BERT

04/22

self – supervised learning

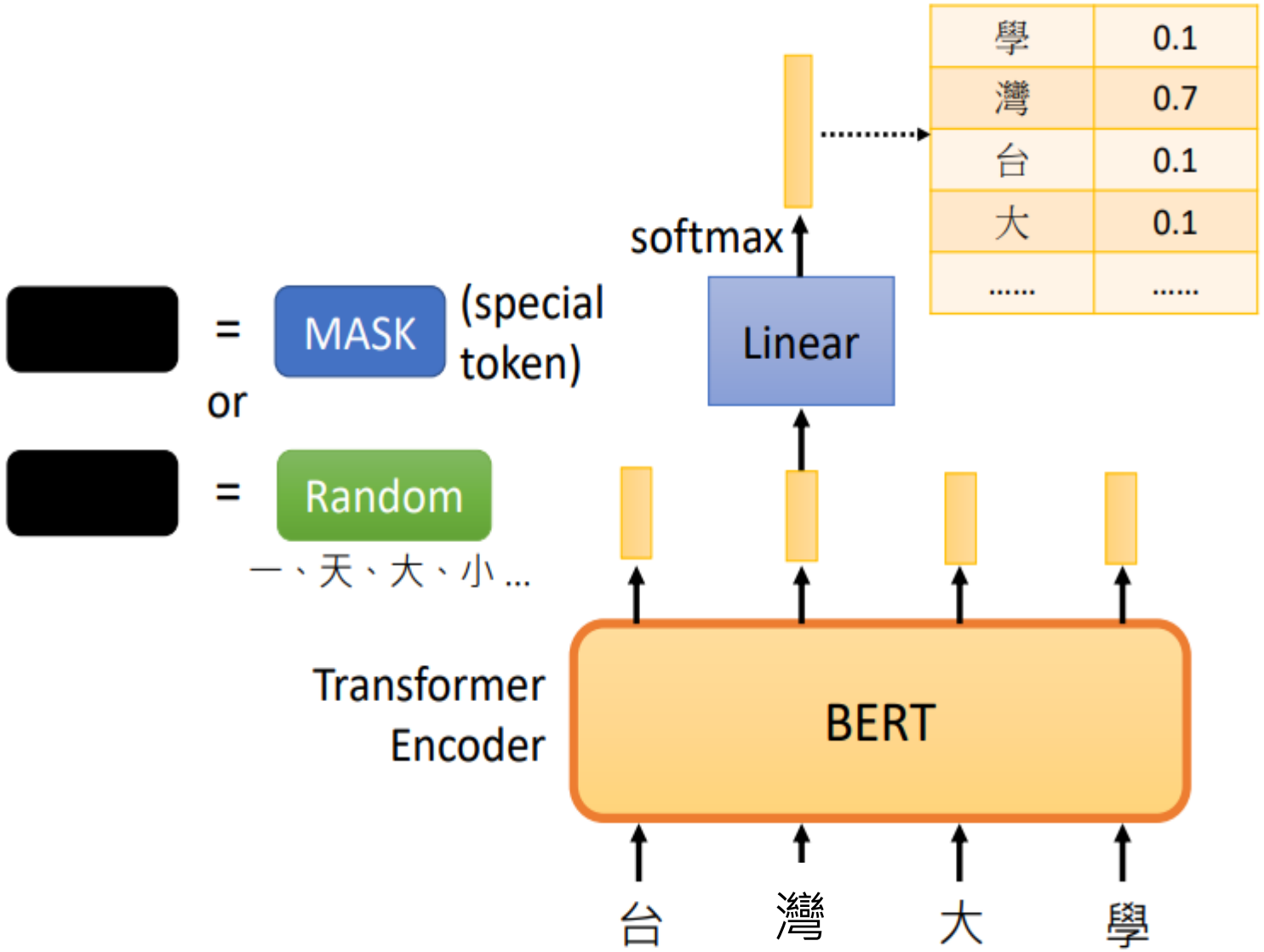
- supervised learning : 每一筆資料都有 labels
- unsupervised learning : 每一筆資料沒有 labels
- self – supervised learning : 介於 unsupervised 和 supervised 之間，拿部分 input 當成 labels，我們希望模型的 output 和 label 越近越好
- self-supervised learning可以算是一種 unsupervised learning，但因為 unsupervised learning 的範圍比較大，有不同的做法，所以將自己的 input 分一部分出來當 label 的叫 self-supervised learning

