

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

ĐỖ THỊ THANH HÀ

**PHƯƠNG PHÁP LỌC CỘNG TÁC VÀ ỨNG DỤNG
TRONG HỆ THỐNG GỢI Ý BÁN HÀNG TRỰC TUYẾN**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN - 2020

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi dưới sự hướng dẫn trực tiếp của PGS.TS. Đoàn Văn Ban.

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.

Học viên

Đỗ Thị Thanh Hà

LỜI CẢM ƠN

Trước hết, tôi xin được tỏ lòng biết ơn và gửi lời cảm ơn chân thành đến thầy PGS.TS Đoàn Văn Ban, người trực tiếp hướng dẫn luận văn, đã tận tình chỉ bảo và hướng dẫn tôi tìm ra hướng nghiên cứu, tiếp cận thực tế, tìm kiếm tài liệu, xử lý và phân tích số liệu,... nhờ đó tôi mới có thể hoàn thành luận văn cao học của mình.

Tôi xin chân thành cảm ơn quý thầy cô Khoa Công nghệ thông tin - trường Đại học Công nghệ Thông tin và Truyền thông, Đại học Thái Nguyên đã truyền đạt cho tôi những kiến thức bổ ích trong suốt hai năm học vừa qua.

Xin chân thành cảm ơn các anh chị em lớp cao học Khoa học máy tính khoá 17 và các bạn đồng nghiệp luôn đồng viên, hỗ trợ tôi trong quá trình học tập và nghiên cứu và thực hiện đề tài.

Cuối cùng, tôi xin gửi đến cha mẹ và những người thân trong gia đình đã hỗ trợ, tạo điều kiện thuận lợi cho tôi trong suốt thời gian qua và đặc biệt trong thời gian tôi theo học khóa thạc sỹ tại trường Đại học Công nghệ Thông tin và Truyền thông, Đại học Thái Nguyên.

Xin chân thành cảm ơn!

Người thực hiện

Đỗ Thị Thanh Hà

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
MỤC LỤC.....	iii
DANH MỤC CÁC BẢNG	iv
DANH MỤC HÌNH VẼ.....	v
MỘT SỐ THUẬT NGỮ	viii
MỞ ĐẦU	1
CHƯƠNG 1. HỆ THỐNG TƯ VẤN VÀ CÁC PHƯƠNG PHÁP LỌC THÔNG TIN.....	5
1.1. Hệ tư vấn.....	5
1.1.1. Giới thiệu hệ thống tư vấn	5
1.1.2. Bài toán tư vấn người dùng	7
1.1.3. Qui trình xây dựng hệ tư vấn	9
1.2. Các phương pháp lọc tin	10
1.2.1. Phân loại hệ thống tư vấn	10
1.2.2. Phương pháp tư vấn dựa trên nội dung	11
1.2.3. Phương pháp tư vấn dựa trên lọc cộng tác	17
1.2.4. Tư vấn dựa trên cách tiếp cận kết hợp.....	23
1.3. Kiến trúc tổng quát của hệ thống lọc thông tin	25
1.4. Ứng dụng của phương pháp lọc tin trong hệ thống tư vấn	26
1.5. Kết luận chương 1	28
CHƯƠNG 2. PHƯƠNG PHÁP THUẬT LỌC CỘNG TÁC DỰA TRÊN SẢN PHẨM ..	29
2.1. Các phương pháp lọc cộng tác.....	29
2.1.1. Giới thiệu về các phương pháp lọc cộng tác	29
2.1.2. Ưu và nhược điểm của phương pháp lọc cộng tác	30
2.2. Lọc cộng tác dựa trên sản phẩm	32
2.3. Các thuật toán tính độ tương tự	34

