Distinguishing Recalled versus Imagined Events in Humans

Background and Methods

The narrative flow of a story, described as "sequentiality", measures the coherence in a story. This coherence can serve as a tool to differentiate between autobiographical memories and imagined events. The foundational paper for this project introduces sequentiality to compare imagined stories and autobiographical tales. Drawing inspiration from the research paper that introduced sequentiality as a metric to measure narrative flow, this project aimed to replicate and validate the major findings depicted in Figure 2 of the paper using HippoCorpus Dataset.

To compute sequentiality, each story's sentences were evaluated for their likelihood given the prior context (history length upto 5). By comparing this contextual likelihood to a topical likelihood (provided by a story summary), the sequentiality of the story was computed. Our implementation deviated slightly from the paper by using the GPT-2 model in lieu of GPT-3, history length 5 in lieu of history length 9 due to computation limitation, anticipating potential variations in the results. Each story's sentences were evaluated for their likelihood given the prior context (sentences before them).

$$c(si,h) = -\frac{1}{|si|}\left[\log p_{LM}(si|T) - \log p_{LM}(si|T, si_{i-h:i-1})\right] \tag{1}$$

In the formula, $c(s_i,h)$ measures the sequentiality of a given segment s_i with respect to its historical context of length h. Here, $|s_i|$ is the length of the segment s_i. The term $p_{LM}(s_i|T)$ is the probability of the segment s_i as predicted by a language model given a topic T , while $p_{LM}(s_i|T, s_{i-h:i-1})$ is the probability of the segment s_i predicted by the same language model but now conditioned not just on the topic T, but also on the h preceding segments $s_{i-h:i-1}$.

Hypotheses

Building upon the preliminary observation from the raw data and considering the inherent structured nature of autobiographical memories:

1. Imagined stories will exhibit higher sequentiality scores than autobiographical stories, hinting at the fact is it is well structured narrative because it is imagined.

2. The sequentiality of autobiographical stories will increase when the memories are recounted several months later hinting the building up of narrative over time.

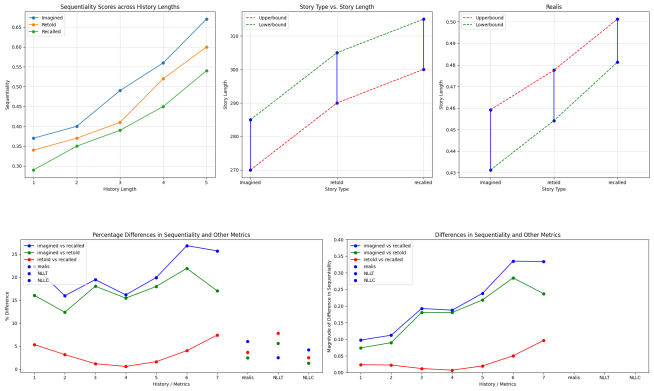
Analysis

The dataset "Hippocorpus," containing 6,854 English short stories about autobiographical and imagined events, was employed. The inclusion of both recalled and imagined narratives in a controlled environment allows for a direct comparison between real-life events and fictional tales using sequentiality. The data for the study was obtained from a dataset comprised of 6,854 English diary-like short stories categorized into imagined, recalled and retold. Package NLTK's Punkt model is used to quantify the lexicons to understand the realis of the events.

Without delving deep into specific figures, a cursory overview of the computed sequentiality scores indicates a trend in alignment with our hypotheses. The imagined narratives generally showcased a higher sequentiality compared to the autobiographical accounts. Also, as anticipated, the sequentiality of recounted autobiographical stories surged compared to the initially narrated versions. These observations, though derived from GPT-2, appear to echo the essence of the results obtained with GPT-3 in the referenced research.

Results

The results successfully captured the trends highlighted in the research paper, albeit with some inconsistencies, which might be attributed to the model difference:



The derived sequentiality showed imagined stories had a slightly higher sequentiality compared to autobiographical ones. Retold narratives also exhibited a greater sequentiality than recalled ones, aligning with the paper's findings. However, the percentage changes and its Magnitude slightly off relatively, this might be due to the LLM used, reduced history sizes and sample space about 3000 stories.

Discussion and Future Plans

In our approach, we made a minor adjustment from the original paper by opting for the GPT-2 model instead of GPT-3, so we expect some differences in the outcomes. The imagined events tend to have more sequentiality, which may be due to the fact it is structered. In Phase 2, we're focusing on: 1. Extracting additional linguistic aspects from the narratives. 2. Creating a model that can differentiate between Imagined and recalled events. We'll gauge the model's effectiveness using a previously defined validation set. By employing advanced NLP using the HippoCorpus Dataset, we can extract linguistic patterns and characteristics such as sentiment, complexity, and nuances. Along with this data, the data collected through Phase 1 can be used to analyze and develop a computational model that can classify imagined and recalled events. This

can be verified using the pre-determined validation set in the corpus dataset.

