

TRANG THÔNG TIN NHỮNG ĐÓNG GÓP MỚI VỀ MẶT HỌC THUẬT VÀ LÝ LUẬN CỦA LUẬN ÁN

Đề tài luận án: Phương pháp lựa chọn thuộc tính và kỹ thuật gom cụm dữ liệu phân loại sử dụng tập thô

Nghiên cứu sinh: Đỗ Sĩ Trường Mã số NCS: 915000393 Khóa: 2015

Chuyên ngành: Khoa Học Máy tính; Mã số ngành: 9480101

Cơ sở đào tạo: Đại học Lạc Hồng

Người hướng dẫn khoa học: PGS.TS. Nguyễn Thanh Tùng

Những đóng góp của luận án về khoa học và thực tiễn:

Về mặt khoa học:

Luận án tập trung vào việc ứng dụng Lý thuyết tập thô nhằm giải quyết hai vấn đề: (1) nghiên cứu thuật toán hiệu quả tìm tập rút gọn thuộc tính trong một bảng quyết định; (2) kỹ thuật gom cụm dữ liệu phân loại cho phép xử lý sự không chắc chắn trong quá trình gom cụm.

Bằng việc nghiên cứu các thuật toán đã được đề xuất bởi các nhà nghiên cứu, tìm ra các thiếu sót, luận án đã đề xuất một thuật toán mới lựa chọn thuộc tính trong một bảng quyết định dựa trên gom cụm. Kết quả thử nghiệm trên các tập dữ liệu thực tế lấy từ kho dữ liệu UCI cho thấy thuật toán đề xuất ACBRC là rất khả quan trong việc làm giảm số thuộc tính trong các bảng quyết định, đồng thời nâng cao được độ chính xác phân lớp.

Bằng việc nghiên cứu các thuật toán cơ sở đã được đề xuất bởi các nhà nghiên cứu, phân tích các thiếu sót, luận án đã đề xuất thuật toán gom cụm dữ liệu phân loại MMNVI theo phương pháp phân cấp. Kết quả thử nghiệm trên các tập dữ liệu thực tế lấy từ kho dữ liệu UCI cho thấy thuật toán MMNVI là một thuật toán ổn định, cho kết quả gom cụm tốt hơn hoặc ít ra là tương đương so với các thuật toán cơ sở. MMNVI là thuật toán có thể được sử dụng thành công trong việc gom cụm dữ liệu phân loại.

Các đóng góp chính trên đây đã được đăng trong hai bài báo trên Journal of Computer Science and Cybernetic. Ngoài các đóng góp chính trình bày trong luận án, nghiên cứu sinh là đồng tác giả của có một số kết quả khác liên quan đến đề tài luận án, bao gồm một bài báo quốc tế và ba báo cáo hội thảo khoa học trong nước.

Về mặt thực tiễn:

Khám phá tri thức từ CSDL là một lĩnh vực khoa học nhằm nghiên cứu để tạo ra những công cụ khai phá những thông tin, tri thức hữu ích, tiềm ẩn mang tính dự đoán trong các CSDL lớn. Tuy nhiên, với tốc độ tăng trưởng nhanh của dữ liệu ngày nay, việc nghiên cứu và ứng dụng các kỹ thuật khai phá dữ liệu còn gặp phải nhiều khó khăn, thách thức.

Lý thuyết tập thô là một công cụ toán học mạnh để xử lý dữ liệu mơ hồ, không chính xác, không đầy đủ và không chắc chắn. Vì vậy, những nghiên cứu và đóng góp của luận án trong lĩnh vực này có thể được ứng dụng thành công trong khám phá tri thức trong cơ sở dữ liệu, cũng như trong học máy, hệ chuyên gia, nhận dạng mẫu và có thể ứng dụng rộng rãi trong thực tiễn.

Đồng Nai, ngày 11 tháng 12 năm 2023

NGƯỜI HƯỚNG DẪN

NGHIÊN CỨU SINH

PGS. TS. NGUYỄN THANH TÙNG

ĐỖ SĨ TRƯỜNG