機率理論與醫療診斷

戴久永

大多數疾病在發作的時候，必然會產生一種或多種症狀。醫生在瞭解病患的病情後，嘗試去推論病人患的是那一種病症，這就是醫療診斷。然而一個常遇到的困擾却是許多病在發作初期所呈現的症狀十分相像，因此醫生必須決定這數種病症中的那一種是使病人生病最可能的原因。雖然這種診斷工作通常是依據醫生本人的判斷與直覺，有些研究人員建議利用機率理論來輔助醫療診斷。

設 $\{d_1, d_2 \cdots\cdots d_k\}$ 爲 k 種疾病，$\{s_1, s_2 \cdots\cdots s_k\}$爲疾病的症狀集合。例如疾病爲痲疹，則顯現的症狀爲發燒和出疹。又如疾病爲肺病，則胸部X光片可以看到某種狀況。

設 S ＝世上所有恰患一種疾病的人的集合。

$A = S$ 中呈現一種或多種症狀的病人。

這裡我們假設S中的病人，有些沒有任何可資辨認症狀。另外，設$D_i = S$中患d_i疾病的人的事件，$i=1, 2\cdots\cdots k$。

我們假設到醫院求診的病人是S中的一樣本點。另一方面，我們認爲病患會到醫院求診，是由於他有患病的症象，因此他是A中一樣本點。既然醫生想要找出生病的最可能原因，他自然想知道以下各機率

$$P(D_i \mid A) \qquad i=1, 2\cdots\cdots k$$

其中$P(D_i \mid A)$表示已知病人有某症狀，是由於d_i引起的機率。假若我們可以得到

$$P(A \mid D_i), i=1, 2\cdots\cdots k \tag{1}$$

和 $$P(D_i), i=1, 2\cdots\cdots k \tag{2}$$

則利用貝氏定理，我們可以計算出

$$P(D_i \mid A) = \frac{P(A|D_i)P(D_i)}{P(A|D_1)P(D_1)+\cdots+P(A|D_k)P(D_k)}$$

⑵式中的機率值可以從能取得的醫療數據資料中估計得出，方法如下：隨機選出恰有 d_1，d_2 …… d_k 疾病之一的病歷多份，例如N份，對於每一種疾病 d_i，決定患該病的人數，設爲N_i。由於 $P(D_i)$ 爲患病 d_i 與所有恰患 $\{d_1, d_2 \cdots\cdots d_k\}$ 病之一的人的次數比的近似值，因此我們可以估計 $P(D_i)$ 如下

$$P(D_i) \simeq \frac{N_i}{N}, \quad i = 1, 2 \cdots\cdots k \tag{3}$$

爲了估計 $D(A \mid D_i)$，利用條件機率的定義

$$P(A \mid D_i) = \frac{P(A \cap D_i)}{P(D_i)} \tag{4}$$

由於⑷式中右端分母已估計如上，我們只需要找出 $P(A \cap D_i)$，做法如下。對於每一種疾病 d_i，由病歷記錄可以發現患疾病 d_i 同時呈現一種或多種可辨認症狀的病患人數，設爲 N_i，則我們可估計 $P(A \cap D_i)$ 如下：

$$P(A \cap D_i) \simeq \frac{N_i}{N} \tag{5}$$

將⑶式和⑸式所得估計值代入⑷式，就可得出條件機率 $P(A \mid D_i)$ 的估計值。我們以例題解說如下

例：假設每種疾病 d_1，d_2，d_3 都會產生以下一種或多種可辨認症狀

s_1 = 沒食慾

s_2 = 胸口發痛

s_3 = 呼吸急促

s_4 = 瞳孔放大

假設依據一項全國性醫院資料調查，結果如表一所示（每萬人患 d_1，d_2，d_3 之一人數）

	患 d_i 人數	患 d_i 同時有 s_1，s_2，s_3，s_4 症狀之一項或多項的人數
疾病 d_1	3750	3000
d_2	2250	2050
d_3	4000	3500

表 一

則由⑶式和上表可得

$$P(D_1) \simeq \frac{3750}{10000} = 0.375$$

$$P(D_2) \simeq \frac{2250}{10000} = 0.225$$

$$P(D_3) \simeq \frac{4000}{10000} = 0.400$$

由表一和⑸式可得

$$P(A \cap D_1) \simeq \frac{3000}{10000} = 0.300$$

$$P(A \cap D_2) \simeq \frac{2050}{10000} = 0.205$$

$$P(A \cap D_3) \simeq \frac{3500}{10000} = 0.350$$

因此，由⑷式可得

$$P(A \mid D_1) = \frac{P(A \cap D_1)}{P(D_1)} \simeq \frac{0.300}{0.375} \simeq 0.800$$

$$P(A \mid D_2) = \frac{P(A \cap D_2)}{P(D_2)} \simeq \frac{0.205}{0.225} \simeq 0.911$$

$$P(A \mid D_3) = \frac{P(A \cap D_3)}{P(D_3)} \simeq \frac{0.350}{0.400} \simeq 0.875$$

$$\sum_{i=1}^{3} P(A \mid D_i)P(D_i)$$
$$= \sum_{i=1}^{3} P(A \cap D_i)$$
$$= 0.300 + 0.205 + 0.350$$
$$= 0.855$$

所以

$$P(D_1 \mid A) \simeq \frac{0.300}{0.855} \approx 0.351$$

$$P(D_2 \mid A) \simeq \frac{0.205}{0.855} \approx 0.240$$

$$P(D_3 \mid A) \simeq \frac{0.350}{0.855} \approx 0.409$$

依據以上計算，在缺乏其他資料的狀況下，數字顯示醫生應以病人患有 d_3 而配方處置。

在一次訪問中，美國Dartmouth College的數學系教授 J.G Kemeny 博士認爲「醫師們隨時都在從事機遇性的預測，我認爲讓一個醫學預料的學生未具備一些機率概念就准許他畢業是不可原諒的錯誤。」或許由以上說明，讀者能體會到Kemeny 教授的言論不無道理。

參考資料

(1) H. R. Warner, A. F. Toronto, L. G. Veasey and R. Stephenson A Mathematical approach to medical diagnosis, *Journal of the American Medical Association* 177 (1961) 177 - 183.

(2) J. E. Overall and C. M. Williams Conditional probability program for diagnosis of thyroid function, *Journal of the American Medical Association* 183 (1963) 307 - 313.

(3) J. E. Overall, D.R. Gorham A pattern probability model for the classification of psychiatric patients, *Behavioral scienee* 8 (1963) 108 - 116.

(4) Howard Antor, Bernard Kolman *Applied finite mathewatice* 2nd ed. Academic press 1978.

(5) Dirk J. Struik *A concise history of mathewatics* 3rd ed. Dover 1917（中譯本「數學史」 吳定遠譯，水牛出版社印行 71年9月）。

(6) *L*. E. Maistrov *Probability theory, a historical sketch* 英譯本由 Samuel Kotz 譯，Accdemic Press 1974.

(7) 戴久永 機率導論 三民書局印行，72年10月

(8) 1983年5月24日 Los Angels Times

— 本文作者任教於交通大學運輸工程與管理學系 —