2.4. Tính toán dự đoán và tư vấn	37
2.4.1. Dự đoán dựa trên trung bình đánh giá sản phẩm lân cận.....	37
2.4.2. Dự đoán dựa trên tổng trọng số (Weighted Sum)	38
2.4.3. Dự đoán dựa trên tổng trọng số với đánh giá trung bình của người dùng	38
2.4.4. Dự đoán dựa trên tổng trọng số với trung bình đánh giá lên sản phẩm	39
2.5. Đánh giá các yếu tố ảnh hưởng đến độ chính xác kết quả tư vấn.....	40
2.5.1. Đánh giá độ tin cậy của thuật toán	40
2.5.2. Các yếu tố ảnh hưởng đến độ chính xác tư vấn	42
2.6. Kết luận chương 2.....	44
CHƯƠNG 3. ỨNG DỤNG PHƯƠNG PHÁP LỌC CỘNG TÁC DỰA TRÊN SẢN PHẨM TRONG HỆ THỐNG GỢI Ý BÁN HÀNG TRỰC TUYẾN	45
3.1. Xây dựng hệ thống gợi ý cho website bán hàng trực tuyến.....	45
3.1.1. Đặc tả hệ thống	45
3.1.2. Môi trường phát triển.....	46
3.1.3. Thiết kế bảng dữ liệu	46
3.2. Ví dụ về lọc cộng tác dựa trên sản phẩm.....	49
3.3. Thuật toán xử lý chính trong hệ thống	53
3.3.1. Thuật toán khách hàng đánh giá sản phẩm.....	53
3.3.2. Thuật toán khách hàng chấm điểm sản phẩm	54
3.3.3. Thuật toán gợi ý sản phẩm cho khách hàng	55
3.4. Các giao diện chính của hệ thống.....	56
3.5. Kết luận chương 3.....	60
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	61
TÀI LIỆU THAM KHẢO.....	62

DANH MỤC BẢNG BIỂU

Bảng 1: Bảng đánh giá người dùng với các sản phẩm.....	33
Bảng 2: Dữ liệu đánh giá các sản phẩm.....	49
Bảng 3: Ma trận utility matrix.....	50
Bảng 4: Giá trị trung bình đánh giá của các item.....	50
Bảng 5: Ma trận chuẩn hóa \hat{Y}	51
Bảng 6: Độ tương tự giữa các item S.....	51
Bảng 7: Ma trận sau khi được dự đoán \hat{y}	52
Bảng 8: Ma trận dự đoán thang điểm 5.....	53

DANH MỤC HÌNH VẼ

Hình 1.1: Một số lĩnh vực ứng dụng của hệ thống gợi ý	6
Hình 1.2. Ví dụ ma trận đánh giá tổng quát.....	8
Hình 1.3: Phân loại hướng tiếp cận xây dựng hệ thống gợi ý.....	11
Hình 1.4. Minh họa phương pháp tư vấn dựa trên lọc nội dung.....	12
Hình 1.5: Minh họa phương pháp tư vấn dựa trên lọc cộng tác	18
Hình 1.6: Quy trình của hệ thống tư vấn dựa trên lọc cộng tác	19
Hình 1.7: Kiến trúc tổng quát của hệ thống lọc thông tin.....	25
Hình 1.8: Hệ thống gợi ý sản phẩm của Amazon	28
Hình 2.1: Tách các sản phẩm cùng được đánh giá và tính toán độ tương tự.....	34
Hình 2.2: Giải thuật lọc cộng tác dựa trên sản phẩm, tính dự đoán.....	37
Hình 2.3: Mô hình hệ thống lọc cộng tác dựa trên sản phẩm	40
Hình 3.1: Giải thuật khách hàng đánh giá sản phẩm	54
Hình 3.2: Khách hàng chấm điểm sản phẩm	55
Hình 3.3: Giải thuật xử lý gợi ý sản phẩm.....	56
Hình 3.4: Giao diện trang chủ.....	58
Hình 3.5: Giao diện trang danh sách sản phẩm.....	58
Hình 3.6: Giao diện gợi ý sản phẩm cho sản phẩm chưa có đánh giá	59
Hình 3.7: Giao diện gợi ý sản phẩm cho sản phẩm đã có đánh giá	59
Hình 3.8: Giao diện gợi ý sản phẩm cho một thành viên cụ thể.....	59

MỘT SỐ THUẬT NGỮ

STT	Tiếng Anh	Tiếng Việt	Giải thích
1	User	Người dùng, người sử dụng	Chỉ những người dùng hệ thống để tìm kiếm lựa chọn sản phẩm
2	Item	Sản phẩm, mục	Chỉ những sản phẩm trên hệ thống như: sản phẩm, phim, ảnh, bản nhạc, trang web, đoạn văn bản, ...
3	Rating	Đánh giá	Chỉ mức độ thích của một người dùng với sản phẩm. Rating có thể có nhiều biểu hiện: như đánh giá thích hoặc không thích, hay đánh giá theo mức độ từ 1-5 đại diện từ không thích đến rất thích.
4	CF	Lọc cộng tác	CF có tên là <i>Neighborhood-based Collaborative Filtering (NBCF)</i> . Khi chỉ nói CF, chúng ta sẽ ngầm hiểu rằng phương pháp được sử dụng là <i>Neighborhood-based</i> .

