

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

MÔNG QUỐC TUẤN

**NGHIÊN CỨU MÔ HÌNH NGƯỜI SỬ DỤNG MỞ TRONG
CÁC HỆ THỐNG GỌI Ý THÔNG TIN THEO NHU CẦU**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN - 2017

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

MÔNG QUỐC TUẤN

**NGHIÊN CỨU MÔ HÌNH NGƯỜI SỬ DỤNG MỞ TRONG
CÁC HỆ THỐNG GỌI Ý THÔNG TIN THEO NHU CẦU**

Chuyên ngành: Khoa học máy tính

Mã số: 60 48 01 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Người hướng dẫn khoa học: TS. NGUYỄN VIỆT ANH

THÁI NGUYÊN - 2017

LỜI CẢM ƠN

Luận văn này được hoàn thành tại Trường Đại học Công nghệ Thông tin và Truyền thông dưới sự hướng dẫn của TS. Nguyễn Việt Anh. Tác giả xin bày tỏ lòng biết ơn tới các thầy cô giáo thuộc Trường Đại học Công nghệ Thông tin và Truyền thông đã tạo điều kiện và giúp đỡ tác giả trong quá trình học tập và làm luận văn tại Trường, đặc biệt tác giả xin bày tỏ lòng biết ơn tới TS. Nguyễn Việt Anh đã tận tình hướng dẫn và cung cấp nhiều tài liệu cần thiết để tác giả có thể hoàn thành luận văn đúng thời hạn.

Xin chân thành cảm ơn anh chị em học viên cao học và bạn bè đồng nghiệp đã trao đổi, động viên và khích lệ tác giả trong quá trình học tập và làm luận văn tại Trường Đại học Công nghệ Thông tin và Truyền thông – Đại học Thái Nguyên.

Thái Nguyên, tháng 5 năm 2017

Học viên

Mông Quốc Tuấn

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn này do chính tôi thực hiện, dưới sự hướng dẫn khoa học của TS. Nguyễn Việt Anh, các kết quả lý thuyết được trình bày trong luận văn là sự tổng hợp từ các kết quả đã được công bố và có trích dẫn đầy đủ, số liệu và kết quả của chương trình thực nghiệm trong luận văn này được tác giả thực hiện là hoàn toàn trung thực, nếu sai tôi hoàn toàn chịu trách nhiệm.

Thái Nguyên, tháng 5 năm 2017

Học viên

Mông Quốc Tuấn

MỤC LỤC

LỜI CẢM ƠN.....	i
LỜI CAM ĐOAN.....	ii
MỤC LỤC	iii
DANH MỤC HÌNH.....	v
DANH MỤC BẢNG BIỂU VÀ ĐỒ THỊ.....	vi
PHẦN MỞ ĐẦU	1
CHƯƠNG 1: KHÁI QUÁT CHUNG VỀ HỆ THỐNG GỢI Ý THÔNG TIN VÀ THƯƠNG MẠI ĐIỆN TỬ	6
1.1 Hệ thống gợi ý thông tin (Recommender Systems).....	6
1.1.1 Khái niệm hệ thống gợi ý thông tin	6
1.1.2 Một số ứng dụng của hệ thống gợi ý thông tin.....	6
1.2 Tổng quát chung về thương mại điện tử.....	7
1.2.1 Thương mại điện tử là gì ?.....	7
1.2.2 Lợi ích của TMĐT	10
1.2.3 Các loại hình ứng dụng TMĐT.....	11
CHƯƠNG 2: HỌC MÁY VÀ CÁC PHƯƠNG PHÁP PHÂN CỤM DỮ LIỆU	16
2.1 Tổng quan về học máy(Machine learning)	16
2.1.1 Học máy là gì?	16
2.2 Các dạng học máy và các thuật toán liên quan	23
2.2.1 Các dạng học máy	23
2.2.2 Thuật toán K-Means và ứng dụng	26
CHƯƠNG 3: MÔ PHỎNG HỆ THỐNG GỢI Ý THÔNG TIN TRONG THƯƠNG MẠI ĐIỆN TỬ	34
3.1 Hướng tiếp cận và kiến trúc hệ thống	34
3.1.1 Hướng tiếp cận.....	34