How to train BERT

- Masking input
- Next sentence prediction

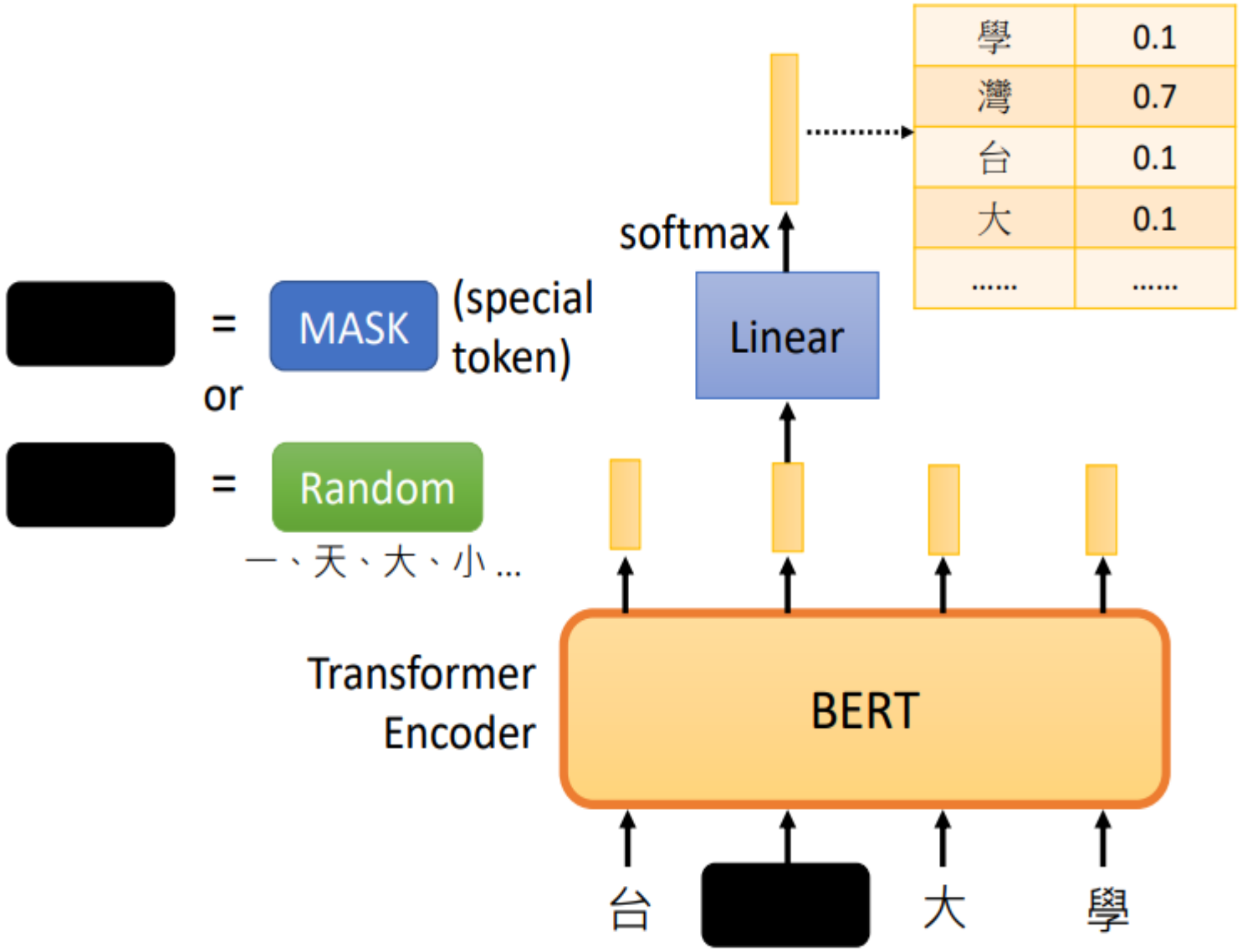
Masking input

- 回顧 transformer encoder ，輸入一排向量，輸出一排相同長度的向量
- 隨機決定將一些文字蓋住或是換成隨機的文字，而蓋住也就是替換成一個非中文字
- 接著輸出的向量做 linear transform，也就是乘一個矩陣，再做 softmax 輸出一個 distribution ，其向量長度包含所有想要處理的文字