MỞ ĐẦU

- **Lí do chọn đề tài**

Sự phát triển mạnh mẽ của thương mại điện tử (E-Commerce) là một trong những tác nhân chính đem lại nhiều lợi ích to lớn cho nền kinh tế toàn cầu. Nhờ có thương mại điện tử, rất nhiều loại hình kinh doanh mới được hình thành, trong đó có mua bán hàng trực tuyến. 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.

Hiện nay, với phương châm phục vụ khách hàng một cách tốt nhất, các hệ thống bán hàng trực tuyến luôn tạo nhiều điều kiện thuận lợi để người mua hàng có thể tiếp cận nhiều mặt hàng cùng lúc. Tuy nhiên, việc trình bày và trang trí quá nhiều các mặt hàng trên một website sẽ gây ra không ít khó khăn cho người mua. Họ khó có thể chọn ra cho mình một sản phẩm ưng ý nhất.

Trong kinh doanh, để khách hàng có thể đến và mua được một sản phẩm ưng ý thì một lời tư vấn, một sự trợ giúp là rất quan trọng. Với phương thức bán hàng truyền thống những lời tư vấn như thế từ một người bán hàng sẽ tạo ra một lợi thế rất lớn cho cửa hàng. Do đó, để phương thức bán hàng qua mạng phát triển thực sự hiệu quả thì bên cạnh các lợi thế vốn có của mình, việc có thêm một “người trợ giúp” là hết sức cần thiết.

Vì vậy, cần thiết phải sử dụng một hệ thống gợi ý (Recommender Systems - RS) đóng vai trò như một người hỗ trợ khách hàng đưa ra các quyết định mua hàng đúng đắn. Bằng cách xác định mục đích và nhu cầu của khách hàng, hệ thống có thể đưa ra tập các gợi ý giúp cho người mua dễ dàng chọn lựa sản phẩm yêu thích hơn. Qua đó hiệu suất của việc mua bán hàng trực tuyến được tăng cao một cách đáng kể [6].

Hệ thống gợi ý được phát triển theo nhiều cách tiếp cận khác nhau. Một trong những hướng tiếp cận được tập trung nghiên cứu và được áp dụng khá

thành công trong nhiều hệ gợi ý đó là phương pháp lọc cộng tác (Collaborative filtering). Thực chất, lọc cộng tác là một hình thức tư vấn tự động bằng cách dựa trên sự tương tự giữa những người dùng hoặc giữa những sản phẩm trong hệ thống và đưa ra dự đoán sự quan tâm của người dùng tới những sản phẩm, hoặc đưa ra gợi ý một sản phẩm mới cho người dùng.

Việc xây dựng hệ thống gợi ý cũng là vấn đề được nhiều nhà khoa học trong nước quan tâm [4, 7]. Trong hệ thống gợi ý, lọc cộng tác là một kỹ thuật được dùng để đánh giá độ quan tâm của người dùng trên các sản phẩm mới (từ mức độ tìm hiểu thuật toán [2] cho đến việc cải tiến [5] cũng như áp dụng trong việc xây dựng website bán hàng trực tuyến [6]). Kỹ thuật này được áp dụng thành công trong nhiều ứng dụng. Trong các hệ thống lọc cộng tác, sở thích của người dùng trên các sản phẩm mới được dự đoán dựa trên dữ liệu về sở thích của người dùng – sản phẩm (hoặc đánh giá của người dùng trên sản phẩm) trong quá khứ.

Các nghiên cứu trong và ngoài nước đã góp phần rất lớn cho việc xây dựng những website bán hàng trực tuyến như website mua sắm trực tuyến Amazon (www.amazon.com) cung cấp cho khách hàng những sản phẩm mà họ có thể quan tâm, công video clip YouTube (www.youtube.com), gợi ý phim của MovieLens (www.movielens.org) [10].

Tuy nhiên, để phát triển mạnh hơn nữa thuật toán lọc cộng tác trong hệ thống gợi ý nói chung hay hệ thống gợi ý sản phẩm nói riêng còn rất nhiều câu hỏi cần lời giải đáp. Cụ thể như: Khi xây dựng hệ thống gợi ý bán hàng trực tuyến, tiếp cận theo hướng dựa trên bộ nhớ (Memory-based) hay dựa trên mô hình (Model-based) hiệu quả hơn? Trong hướng tiếp cận theo Memory-based thì nên tư vấn dựa trên người dùng (User based) hay sản phẩm (Items based)? Nếu tiếp cận theo hướng Model-based thì nên sử dụng mô hình mạng Nơ ron hay Mạng Bayessian? ... Trong mỗi cách tiếp cận, sử dụng độ đo nào để nắm bắt chính xác việc khách