3.1.2 Kiến trúc hệ thống.....	35
3.2 Thiết kế và cài đặt chi tiết các thành phần hệ thống	38
3.2.1 Phân nhóm đối tượng bằng phương pháp học bán giám sát.....	38
3.2.2 Huấn luyện mạng nơ ron để xây dựng hàm khoảng cách.....	43
3.2.3 Đánh giá mức độ hiệu quả	49
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....	54
TÀI LIỆU THAM KHẢO	55

DANH MỤC HÌNH

Hình 1.1: Ví dụ về giao diện một hệ thống thương mại điện tử	3
Hình 1.2: Sơ đồ chu trình hệ thống TMĐT B2C	14
Hình 2.1: Sơ đồ tổng quát về học máy.....	16
Hình 2.2: Sơ đồ các lớp trí tuệ nhân tạo.....	18
Hình 2.3: Mô phỏng khái quát về phân cụm dữ liệu	20
Hình 2.4: Mô phỏng dữ liệu sau khi đã được phân cụm.....	22
Hình 2.5: Mô hình thuật toán học có giám sát.....	23
Hình 2.6: Mô phỏng tâm của các cụm được tính toán.....	27
trong thuật toán K-Means.....	27
Hình 2.7: Mô tả thuật toán K-Means	29
Hình 3.1: Gợi ý đối tượng tương tự	34
Hình 3.2: Sơ đồ luồng hệ thống	35
Hình 3.3: Mô hình khoảng cách đến tâm cụm của tập dữ liệu mẫu	44
Hình 3.4: Mô hình mạng nơ ron để huấn luyện hàm khoảng cách.....	45
Hình 3.5: Quá trình phân cụm các đối tượng.....	48
Hình 3.6: Đánh giá mức độ hiệu quả	49
Hình 3.7: Giao diện tổng quan hệ thống khi truy cập.....	50
Hình 3.8: Giao diện tổng quan hệ thống khi ở trạng thái Online Mode	50
Hình 3.9: Giao diện chi tiết sản phẩm khi truy cập	51
Hình 3.10: Những sản phẩm tương tự đã được gợi ý trong hệ thống.....	51
Hình 3.11: Đăng nhập vào Offshore mode trên hệ thống.....	52
Hình 3.12: Tổng quan hệ thống quản lý sản phẩm	52
Hình 3.13: Lựa chọn số cụm để phân cụm cho thuật toán K-Means.....	53
Hình 3.14: Chi tiết quản lý thông tin cho từng sản phẩm	53

DANH MỤC BẢNG BIỂU VÀ ĐỒ THỊ

Bảng 1.1: Các loại hình TMĐT.....	11
Bảng 3.1: Mô tả cấu trúc bảng lưu trữ hành vi người sử dụng.....	37
Bảng 3.2: Ví dụ lưu trữ hành vi người sử dụng	38
Bảng 3.3: Các hàm khoảng cách.....	41

PHẦN MỞ ĐẦU

Trong xã hội ngày nay, con người không những cần nắm bắt nhiều thông tin hơn, mà còn phải nhanh hơn. Internet là một trong những phương tiện quan trọng giúp con người có thể tiếp cận thông tin nhanh nhất. Một trong những tác dụng lớn của Internet trong thập kỷ vừa qua là Thương mại điện tử. Thương mại điện tử ra đời mở ra một kỉ nguyên mới trong thời kì thương mại trên Internet. Một trong những lợi thế lớn nhất của thương mại điện tử chính là khả năng cung cấp cho khách hàng mối liên hệ linh hoạt và mang tính cá nhân hóa.

