Spelling convention sensitivity in neural language models

Introduction:

This report delves into T5 language models' treatment of spelling variations. The study aims to decipher patterns, biases, and areas for improvement in language generation.

Methodology:

Using templates like "My preferred words are ..., ..., and tree," the study evaluates T5's response to spelling variations in adjacent and non-adjacent conditions. Metrics include mean conditional probabilities, prediction accuracy, and mutual information.

Key Findings:

T5 shows a notable preference for spelling consistency, especially in adjacent conditions. Biases from training data, favoring US English forms, impact the model's behavior.

Finetuning and Future Challenges:

Attempts to enhance spelling consistency through finetuning on a synthetic corpus yielded intriguing results. Although overall consistency did not improve significantly, the drop-off in non-adjacent conditions was mitigated.

Takeaways and Future Avenues:

The study highlights the need for nuanced training data curation and suggests future research across languages. It paves the way for refining language model behavior and exploring ethical considerations.

Conclusion:

In summary, this research unravels T5 language models' approach to spelling consistency, contributing valuable insights to the field of natural language processing.