Event Reasoning with Explicit Time and Space

1st Ananth Rangarajan
1225285633
arangar9@asu.edu

2nd Tanmai Mukku
1229609006
tmukku@asu.edu

3rd Nikhil Bodela
1229975320
nbodela@asu.edu

4th Devadutt Sanka
1225362138
dsanka@asu.edu

5th Roop Sumanth Gundu
1226662413
rgundu2@asu.edu

***Abstract*—In the realm of Natural Language Processing (NLP), the comprehension and interpretation of textual data are pivotal. One fundamental aspect of this comprehension involves the analysis of events, which are specific occurrences or incidents that transpire at distinct times and locations. These events are intimately intertwined with temporal and spatial information, where the former relates to the timing of an event, and the latter pertains to the physical space where it unfolds. For an intelligent NLP system to truly understand and reason about text, it must be equipped with the capabilities to decipher and manipulate these temporal and spatial intricacies. In this context, we propose an NLP task that merges event reasoning with the precise handling of temporal and spatial changes. This task revolves around the analysis of short stories that feature various events, each accompanied by explicit temporal and spatial details. The challenge lies in successfully answering questions that demand adept temporal and spatial reasoning, thus pushing the boundaries of natural language understanding and cognitive reasoning in artificial intelligence.**

***Index Terms*—Huggingface, Transformers, BERT, RoBERTa**

I. Problem Statement

The problem we aim to address in this research pertains to the nuanced and intricate nature of event comprehension in natural language text. Events, being the building blocks of narratives, are central to understanding the dynamics of a story or description. These events are not isolated occurrences; they are intricately linked with both temporal and spatial dimensions. Temporal information encapsulates the when of an event, signifying the exact moment in time when it unfolds. Spatial information, however, encapsulates the where, signifying the specific location or space where an event takes place. The challenge arises in that, in order to fully grasp the context and meaning of a text, an intelligent NLP system must not only identify these events but also possess the ability to reason about their temporal and spatial attributes.

In the current NLP landscape, the capabilities of large language models (LLMs) such as GPT-4 and ChatGPT have demonstrated a significant stride in reasoning about the temporal and spatial information embedded in textual contexts. These models, boasting extensive architectures and a plethora of hyperparameters, have exhibited a remarkable ability to discern and comprehend the intricate details of events in a text. Their capacity to perform temporal and spatial reasoning is rooted in their capability to capture subtle nuances, contextual relationships, and temporal sequences, all of which are essential for a comprehensive understanding of narratives. Conversely, smaller models like the vanilla BERT and RoBERTa, while undoubtedly effective in various NLP tasks, often fall short in terms of their capacity to learn and reason about temporal and spatial intricacies within the text. The primary limitation lies in the sheer difference in the scale of architecture and the number of hyperparameters used during the training of these models. LLMs, like GPT-4 and ChatGPT, are designed with a more extensive focus on context modeling, enabling them to grasp the underlying temporal and spatial dynamics within events.

The primary objective of this project is to enhance the temporal and spatial reasoning capabilities of smaller language models, such as vanilla BERT and RoBERTa. To achieve this goal, the project focuses on the crucial step of training these models with an extensive and diverse data corpus. By exposing these smaller models to a larger and more varied set of textual contexts, we aim to empower them with the capacity to effectively reason about temporal and spatial information. This targeted training approach seeks to bridge the gap in performance between smaller models and their larger counterparts, such as GPT-4 and ChatGPT, ensuring that even the more compact NLP models can adeptly handle temporal and spatial intricacies within textual data.

II. Approach

In this research project, we outline a systematic approach designed to empower smaller language models, specifically BERT/RoBERTa, with the ability to proficiently reason about temporal and spatial information within natural language text. To begin, we employ GPT-4, a state-of-the-art large language model, to generate topics that inherently involve temporal and spatial dimensions. These topics serve as the foundation for the subsequent steps in our approach. In the next phase, we prompt GPT-4 to create succinct narratives, each encompassing 5 to 6 lines, which describe events while providing explicit temporal and spatial details. Additionally, we instruct GPT-4 to craft questions, generate answer choices, provide the correct answer, and furnish a reasoning component for each generated narrative, thus creating a comprehensive dataset.

This dataset contains approximately 3000 data points, with each entry consisting of a narrative, a corresponding question, multiple answer choices, the correct answer, and reasoning. By encompassing a wide array of topics and scenarios, this dataset ensures the diversity and robustness required for effective training. To validate the quality and alignment of the generated data with the temporal and spatial objectives of the project, we subject a random 10% sample of the dataset to manual annotation, which serves as a critical quality control measure, and also as a validation dataset. Once we ascertain the fidelity of the dataset, we proceed to train smaller language models such as BERT and RoBERTa using this extensive corpus, aiming to equip them with the capability to comprehend and reason about temporal and spatial intricacies within textual contexts. Subsequently, the trained models undergo a fine-tuning process, refining their accuracy and performance in tasks associated with temporal and spatial reasoning. This iterative step is vital to ensure that the models reach a level of proficiency aligned with the objectives of the project.

