

機器學習 HW#4 Report
資管碩一 R10725018 郭宇雋

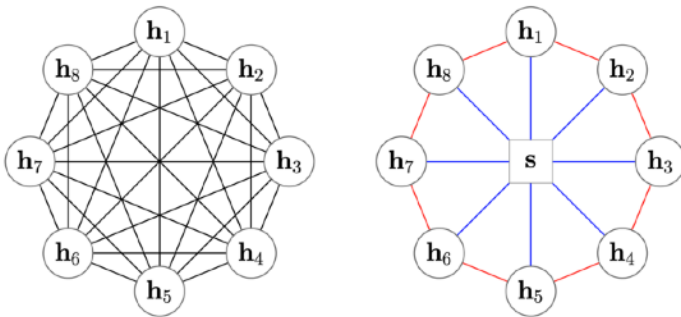
1. Make a brief introduction about a variant of Transformer. (2 pts)

StarTransformer :

將原始 Transformer 的全連接改成星狀拓撲結構，以減少運算量，將時間複雜度降至 $O(N)$ ，但仍保留必要的資訊提取方式。下圖左邊是原始的 Transformer，而右邊是 StarTransformer。

StarTransformer 由一個中繼節點與多個衛星節點組成，每個衛星節點表示的就是原始序列中某個位置的特徵，而中繼節點則是額外創建出的虛擬節點，不存在於原始序列中，其扮演「收集與分發資訊到每個衛星節點」的虛擬中繼站角色。

StarTransformer 的特色是鄰居節點可以直接交互，而非鄰居節點需要至少兩個 Hops 才能交互，成本較高，因此 StarTransformer 比原始的 Transformer 更容易學習到「局部資訊」。



2. Briefly explain why adding convolutional layers to Transformer can boost performance. (2 pts)

Transformer 中的 Self-attention 機制，能夠很好的捕捉序列中遠距離特徵的相關性，然而他在局部特徵提取的能力上較不突出；反觀 Convolutional Layer 的 Kernel 機制，雖說無法提取序列中遠距離的特徵相關性，但他有非常好的局部細微特徵提取能力。因此在 Transformer 的架構中加入 Convolutional Layer 可以補足 Transformer 在局部特徵提取能力上的不足，進而提升整體模型表現。