Learning by Prediction and Integration: Human-inspired Approaches for Natural Language Understanding

Vladimir Araujo

Abstract

Giving machines the skills to represent and understand natural language for applications in the real world presents a significant challenge in the area of Natural Language Processing. Pre-trained language models based on neural networks have recently achieved outstanding performance in several natural language understanding tasks. Although effective, these models lack the ability humans possess to text comprehension. For example, as we read, we can anticipate what content may come next or use prior knowledge to better understand a passage.

We hypothesize that current language models could benefit from human language processing mechanisms. In this work, we investigate and propose different approaches to improve current language models, drawing inspiration from prediction and integration theories of human language comprehension. Our contributions show that augmenting models with human mechanisms leads to improvements in natural language understanding across various tasks.

First, we extend the architecture of pre-trained language models with insights from predictive coding theory. We demonstrate that introducing bottom-up and top-down computation to predict future sentences in latent space in the neural networks improves sentence and discourse-level representations. We also validate the generalization of our models in Spanish by using benchmarks and pre-trained models developed by us.

Second, we adapt and propose memory population methods for pre-trained language models under the lifelong learning with episodic memory paradigm. We show that a process that samples the entire data distribution works well enough to integrate previous knowledge and prevent forgetting in the neural network. Furthermore, we also found that some tasks benefit more from selective-based population methods.

Third, extending our second contribution, we propose a method to deal with the stability-plasticity dilemma that occurs in lifelong learning with memory. We show that entropy can be used as a plasticity factor to decide how much a layer in a neural network must be modified according to the current input. We found that this process not only improves performance on text and image classification tasks but also promotes efficiency in the model.

Fourth, we propose a method that incorporates prediction and integration ideas to solve the question answering task in a data stream. Specifically, our approach leverages cross-attention mechanisms to integrate information into external memory. This memory is supported by anticipation and rehearsal pretext tasks. We show that our model can improve the memorization of text-based and video-based sequences.

In summary, we present several approaches that follow ideas from human language processing, demonstrating that human inspiration can be a way to improve the current state of language models based on neural networks. By including human-based mechanisms, we bolster or add some abilities that models do not have and are key to obtaining human-like language processing.