

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

ĐINH ĐỨC LONG

**KỸ THUẬT DATAMINING ĐỂ KHUYẾN NGHỊ
KHÁCH HÀNG TRONG HỆ THỐNG BI
(BUSINESS INTELLIGENCE)**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN - 2015

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

ĐINH ĐỨC LONG

KỸ THUẬT DATAMINING ĐỂ KHUYẾN NGHỊ
KHÁCH HÀNG TRONG HỆ THỐNG BI
(BUSINESS INTELLIGENCE)

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

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

HƯỚNG DẪN KHOA HỌC: TS HOÀNG ĐỖ THANH TÙNG

THÁI NGUYÊN - 2015

LỜI CAM ĐOAN

Luận văn là kết quả nghiên cứu và tổng hợp các kiến thức mà học viên đã thu thập được trong quá trình học tập tại trường Đại học Công nghệ Thông tin và Truyền thông - Đại học Thái Nguyên, dưới sự hướng dẫn, giúp đỡ của các thầy cô và bạn bè đồng nghiệp, đặc biệt là sự hướng dẫn, giúp đỡ của TS Hoàng Đỗ Thanh Tùng - Viện Công nghệ Thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Em xin cam đoan luận văn không phải là sản phẩm sao chép của bất kỳ tài liệu khoa học nào.

Thái Nguyên, ngày 30 tháng 6 năm 2015

Học viên

Đinh Đức Long

LỜI CẢM ƠN

Em xin gửi lời cảm ơn tới Trường Đại học Công Nghệ Thông Tin và Truyền thông - ĐHTN, Viện Công nghệ Thông tin - Viện Hàn lâm Khoa học và Công nghệ Việt Nam, nơi các Thầy cô đã tận tình truyền đạt các kiến thức quý báu cho em trong suốt quá trình học tập. Xin cảm ơn Ban chủ nhiệm khoa và các cán bộ khoa đã tạo điều kiện tốt nhất cho chúng em học tập và hoàn thành đề tài tốt nghiệp của mình.

Đặc biệt, em xin gửi lời cảm ơn sâu sắc nhất tới TS Hoàng Đỗ Thanh Tùng, người đã trực tiếp hướng dẫn, giúp đỡ để em hoàn thành luận văn của mình.

Mặc dù đã hết sức cố gắng hoàn thành luận văn với tất cả sự nỗ lực của bản thân, nhưng luận văn vẫn còn những thiếu sót. Kính mong nhận được những ý kiến đóng góp của quý thầy, cô và bạn bè đồng nghiệp.

Em xin chân thành cảm ơn!

Thái Nguyên, ngày 30 tháng 6 năm 2015

Học viên

Đinh Đức Long

MỤC LỤC

LỜI CAM ĐOAN.....	i
LỜI CẢM ƠN.....	iv
MỤC LỤC.....	v
DANH MỤC CÁC TỪ VIẾT TẮT.....	viii
DANH MỤC CÁC HÌNH VẼ.....	ix
MỞ ĐẦU.....	1
I. ĐẶT VẤN ĐỀ.....	1
II. ĐỐI TƯỢNG VÀ PHẠM VI NGHIÊN CỨU.....	4
III. Ý NGHĨA KHOA HỌC VÀ THỰC TIỄN CỦA ĐỀ TÀI.....	4
IV. PHƯƠNG PHÁP NGHIÊN CỨU.....	4
CHƯƠNG 1: TÌM HIỂU HỆ THỐNG BUSINESS INTELLIGENCE.....	5
1.1 Business Intelligence (BI) là gì ?.....	5
1.2 Vai trò của Data Mining trong hệ thống BI.....	7
1.2.1 Khai phá dữ liệu(Data Mining - DM).....	8
1.2.2 Khám phá tri thức trong CSDL (Knowledge Discovery in Database - KDD).....	9
1.2.3 Vai trò của DM trong hệ thống BI.....	12
1.3 Hệ thống khuyến nghị khách hàng.....	16
1.3.1 Ma trận khả dụng.....	16
1.3.2 Các ứng dụng của hệ thống khuyến nghị.....	18
1.3.3 Xây dựng ma trận khả dụng.....	19

1.4. Kết luận chương 1.....	19
CHƯƠNG 2. KHAI PHÁ DỮ LIỆU TRONG HỆ THỐNG BI	20
2.1 Giới thiệu một số kỹ thuật khai phá dữ liệu dùng trong BI	20
2.1.1 Phân cụm.....	20
2.1.2 Luật kết hợp	21
2.1.3 Lý thuyết luật kết hợp	22
2.1.4 Thuật toán Apriori sinh luật kết hợp	23
2.2 Hệ thống khuyến nghị dựa trên nội dung.....	26
2.2.1 Hồ sơ hàng hóa.....	26
2.2.2 Khám phá đặc điểm của các dữ liệu	27
2.2.3 Lây đặc điểm của mặt hàng từ thẻ (Tag).....	29
2.2.4 Trình bày hồ sơ hàng hóa.....	30
2.2.5 Hồ sơ người dùng.....	32
2.2.6 Khuyến nghị sản phẩm cho người dùng dựa trên nội dung.....	33
2.2.7 Các thuật toán phân lớp.....	35
2.3. Lọc cộng tác (collaborative filtering).....	38
2.3.1 Đo độ tương đồng	38
2.3.2 Tính đối ngẫu của sự tương đồng.....	42
2.3.3 Phân cụm những người dùng và các mặt hàng	45
2.4 Kết luận chương 2.....	47
CHƯƠNG 3: ỨNG DỤNG TRIỂN KHAI THỬ NGHIỆM HỆ THỐNG	
TƯ VẤN CHỌN PHIM	48