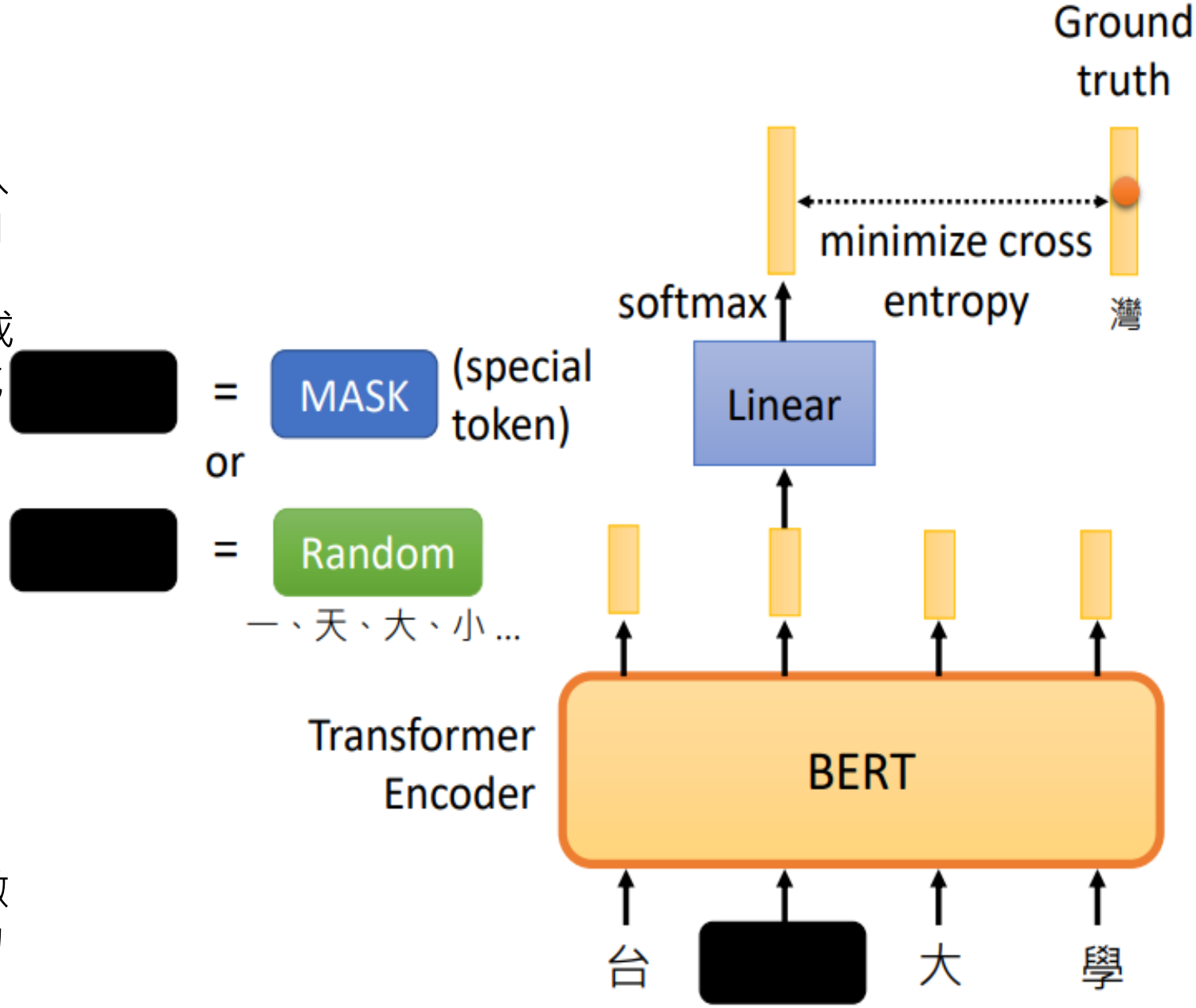
Masking input

- 回顧 transformer encoder ，輸入一排向量，輸出一排相同長度的向量
- 隨機決定將一些文字蓋住或是換成隨機的文字，而蓋住也就是替換成一個非中文字
- 接著輸出的向量做 linear transform，也就是乘一個矩陣，再做 softmax 輸出一個 distribution ，其向量長度包含所有想要處理的文字