Phase 2 : Introduction

Building on the foundational work of quantifying narrative flow in imagined versus autobiographical stories in Phase 1, Phase 2 of our project aims to derive additional linguistic features and implement classification models to distinguish between recalled and imagined events. This phase is crucial for understanding the nuanced differences in how humans construct and recall memories versus fabricate stories.

Method

Pre-Determined Validation Set

In selecting the validation set, these linguistic features and sequentiality probabilities were carefully considered to ensure a representative mix of imagined and recalled narratives, and by employed Linguistic features. This approach allows for a more nuanced understanding of how these features correlate with and influence the classification of different memory types. The Training, Testing and Validation sets are carefully selected by stratified split in 70:15:15 proportions as opposed to randomized split. The stratified split ensures the generality, overall comprehensiveness and avoid biased prediction for tackling overfitting. Additionally, the features were standardized, and linguistic features were added to enhance the model's ability to capture complex relationships.

Linguistic Features Characteristics

In the Phase 2 analysis, various linguistic features played crucial role for differentiating between imagined and recalled narratives (including retold events) on top of characteristic data provided within the Hippocorpus dataset. Features such as *word count* and *sentence count* often indicate more detailed descriptions in recalled stories, while imagined stories may exhibit longer word lengths and greater *lexical diversity*, reflecting creative freedom. Imagined narratives might also show higher *punctuation counts* and *average sentence lengths*, along with more *sensory details*, to enhance vividness. In contrast, recalled stories typically have higher counts of *first-person pronouns* and *past tense verbs*, emphasizing personal experience and past events. *Emotional word* usage is likely more pronounced in recalled stories due to their personal nature. The use of *dialogue tags* are more frequent in imagined stories. Additionally, sequentiality probabilities, calculated using models like GPT-2, help distinguish these narratives, with imagined stories showing higher sequentiality, indicating a smoother narrative flow.

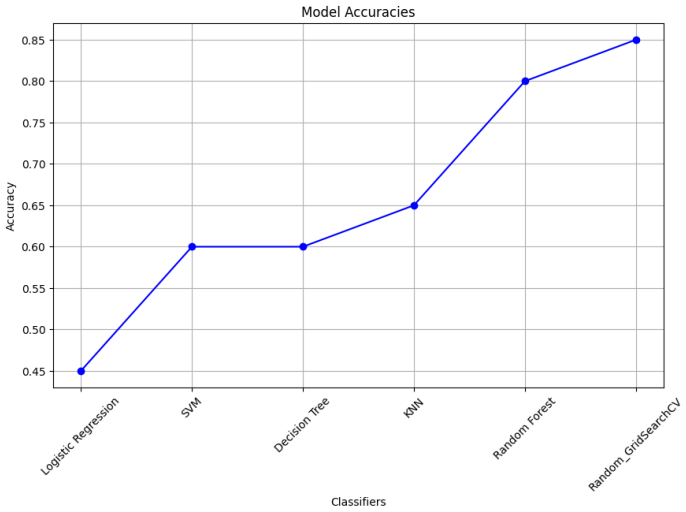
Classifying: Recalled Vs Imagined Narratives

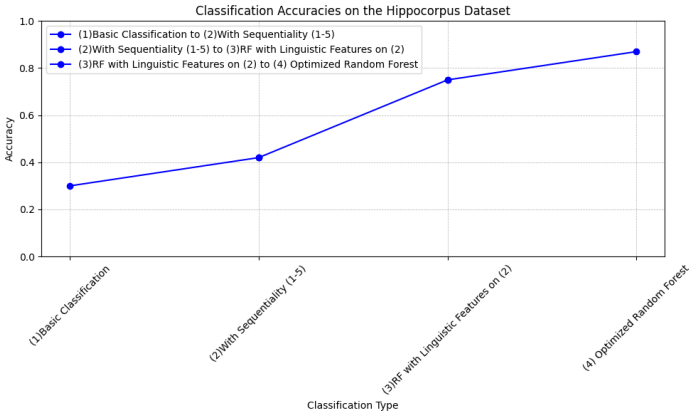
Machine learning algorithms are trained to identify patterns unique to each narrative type by leveraging a range of linguistic features and sequentiality probabilities. This classification process involves feeding these features into models like Random Forest, SVM, and Logistic Regression, which then learns to predict each story's memory type (imagined or recalled). After Identifying the best classification model, the model is subjected to further enhancements based on the hyperparameters and modern machine learning techniques. The effectiveness of this classification is measured by the accuracy of the models in correctly identifying the type of narrative based on the given linguistic characteristics.

