

**BỘ GIÁO DỤC
VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM**

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



Cù Nguyên Giáp

**NGHIÊN CỨU PHÁT TRIỂN HỆ TƯ VẤN NHÓM
THEO TIẾP CẬN MỜ TRỰC CẢM
VÀ TÍCH PHÂN CHOQUET**

LUẬN ÁN TIẾN SĨ MÁY TÍNH

Hà Nội – 2025

BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Cù Nguyên Giáp

NGHIÊN CỨU PHÁT TRIỂN HỆ TƯ VẤN NHÓM
THEO TIẾP CẬN MỜ TRỰC CẢM
VÀ TÍCH PHÂN CHOQUET

LUẬN ÁN TIẾN SĨ MÁY TÍNH

Ngành: Khoa học máy tính

Mã số: 9 48 01 01

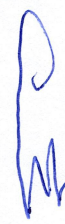
Xác nhận của Học viện
Khoa học và Công nghệ

KT. GIÁM ĐỐC
PHÓ GIÁM ĐỐC



Nguyễn Thị Trung

Người hướng dẫn 1
(Ký, ghi rõ họ tên)


Lê Hoàng Sơn

Người hướng dẫn 2
(Ký, ghi rõ họ tên)


Nguyễn Như Sơn

Hà Nội – 2025

MỤC LỤC

Danh mục các thuật ngữ.....	IV
Bảng các ký hiệu, từ viết tắt.....	V
Danh sách hình vẽ.....	IX
CAM KẾT	X
LỜI CẢM ƠN.....	XI
MỞ ĐẦU	1
1. Tính cấp thiết của nghiên cứu	1
2. Mục tiêu nghiên cứu của luận án.....	4
3. Đối tượng và Phương pháp nghiên cứu của luận án.....	5
4. Bố cục của luận án	6
Chương 1. TỔNG QUAN VỀ HỆ TƯ VẤN NHÓM.....	8
1.1. Giới thiệu về Hệ tư vấn nhóm.....	8
1.1.1. Hệ tư vấn đơn người dùng	8
1.1.2. Hệ tư vấn nhóm.....	10
1.1.3. Các cách tiếp cận xây dựng HTVN	12
1.1.4. HTVN sử dụng cách tiếp cận hợp khuyến nghị.....	13
1.1.5. Các phương pháp đánh giá Hệ tư vấn nhóm.....	17
1.2. Tổng quan nghiên cứu về HTVN.....	19
1.2.1. Tổng quan nghiên cứu về độ công bằng trong HTVN.....	20
1.2.2. Tổng quan hệ tư vấn nhóm động	25
1.2.3. Hệ tư vấn nhóm theo tiếp cận tính toán mờ.....	28
1.3. Giới thiệu về lý thuyết tập mờ trực cảm	30
1.3.1. Khái quát về Tập mờ trực cảm.....	30
1.3.2. Khoảng cách và độ tương tự cho tập mờ trực cảm	32
1.3.3. Phép toán trung bình với các số mờ trực cảm.....	34
1.4. Giới thiệu về tích phân Choquet.....	35
1.4.1. Định nghĩa hàm dung lượng	35
1.4.2. Tích phân Choquet.....	36
1.5. Kết luận	38

