

KHAI PHÁ DỮ LIỆU

(DATA MINING)

Giáo trình điện tử

haitruong@hcmut.edu.vn

haitruongcse.wordpress.com

Nội dung

- Chương 1: Tổng quan về khai phá dữ liệu
- Chương 2: Giải thuật Gradient Descent
- Chương 3: Các vấn đề tiền xử lý dữ liệu
- Chương 4: Hồi qui dữ liệu
- Chương 5: Phân loại dữ liệu
- Chương 6: Gom cụm dữ liệu
- Chương 7: Khai phá luật kết hợp
- Chương 8: Phát triển ứng dụng khai phá dữ liệu
- Chương 9: Các đề tài nghiên cứu trong khai phá dữ liệu

Tài liệu tham khảo

- [1] Jiawei Han, Micheline Kamber, “Data Mining: Concepts and Techniques”, Second Edition, Morgan Kaufmann Publishers, 2006.
- [2] David Hand, Heikki Mannila, Padhraic Smyth, “Principles of Data Mining”, MIT Press, 2001.
- [3] David L. Olson, Dursun Delen, “Advanced Data Mining Techniques”, Springer-Verlag, 2008.
- [4] Graham J. Williams, Simeon J. Simoff, “Data Mining: Theory, Methodology, Techniques, and Applications”, Springer-Verlag, 2006.
- [5] ZhaoHui Tang, Jamie MacLennan, “Data Mining with SQL Server 2005”, Wiley Publishing, 2005.
- [6] Oracle, “Data Mining Concepts”, B28129-01, 2008.
- [7] Oracle, “Data Mining Application Developer’s Guide”, B28131-01, 2008.

Môn học trước

- Phân tích - Thiết kế Giải thuật (501044)
- Hệ Cơ sở dữ liệu (503002)
- Trí tuệ nhân tạo (505004)

Hiểu biết - Kỹ năng đạt được

- Hiểu các bước trong quá trình khám phá tri thức
- Mô tả được các khái niệm cơ bản, công nghệ, và ứng dụng của khai phá dữ liệu
- Giải thích được các tác vụ khai phá dữ liệu phổ biến như hồi qui, phân loại, gom cụm, và khai phá luật kết hợp
- Nhận dạng được các vấn đề về dữ liệu trong giai đoạn tiền xử lý cho các tác vụ khai phá dữ liệu
- Hiểu cách sử dụng khai phá dữ liệu để có được các quyết định tốt hơn
- Sử dụng được các giải thuật và công cụ khai phá dữ liệu để phát triển ứng dụng khai phá dữ liệu
- Được chuẩn bị về kiến thức để có thể nghiên cứu trong lĩnh vực khai phá dữ liệu

Đánh giá

- Bài tập lớn + Thuyết trình : 40%
- Giữa kỳ : 20%
- Thi cuối kỳ : 40%

Yêu cầu đối với sinh viên

- Sinh viên nên có mặt tại lớp hơn 75%.
- Sinh viên nên đọc trước tài liệu tham khảo cho mỗi chương.
- Sinh viên nên làm các bài tập của mỗi chương.
- Sinh viên nên tham khảo thêm các tài liệu học tập khác, đặc biệt từ nguồn Internet.

Bài tập lớn - thuyết trình

- Sinh viên làm việc nhóm gồm 4-5 thành viên.
- Sinh viên nhận đề tài vào tuần thứ 1.
- Sinh viên chốt đề tài và bắt đầu thực hiện từ tuần thứ 2.

Hỏi & Đáp ...