Contribution ID: **194** Type: **Invited Speaker**

Towards Explainable Natural Language Modeling with Quantum Tensor Networks

Deep Learning (DL) has empowered computers with superior performance in modern Natural Language Processing (NLP) tasks, such as sentiment analysis and machine translation. Even for texts with long-range correlations such as sequences of characters in Wikipedia, DL can effectively express the power-law decay in the mutual information between two distant characters. Despite empirical successes, its intrinsic non-linearity complicates the analysis of algorithmic behaviours. Which network architectures and how many parameters are essential to reproduce long-range correlations are important yet theoretically challenging questions to tackle. Here, we attempt to provide systematic answers through the mapping between DL and its matrix product state (MPS) counterpart. By recasting DL as MPS, we show that the number of parameters required to achieve high performance in sentiment analysis, and to reproduce power- law decay in the mutual information in Wikipedia texts, can be efficiently extracted from the entanglement entropy in the dual MPS. Our work utilises tools in many-body quantum physics to resolve explainability issues of NLP, and more generally of sequence modelling.

Primary author: CHOTIBUT, thiparat (Chulalongkorn University)

Presenter: CHOTIBUT, thiparat (Chulalongkorn University)

Track Classification: Statistical and Theoretical Physics