GPT-2: A Novel Language Model to Analyze Patterns in Sentence Predictability

Santiago Noriega, Victoria Sharpe, Gina Kuperberg (Tufts University Department of Psychology)

Background

- A **language model (LM)** is a probabilistic model that can predict the next word in a sequence of text when provided with the preceding context. LMs are trained on a library of text and different LMs follow different strategies to "learn" the probability of upcoming words.
- **Ngrams** are an example of a LM widely used to predict the probability of a word given its context as the *n* number of preceding words. Trigrams, for example, predict the probability of a third word given its two preceding words as context.
- **Cloze** is a measure of predictability derived from answers by participants who are asked to complete a sentence with what they think the most likely continuation is.
- **GPT-2** is a transformers-based machine learning LM developed by Open-AI that can retain a memory of the preceding context and is trained on a library of over 40GB of text.
- Here, we analyze the potential to use GPT-2 as a more flexible and optimal measure of predictability compared to traditional Cloze and Ngram measures.

Results

- For our experiments, we used previously gathered cloze measures from four-sentence sets and compared those results to the predictability measures obtained from GPT-2 and Ngram measures.
- **Fig 1:** GPT-2 Trigrams Vs. Ngram Probabilities Across Sentences.
  - The average probability of words by their sentence position were measured and compared using GPT-2 and Ngram techniques.
  - The data were strongly correlated (r = 0.81).
  - The data followed similar patterns across the distribution, showing peaks and troughs at very similar positions.
- **Fig 2:** GPT-2 Vs. Human Entropy Measures Across Sentences.
  - Entropy is a measure often associated with information density that is derived from word probability measures.
  - The data were strongly correlated (r = 0.80).
  - These data suggest that GPT-2 measures can be used as a proxy for patterns in predictability-derived cloze measures.

Fig 1.

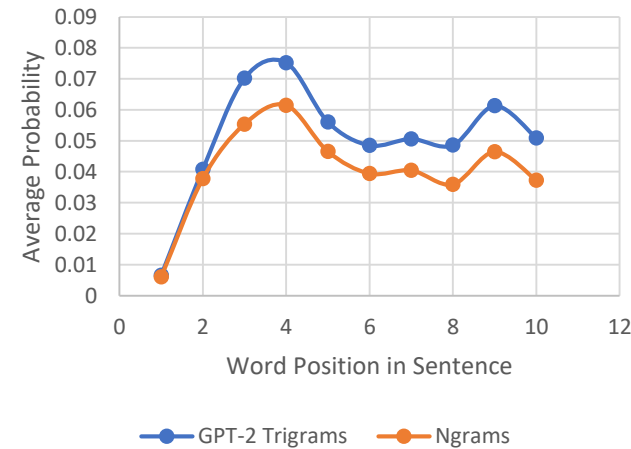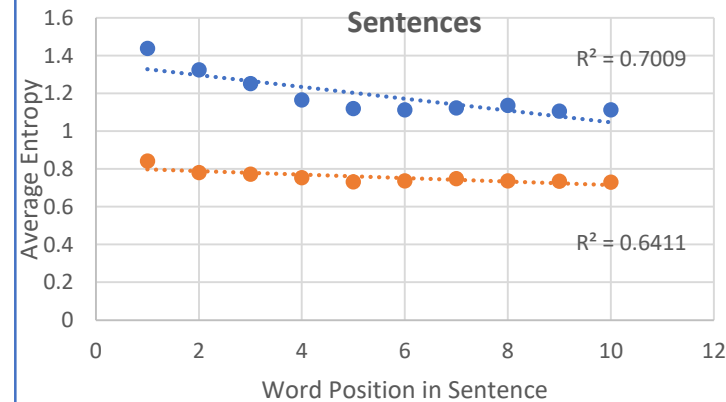GPT-2 Trigrams Vs Ngram Probabilities Across Sentences

— GPT-2 Trigrams — Ngrams

Fig 2.

GPT-2 Vs. Human Entropy Measures Across Sentences

$R^2 = 0.7009$

$R^2 = 0.6411$

Discussion & Future Directions

- In probability and entropy measures, GPT-2 was a strong predictor for the patterns and magnitudes for results obtained from Ngram and cloze techniques.
- GPT-2 can be manipulated to alter the amount of context that it considers, making it a more flexible model than Ngrams that could also consider preceding context beyond the previous two words.
- Although obtaining cloze measures is still the golden standard for measuring predictability, it is a time-consuming and expensive procedure because it requires the recruitments of participants. GPT-2, on the other hand, can be used for any text in a much more economic and timely manner.

Replication Studies:
- Replication studies are often overlooked in Psychology, and various new findings are often unchallenged.
- GPT-2 is reusable in nature since it can be written into a computer program that could be applied to any text.
- GPT-2 could be used as a tool to quickly yet reliably assess findings from previous studies on language predictability.
- Efforts are currently ongoing at the Kuperberg Lab to perform replication studies on past studies that propose patterns in the spread of information density as entropy throughout discourse (Xu, Reitter 2018).

Analyzing Predictability in Schizophrenia:
- The speech of Schizophrenia patients is often characterized as being unpredictable and tangential.
- We plan to use GPT-2 to analyze changes in entropy across different levels of context in transcribed interviews with Schizophrenia patients.
- This would be an important step to better understand underlying neurological processes of Schizophrenia and has the potential to e used as a diagnostic tool.

Acknowledgments