Vanilla Self-attention

Self-attention with Sequence Complementor

Multi-head Self-attention $\times n$

Complementary Sequence

Multi-head Self-attention $\times n$

Richness of Representations (↑)

Forecasting Error (↓)

■ *w/o Seq. Comp.*   ■ *w/ Seq. Comp.*