

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

LÊ THANH HUYỀN

PHƯƠNG PHÁP LỌC CỘNG TÁC VÀ ỨNG DỤNG  
TRONG HỆ THỐNG TIN TƯ VẤN

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**

**LÊ THANH HUYỀN**

**PHƯƠNG PHÁP LỌC CỘNG TÁC VÀ ỨNG DỤNG  
TRONG HỆ THỐNG TIN TƯ VẤN**

**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: PGS.TS ĐOÀN QUANG BAN**

**THÁI NGUYÊN - 2015**

## LỜI CAM ĐOAN

Tôi xin cam đoan luận văn này của tự bản thân tôi tìm hiểu, nghiên cứu dưới sự hướng dẫn của PGS. TS Đoàn Văn Ban. Các chương trình do chính bản thân tôi lập trình, các kết quả là hoàn toàn trung thực. Các tài liệu tham khảo được trích dẫn và chú thích đầy đủ.

**Tác giả**

**Lê Thanh Huyền**

## LỜI CẢM ƠN

Tôi xin bày tỏ lời cảm ơn chân thành tới tập thể các Thầy cô Viện Công nghệ thông tin - Viện Hàn Lâm Khoa học và công nghệ Việt Nam, các Thầy cô giáo Đại học Công nghệ thông tin và truyền thông - Đại học Thái Nguyên đã dạy dỗ chúng tôi trong suốt quá trình học tập chương trình cao học tại trường.

Đặc biệt tôi xin bày tỏ lòng biết ơn sâu sắc tới Thầy giáo PGS.TS Đoàn Văn Ban đã quan tâm, định hướng và đưa ra những góp ý, gợi ý, chỉnh sửa quý báu cho tôi trong quá trình làm luận văn tốt nghiệp. Cũng như bạn bè, đồng nghiệp, gia đình và người thân đã quan tâm giúp đỡ, chia sẻ với tôi trong suốt quá trình làm luận văn tốt nghiệp.

Dù đã cố gắng nhưng chắc chắn sẽ không tránh khỏi những thiếu sót vì vậy rất mong nhận được sự đóng góp ý kiến của các Thầy, Cô và các bạn để luận văn được hoàn thiện hơn.

Tôi xin chân thành cảm ơn!

*Thái Nguyên, tháng 9 năm 2015*

**Lê Thanh Huyền**

## MỤC LỤC

	Trang
LỜI CAM ĐOAN.....	i
LỜI CẢM ƠN.....	ii
MỤC LỤC.....	iii
DANH MỤC CÁC CHỮ VIẾT TẮT.....	vi
DANH MỤC CÁC BẢNG.....	vii
DANH MỤC CÁC HÌNH.....	viii
<b>MỞ ĐẦU</b> .....	<b>1</b>
<b>Chương 1: PHƯƠNG PHÁP LỌC TIN</b> .....	<b>4</b>
1.1. Các phương pháp lọc thông tin.....	4
1.1.1. Phương pháp lọc tin theo nội dung.....	4
1.1.1.1 Bài toán lọc theo nội dung.....	4
1.1.1.2 Các phương pháp lọc theo nội dung.....	5
1.1.2. Phương pháp lọc tin theo cộng tác.....	6
1.1.2.1 Bài toán lọc cộng tác.....	6
1.1.2.2 Các phương pháp lọc cộng tác.....	7
1.1.3. Phương pháp lọc tin kết hợp.....	11
1.1.3.1 Bài toán lọc kết hợp.....	11
1.1.3.2 Các phương pháp lọc kết hợp.....	12
1.1.4. Ứng dụng của các phương pháp lọc tin.....	13
1.2. Hệ thống thông tin tư vấn.....	13
1.2.1. Kiến trúc tổng quan của hệ thống lọc thông tin.....	13
1.2.2. Lọc thông tin và các hệ tư vấn.....	15
<b>Chương 2: MỘT SỐ PHƯƠNG PHÁP LỌC CỘNG TÁC</b> .....	<b>17</b>
2.1. Lọc cộng tác dựa trên sản phẩm.....	17
2.1.1. Thuật toán tính độ tương tự.....	19

