

Recent Progress on State Space Language Modeling

Sasha Rush, Cornell University and Hugging Face

Transformers remain the core architecture of foundation models. However, many have noted practical limitations in this architecture class as models grow in size. Recent progress on alternative architectures based on state space models (SSMs) has sparked interest that these models could provide a variant with similar modeling ability and more favorable computational properties. However, this architecture also opens up questions about whether we are correctly measuring the capabilities of language models, particularly in terms of long-range memory. This talk will consist of three parts: 1) A short tutorial on the current state-of-the-art in state space language models, known as Mamba, 2) A summary of recent work on the use of SSMs in language modeling, 3) A discussion of the SSM properties in related areas such as image generation.