ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

Cơ sở trí tuệ nhân tạo

Ôn tập: Tập hợp, phương pháp đếm và xác suất

Nguyễn Ngọc Đức

2021

Nội dung

Tập hợp

Tập hợp I

- *Tập hợp* là một khái niệm chỉ nhóm đối tượng nào đó mà chúng ta quan tâm
- Đóng vai trò chính trong việc đánh giá một mô hình xác suất

Định nghĩa

Một tập hợp là một tổ hợp các phần tử **không có thứ tự** và **riêng biệt**

- Khi phần tử x thuộc tập hợp A ta ký hiệu $\boldsymbol{x} \in \boldsymbol{A}$, ngược lại $\boldsymbol{x} \notin \boldsymbol{A}$

Tập hợp II

Lực lượng của một tập hợp

Số phần tử của tập hợp A được gọi là *lực lượng của tập hợp A*, ký hiệu $|A|$

- Cách xác định một tập hợp:
  1. Liệt kê: $A = \{1, 2, 3, a, b\}$
  2. Tính chất: $B = \{n \in \mathbb{N} \mid n \bmod 3 = 0\}$

Tập hợp III

- Quan hệ:
  1. *Bao hàm.* $A \subset B \Leftrightarrow \forall x, x \in A \to x \in B$
  2. *Bằng nhau.* $A = B \Leftrightarrow A \subset B \wedge B \subset A$
  3. *Hợp (union).* $A \cup B = \{x \mid x \in A \vee x \in B\}$
  4. *Giao (intersection).* $A \cap B = \{x \mid x \in A \wedge x \in B\}$
  5. *Hiệu.* $A \backslash B = \{x \mid x \in A \wedge x \notin B\}$
  6. *Tập bù (complement).* Nếu $A \subset U$, $A^c = U \backslash A$ là tập bù của A trong U

Phương pháp đếm

Phương pháp đếm I

Cho tập hợp A gồm n phần tử:

- *Chỉnh hợp (Permutation).* Mỗi cách sắp đặt có thứ tự k phần tử của A được gọi là một chỉnh hợp chập k của n

$$P_n^k = n \times (n-1) \times (n-2) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!}$$

- *Tổ hợp (Combination).* Mỗi **tập con** k phần tử của A là một tổ hợp chập k của A

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Phương pháp đếm II

- *Chỉnh hợp của đa tập hợp (Permutation of multisets)*. Một cách sắp xếp có thứ tự cho n phần tử từ đa tập hợp B với k phần tử khác nhau được gọi là chỉnh hợp chập n của đa tập hợp B.

$$P_n^* = \frac{n!}{\prod_{i=1}^{k} n_i!} \left(\text{với } n_i \text{ là số phần tử loại } i, \sum_{i=1}^{k} n_i = n \right)$$

- Ví dụ: Có bao nhiêu chuỗi kí tự khác nhau bằng cách sắp xếp các chữ cái của từ SUCCESS?

Phương pháp đếm III

- *Tổ hợp của đa tập hợp (Combination of multisets).* Một cách chọn **đa tập hợp** k phần tử từ một đa tập hợp B với n phần tử được gọi là đa tập hợp chập k của B

$$K_n^k = C_{n+k-1}^k$$

- Ví dụ: Tìm số nghiệm nguyên không âm của phương trình

$$x_1 + x_2 + x_3 = 12$$

Xác suất

Xác suất

- Xác suất là cơ hội một sự kiện diễn ra
- Biểu diễn toán học một không gian xác suất
  1. **Không gian mẫu:** tập kết quả của một thử nghiệm
  2. **Các sự kiện:** mỗi sự kiện là một tập con của không gian mẫu
  3. **Hàm xác suất:** chỉ định cơ hội một sự kiện xảy ra

Ví dụ

Tung đồng xu 2 lần

- **Không gian mẫu:** $\Omega = \{hh, ht, th, tt\}$
- Nếu A là sự kiện có ít nhất một đồng xu có mặt ngửa: $A = \{hh, ht, th\}$
- Nếu B là sự kiện xuất hiện cả mặt úp và mặt ngửa:

