World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:18, No:11, 2024

Graph Neural Networks and Rotary Position Embedding for Voice Activity Detection

Authors : YingWei Tan, XueFeng Ding

Abstract : Attention-based voice activity detection models have gained significant attention in recent years due to their fast training speed and ability to capture a wide contextual range. The inclusion of multi-head style and position embedding in the attention architecture are crucial. Having multiple attention heads allows for differential focus on different parts of the sequence, while position embedding provides guidance for modeling dependencies between elements at various positions in the input sequence. In this work, we propose an approach by considering each head as a node, enabling the application of graph neural networks (GNN) to identify correlations among the different nodes. In addition, we adopt an implementation named rotary position embedding (RoPE), which encodes absolute positional information into the input sequence by a rotation matrix, and naturally incorporates explicit relative position information into a self-attention module. We evaluate the effectiveness of our method on a synthetic dataset, and the results demonstrate its superiority over the baseline CRNN in scenarios with low signal-to-noise ratio and noise, while also exhibiting robustness across different noise types. In summary, our proposed framework effectively combines the strengths of CNN and RNN (LSTM), and further enhances detection performance through the integration of graph neural networks and rotary position embedding.

Keywords : voice activity detection, CRNN, graph neural networks, rotary position embedding
Conference Title : ICSLP 2024 : International Conference on Spoken Language Processing
Conference Location : Bangkok, Thailand
Conference Dates : November 25-26, 2024