CS 598 - Distinguish recalled versus imagined events in humans

Satyabrat Bhol
Computer Science Department
Rutgers University
Piscataway, NJ, USA
Email: sb2311@scarletmail.rutgers.edu

Abstract

Understanding narrative progression through accumulated life experiences and acquired knowledge empowers individuals to craft cohesive stories. In this paper, the authors have introduced "sequentiality" as a metric to quantify disparities between autobiographical and imagined narratives. Phase I of the task revolves around employing the advanced large language model GPT-2 to distinguish between remembered experiences and fictional tales. The results indicate higher sequentiality in imagined narratives, along with heightened sequentiality in recalled stories. The final phase of the task involved analyzing linguistic features in Imagined and Recalled stories. Attributes such as Temporal Words, Speculative Sentences, Named-Entities, Modal Verbs, Bi-grams, Cause Words, etc., were scrutinized through this linguistic approach. The sequentiality value obtained from the initial phase was used as one of the features. Statistical techniques using ANOVA and Kendall correlation were utilized to determine the best-performing features. Ultimately, classifier models were employed to distinguish between Imagined and Recalled narratives, achieving a peak accuracy of 65%.

Background

The paper delves into how life experiences and knowledge shape expectations in narratives, introducing "sequentiality" to measure event flow. It employs GPT-3 to analyze autobiographical and imagined stories, distinguishing between remembered experiences and fictional tales. The findings reveal higher sequentiality in imagined stories, and increased sequentiality in autobiographical stories recalled months later. The study also links lower sequentiality to a higher proportion of major events. This research highlights the potential of computational analyses like sequentiality in understanding language generation and memory's role in narrative formation using extensive autobiographical and imagined story datasets.

Methods

The sequentiality measure for each sentence (s_i) within a story focused on a particular topic (T) is determined by computing the discrepancy in negative log-likelihood (NLL) based on both contextual and topic-driven models. The dataset was categorized into three groups according to memory type: Imagined, Recalled, and Retold. To ensure data integrity, duplicate records were removed based on their assignment IDs, resulting in distinct records (imagined, recalled, and retold) pertaining to the same story type or topic.

In the initial attempt to calculate the negative log-likelihood, the BERT Next Sentence Prediction model was employed. However, it was observed that the logits provided by BERT were actually classification logits (yes or no - list tensors) rather than representing sentence probability.

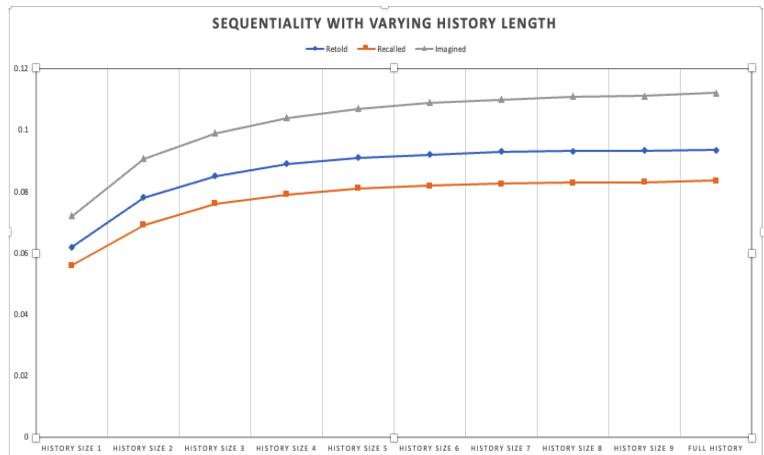
To address this, the OpenAI GPT-2 model was utilized to calculate the probability of each sentence considering the given topic or historical context. The probability of each individual word (token) was computed, and the log-likelihood of each value was summed to obtain the overall sentence probability. The probability of individual tokens was determined based on previously provided tokens or words. For the topic model, only the story's topic was used to calculate the topic-driven probability. For the contextual model, depending on the historical context size ($1, 2, 3$, and so on to the full history), only preceding sentences of that particular size were taken into account. The source code for this project - Sequentiality of stories

Results

The overall sequentiality of the entire story was computed as the average sequentiality across its sentences. Similarly, the collective sequentiality for a specific memory type was calculated as the mean of the sequentiality records corresponding to that memory type.