Chương 2. TĂNG CƯỜNG ĐỘ CÔNG BẰNG KHUYẾN NGHỊ CỦA HỆ TƯ VẤN NHÓM VỚI ĐỘ ĐO MỜ.....	40
2.1. Mở đầu	40
2.2. Đề xuất hệ tư vấn nhóm nâng cao tính công bằng với độ đo mờ.....	43
2.2.1. Phép toán tổng hợp có trọng số và phép toán tổng hợp dựa trên tích phân Choquet	43
2.2.2. Vấn đề tăng cường sự công bằng của hệ tư vấn nhóm dựa trên độ đo mờ.....	46
2.3. Mô hình Hệ tư vấn nhóm với pha đồng thuận sử dụng tích phân Choquet	48
2.3.1. Đề xuất mô hình Hệ tư vấn nhóm với pha đồng thuận dựa trên tích phân Choquet	48
2.3.2. Thuật toán đề xuất	57
2.3.3. Độ phức tạp thuật toán pha đồng thuận.....	59
2.4. Kết quả thực nghiệm và bàn luận	60
2.4.1. Mục tiêu thực nghiệm	60
2.4.2. Dữ liệu thực nghiệm	61
2.4.3. Phương thức đánh giá	62
2.4.4. Kết quả và bàn luận	63
2.5. Kết luận chương 2	73
Chương 3. HỆ TƯ VẤN NHÓM ĐỘNG THEO TIẾP CẬN TÍNH TOÁN MỜ TRỰC CẢM ĐẢM BẢO TÍNH CÔNG BẰNG	74
3.1. Mở đầu	74
3.2. Đề xuất mô hình Hệ tư vấn nhóm động trên tập mờ trực cảm	76
3.2.1. Đề xuất mô hình Hệ tư vấn nhóm động trên tập mờ trực cảm	76
3.2.2. Phép toán đồng thuận cho HTVN động trên tập mờ trực cảm	83
3.2.3. Các chiến lược tổng hợp đánh giá khác cho pha đồng thuận.....	85
3.3. Giải mã của giải thuật	87
3.3.1. Giải mã của thuật toán Hệ tư vấn nhóm động trên tập mờ trực cảm	87
3.3.2. Độ phức tạp của thuật toán	88
3.3.3. Phương pháp học tham số của thuật toán	90
3.4. Thực nghiệm thuật toán.....	92
3.4.1. Mục tiêu thử nghiệm	92
3.4.2. Dữ liệu thực nghiệm	92

3.4.3. Phương thức đánh giá	92
3.4.4. Kết quả và bàn luận	97
3.5. Kết luận chương 3	106
KẾT LUẬN.....	108
DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ	110
TÀI LIỆU THAM KHẢO	111

Danh mục các thuật ngữ

Thuật ngữ tiếng Việt	Thuật ngữ tiếng Anh
<i>Người dùng giả định</i>	<i>Pseudo user</i>
<i>Pha đồng thuận</i>	<i>Consensus phase</i>
<i>Chiến lược cộng</i>	<i>Additive utilitarian strategy</i>
<i>Chiến lược tích</i>	<i>Multiple strategy</i>
<i>Chiến lược ít thiệt thòi nhất</i>	<i>Least misery strategy</i>
<i>Chiến lược ưu thế</i>	<i>Most pleasure strategy</i>
<i>Chiến lược biểu quyết phê duyệt</i>	<i>Approval voting strategy</i>
<i>Chiến lược công bằng</i>	<i>Fairness strategy</i>
<i>Chiến lược Copeland Rule</i>	<i>Copeland rule strategy</i>
<i>Chiến lược chỉ số Borda</i>	<i>Borda count strategy</i>
<i>Chiến lược dựa trên tích phân Choquet</i>	<i>Choquet integral based strategy</i>
<i>Độ đo phân lớp</i>	<i>Classification metric</i>
<i>Độ đo sai số</i>	<i>Error metric</i>
<i>Độ đo công bằng</i>	<i>Fairness metric</i>
<i>Hàm dung lượng</i>	<i>Capacity function</i>
<i>Hệ tư vấn</i>	<i>Recommender systems</i>
<i>Hệ tư vấn nhóm</i>	<i>Group recommender systems</i>
<i>Lọc cộng tác</i>	<i>Collaborative filtering</i>
<i>Lọc dựa trên nội dung</i>	<i>Content-based filtering</i>
<i>Suy hao theo thời gian</i>	<i>Time decay</i>
<i>Sự thỏa mãn của người dùng</i>	<i>User's satisfaction</i>
<i>Sự yêu thích của người dùng</i>	<i>User's preference</i>
<i>Tính đồng thuận và sự công bằng</i>	<i>Consensus and Fairness</i>
<i>Độ đo sự yêu thích mức m</i>	<i>m-proportional</i>
<i>Độ đo không ghen tị mức m</i>	<i>m-envy-freeness</i>
<i>Cách tiếp cận dựa trên cơ sở gom nhóm</i>	<i>Bining-based approach</i>
<i>Phương pháp cập nhật trực tuyến</i>	<i>Online updating approach</i>
<i>Cách tiếp cận dựa trên tính động</i>	<i>Dynamic-based approach</i>
<i>Tập hợp sắp thứ tự một phần</i>	<i>Partially ordered set (poset)</i>

