

# NLP Lab: Unit 11

## EM 演算法 (2)

# Estimating Word Translation Probability Model

張俊盛 Jason S. Chang

National Tsing Hua University

10820ISA 562100

自然語言處理實作 Natural Language Processing Lab

2020-0512 Google Hangout 網路直播

有助教在 台達館107室

# Expectation and Maximization Algorithm

- A. P. Dempster, N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, 39 (Ser B):1–38.
  - 統計模型產生的資料（有隱藏變數＝未加注釋的資料）
  - 反覆求解以得到統計模型中參數的最大似然估計（maximum likelihood estimates, MLE）
  - EM 演算法交替地反覆執行兩個步驟：
    - 期望值步驟（expectation step, E step）
      - 運用模型估計參數，決定隱藏變數的機率分布
      - 運用隱藏變數機率，計算變數條件下之事件期望值
    - 最大似然估算步驟（maximization step, M step）
      - 運用 E Step 的事件期望值，估計模型參數值
      - 採最大似然估計值（MLE）

# 運用 EM 演算法的實例

- 平行語料庫之詞彙對應
- 統計模型：詞彙機率  $P(c | e)$
- 觀察：英文句、中文句（詞彙機率模型所產生之事件）
  - support<sub>0</sub> vector<sub>1</sub> machine<sub>2</sub> 支<sub>0</sub>撐<sub>1</sub>向<sub>2</sub>量<sub>3</sub>機<sub>4</sub>
- 隱藏變數：英文字到中文字的對應—  $a_i = j$ 
  - $a_0 = 01, a_1 = 23, a_2 = 4, P(a) = P(\text{支撐}|\text{support}) P(\text{向量}|\text{vector}) P(\text{機}|\text{machine})$
  - $a_0 = 0, a_1 = 12, a_2 = 34, P(a) = P(\text{支}|\text{support}) \dots$ （還有很多可能）
- 隱藏變數條件下之事件
  - $a = \text{support 支撐、vector 向量、machine 機}$
  - 以上的  $(e, c)$  事件，經過  $P(a)$  加權後，可用於估計  $P(c | e)$
- E Step: (i) 窮舉所有可能的  $a$  (ii) 用  $P(c | e)$  估計  $P(a)$  (iii) 估算加權的  $(e, c)$
- M Step: (i) 取出  $\text{count}(e, c), \text{count}(e)$  的期望值 (ii) 估計  $P(c | e)$

# 運用 EM 演算法——以雙語術語詞彙對應為例

- 中、英詞彙之間的翻譯機率模型（就像兩枚加重銅板 A, B）
  - $P(c | e)$ （就像  $P_A(H), P_B(H)$ ）
- 此一模型產生了 10 萬組計算機雙語名詞，e.g.,
  - $P(\text{平行度} | \text{degree of parallelism})$  涉及隱藏的對應  $a$  以及其機率  $P(a)$
  - 也就是  $P(\text{平行度} | \text{degree of parallelism})$  等於以下的總合：
    - $P(a = \text{EMPTY (of) 平行 (parallelism) 度 (degree)}) = P(\text{度} | \text{degree}) P(\text{EMPTY} | \text{of}) P(\text{平行} | \text{parallelism})$
    - $P(a = \text{EMPTY (degree) 平行 (of) 度 (parallelism)}) = P(\text{EMPTY} | \text{degree}) P(\text{平行} | \text{of}) P(\text{度} | \text{parallelism})$
    - ... 共  $3 \times 2 \times 1$ ——三英文詞重排（permutation）對應到 EMPTY+平行+度
- 兩個銅板模型中，不知道用 A 或 B 投擲，隱藏變數有兩個可能值 A, B
- 在雙語術語對應（平行度, degree of parallelism）隱藏變數  $a$  有 6 個可能值