Ví dụ

Tung đồng xu 2 lần

- **Không gian mẫu:** $\Omega = \{hh, ht, th, tt\}$
- Nếu A là sự kiện có ít nhất một đồng xu có mặt ngửa: $A = \{hh, ht, th\}$
- Nếu B là sự kiện xuất hiện cả mặt úp và mặt ngửa: $B = \{ht, th\}$
- Hàm xác suất $\mathbb{P}(\cdot)$, nhận một sự kiện là tham số đầu vào và trả về giá trị số thực trong khoảng $[0, 1]$
- **Lưu ý:**
  - $\mathbb{P}(\emptyset) = 0$
  - $\mathbb{P}(\Omega) = 1$

Phép thử ngẫu nhiên

Tung xúc sắc I

Tung 2 con xúc sắc cân bằng, độc lập, 6 mặt; tìm xác suất tổng 2 mặt là chẵn

1. Không gian mẫu: $\Omega = \{1, ..., 6\}^2$
2. Mỗi phần tử trong không gian mẫu có dạng (i, j) với $i, j \in \{1, ..., 6\}$
3. Sự kiện quan tâm: $A = \{(i, j) \mid i + j \text{ chẵn}\}$
4. Hàm xác suất: đối xứng vì mỗi kết quả thuộc Ω đều có xác suất như nhau. Với $B \subset \Omega$:
$$\mathbb{P}(B) = \frac{|B|}{|\Omega|}$$

Tung xúc sắc II

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Hình 1: Các kết quả có thể xảy ra khi thảy 2 con xúc sắc

Ước tính $\mathbb{P}(A) = 0.5$

So khớp mật khẩu I

Giả sử một hệ thống bảo mật có mật khẩu dài 8 ký tự. Mỗi ký tự là 1 trong 62 ký tự sau:

- "a "z"
- "A "Z"
- "0 "9"

$\Rightarrow |\Omega| = 62^8$

- Giả sử hàm xác suất đối xứng, tỉ lệ một hacker đoán đúng mật khẩu: 62^{-8}. Khá bảo mật?

So khớp mật khẩu II

- Khi nhập mật khẩu đúng 1 hoặc nhiều ký tự, một sự kiện sẽ được ghi lại
- Ví dụ nếu passwords là 3xyZu4vN, một lần nhập mật khẩu là **3**5xyZ**4vN**; 4 ký tự trùng khớp, một sự kiện được ghi lại
- Rõ ràng cơ hội đoán trúng mật khẩu là rất thấp nhưng trong hệ thống bảo mật (hư cấu và quá đơn giản) này vẫn tồn tại một lỗ hổng bảo mật: hacker sẽ thử làm quá tải hệ thống log bằng các cuộc tấn công ngẫu nhiên
- Xác suất một sự kiện đăng nhập A được ghi lại $\mathbb{P}(A)$?

So khớp mật khẩu III

- Xác suất cuộc tấn công không được ghi lại (phần bù): $A^c = \Omega \backslash A$
- $|A^c| = 61^8 \Rightarrow \mathbb{P}(A^c) = \dfrac{61^8}{62^8} \approx 0.87802$
- $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) \approx 0.12198$
- Vậy nếu thực hiện 10^7 lượt đăng nhập sẽ có khoảng 1.2 triệu lượt ghi!!!

Bài toán sinh nhật I

Tìm xác suất một cặp người trong một căn phòng có chung ngày sinh nhật Quan sát:

- Nếu có 366 người trong phòng $\Rightarrow$ nguyên lý chuồng bồ câu
- Vậy nếu ít hơn?

Bài toán sinh nhật II

- Điều thú vị là trên thực tế nếu xem xét khoảng 50 người thì gần như là có 2 người có chung ngày sinh và với 23 người thì có tỉ lệ khoảng 50%
- Tỉ lệ nổi tiếng này đạt được danh hiệu **nghịch lý sinh nhật (the birthday paradox)**
- Tuy nhiên chúng ta chỉ nên gọi là bài toán sinh nhật

Bài toán sinh nhật III

Mô hình hóa bài toán:

- Xét n người trong một căn phòng, giả sử ngày sinh của mọi người tuân theo phân phối đều thuộc tập $\{1, ..., 365\}$
- Không gian mẫu Ω được biểu diễn với các biến $(x_1, ..., x_n)$ với $x_i \in \{1, ..., 365\} \Rightarrow |\Omega| = 365^n$
- Tập sự kiện $A = \{(x_p, ..., x_q) \mid \exists i, j \in [p, q] \wedge i \neq j \wedge x_i = x_j\}$

