

Lab01 作業

目的：了解 Language Model 的理論及應用

作業：套用 LM 來判斷句子是否可能有文法上的錯誤，藉由比較兩者的 LM 機率，可以發現第二個句子比較可能為正確的。

Sentence1: He is looking a new job. (-30.935284159760272)

Sentence2: He is looking for a new job. (-30.381592816579083)

內容：

- 在此項作業中，LM 請以 bigram 去計算 bigram 詞組出現的機率，並且計算各個句子的機率。
- 因為有些 bigram 不存在於文本 (big.txt) 中，會導致分母為 0 (或 $\log(0)$)，需要使用 add-one smoothing 解決此情況。
- 以防 underflow，使用 log 將相乘轉換為相加。

加分題：

- 使用 good turning 取代 add-one smoothing，並向助教解釋實作過程。

測試：

- Are you interested about your offer for Marketing Assistant. (-61.008203896597465)
- Are you interested in your offer for Marketing Assistant (-55.930040211186764)
- The watch is nice but to expensive. (-42.00491883687611)
- The watch is nice but too expensive. (-34.879363033606204)