# 如果知道隱藏變數

- 如果知道隱藏變數
  - degree of parallelism 平行 度  $P(a = \text{EMPTY (of) 平行 (parallelism) 度 (degree)}) = 1$
  - degree of freedom 自由 度  $P(a = \text{EMPTY (of) 自由 (freedom) 度 (degree)}) = 1$
  - master degree 碩士 學位  $P(a = \text{碩士 (master) 學位 (degree)}) = 1$
  - master slave system 主 從 系統  $P(a = \text{主 (master) 從 (slave) 系統 (system)}) = 1$
  - master of science 科學碩士  $P(a = \text{EMPTY (of) 科學 (science) 碩士 (master)}) = 1$
- 最大似然估計（頻率估計機率可以最大化樣本的機率值）
  - $P(\text{度} | \text{degree}) = 2/3 = 0.67$        $P(\text{學位} | \text{degree}) = 1/3 = 0.33$
  - $P(\text{平行} | \text{parallelism}) = 1.0$
  - $P(\text{碩士} | \text{master}) = 0.67$        $P(\text{主} | \text{master}) = 0.33$
  - $P(\text{從} | \text{slave}) = 1.0$
  - $P(\text{模型} | \text{model}) = 1.0$
  - $P(\text{科學} | \text{science}) = 1.0$
  - $P(\text{EMPTY} | \text{of}) = 1.0$

# E Step - 1：用模型估計隱藏變數的機率

- 在起始時，機率為平均分布 aligns 的各個機率值 = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)

0.17-- EMPTY (degree) 平行 (of) 度 (parallelism)

0.17-- EMPTY (degree) 平行 (parallelism) 度 (of)

0.17-- EMPTY (of) 平行 (degree) 度 (parallelism)

0.17-- EMPTY (of) 平行 (parallelism) 度 (degree)

0.17-- EMPTY (parallelism) 平行 (degree) 度 (of)

0.17-- EMPTY (parallelism) 平行 (of) 度 (degree)

- 之後 aligns 的各機率，必須以  $P(c | e)$  估計，並正規化

- 第一組：

0.11-- EMPTY (degree) 平行 (of) 度 (parallelism)

0.21-- EMPTY (degree) 平行 (parallelism) 度 (of)

0.16-- EMPTY (of) 平行 (degree) 度 (parallelism)

0.32-- EMPTY (of) 平行 (parallelism) 度 (degree)

0.11-- EMPTY (parallelism) 平行 (degree) 度 (of)

0.11-- EMPTY (parallelism) 平行 (of) 度 (degree)

- 第二組：

0.11-- EMPTY (degree) 自由 (of) 度 (freedom)

0.21-- EMPTY (degree) 自由 (freedom) 度 (of)

0.16-- EMPTY (of) 自由 (degree) 度 (freedom)

0.32-- EMPTY (of) 自由 (freedom) 度 (degree)

0.11-- EMPTY (freedom) 自由 (degree) 度 (of)

0.11-- EMPTY (freedom) 自由 (of) 度 (degree)

# E Step - 1：用模型估計隱藏變數的機率

- 第三組：

0.62-- 碩士 (master) 學位 (degree)

0.37-- 碩士 (degree) 學位 (master)

- 第四組：

0.17-- 主 (master) 從 (slave) 系統 (system)

0.17-- 主 (master) 從 (system) 系統 (slave)

0.17-- 主 (slave) 從 (master) 系統 (system)

0.17-- 主 (slave) 從 (system) 系統 (master)

0.17-- 主 (system) 從 (master) 系統 (slave)

0.17-- 主 (system) 從 (slave) 系統 (master)

- 第五組：

0.06-- EMPTY (master) 科學 (of) 碩士 (science)

0.06-- EMPTY (master) 科學 (science) 碩士 (of)

0.19-- EMPTY (of) 科學 (master) 碩士 (science)

0.47-- EMPTY (of) 科學 (science) 碩士 (master)

0.06-- EMPTY (science) 科學 (master) 碩士 (of)

0.16-- EMPTY (science) 科學 (of) 碩士 (master)



- 事件期望值儲存為 eccount (用 dictionary 的資料結構)

初始化  $\text{eccount} = \{\}$

累加  $\text{eccount}[(e, c)] += P(a)$



# E Step - 2 : 計算期望值 (累加變數機率 x 事件)

- 第一組 :

0.17-- EMPTY (degree) 平行 (of) 度 (parallelism)

0.17-- EMPTY (degree) 平行 (parallelism) 度 (of)

0.17-- EMPTY (of) 平行 (degree) 度 (parallelism)

