

Self-Attention

- Goal: Aggregate arbitrary-length input to fixed-size vector.
Allow data-driven, trainable aggregation.