Bài toán sinh nhật IV

- Phần bù $|A^c| = 365 \cdot 364 \cdot \ldots \cdot (365 - n + 1) = \dfrac{365!}{(365-n)!}$
- Vậy $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \dfrac{|A^c|}{|\Omega|} = 1 - \dfrac{365!}{(365-n)! \cdot 365^n}$
- Với $n = 23$, $\mathbb{P}(A) \approx 0.5073$ và với $n = 50$, $\mathbb{P}(A) \approx 0.9704$

Xác suất và quan hệ tập hợp

Nguyên lý bù trừ

Xét 2 sự kiện A và B. Xác suất xảy ra một trong 2 sự kiện A và B:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Xét n sự kiện $C_i, ..., C_n$. Xác suất 1 trong n sự kiện xảy ra:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} C_i\right) = \sum_{\emptyset \neq J \subseteq \{1,...,n\}} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{j \in J} A_j\right)$$

Ví dụ: Bài toán trao quà I

- Có n gói quà được trao tặng cho n sinh viên tương ứng
- Tuy nhiên khi phát quà các gói quà này thường được phát ngẫu nhiên
- Xác suất sinh viên nhận được đúng gói quà của mình $\frac{1}{n!}$
- Xác suất tất cả gói quà được trao sai?

Ví dụ: Bài toán trao quà II

- Gọi A_i là sự kiện gói quà thứ i được trao đúng cho sinh viên thứ i
- Sự kiện cần tìm $B = A_1^c \cap A_2^c \cap ... \cap A_n^c$
- Luật De Morgan $B^c = A_1 \cup A_2 \cup ... \cup A_n$

Ví dụ: Bài toán trao quà III

- Xác suất trao đúng quà cho k sinh viên: $p_k = \dfrac{(n-k)!}{n!}$
- Xác suất tất cả gói quà được trao sai:

$$\begin{aligned}\mathbb{P}(B) = 1 - \mathbb{P}(B^c) &= 1 - \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} p_k \\ &= 1 - \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}\end{aligned}$$

Xác suất có điều kiện

- Nếu 2 sự kiện A và B độc lập với nhau. Xác suất 2 sự kiện đồng thời xảy ra:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$$

- A và B độc lập với nhau không đồng nghĩa với A và B không thể đồng thời xảy ra:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B) \neq \mathbb{P}(\emptyset) = 0$$

- Xác suất xảy ra sự kiện A **khi** xảy ra sự kiện B:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{1}$$

Định lý Bayes

- Nguyên lý Bayes hay định lý Bayes đơn giản là một suy diễn của công thức 1:

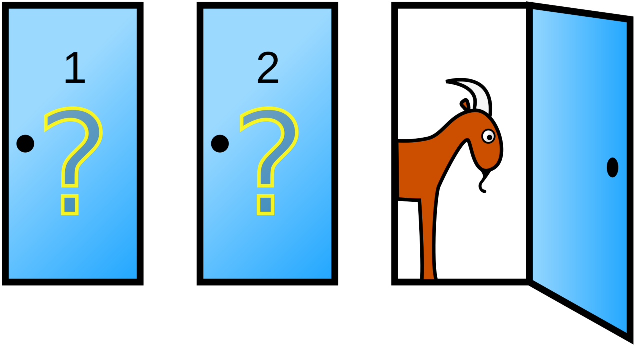
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

- **Tuy nhiên công thức này rất quan trọng trong thống kê**

$$\underbrace{\mathbb{P}(w|\mathcal{D})}_{Posterior} = \frac{1}{\underbrace{\mathbb{P}(\mathcal{D})}_{Normalization}} \overbrace{\mathbb{P}(\mathcal{D}|w)}^{\text{Likelihood}} \overbrace{\mathbb{P}(w)}^{Prior}$$

Bài toán Monty Hall I

- Bài toán Monty Hall là một bài toán nổi tiếng được đặt ra và giải vào năm 1975 bởi nhà toán học Steve Selvin



Bài toán Monty Hall II

- Gọi A_i là sự kiện giải thưởng nằm ở ô i, B_i là sự kiện cửa i được mở bởi Monty
- Nếu người chơi chọn cửa 1, Monty mở cửa 2. Ta có:

$$\mathbb{P}(A_1 \mid B_2) = \frac{\mathbb{P}(B_2 \mid A_1)\mathbb{P}(A_1)}{\mathbb{P}(B_2)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

$$\mathbb{P}(A_3 \mid B_2) = \frac{\mathbb{P}(B_2 \mid A_3)\mathbb{P}(A_3)}{\mathbb{P}(B_2)} = \frac{1 \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$