

② バイズの定理の応用例: 迷惑メールの判別

現在インターネット上には多量の迷惑メールが流出している。これは受信側でフィルタリング、すなわちある種のルールに従い迷惑メールか否かを判別し、迷惑メールならゴミ箱(のふなご)に捨てる、ということが行われている。

迷惑メールの判別には、

- ある種のキーワードが含まれている
- メールが書かれている文字コード (中国語、ロシア語等)
- メールの中身に特定の単語が含まれている
- メールの中身に特定の単語や添付ファイルの種類

などの条件が用いられる。これらの条件から、果てはメールが迷惑メールである確率をバイズの定理から計算し、その値に基づき捨てるか捨てないかを決定する バイズ判別 という方法がある。

ごく単純な例でその原理を示す。

例 ある20%のメールのうち、

- ① 30%が正常で、残りは迷惑メール
- ② 迷惑メールの8割は「base64エンコード」形式
- ③ 正常メールの「base64エンコード」形式は5%

であることがあらかじめわかっているとすると、

このとき、新原理に「base64エンコード」形式のメールがどれほどあるかを、迷惑メールである確率を求めよう。

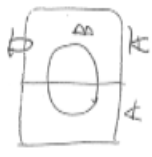
- ① ② ③ は、メールを多く受けとることによって徐々にその精度が上がるというので、使えばよくなる (ソフトに学習させること)。重作がよくなる、という。

答 } A = 迷惑メール (\bar{A} = 正常メール)
 B = 「base64エンコード」形式のメール
 とする。求める確率は $P_B(A)$

① $\Rightarrow P(A) = \frac{7}{10}, P(\bar{A}) = \frac{3}{10}$ — ④

② $\Rightarrow P_A(B) = \frac{8}{10}$ — ⑤

③ $\Rightarrow P_{\bar{A}}(B) = \frac{5}{100}$ — ⑥



$P_B(A) = \frac{P(B \cap A)}{P(B)}$

$P(B) = P(B \cap A) + P(B \cap \bar{A})$

$= P(A)P_A(B) + P(\bar{A})P_{\bar{A}}(B) \leftarrow \text{④, ⑤, ⑥}$

$= \frac{7}{10} \times \frac{8}{10} + \frac{3}{10} \times \frac{5}{100} = \frac{560 + 15}{1000} = \frac{575}{1000}$

$\therefore P(B \cap A) = P(A)P_A(B) = \frac{7}{10} \times \frac{8}{10} = \frac{56}{100}$

$\therefore P_B(A) = \frac{\frac{56}{100}}{\frac{575}{1000}} = \frac{560}{575} = 97.4\%$

(高い確率で、迷惑メールの可能性が高い)