

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO ĐỒ ÁN MÔN HỌC
TOÁN ỨNG DỤNG VÀ THỐNG KÊ
Chủ đề: Mô hình Markov ẩn

Giảng viên lý thuyết: PGS. TS. Nguyễn Đình Thúc
Giảng viên hướng dẫn thực hành:

- Thầy Nguyễn Văn Quang Huy
- Cô Võ Nam Thục Đoan

Lớp: 20TN

Thành viên thực hiện:

- 20120131 – Nguyễn Văn Lộc
- 20120536 – Võ Trọng Nghĩa
- 20120572 – Nguyễn Kiều Minh Tâm

THÀNH PHỐ HỒ CHÍ MINH, THÁNG 4-5 NĂM 2022

Lời nói đầu

1 Lý thuyết mô hình Markov ẩn

Mô hình Markov ẩn (Hidden Markov model – HMM) là một mô hình máy học cổ điển thông dụng trong việc xử lý chuỗi.

1.1 Các thành phần của một mô hình Markov ẩn là gì? Chúng khác gì với mô hình Markov?

Một mô hình Markov ẩn có cấu tạo như sau^{1,2}:

- $Q = q_1 q_2 \dots q_N$: tập hợp N trạng thái (states).
- $A = a_{11} \dots a_{ij} \dots a_{NN}$: ma trận xác suất chuyển (transition probability matrix) A , a_{ij} là xác suất chuyển từ trạng thái i sang trạng thái j , $\sum_{j=1}^N a_{ij} = 1, \forall i$.
- $O = o_1 o_2 \dots o_T$: chuỗi các quan sát (observations), lấy từ bộ từ vựng (vocabulary) $V = v_1, v_2, \dots, v_V$.
- $B = b_i(o_t)$: ma trận xác suất phụ thuộc trạng thái (emission probabilities), thể hiện xác suất một quan sát o_t được tạo thành từ trạng thái i .
- $\pi = \pi_1, \pi_2, \dots, \pi_N$: phân phối xác suất ban đầu theo trạng thái, có nghĩa là π_i thể hiện xác suất xích Markov bắt đầu ở trạng thái i . Một số trạng thái j có thể có $\pi_j = 0$, do chúng không thể là trạng thái ban đầu của xích Markov.

So với xích Markov (Markov chains), mô hình Markov ẩn có thêm 2 thành phần là O – chuỗi các quan sát và B – ma trận xác suất phụ thuộc trạng thái.

1.2 Các giả thiết (assumptions) đặt ra cho mô hình Markov ẩn là gì? Tìm ví dụ các bài toán mà các giả thiết này hợp lý và bất hợp lý.

Các giả thiết (assumptions) đặt ra cho mô hình Markov ẩn.

Một mô hình Markov ẩn có 2 giả thiết chính:

- Xác suất của một trạng thái cụ thể chỉ phụ thuộc vào trạng thái ngay trước đó (Markov Assumptions):

$$P(q_i | q_1 \dots q_{i-1}) = P(q_i | q_{i-1}). \quad (1)$$

Tài liệu tham khảo

- [1] Nguyễn Tiến Dũng. *Mô hình Markov ẩn*. URL: <https://jurgendn.github.io/mathematics/2020/06/16/Hidden-Markov-Model>.
- [2] Daniel Jurafsky and James H. Martin. “Hidden Markov model”. In: *Speech and Language Processing*. 3rd ed. Stanford, 2021. Chap. Appendix A. URL: [https://web.stanford.edu/%5Csim\\$jurafsky/slp3/A.pdf](https://web.stanford.edu/%5Csim$jurafsky/slp3/A.pdf).