREAM Model, T5-BASE, and BERT Fine-Tuning for the CommonSense Question Answering (QA) task. The DREAM-FLUTE approach harnesses the combined strengths of these models to boost performance in the QA domain. The strategy is executed through the following modules:

- DREAM-FLUTE SYSTEM 1
  - Model Selection The BERT model, a transformer-based model known for its effectiveness in natural language understanding tasks, is selected for fine-tuning on the COS-E dataset.
  - Data-set Preparation The COS-E dataset is prepared for training by preprocessing the input data. This involves tokenizing the context, question, and choices, and encoding them into a format suitable for the BERT model.
  - Fine-tuning Process BERT is fine-tuned on the COS-E dataset. During fine-tuning, the model learns to predict the correct answer from multiple choices based on the provided context and question. The training process adjusts the model's weights to minimize the loss, improving its ability to understand and answer common sense questions accurately.
  - Objective The primary objective of System 1 is to improve the model's ability to correctly select the right answer from the multiple choices given the context and the question. This fine-tuning process enhances the model's baseline performance on the CommonSense QA task.
- DREAM-FLUTE SYSTEM 2
  - Model Selection The T5-BASE model, a versatile text-to-text transformer model, is used in System 2. T5-BASE is known for its ability to generate coherent and contextually relevant text.
  - Elaboration Generation The DREAM model is used to generate elaborations on the questions. These elaborations provide additional context, detail, and insights into the question, helping the model better understand the underlying common sense reasoning required to answer the question accurately.
  - Integration with T5-BASE The generated elaborations are fed into the T5-BASE model along with the original context, question, and choices. This combined input helps the model to utilize the additional context provided by the elaborations to make more precise predictions.
  - Fine-Tuning with Elaboration T5-BASE is fine-tuned on the dataset with these enhanced inputs. The model learns to generate explanations that are not only accurate but also detailed and contextually relevant, improving the overall precision of the answers.
  - Objective The goal of System 2 is to improve the precision of the model's predictions by providing detailed elaborations that enrich the input data. This leads to a deeper understanding of the questions and better performance on the CommonSense QA task.

III. Experiments

A. Data Description

The COS-E (Common Sense Explanation) dataset is designed to evaluate and improve models' ability to perform common sense reasoning with the addition of natural language explanations. Here are the key aspects of the dataset:

B. Experimental Design

This section outlines the tests to evaluate the Dream-Flute method's performance, detailing the datasets setup, computational resources, and performance metrics. Results are analyzed within the NLP context, supported by tables.

1) **Examples**: Each example in the COS-E dataset comprises a context, a question, a set of multiple-choice answers, and an explanation for the correct answer.
2) **Fields**:
   - **Context**: A brief scenario or passage that provides background information.
   - **Question**: A question related to the context, requiring common sense reasoning to answer.
   - **Choices**: A list of multiple-choice answers (typically 5), from which the correct answer needs to be selected.
   - **AnswerKey**: The correct answer indicated by a label such as 'A', 'B', 'C', 'D', or 'E'.
   - **Explanation**: A textual explanation providing the rationale behind the correct answer.

1) Purpose: The primary goal of the COS-E dataset is to push the boundaries of natural language understanding models, particularly in the domain of common sense reasoning. By providing not only the correct answer but also a detailed explanation, the dataset encourages the development of models that can mimic human-like reasoning processes and justify their answers.

2) Usage:

1) **Training**: The dataset is used to train models on understanding context, formulating answers, and providing explanations for their choices.
2) **Evaluation**: It serves as a benchmark for evaluating how well models can integrate common sense knowledge into their reasoning processes and generate meaningful explanations.

3) Example:

- **Context**: "Sam had a busy day at work and forgot to eat lunch."
- **Question**: "What might Sam be feeling at the end of the day?"
- **Choices**: ['A. Happy', 'B. Hungry', 'C. Energetic', 'D. Relaxed', 'E. Indifferent']
- **AnswerKey**: 'B'

- **Explanation**: "Sam is likely feeling hungry because he forgot to eat lunch."

By incorporating both the reasoning process and the explanation, the COS-E dataset provides a comprehensive framework for developing and evaluating advanced natural language understanding models.

C. Execution Times

The experiment was designed to generate elaborations from Dream on the COS-E dataset to improve commonsense understanding. T5-base and BERT models were then fine-tuned on the COS-E dataset. Utilizing Google's RTX 6000 GPU, each pre-training and fine-tuning cycle took around one hour, but generating from the Dream model took around six hours. Model performance was evaluated using the accuracy@0 metric.

1) Results and Analysis:

Model	**Accuracy@0**
T5-base	0.3604
T5-base with elaboration	0.4046
BERT	0.2039
BERT with elaboration	0.2048

The results indicate that elaboration is effective for the T5 model and increases the accuracy, but there is no significant change for the BERT model. Further tests are needed on other models and datasets to draw more comprehensive conclusions.

2) Code Availability: The implementation code for our CharBERT extensions is available on: https://github.com/hamedanisina/figurative-language-understanding

REFERENCES

[1] I.-T. Sorodoc, L. Aina, and G. Boleda, "Challenges in including extra-linguistic context in pre-trained language models," in *Tafreshi S, Sedoc J, Rogers A, Drozd A, Rumshisky A, Akula A, editors. The Third Workshop on Insights from Negative Results in NLP (Insights 2022): proceedings of the Workshop; 2022 May 26; Dublin, Ireland.[Stroudsburg]: Association for Computational Linguistics; 2022. p. 134-8.* ACL (Association for Computational Linguistics), 2022.

[2] X. Yu, F. Dalvi, N. Durrani, and H. Sajjad, "Latent concept-based explanation of nlp models," *arXiv preprint arXiv:2404.12545*, 2024.

[3] N. F. Liu, M. Gardner, Y. Belinkov, M. E. Peters, and N. A. Smith, "Linguistic knowledge and transferability of contextual representations," *arXiv preprint arXiv:1903.08855*, 2019.

[4] N. Maheswaranathan and D. Sussillo, "How recurrent networks implement contextual processing in sentiment analysis," *arXiv preprint arXiv:2004.08013*, 2020.

[5] R. Tikochinski, A. Goldstein, Y. Meiri, U. Hasson, and R. Reichart, "Incremental accumulation of linguistic context in artificial and biological neural networks," *bioRxiv*, pp. 2024–01, 2024.

[6] J.-F. Lavallee and K. W. Smith, "Processing natural language text with context-specific linguistic model," Oct. 23 2018, uS Patent 10,108,603.

[7] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of machine learning research*, vol. 21, no. 140, pp. 1–67, 2020.

[8] Y. Gu, B. D. Mishra, and P. Clark, "Dream: Improving situational qa by first elaborating the situation," *arXiv preprint arXiv:2112.08656*, 2021.

[9] T. Chakrabarty, A. Saakyan, D. Ghosh, and S. Muresan, "Flute: Figurative language understanding through textual explanations," *arXiv preprint arXiv:2205.12404*, 2022.

[10] Y. Gu, Y. Fu, V. Pyatkin, I. Magnusson, B. D. Mishra, and P. Clark, "Just-dream-about-it: Figurative language understanding with dream-flute," *arXiv preprint arXiv:2210.16407*, 2022.

[11] N. F. Rajani, B. McCann, C. Xiong, and R. Socher, "Explain yourself! leveraging language models for commonsense reasoning," *arXiv preprint arXiv:1906.02361*, 2019.