ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

BÙI VĂN THẮNG

LUẬT KẾT HỢP MỜ VÀ ỨNG DỤNG ĐỐI VỚI MỘT SỐ BÀI TOÁN DỰ BÁO

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN - 2014

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

BÙI VĂN THẮNG

LUẬT KẾT HỢP MỜ VÀ ỨNG DỤNG ĐỐI VỚI MỘT SỐ BÀI TOÁN DỰ BÁO

Chuyên ngành: KHOA HỌC MÁY TÍNH
Mã số: 60 48 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

HƯỚNG DẪN KHOA HỌC: TS. VŨ VINH QUANG

THÁI NGUYÊN - 2014

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi dưới sự hướng dẫn trực tiếp của **TS. Vũ Vinh Quang**.

Mọi trích dẫn sử dụng trong báo cáo này đều được ghi rõ nguồn tài liệu tham khảo theo đúng qui định.

Mọi sao chép không hợp lệ, vi phạm quy chế đào tạo, hay gian trá, tôi xin chịu hoàn toàn trách nhiệm.

Thái Nguyên, ngày27tháng8năm 2014

Tác giả

Bùi Văn Thắng

MỤC LỤC

CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

CNTT	Công nghệ thông tin
CSDL	Cơ sở dữ liệu
KPDL	Khai phá dữ liệu
KDD	Knowledge Discovery in Databases
ItemSet	Tập mục
Item	Mục

DANH MỤC BẢNG

DANH MỤC HÌNH VẼ

MỞ ĐẦU

1. Đặt vấn đề

Khai phá dữ liệu là một lĩnh vực nghiên cứu quan trọng trong lý thuyết về cơ sở dữ liệu, có nhiều ứng dụng trong đời sống xã hội. Mục đích chính là nhằm phát hiện những thông tin mới, các luật mới từ cơ sở dữ liệu đã có hay một cách tổng quát hơn là từ các kho dữ liệu. Rất nhiều lĩnh vực ứng dụng trong thực tiễn đều sử dụng công cụ khai phát dữ liệu và tìm kiếm tri thức. Trong lý thuyết về khai phá dữ liệu, khai phá luật kết hợp đang được quan tâm nghiên cứu nhiều trên thế giới. Một số hướng nghiên cứu đã và đang được các chuyên gia công nghệ thông tin tập chung nghiên cứu là: nghiên cứu thiết kế các hệ mờ cho các ứng dụng cụ thể như hệ trợ giúp quyết định, hệ điều khiển dựa trên hệ tri thức luật, hệ phân loại dựa trên hệ tri thức luật, hệ phân loại dựa trên lập luận dựa trên hệ luật ứng dụng trong các lĩnh vực như: kinh doanh, thị trường chứng khoán và dự đoán thị trường, công nghệ sinh học, giáo dục và đào tạo,…

Một số hướng nghiên cứu trong khai phá dữ liệu

- **Luật kết hợp nhị phân**: Đây là hướng nghiên cứu đầu tiên của luật kết hợp. Thuật toán tiêu biểu là Apriori.
- **Luật kết hợp có thuộc tính số và thuộc tính hạng mục**: Nghiên cứu các hệ CSDL có thuộc tính số hoặc thuộc tính hạng mục bằng cách rời rạc hóa dữ liệu cho thuộc tính số để chuyển chúng về thuộc tính nhị phân.
- **Luật kết hợp mờ**: Phương pháp rời rạc hóa dữ liệu có thuộc tính số và thuộc tính hạng mục gặp phải vấn đề“điểm biên gãy”. Để khắc phục điều này, các nhà nghiên cứu đề xuất sử dụng lý thuyết tập mờ và xây dựng các luật kết hợp dạng mờ.
- **Luật kết hợp có trọng số**: Sử dụng phương pháp tính độ hỗ trợ cho các tập mục dựa trên trọng số của các tập mục.

Ngoài ra, còn một số hướng nghiên cứu: khai phá luật kết hợp song song, khai phá luật kết hợp nhiều mức, luật kết hợp tiếp cận theo hướng tập thô,…

Luận văn tập trung nghiên cứu vào khai phá **Luật kết hợp mờ và ứng dụng đối với bài toán dự báo**.

2. Hướng nghiên cứu của đề tài

- Nghiên cứu lý thuyết tập mờ.
- Nghiên cứu khai phá dữ liệu và khai phá dữ liệu mờ trên CSDL. Tìm hiểu một số thuật toán trong khai phá dữ liệu: Apriori mờ, thuật toán FP Growth, thuật toán biểu diễn dữ liệu giao dịch mờ dựa trên cây FP-Tree.
- Cài đặt thử nghiệm một số thuật toán khai phá dữ liệu mờ và thử nghiệm trên một số bộ dữ liệu.Đánh giá kết quả sau khi thử nghiệm.

3. Đối tượng nghiên cứu

- Nghiên cứu phương pháp luận cho phép phát hiện tri thức dạng luật mờ, như luật kết hợp mờ, luật mờ với thuộc tính có trọng số,… từ các kho dữ liệu.
- Cơ sở lý thuyết của việc nghiên cứu lập luận xấp xỉ dựa trên lý thuyết tập mờ, phương pháp tính toán các thông tin mờ, đánh giá các phương pháp để lấy quyết định.
- Ứng dụng luật kết hợp mờ đối với một số bài toán dự báo.

4. Kết quả đạt được

- Tìm hiểu thuật toán nén dữ liệu giao dịch mờ dựa trên cây FP Tree, khai phá tập phổ biến dựa trên cây đã xây dựng. Đây là hướng nghiên cứu mới, giúp làm giảm thời gian khai phá tập phổ biến rất nhiều so với thuật toán Apriori mờ.
- Cài đặt thử nghiệm thuật toán Apriori mờ và thuật toán khai phá luật kết hợp mờ dựa trên thuật toán Fp-Growth.
- Thử nghiệm hai thuật toán trên với một số bộ dữ liệu, so sánh các kết quả đã thu được sau khi thử nghiệm.

5. Bố cục của luận văn

Phần mở đầu

Chương 1: Một số kiến thức cơ bản về khai phá dữ liệu

Chương 2: Khai phá luật kết hợp mờ

Chương 3: Ứng dụng khai phá dữ liệu trong mô hình dự báo

Kết luận

Tài liệu tham khảo