3.1 Bài toán.....	48
3.2 Xây dựng hệ tư vấn phim.....	50
3.2.1 Chuẩn bị dữ liệu.....	50
3.2.3 Thiết kế hệ thống.....	54
3.2.2 Lựa chọn giải pháp.....	56
3.3 Kết luận chương 3.....	62
KẾT LUẬN VÀ KIẾN NGHỊ.....	64
TÀI LIỆU THAM KHẢO.....	65

DANH MỤC CÁC TỪ VIẾT TẮT

Từ viết tắt	Viết đầy đủ	Ý nghĩa
BI	Business Intelligence	Hệ thống trí tuệ doanh nghiệp
DSS	Decision Support Systems	Hệ thống hỗ trợ ra quyết định
DM	Data Mining	Khai phá dữ liệu
IMDB	Internet Movies DataBase	Dữ liệu các bộ phim trên internet
KDD	Knowledge Discovery in Database	Khám phá tri thức trong cơ sở dữ liệu
OLAP	On – Line Analytical Processing	Phân tích dữ liệu trực tuyến đa chiều
RS	Recommender System	Hệ thống khuyến nghị

DANH MỤC CÁC HÌNH VẼ

Hình 1.1. Nguồn gốc của khai thác dữ liệu.....	9
Hình 1.2. Các bước trong qui trình khám phá tri thức trong CSDL	10
Hình 1.3. Các bước của quá trình khai phá dữ liệu.....	11
Hình 1.4. Vai trò của khai thác dữ liệu và khám phá tri thức trong 3 lĩnh vực chính của một doanh nghiệp.	13
Hình 1.5. Vai trò của DM và KDD và các lĩnh vực quan tâm của DN.....	15
Hình 1.6. Một ma trận khả dụng đại diện cho việc đánh giá	16
các bộ phim theo thang từ 1-5.....	16
Hình 2.1. Một cây quyết định.....	38
Hình 2.2. Ma trận khả dụng được gợi ý trong hình 1.6	39
Hình 2.3. Độ khả dụng 3, 4 và 5 được thay thế bằng 1,	41
trong khi các đánh giá 1 và 2 bị loại bỏ	41
Hình 2.4. Ma trận khả dụng được gợi ý trong hình 2.2	42
Hình 2.5. Ma trận khả dụng cho người dùng và cụm các mặt hàng	45
Hình 3.1. Biểu đồ hoạt động hệ thống tư vấn phim.....	50
Hình 3.2. Kiến trúc hệ tư vấn phim.....	54

MỞ ĐẦU

I. ĐẶT VẤN ĐỀ

Trong những năm gần đây, sự phát triển của thương mại điện tử (E-Commerce) đã đem lại nhiều lợi ích to lớn cho nền kinh tế toàn cầu. Thông qua thương mại điện tử, nhiều loại hình kinh doanh mới được hình thành, trong đó có mua bán hàng trên mạng. Với hình thức mới này, người tiêu dùng có thể tiếp cận với hàng hóa một cách dễ dàng và nhanh chóng hơn rất nhiều so với phương thức mua bán truyền thống trong môi trường cạnh tranh ngày càng tăng, các doanh nghiệp/tổ chức (DN/TC) đã nhận ra rằng để có thể thành công cũng như có được những kết quả tốt trong kinh doanh thì vấn đề nhận ra các xu hướng và cơ hội của thị trường là rất quan trọng, từ đó đáp ứng nhanh cho các nhu cầu của khách hàng mới. Một cách ngắn gọn hơn mục tiêu chính của các doanh nghiệp là hướng tới mục tiêu của các khách hàng của mình.

Ngày nay việc lưu trữ, xử lý dữ liệu để tổng hợp thông tin và hỗ trợ ra quyết định đã trở nên phổ biến đối với nhiều doanh nghiệp/tổ chức có nhiều giải pháp cho vấn đề này trong đó Business Intelligence (BI – giải pháp quản trị doanh nghiệp thông minh hay hệ thống trí tuệ doanh nghiệp) là một giải pháp tiêu biểu được nhiều DN/TC lựa chọn cho mục đích quản lý và điều hành các hoạt động của mình. Ở các nước phát triển, thuật ngữ Business Intelligence (BI) tạm dịch là giải pháp kinh doanh thông minh hay hệ thống trí tuệ doanh nghiệp không còn mới mẻ, tuy nhiên ở Việt Nam chúng ta lĩnh vực này vẫn đang ở mức sơ khai. Vậy BI là gì?

Business Intelligence (BI)

Có rất nhiều định nghĩa cũng như các quan điểm khác nhau về BI, mỗi định nghĩa đề cập đến một đặc trưng nổi bật của hệ thống BI nhưng chung qui lại tất cả đều đề cập đến khả năng hỗ trợ ra quyết định một cách hiệu quả hay BI còn được gọi là hệ thống hỗ trợ ra quyết định (Decision Support Systems