

1.

Make a brief introduction about one of the variant of Transformer, and use a image of the structure of the model to help explain (2 pts)

Transformer-XL 是為了進一步提升 Transformer 建模長期依賴的能力。它的核心算法包含兩部分：片段遞歸機制（segment-level recurrence）和相對位置編碼機制（relative positional encoding）。

片段遞歸：

和 Transformer 一樣，Transformer-XL 在訓練的時候也是以固定長度的片段的形式進行輸入的，不同的是 Transformer-XL 的上一個片段的狀態會被緩存下來然後在計算當前段的時候再重複使用上個時間片的隱層狀態。因為上個片段的特徵在當前片段進行了重複使用，這也就賦予了 Transformer-XL 建模更長期的依賴的能力。

相對位置編碼：

表明對於所有的 query 位置對應的 query(位置)向量是相同的。即無論 query 位置如何，對不同詞的注意偏差都保持一致。

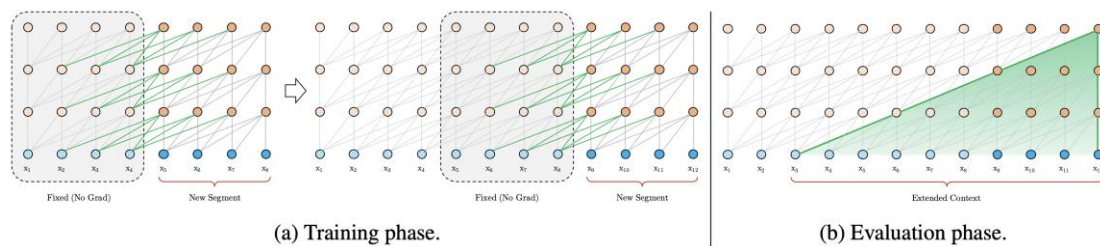


Figure 2: Illustration of the Transformer-XL model with a segment length 4.

2.

Briefly explain what's the advantages of this variant under certain situations. (2 pts)

Transformer-XL 帶來的提升包括：1. 捕獲長期依賴的能力；2. 解決了上下文碎片問題（context segmentation problem）；3. 提升模型的預測速度和準確率。