



Học máy

Bởi:

Wiki Pedia

Học máy, có tài liệu gọi là Máy học, (tiếng Anh: machine learning) là một lĩnh vực của trí tuệ nhân tạo liên quan đến việc phát triển các kỹ thuật cho phép các máy tính có thể "học". Cụ thể hơn, học máy là một phương pháp để tạo ra các chương trình máy tính bằng việc phân tích các tập dữ liệu. Học máy có liên quan lớn đến thống kê, vì cả hai lĩnh vực đều nghiên cứu việc phân tích dữ liệu, nhưng khác với thống kê, học máy tập trung vào sự phức tạp của các giải thuật trong việc thực thi tính toán. Nhiều bài toán suy luận được xếp vào loại bài toán NP-khó, vì thế một phần của học máy là nghiên cứu sự phát triển các giải thuật suy luận xấp xỉ mà có thể xử lý được.

Học máy có tính ứng dụng rất cao bao gồm máy truy tìm dữ liệu, chẩn đoán y khoa, phát hiện thẻ tín dụng giả, phân tích thị trường chứng khoán, phân loại các chuỗi DNA, nhận dạng tiếng nói và chữ viết, dịch tự động, chơi trò chơi và cử động rô-bốt (robot locomotion).

Tương tác với con người

Một số hệ thống học máy nỗ lực loại bỏ nhu cầu trực giác của con người trong việc phân tích dữ liệu, trong khi các hệ thống khác hướng đến việc tăng sự cộng tác giữa người và máy. Không thể loại bỏ hoàn toàn tác động của con người vì các nhà thiết kế hệ thống phải chỉ định cách biểu diễn của dữ liệu và những cơ chế nào sẽ được dùng để tìm kiếm các đặc tính của dữ liệu. Học máy có thể được xem là một nỗ lực để tự động hóa một số phần của phương pháp khoa học. Một số nhà nghiên cứu học máy tạo ra các phương pháp bên trong các khuôn khổ của thống kê Bayes.

Các loại giải thuật

Các thuật toán học máy được phân loại theo kết quả mong muốn của thuật toán. Các loại thuật toán thường dùng dùng bao gồm:

- Học có giám sát -- trong đó, thuật toán tạo ra một hàm ánh xạ dữ liệu vào tới kết quả mong muốn. Một phát biểu chuẩn về một việc học có giám sát là bài toán phân loại: chương trình cần học (cách xấp xỉ biểu hiện của) một hàm ánh

xạ một vector $[X_1, X_2, \dots, X_n]$ tới một vài lớp bằng cách xem xét một số mẫu dữ_liệu - kết_quả của hàm đó.

- Học không giám sát -- mô hình hóa một tập dữ_liệu, không có sẵn các ví dụ đã được gắn nhãn.
- Học nửa giám sát -- kết hợp các ví dụ có gắn nhãn và không gắn nhãn để sinh một hàm hoặc một bộ phân loại thích hợp.
- Học tăng cường -- trong đó, thuật toán học một chính sách hành động tùy theo các quan sát về thế giới. Mỗi hành động đều có tác động tới môi trường, và môi trường cung cấp thông tin phản hồi để hướng dẫn cho thuật toán của quá trình học.
- Chuyển đổi -- tương tự học có giám sát nhưng không xây dựng hàm một cách rõ ràng. Thay vì thế, cố gắng đoán kết quả mới dựa vào các dữ_liệu huấn luyện, kết quả huấn luyện, và dữ_liệu thử nghiệm có sẵn trong quá trình huấn luyện.
- Học cách học -- trong đó thuật toán học thiên kiến quy nạp của chính mình, dựa theo các kinh nghiệm đã gặp.

Phân tích hiệu quả các thuật toán học máy là một nhánh của ngành thống kê, được biết với tên lý thuyết học điện toán.

Các chủ đề về học máy

Danh sách các chủ đề của môn học này:

- Mô hình hóa các hàm mật độ xác suất điều kiện: hồi quy và phân loại
 - Mạng nơ-ron
 - Cây quyết định
 - Lập trình biểu thức gen
 - Lập trình di truyền
 - Hồi quy quá trình Gauss
 - Phân tích biệt thức tuyến tính
 - k láng giềng gần nhất
 - Độ dài thông điệp tối thiểu
 - Cảm tri nguyên
 - Hàm cơ sở xuyên tâm
 - Máy hỗ trợ vector
- Mô hình hóa các hàm mật độ xác suất qua các mô hình phát sinh:
 - Thuật toán cực đại kì vọng
 - Các mô hình đồ họa gồm mạng Bayes và mạng Markov
 - Ánh xạ topo phát sinh
- Các kỹ thuật suy luận xấp xỉ đúng:
 - Chuỗi Markov phương pháp Monte Carlo
 - Phương pháp biến thiên

- Tối ưu hóa: hầu hết các phương pháp trên đều sử dụng tối ưu hóa hoặc là các thể hiện của các thuật toán tối ưu hóa.

Xem thêm

- Trí tuệ nhân tạo
- Trí tuệ điện toán
- Khai phá dữ liệu
- Nhận dạng mẫu
- Các ấn bản quan trọng trong học máy (khoa học máy tính)
- Các ấn bản quan trọng trong học máy (thống kê)
- rô-bốt tự hành
- Lập trình suy diễn logic