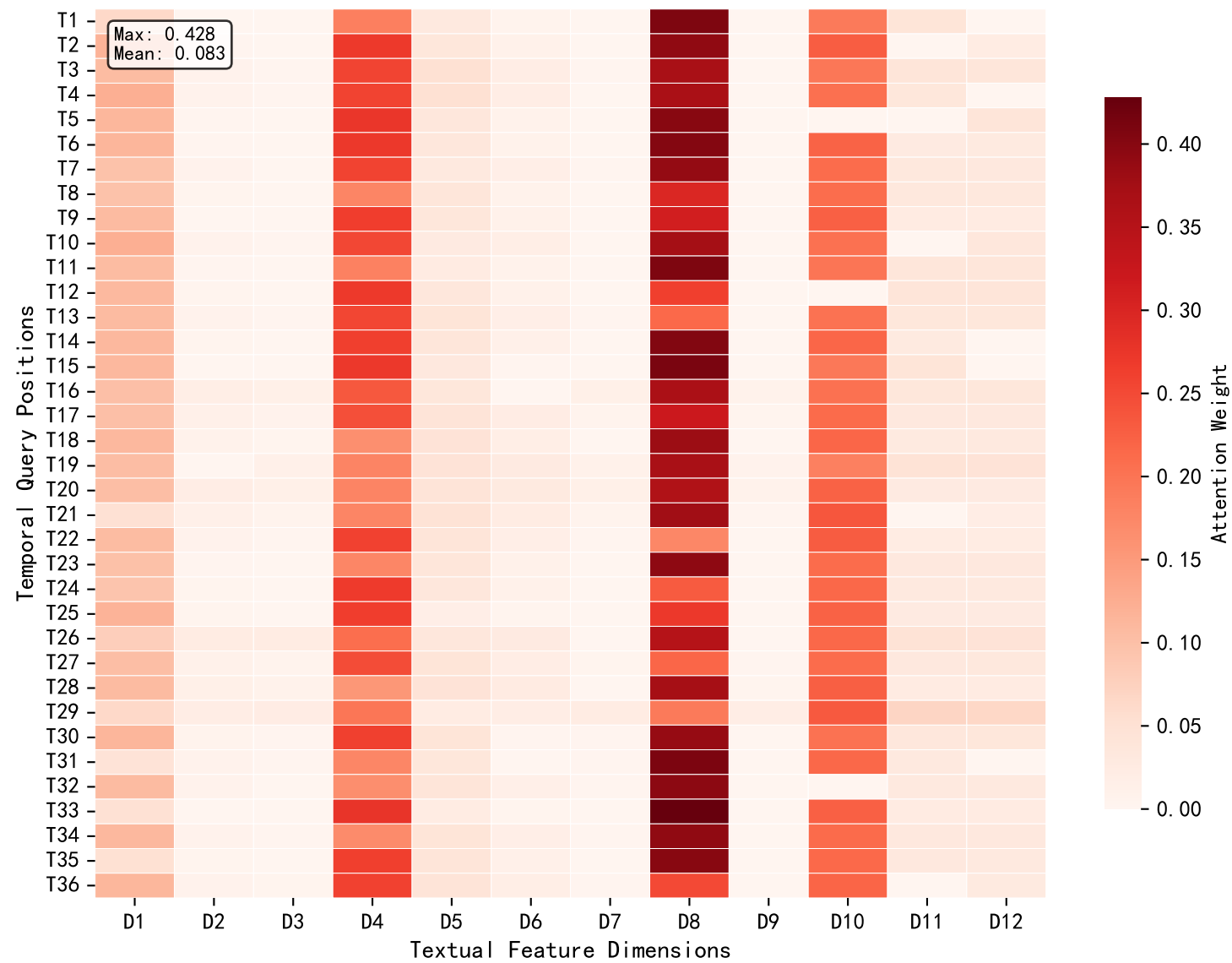
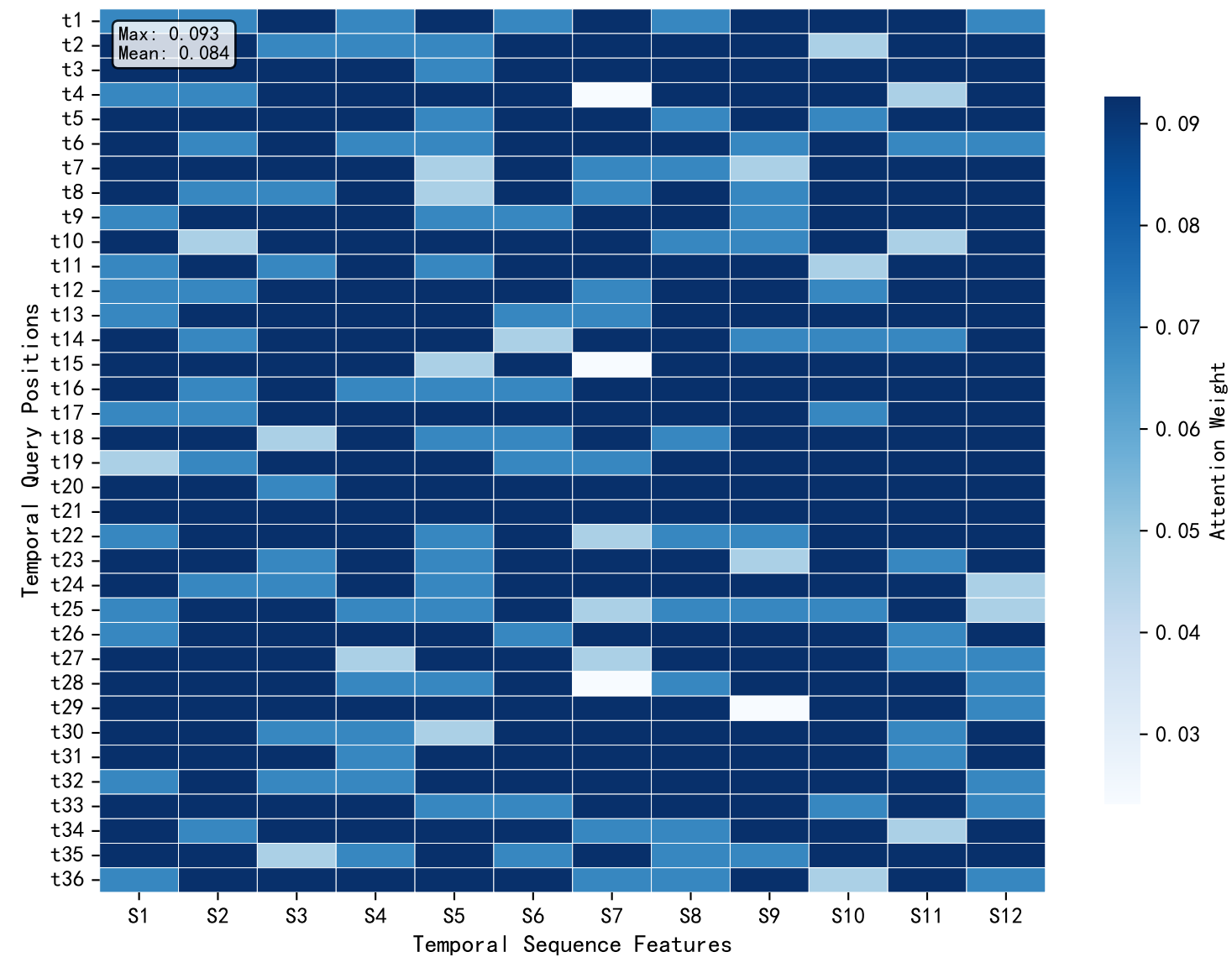


Cross-Modal Attention Mechanisms (Training Epoch 3)



(a) Model Attention to Textual Features



(b) Model Attention to Temporal Sequence