Trên quan điểm của người sử dụng luôn có xu hướng muốn tìm được sản phẩm và dịch vụ thích hợp nhất đối với nhu cầu và sở thích của bản thân, nhưng mất càng ít thời gian tìm kiếm càng tốt, và với các thao tác càng đơn giản càng tốt. Trên quan điểm của những người thiết kế hệ thống và những nhà cung cấp dịch vụ, vấn đề đặt ra là làm sao xây dựng được các chiến lược kinh doanh và các giải pháp kỹ thuật tích hợp cho việc cung cấp các sản phẩm và dịch vụ đến cho các khách hàng tiềm năng. Các chiến lược kinh doanh tốt sẽ giúp mang lại hiệu quả đầu tư và tăng lợi nhuận. Hai mục tiêu này (của người sử dụng và của nhà cung cấp dịch vụ) có thể đạt được bằng cách cung cấp các hỗ trợ cho người sử dụng trong việc ra quyết định.

Tuy vậy, không phải hoàn toàn các website thương mại điện tử đều có thể đáp ứng được tất cả nhu cầu của người dùng và có thể giúp họ tìm kiếm được chính xác những sản phẩm mà họ cần mua. Lượng sản phẩm quá lớn, khiến người sử dụng không nhận được những thông tin cần thiết khi họ sử dụng công cụ tìm kiếm của sản phẩm. Phải duyệt qua tất cả các kết quả của quá trình tìm kiếm là công việc mệt mỏi đối với người dùng. Trong những năm gần đây, hệ thống gợi ý (recommender system) được biết đến như là một sự phát triển quan trọng trong việc giúp người dùng đối mặt

với sự bùng nổ thông tin. Hệ thống này được ứng dụng trong nhiều lĩnh vực như thương mại điện tử với Amazon, Netflix, Ebay trong lĩnh vực giải trí với MovieLens, Last.fm, Film-Conseil; trong lĩnh vực khác như tin tức trực tuyến Netnews,...

Hệ gợi ý (recommender systems) là một dạng của hệ hỗ trợ ra quyết định, cung cấp giải pháp mang tính cá nhân hóa mà không phải trải qua quá trình tìm kiếm phức tạp. Hệ gợi ý học từ khách hàng và gợi ý các sản phẩm tốt nhất trong số các sản phẩm phù hợp. Sự phát triển nhanh chóng của thương mại điện tử, sự bùng nổ thông tin khiến cho việc tìm kiếm sản phẩm thích hợp để mua của khách hàng khó khăn hơn.

Hiện nay, việc áp dụng hệ gợi ý vào các website thương mại điện tử là điều tất yếu nhằm tiết kiệm thời gian, công sức và chi phí cho khách hàng, giúp họ tìm ra sản phẩm ưng ý nhất để mua. Hệ gợi ý sử dụng các tri thức về sản phẩm, các tri thức của chuyên gia hay tri thức khai phá học được từ hành vi của người tiêu dùng để đưa ra các gợi ý về sản phẩm mà họ thích trong hàng ngàn hàng vạn sản phẩm có trong hệ thống. Các website thương mại điện tử, ví dụ như sách, phim, nhạc, báo,... sử dụng hệ thống gợi ý để cung cấp các thông tin giúp cho người sử dụng quyết định sẽ lựa chọn sản phẩm nào.

Các sản phẩm được gợi ý dựa trên số lượng sản phẩm đó đã được bán, dựa trên các thông tin người sử dụng, dựa trên sự phân tích hành vi mua hàng trước đó của người sử dụng để đưa ra các dự đoán về hành vi mua hàng trong tương lai của chính khách hàng đó. Các dạng gợi ý gồm: Gợi ý sản phẩm tới người tiêu dùng, các sản phẩm mang tính cá nhân hóa, tổng kết các ý kiến cộng đồng, và cung cấp các chia sẻ, các phê bình, đánh giá mang tính cộng đồng liên quan tới yêu cầu, mục đích của người sử dụng đó.