0.17-- EMPTY (of) 平行 (parallelism) 度 (degree)

0.17-- EMPTY (parallelism) 平行 (degree) 度 (of)

0.17-- EMPTY (parallelism) 平行 (of) 度 (degree)

- 第二組 :

0.17-- EMPTY (degree) 自由 (of) 度 (freedom)

0.17-- EMPTY (degree) 自由 (freedom) 度 (of)

0.17-- EMPTY (of) 自由 (degree) 度 (freedom)

0.17-- EMPTY (of) 自由 (freedom) 度 (degree)

0.17-- EMPTY (freedom) 自由 (degree) 度 (of)

0.17-- EMPTY (freedom) 自由 (of) 度 (degree)

- 第三組 :

0.50-- 碩士 (master) 學位 (degree)

0.50-- 碩士 (degree) 學位 (master)

$$\text{EMPTY (degree)} = 0.17 + 0.17 + 0.17 + 0.17$$

$$\text{EMPTY (of)} = 0.17 + 0.17 + 0.17 + 0.17$$

$$\text{EMPTY (parallelism)} = 0.17 + 0.17$$

$$\text{平行 (of)} = 0.17 + 0.17$$

$$\text{平行 (parallelism)} = 0.17 + 0.17$$

$$\text{平行 (degree)} = 0.17 + 0.17$$

$$\text{度 (parallelism)} = 0.17 + 0.17$$

$$\text{度 (of)} = 0.17 + 0.17 + 0.17 + 0.17$$

$$\text{度 (degree)} = 0.17 + 0.17 + 0.17 + 0.17$$

$$\text{度 (freedom)} = 0.17 + 0.17$$

$$\text{自由 (degree)} = 0.17 + 0.17$$

$$\text{自由 (of)} = 0.17 + 0.17$$

$$\text{自由 (freedom)} = 0.17 + 0.17$$

$$\text{碩士 (master)} = 0.50 + 0.17 + 0.17$$

$$\text{學位 (degree)} = 0.50$$

$$\text{碩士 (degree)} = 0.50$$

# E Step - 1 : 計算變數機率 x 事件

- 第一組：

0.11-- EMPTY (degree) 平行 (of) 度 (parallelism)

0.21-- EMPTY (degree) 平行 (parallelism) 度 (of)

0.16-- EMPTY (of) 平行 (degree) 度 (parallelism)

0.32-- EMPTY (of) 平行 (parallelism) 度 (degree)

0.11-- EMPTY (parallelism) 平行 (degree) 度 (of)

0.11-- EMPTY (parallelism) 平行 (of) 度 (degree)

- 第二組：

0.11-- EMPTY (degree) 自由 (of) 度 (freedom)

0.21-- EMPTY (degree) 自由 (freedom) 度 (of)

0.16-- EMPTY (of) 自由 (degree) 度 (freedom)

0.32-- EMPTY (of) 自由 (freedom) 度 (degree)

0.11-- EMPTY (freedom) 自由 (degree) 度 (of)

0.11-- EMPTY (freedom) 自由 (of) 度 (degree)

- 第三組：

0.62-- 碩士 (master) 學位 (degree)

0.37-- 碩士 (degree) 學位 (master)

EMPTY (degree) = 0.11+0.21+ 0.11+0.21

EMPTY (of) = 0.16+0.32+0.16+0.32

EMPTY (parallelism) = 0.11+0.11

平行 (of) = 0.11+0.11

平行 (parallelism) = 0.21+0.32

平行 (degree) = 0.16+0.11

度 (parallelism) = 0.11+0.16

度 (of) = 0.21+0.11+0.21+0.11

度 (degree) = 0.32+0.11+ 0.32+0.11

度 (freedom) = 0.11+0.16

自由 (degree) = 0.16+0.11

自由 (of) = 0.11+0.11

自由 (freedom) = 0.21+0.32

碩士 (master) = 0.62

學位 (degree) = 0.62

碩士 (degree) = 0.37

# M Step

- 以事件期望值（次數）估計模型參數（最大似然估計）
- E Step 中事件期望值儲存為 eccount (dictionary)
- $P(c | e) = \text{eccount}[(e, c)] / \text{count}(e)$