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

NGUYỄN ĐÔNG HUY

MỘT SỐ KỸ THUẬT PHÂN CỤM DỮ LIỆU VÀ ỨNG DỤNG PHÂN LOẠI KHÁCH HÀNG SỬ DỤNG DỊCH VỤ VIỄN THÔNG

LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

Thái Nguyên - 2014

LỜI CẢM ƠN

Trong quá trình làm luận văn, bản thân em đã nhận được nhiều sự giúp đỡ chỉ bảo tận tình của các thầy cô giáo, sự giúp đỡ, tạo điều kiện của gia đình, bạn bè để hoàn thành khóa luận đúng tiến độ.

Em xin trân trọng cảm ơn thầy giáo TS. Nguyễn Huy Đức đã trực tiếp hướng dẫn nhiệt tình, chỉ bảo cặn kẽ trong quá trình làm luận văn.

Em cũng xin gửi lời cám ơn chân thành tới Ban lãnh đạo nhà trường, các cán bộ giảng viên của trường Đại học Công nghệ Thông tin và Truyền thông – Đại học Thái Nguyên đã tạo điều kiện thuận lợi để em hoàn thành tốt khóa luận.

Học viên

Nguyễn Đông Huy

Số hóa bởi Trung tâm Học liệu – Đại học Thái Nguyên http://www.lrc-tnu.edu.vn/

LỜI CAM ĐOAN

Em xin cam đoan những kiến thức trình bày trong luận văn này là do em tìm hiểu, nghiên cứu và trình bày lại theo cách hiểu của em. Trong quá trình làm luận văn em có tham khảo các tài liệu liên quan và đã ghi rõ nguồn tài liệu tham khảo đó. Phần lớn những kiến thức do em trình bày trong luận văn này chưa được trình bày hoàn chỉnh trong bất cứ tài liệu nào.

Thái Nguyên, ngày 10 tháng 4 năm 2014

Học viên

Nguyễn Đông Huy

Số hóa bởi Trung tâm Học liệu – Đại học Thái Nguyên http://www.lrc-tnu.edu.vn/

MỤC LỤC

Số hóa bởi Trung tâm Học liệu – Đại học Thái Nguyên http://www.lrc-tnu.edu.vn/

DANH SÁCH HÌNH VẼ

DANH SÁCH BẢNG BIỂU

DANH MỤC CÁC TỪ VIẾT TẮT

Stt	Viết tắt	Cụm từtiếngAnh	Cụm từtiếngViệt
1	CNTT	InformationTechnology	Côngnghệthôngtin
2	CSDL	Database	Cơ sởdữliệu
3	KDD	KnowledgeDiscovery inDatabase	Khámphá trithứctrongcơ sởdữliệu
4	KPDL	Datamining	Khaiphá dữliệu
5	KPVB	TextMining	Khaiphávănbản
6	PCDL	DataClustering	Phâncụmdữ liệu

LỜIMỞĐẦU

Trongnhữngnămgầnđâycùngvớipháttriểnnhanhchóngcủakhoahọckỹthuậtlà sựbùngnổvềtrithức.Khodữliệu,nguồntrithứccủanhânloạicũngtrởnênđồsộ,vôtậnlà mchovấnđềkhaitháccácnguồntrithứcđóngàycàngtrởnênnóngbỏngvàđặtratháchthứ clớnchonềncôngnghệthôngtinthếgiới.

Đốivớimộtdoanhnghiệpthôngtin diđộngviệcpháttriểnthuêbao mới đểkiếmtìmlợinhuậnvàothờiđiểmhiệntạiđãkhôngcònđemlạihiệuquả.Thayvàođólà mộtphươngánkinhdoanhtiếnđếnpháttriểnchấtlượngdịchvụ vàcungcấpthêmnhiềudịchvụgiátrịgiatăng.Tuynhiêncácdịchvụtruyềnthốngnhưtho ại,nhắntinvẫncóthểđemlạinguồnlợinhuậncaohơnnếukíchthíchđượcnhucầusửdụng củakháchhàng. Để thực hiện được điều đó, các doanh nghiệp phải không ngừng giữ vững được khách hàng hiện có mà còn phải đưa ra được các chiến lược phát triển kinh doanh dài hạn, phân loại được các nhóm khách hàng đang sử dụng để từ đó có chính sách phân khúc thị trường hợp lý. Vì vậy, em dựa vào thực trạng như trên và kết hợp với kỹ thuật phân cụm trong khai phá dữ liệu để thực hiện đề tài: *"Một số kỹ thuật phân cụm dữ liệu và ứng dụng phân loại khách hàng sử dụng dịch vụ Viễn thông"*

Bố cục luận văn gồm 3 chương:

Chương 1: Trình bày một cách tổng quan các kiến thức cơ bản về khai phá dữ liệu và phát hiện tri thức, các kỹ thuật phân cụm trong khai phá dữ liệu.

Chương 2: Giới thiệu một số dữ liệu phân cụm phổ biến thường được sử dụng trong khai phá dữ liệu và phát hiện tri thức.

Chương 3: Sử dụng kỹ thuật phân cụm để ứng dụng vào phân loại khách hàng sử dụng dịch vụ viễn thông. Trong chương này cũng trình bày chương trình mô phỏng áp dụng kỹ thuật phân cụm để phân loại sử dụng dịch vụ Viễn thông.

Phần kết luận của luận văn tổng kết lại những vấn đề đã nghiên cứu, đánh giá kết quả nghiên cứu, hướng phát triển của đề tài.

CHƯƠNG 1: TỔNG QUAN VỀ KHAI PHÁ DỮ LIỆU