Sequentiality was assessed across a range of historical context sizes, varying from history size 1 to the entire history length, for all three memory types. Upon comparing sequentiality across different story types, involving a total of $N = 2788$ unique topics, it was evident that imagined stories exhibited a notably higher degree of sequentiality compared to autobiographical memories. Additionally, it was observed that retold autobiographical stories demonstrated higher sequentiality than freshly recalled memories.

There was little variation observed in the contextual likelihood (NLL_C). However, when comparing imagined story sentences to autobiographical story sentences, a noticeable increase in the negative log-likelihood for topic-driven models (NLL_T) was evident.



Phase II plans

Idea 1 The Hippocorpus dataset includes an event-annotation subset for all three distinct story types. In this subset, every sentence in a story was evaluated by eight crowdworkers to determine if it conveyed a major or minor event, and whether the event identified was anticipated or unexpected. This event-annotation process aims to capture the emotional essence of the story and, consequently, could assist in discerning between imagined and recalled events.

Idea 2 In narrative composition, imagined stories often present diverse sentence lengths and intricate structures to evoke intrigue. Conversely, recalled stories tend to feature simpler sentence structures and consistent lengths. Analyzing story length and the presence of real-life, non-hypothetical references (realis event words) can serve as effective markers to distinguish between imagined and recalled events.

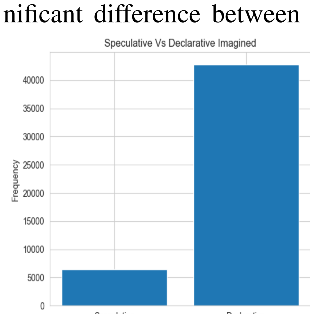
Idea 3 In storytelling, imagined stories favor future or conditional verbs and third-person pronouns, distancing the narrator from the events. Autobiographical stories, in contrast, employ past/present verbs and first-person pronouns for personal connection. Part-of-Speech (POS) tagging would help in classifying these narratives. Imagined stories feature fictional character interactions, while real conversations and memories characterize autobiographical accounts. Named Entity Recognition (NER) may aid in distinguishing events by identifying proper nouns and named entities, thus classifying between imagined and recalled events.

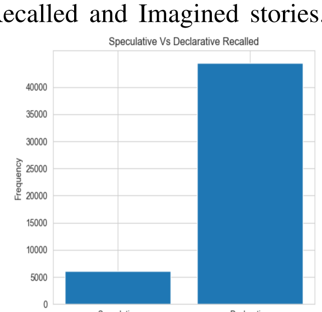
Background and Methods

The initial idea from the midterm report, which incorporated event annotation features, was reconsidered due to a limited dataset of only 200 records, thus discontinue its use in classifier construction. The second idea involved a comprehensive analysis of the intricate structures inherent in Imagined and Recalled stories. Key features explored in this context included Temporal Words, Speculative Sentences, and Named Entities (realis event words). The third idea centered around the utilization of Part-of-Speech tagging, a grammatical classification method encompassing verbs, nouns, prepositions etc. Features such as Modal Verbs, Prepositions, Cause Words, etc., were investigated through this linguistic lens. In addition to these primary ideas, an exploration of several other features was conducted, encompassing common Bi-grams and emotional cues. The culmination of this process involved a thorough feature analysis, aimed at scrutinizing attributes and eliminating redundant features. Code

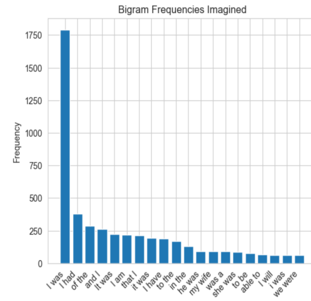
Results

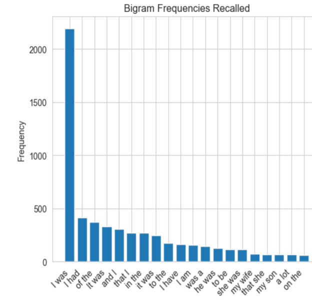
Critical features that significantly influenced the classification task include Temporal Words, Named Entities, Sequentiality, Emotional cues, Speculative Sentences, and Preposition probability. Temporal words, indicating time-related terms, helped in conveying the temporal aspects or realis events within the narrative. Same can be said about the frequency of Named entities, which also refers to realistic event types. In fact, it exhibited a high correlation with Temporal Words, as observed in the Kendall Rank coefficient heat map. Speculative Sentences, denoting conjecture and imagination, provided valuable cues distinguishing factual content from hypothetical elements. Despite the prevalence of declarative sentences outweighing speculative ones in both cases, the calculated probability of speculative sentences revealed a significant difference between Recalled and Imagined stories.