Project URL link here

Results

The autobiographical events are subjected to distinguishable events, retold and recalled. However, The Result captures the accuracy by combining retold narratives as recalled events. The Results successfully captures the trend by comparing accuracy's of different classifications algorithm and give insights into how human's congnitive abilities captures the narratives memory type and GPT2's inconsistency in accuracy (in lieu GPT2 Vs. GPT3, and with 3500 entries)





Analyzing the Hippocorpus dataset using various machine learning classifiers has yielded significant insights into classifying narratives as either imagined or recalled. (Figure: Classifiers) The initial classification, based solely on the dataset

without additional features, achieved an accuracy of 0.3, which contains characteristics of a narrative (Stressful, distracted, draining, etc.). This baseline performance set the stage for further enhancements.

Subsequent classifications incorporated more sophisticated features. The introduction of Sequentiality scores (ranging from 1 to 5) from Phase 1, which measure the narrative flow based on preceding sentences from 1 through 5 sentences and story topics, improved the accuracy to 0.42, using the Support Vector Classification Model. This increase underscored the importance of narrative structure in differentiating between imagined and recalled stories. However, the most notable leap in accuracy was observed when linguistic features were added to the Sequentiality scores and by using Random Forest Classification Model. The Linguistic features included word count, sentence structure, linguistic diversity, and other narrative elements, elevating the accuracy to 0.75. This substantial improvement highlighted the critical role of linguistic nuances in distinguishing between different types of narratives.

The final phase involved optimization of the classification process, mainly focusing on the Random Forest classifier. This optimization, likely involving hyperparameter tuning and feature selection, further increased the accuracy to 0.87. The Figure: Model Accuracy's illustrates this progression, with the Random Forest classifier achieving the highest accuracy among the tested models, including Logistic Regression, SVM, Decision Tree, and KNN.

The implementation of this project has successfully provided valuable insights into human cognitive abilities, particularly in perceiving and classifying narratives as either imagined or recalled. By employing advanced machine learning techniques and analyzing linguistic features in conjunction with narrative flow metrics, the study has demonstrated a significant capability in distinguishing between these two types of narratives. This approach underscores the nuanced nature of human storytelling and memory recall and highlights the potential of computational methods in understanding complex cognitive processes. Furthermore, it is essential to acknowledge that the current study, is limited due to computational abilities, which presents a brighter scope for even greater accuracy and more comprehensive understanding of the intricacies involved in how humans perceive and categorize their narratives.

Discussions

The progression from Phase 1 to Phase 2 in our project has been a journey of deepening our understanding of the cognitive processes involved in human storytelling and memory recall. The introduction of additional linguistic features and various classification models in Phase 2 have significantly advanced our ability to distinguish between recalled and imagined narratives. This phase has been instrumental in unraveling the nuanced differences in narrative construction, whether stemming from memory or imagination. The strategic selection of a pre-determined validation set, incorporating both linguistic features and sequentiality probabilities, has ensured a comprehensive and representative analysis, enhancing the validity of our findings.

The results obtained in Phase 2 have been illuminating. Machine learning algorithms, such as Random Forest, SVM, and Logistic Regression, have allowed for a nuanced classification of narratives, leveraging the rich linguistic features extracted from the Hippocorpus dataset. Even with smaller LLM model such as GPT2, as opposed to the research paper usage of GPT3, the model can distinguish between Fabricated events and Real events with 87% accuracy.

The accuracy achieved in these classifications, particularly with the optimized Random Forest model, is a testament to the effectiveness of combining linguistic analysis with narrative flow metrics. However, it is noteworthy that the study's scope was limited to 3,500 records and employed GPT-2 due to computational constraints and budgetary limitations. However, attempts to use ensemble techniques or Voting classifier techniques to increase accuracy resulted in inconsistency and overfitting, bringing the overall accuracy to 0.75. As the phase 2 results provided favorable outcome, it can be hypothetically theorized, that this result could be potentially avoided by using complete dataset (7000 entries) and most advanced Large Language Models (LLMs). This limitation suggests that the potential for even greater accuracy and consistency exists. The research can be extended by utilizing the complete dataset and leveraging the additional capabilities of GPT-3 or the most advanced LLM such as GPT-4. That could further refine the classification process, potentially leading to more precise, insightful results, and thoughtful practical applications.

Looking forward, there are several avenues for future enhancement of this research. One key area is the exploration of *Ensemble techniques* or *Voting classifier* for combining the strength of selected classification algorithm, or advanced machine learning models that could offer more robust classification without the risk of overfitting. Another promising direction is the application of this research in real-world scenarios, such as in forensic settings or therapeutic contexts, where understanding the nature of narratives can be crucial. The potential of this research to contribute to our understanding of human cognition and storytelling is vast, and the advancements made in Phase 2 lay a solid foundation for continued exploration and discovery in this fascinating field.

References

Sap, M., Jafarpour, A., Yejin, C., Smith, N., Pennebaker, J., & Horvitz, E. (2022, 11). Quantifying the narrative flow of imagined versus autobiographical stories. *Proceedings of the National Academy of Sciences of the United States of America*, *119*, e2211715119. doi: 10.1073/pnas.2211715119