2.1.1.1 Độ tương tự Cosine.....	19
2.1.1.2 Độ tương tự tương quan.....	20
2.1.1.3 Độ tương tự Cosine điều chỉnh.....	21
2.1.2. Tính toán dự đoán và tư vấn .....	23
2.1.2.1 Công thức dự đoán dựa trên trung bình đánh giá sản phẩm lân cận ...	23
2.1.2.2 Công thức dự đoán dựa trên tổng trọng số. ....	24
2.1.2.3 Công thức dự đoán dựa trên tổng trọng số với đánh giá trung bình của người dùng .....	25
2.1.2.4 Công thức dự đoán dựa trên tổng trọng số với trung bình đánh giá lên sản phẩm .....	26
2.1.3. Thuật toán lọc cộng tác dựa trên sản phẩm .....	27
2.1.3.1 Độ tương tự Cosine.....	28
2.1.3.2 Độ tương tự Cosine điều chỉnh.....	28
2.1.3.3 Dự đoán dựa trên trung bình đánh giá sản phẩm lân cận .....	29
2.1.3.4 Dự đoán dựa trên tổng trọng số .....	29
2.1.3.5 Dự đoán dựa trên tổng trọng số với trung bình đánh giá lên người dùng.	30
2.1.4. Đánh giá các yếu tố ảnh hưởng đến độ chính xác kết quả tư vấn .....	31
2.1.4.1 Đánh giá chất lượng của hệ thống tư vấn .....	31
2.1.4.2 Các yếu tố ảnh hưởng đến độ chính xác tư vấn.....	31
2.2. Lọc cộng tác dựa trên mô hình đồ thị.....	32
2.2.1. Phương pháp biểu diễn đồ thị.....	32
2.2.2. Phương pháp dự đoán trên đồ thị người dùng - sản phẩm .....	34
2.2.2.1. Tách đồ thị Người dùng-Sản phẩm thành các đồ thị con .....	35
2.2.2.2. Phương pháp dự đoán trên đồ thị có trọng số dương $G^+$ .....	37
2.2.2.3. Phương pháp dự đoán trên đồ thị các cạnh có trọng số âm $G^-$ .....	39
2.2.2.4. Phương pháp dự đoán theo tất cả đánh giá.....	41
2.3. Lọc cộng tác dựa vào lọc đồng huấn luyện .....	43

2.3.1. Mô tả thuật toán đồng huấn luyện .....	44
2.3.2. Thuật toán lọc cộng tác bằng phương pháp đồng huấn luyện theo người dùng .....	44
2.3.3 Lọc cộng tác bằng phương pháp đồng huấn luyện theo sản phẩm.....	46
<b>Chương 3: XÂY DỰNG HỆ THỐNG TIN TƯ VẤN SẢN PHẨM SỮA DÀNH CHO NGƯỜI TIÊU DÙNG .....</b>	<b>51</b>
3.1. Phát biểu bài toán.....	51
3.2. Phân tích thiết kế hệ thống tư vấn sản phẩm sữa.....	51
3.2.1. Phân tích các yêu cầu.....	51
3.2.2. Thiết kế hệ thống tư vấn sản phẩm sữa. ....	53
3.3. Xây dựng chương trình ứng dụng.....	54
3.4. Kết luận. ....	56
<b>KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....</b>	<b>57</b>
<b>TÀI LIỆU THAM KHẢO .....</b>	<b>57</b>

## DANH MỤC CÁC CHỮ VIẾT TẮT

KÝ HIỆU	DIỄN GIẢI
IF	Information Filtering (Lọc thông tin)
IR	Information Retrieval (Truy vấn thông tin)
RS	Recommender System (Hệ thống tư vấn)
u	User (Người dùng)
i	Item (Sản phẩm)



## DANH MỤC CÁC BẢNG

	Trang
Bảng 1.1. Ví dụ về ma trận đánh giá của lọc cộng tác.....	7
Bảng 2.1. Bảng đánh giá người dùng với các sản phẩm.....	18
Bảng 2.2. Bảng tính độ tương tự theo công thức Cosine.....	20
Bảng 2.3. Bảng tính độ tương tự theo công thức tương quan.....	21
Bảng 2.4. Bảng tính độ tương tự theo công thức Cosine điều chỉnh.....	22
Bảng 2.5. Bảng dự đoán và tư vấn theo phương pháp tính trung bình dự đoán.	24
Bảng 2.6. Bảng dự đoán và tư vấn theo phương pháp Weigth Sum.....	25
Bảng 2.7. Bảng dự đoán và tư vấn theo phương pháp tổng trọng số với đánh giá trung bình của người dùng và sử dụng độ tương tự Ajusted Cosine. ....	26
Bảng 2.8. Bảng dự đoán và tư vấn theo phương pháp tổng trọng số với đánh giá trung bình sản phẩm và sử dụng độ tương tự Ajusted Cosine. ....	27
Bảng 2.9. Ma trận đánh giá R .....	33
Bảng 2.10. Ma trận X biểu diễn đánh đồ thị Người dùng- Sản phẩm .....	33
Bảng 2.12. Ma trận $X^+$ biểu diễn các đánh giá thích hợp .....	36
Bảng 2.12. Ma trận $X^-$ biểu diễn các đánh giá không thích hợp.....	36
Bảng 2.13: Người dùng và sản phẩm.....	48
Bảng 2.14: Bảng giá trị đánh giá theo người dùng. ....	48
Bảng 2.15: Bảng giá trị đánh giá theo sản phẩm .....	49

## DANH MỤC CÁC HÌNH

	Trang
Hình 1.1. Kiến trúc tổng quát của hệ thống lọc thông tin.....	14
Hình 2.1. Mô hình hệ thống lọc cộng tác dựa trên sản phẩm.....	31
Hình 2.2. Đồ thị người dùng - sản phẩm.....	34
Hình 2.3. Đồ thị G biểu diễn cách đánh giá thích hợp.....	36
Hình 2.4. Đồ thị G biểu diễn cách đánh giá không thích hợp.....	37
Hình 3.3: Giao diện chương trình dự đoán sản phẩm sữa. ....	55
Hình 3.4: Người dùng đăng nhập vào hệ thống.....	55
Hình 3.5: Hệ thống lọc cộng tác dựa vào bộ nhớ .....	55
Hình 3.6: Hệ thống lọc cộng tác dựa vào đồ thị .....	56