

Trường điều kiện ngẫu nhiên - Conditional Random Field (CRF)

CRF là mô hình giúp xử lý bài toán về chuỗi tương đối giống với mô hình HMM. Mô hình CRF có tất cả ưu điểm của HMM và tránh được vấn đề thiên vị cho nhãn của mô hình cực đại hóa entropy Markov (MEMM) ([mccallum2000maximum]).

Mục đích của mô hình là học hàm ánh xạ $x_s \rightarrow y_s$. Tuy nhiên mỗi đầu ra y_s không độc lập với nhau. Mô hình CRF có khả năng dự đoán vector đầu ra $y = y_0, y_1, \dots, y_t$ thông qua tính toán xác suất có điều kiện của các biến ngẫu nhiên đưa ra trong vector quan sát là $x = x_0, x_1, \dots, x_t$ ([sutton2012introduction]). Mô hình CRF kết hợp phân loại phân biệt và mô hình xóa bằng đồ thị. Nhờ thế mà có thể mô hình hóa các đặc trưng một cách gọn gàng và sử dụng lượng lớn các thông tin đầu vào cho việc dự đoán. ([sutton2012introduction])

Mô hình CRF cũng được áp dụng cho nhiều lĩnh vực. Ví dụ như việc xử lý ngôn ngữ hay tiếng nói như bài toán gán nhãn từ ([ghosh2016part]), NER ([seker2017extending]), trích xuất thông tin ([ebersbach2016wrappers]) hay giải quyết mập mờ về ngữ pháp ([radziszewski2013tiered]). Trong Tin sinh học, các ứng dụng CRF có thể kể đến là tìm ra liên kết protein ([morales2018protein]), khám phá và dự đoán cấu trúc RNA [johansen2017deep].

Để mô tả mô hình CRF bằng đồ thị, gọi $G = (V, E)$ là một đồ thị. V biểu diễn tập đỉnh, E biểu diễn tập cạnh nối trực tiếp giữa 2 đỉnh. Mỗi đỉnh ν được gán một biến Y_ν , thế nên $Y = (Y_\nu)_{\nu \in V}$, với mọi Y_ν tuân theo tính chất của Markov. Nhờ vậy mà cả mô hình đồ thị có thể mô tả bằng biểu thức sau: ([lafferty2001conditional])

$$P(Y_\nu | X, Y_w, w \neq \nu) = P(Y_\nu | X, Y_w, w \sim \nu)$$

Trong đó, $w \sim \nu$ có nghĩa rằng w và ν kết nối với nhau trong G .

Điều cần chú ý trong CRF là việc mô phỏng CRF thì lại dùng đồ thị vô hướng. Vì thế, trong xử lý ngôn ngữ tự nhiên, các nhà nghiên cứu đã sử dụng

chuỗi tuyến tính CRF [yu2019named].

Hình 0.1: Cấu trúc của mô hình chuỗi tuyến tính CRF

Với bài toán nhận diện tên thực thể cũng thường sử dụng chuỗi tuyến tính CRF. Cấu trúc của chuỗi được nêu ra ở trên. Các đỉnh được gộp vào cấu trúc chuỗi tuyến tính, và chuỗi các đỉnh này tạo ra tương ứng với chuỗi Y . Khi cho một biến ngẫu nhiên $X = x_1, x_2, \dots, x_n$ cùng $Y = y_1, y_2, \dots, y_n$ với cấu trúc chuỗi tuyến tính thì phân phối xác suất có điều kiện của biến ngẫu nhiên Y là $P(Y|X)$ cũng tuân theo tính chất Markov.

$$P(y_i|x, y_1, y_2, \dots, y_n) = P(y_i|x, y_{i-1}, y_{i+1})$$

Vì thế, với chuỗi quan sát X , ta có thể sử dụng phân phối $P(Y|X)$ của biến ngẫu nhiên Y để dự đoán chuỗi Y [wallach2004conditional]