III. Individual Contributions

Ananth - contributing significantly to various aspects of the endeavor. Instrumental in generating a diverse array of topics using GPT-4, establishing the foundation for subsequent research. Generated a substantial dataset comprising 1000 data points, with each data point encompassing a comprehensive narrative, questions, answer choices, answers, and reasoning components. This extensive dataset spanned a wide spectrum of topics, including Natural and Environmental Events, Sports and Athletic Events, Entertainment and Media Events, Travel and Tourism Events, and Population and Demographic Events, with 200 data points dedicated to each category. Researched the most suitable smaller models for this research, ultimately recommending the utilization of Longformers [6] or AutoModelForMultipleChoice [7] models for the task at hand. Performed essential preprocessing techniques on a sample dataset, aligning it with the suitable format required for training the transformer models.

Tanmai - I've generated a diverse dataset using OpenAI's GPT-4 model. This dataset includes stories from various domains such as history, social movements, culture, religion, and arts. I've cleaned and preprocessed this data, ensuring it's stored in a JSON format for ease of future use. Rigorous validation was done to keep only relevant samples. Parallelly, I've begun a thorough literature review, exploring significant NLP conference papers like ACL, NAACL, and EMNLP, to gain insights into event reasoning. My next step involves training models like BERT and RoBERTa with our dataset to improve their understanding of events based on spatiotemporal information.As a starting point, I'm exploring hyperparameter tuning and considering the integration of attention mechanisms to better capture temporal relationships in the data.

Devadutt Sanka - I used the OpenAI API to generate a dataset comprising nearly 1000 data points, incorporating around 20 diverse event types such as Health and Medical, Educational and Academic, Arts and Entertainment, Cultural and Heritage events, and many more. This was achieved through the utilization of varied prompts for questions, answers, and reasoning. I ensured the data's uniqueness, extracted the necessary components, and formatted everything in JSON. My responsibilities also included closely working with my colleagues, Ananth and Tanmai, to seamlessly integrate our individual contributions and create a synergistic work environment. Looking ahead, my focus will be on refining our dataset and commencing the model training phase. I plan to apply my expertise in NLP and machine learning to drive our project toward its research goals, building on our strong foundation and clear future vision. I am fully committed to contributing to the ongoing success and innovation of our project.

References

[1] Zhou, Ben, Richardson, Kyle, Ning, Qiang, Khot, Tushar, Sabharwal, Ashish, and Roth, Dan. Temporal Reasoning on Implicit Events from Distant Supervision. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1361–1371, Online, June 2021. Association for Computational Linguistics. https://aclanthology.org/2021.naacl-main.107. DOI: 10.18653/v1/2021.naacl-main.107.

[2] Han, Rujun, Hsu, I-Hung, Sun, Jiao, Baylon, Julia, Ning, Qiang, Roth, Dan, and Peng, Nanyun. *ESTER*: A Machine Reading Comprehension Dataset for Reasoning about Event Semantic Relations. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 7543–7559, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. https://aclanthology.org/2021.emnlp-main.597. DOI: 10.18653/v1/2021.emnlp-main.597.

[3] Spiliopoulou, Evangelia, Pagnoni, Artidoro, Bisk, Yonatan, and Hovy, Eduard. *EvEntS ReaLM*: Event Reasoning of Entity States via Language Models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 1982–1997, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics. https://aclanthology.org/2022.emnlp-main.129. DOI: 10.18653/v1/2022.emnlp-main.129.

[4] Mirzaee, Roshanak, Rajaby Faghihi, Hossein, Ning, Qiang, and Kordjamshidi, Parisa. *SPARTQA*: A Textual Question Answering Benchmark for Spatial Reasoning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4582–4598, Online, June 2021. Association for Computational Linguistics. https://aclanthology.org/2021.naacl-main.364. DOI: 10.18653/v1/2021.naacl-main.364.

[5] Zhang, Li, Lyu, Qing, and Callison-Burch, Chris. Reasoning about Goals, Steps, and Temporal Ordering with WikiHow. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 4630–4639, Online, November 2020. Association for Computational Linguistics. https://aclanthology.org/2020.emnlp-main.374. DOI: 10.18653/v1/2020.emnlp-main.374.

[6] Huggingface Longformers model

[7] Huggingface AutoModel for Multiple Choice.