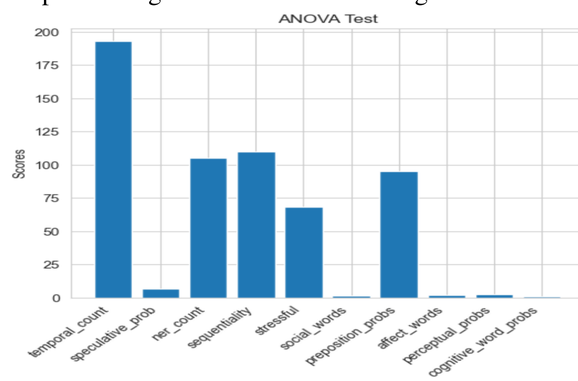


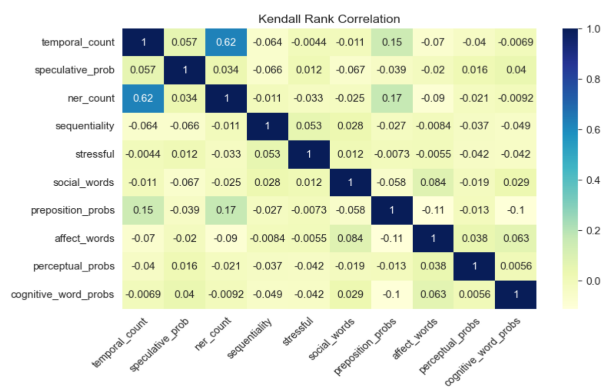
Prepositions, offering insights into spatial and temporal dimensions, exhibited a slight correlation with Temporal Words and Named Entities, further contributing to the effectiveness of the classification. The consideration of Sequentiality, a fundamental aspect of the phase 1 task, remained a crucial feature in this analysis. Emotional cues, gauging the speaker's stress level, provided nuanced insights into the authenticity of emotions expressed in the story. While features like cause words, social words, and affect words contributed marginally to the classification, others such as common Bi-grams and verbs were found to be non-influential and were consequently discarded.





An ANOVA test was conducted to identify the most impactful features for the classification. Multiple models, including XGBoost, SVM, Random Forests, and Logistic Classifier, were fine-tuned using a validation set with a 70-15-15 split. Among these, the Random Forest classifier demonstrated the highest accuracy at 65%, outperforming other models that ranged between 62-65%.





Discussions

Examining the unique elements and patterns within the structure, which entails a deep dive into the organization, syntax, and overall composition of both Imagined and Recalled stories, formed the foundation for this classification. Employing Part-of-Speech tagging (involves investigating Verbs and Prepositions) played a pivotal role in this process. A comprehensive analysis, incorporating ANOVA and Kendall correlation, enabled the identification of influential elements before inputting them into a classifier. While numerous other linguistic features were explored (though not detailed in the report), many proved to be indistinguishably intertwined in both story types, rendering them irrelevant.

References

A Gilboa, A. M., RS Rosenbaum. (2018). *Autobiographical memory: From experiences to brain representations* (Vol. 110; Tech. Rep.).

Alebachew Chiche, B. Y. (2022). Part of speech tagging: a systematic review of deep learning and machine learning approaches. *Journal of Big Data*.

Bartlett, F. (1932). Remembering: A study in experimental and social psychology. Cambridge University Press.

Jacob Devlin, K. L. K. T., Ming-Wei Chang. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *CoRR*, *abs/1810.04805*.

MA Conway, S. W. H. S., CW Pleydell-Pearce. (2003). Neurophysiological correlates of memory for experienced and imagined events. *Neuropsychologia*, *41*, 334–340.

Paskin, M. (2001). Grammatical bigrams. *Advances in Neural Information Processing Systems 14*.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners*.

R Reichardt, P. S., B Polner. (2020). Novelty manipulations, memory performance, and predictive coding: The role of unexpectedness. *Front. human neuroscience*, *14*, 152.

Sap, M., & Jafarpour, A. (2022). Quantifying the narrative flow of imagined versus autobiographical stories..