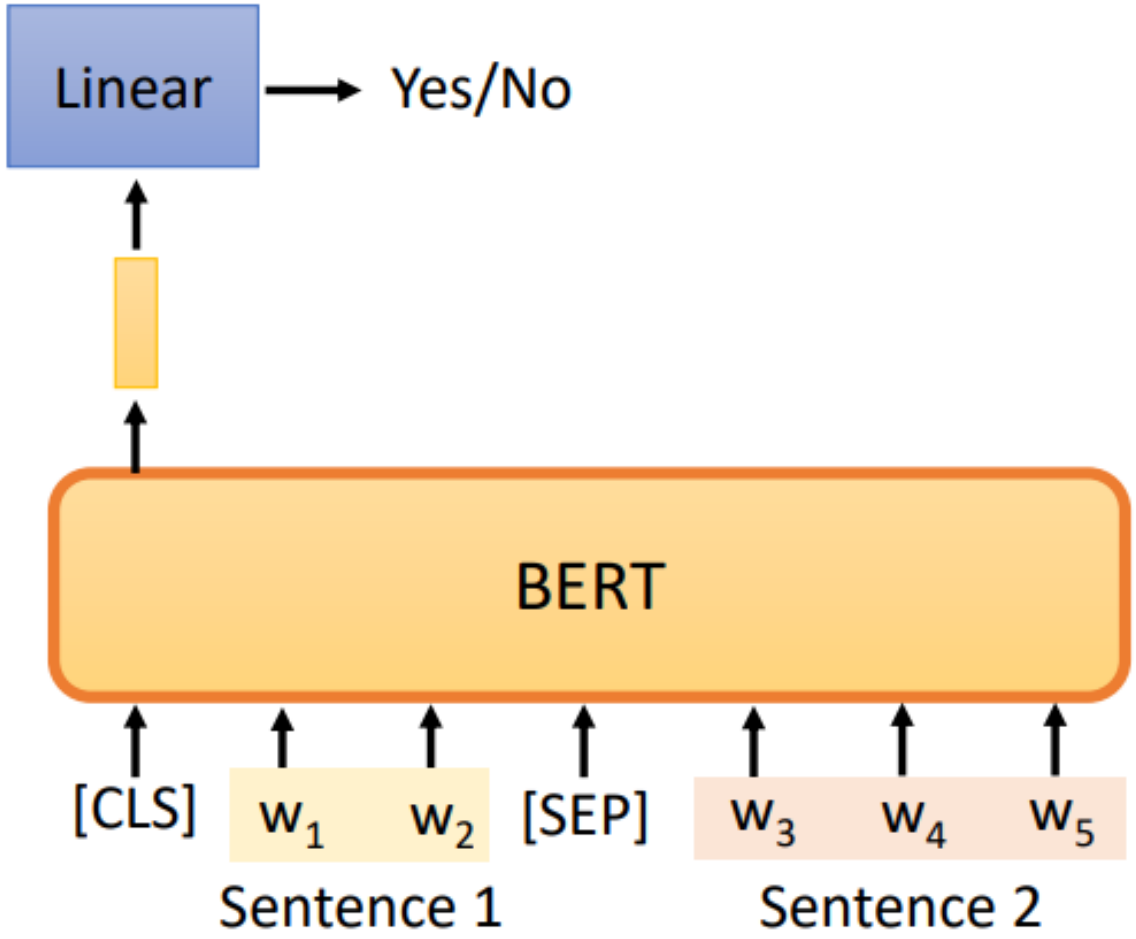
Masking input

- 回顧 transformer encoder ，輸入一排向量，輸出一排相同長度的向量
- 隨機決定將一些文字蓋住或是換成隨機的文字，而蓋住也就是替換成一個非中文字
- 接著輸出的向量做 linear transform，也就是乘一個矩陣，再做 softmax 輸出一個 distribution ，其向量長度包含所有想要處理的文字
- 最後將機率最高的單字跟 ground truth 算 cross entropy，就像是做分類一樣，類別數量為所有關注的字一樣多



Next sentence prediction

- 從 input 裡面隨機挑兩個句子，中間加入 [sep] 分隔，開頭加入 [CLS] 表示起點
- 丟入 BERT 中去看兩個句子是否相接
- 但之後的論文發現效果不太好，因為單看兩個句子是否相接可能太簡單了，之後有人改成判斷兩個句子順序是否相反



BERT

- 看似 BERT 只會做填空題，但實際上他可以運用在很多任務上，透過 fine-tune 去微調

- For example : sentiment analysis

  輸入一個 sequence ，輸出是正面還是負面，通常在做 fine-tune

  使用一點有 label 的資料去 fine-tune