Bảng các ký hiệu, từ viết tắt

Ký hiệu, từ viết tắt	Diễn giải
r_{ui}	Đánh giá của người dùng u với sản phẩm i
$pref(g, i)$	Đánh giá của nhóm người dùng g với sản phẩm i
$fairness_{\text{"label"}}(g, i)$	Độ đo công bằng có tên là "label" cho khuyến nghị nhóm
A	Tập mờ trực cảm A trên không gian nền X
$\mu_A : X \rightarrow [0, 1]$	Hàm thuộc
$\nu_A : X \rightarrow [0, 1]$	Hàm không thuộc
$\mu_A(x)$	Giá trị độ thuộc
$\nu_A(x)$	Giá trị độ không thuộc
$\pi_A(x)$	Giá trị độ do dự
$sim_{\cosine}(A, B)$	Độ tương tự <i>cosine</i> giữa hai tập mờ trực cảm
$wsim_{\cosine}(A, B)$	Độ tương tự <i>cosine</i> có trọng số giữa hai tập mờ trực cảm
$IFAW(A, W)$	Giá trị trung bình số học có trọng số của một tập giá trị mờ trực cảm.
$IFB^{p,q}(A)$	Giá trị trung bình Bonferroni của một tập giá trị mờ trực cảm.
2^M	Không gian thể hiện tất cả các tổ hợp của M điều kiện
$\xi(A)$	Hàm dung lượng
$CQ_{\xi}(r)$	Tích phân Choquet cho vector r với hàm dung lượng ξ
$\omega(u_i)$	Mức độ tích cực của người dùng u_i trong hệ thống, được tính theo số lượng sản phẩm mà người dùng đã đánh giá
$\omega'(u_i)$	Mức độ “hứng thú” của người dùng u_i trong hệ thống
$pref^{CQ}(g, i)$	Đánh giá của nhóm người dùng g với sản phẩm i theo cơ chế đồng thuận dựa trên tích phân Choquet
t_{ui}	Thời điểm người dùng u đưa ra đánh giá với sản phẩm i
a_{uj}	Đánh giá của người dùng u với sản phẩm i thể hiện bằng giá trị mờ trực cảm
$\bar{a}_u$	Đánh giá trung bình của người dùng u với các sản phẩm đã được người dùng này đánh giá

$\tilde{a}_{ui}$	Độ khác biệt giữa đánh giá của người dùng u cho sản phẩm i với mức độ đánh giá trung bình của người dùng này.
a^g_{ui}	Đánh giá của người dùng u với sản phẩm i khi u tham gia trong nhóm người dùng g thể hiện bằng giá trị mờ trực cảm
$\hat{a}_{ui}$	Giá trị đánh giá thật của người dùng u với sản phẩm i thể hiện bằng giá trị mờ trực cảm
<i>AUS</i>	Chiến lược cộng
<i>AVS</i>	Chiến lược biểu quyết phê duyệt
<i>BCS</i>	Chiến lược chỉ số Borda
<i>CC</i>	Catalog coverage
<i>CIS</i>	Chiến lược đồng thuận sử dụng tích phân Choquet
<i>CIS_CF1</i>	Chiến lược đồng thuận sử dụng tích phân Choquet sử dụng hàm dung lượng thứ nhất được đề xuất.
<i>CIS_CF2</i>	Chiến lược đồng thuận sử dụng tích phân Choquet sử dụng hàm dung lượng thứ hai được đề xuất.
<i>CRS</i>	Chiến lược Copeland Rule
<i>FS</i>	Chiến lược công bằng
<i>GC</i>	Group coverage
<i>GINI</i>	Chỉ số Gini
<i>HTTT</i>	Hệ thống thông tin
<i>HTV</i>	Hệ tư vấn
<i>HTVN</i>	Hệ tư vấn nhóm
<i>HTVND</i>	Hệ tư vấn nhóm động
<i>IF_AUS</i>	Chiến lược cộng trên tập mờ trực cảm
<i>IF_MS</i>	Chiến lược tích trên tập mờ trực cảm
<i>IF_LMS</i>	Chiến lược ít thiệt thòi nhất trên tập mờ trực cảm
<i>IF_MPS</i>	Chiến lược ưu thế trên tập mờ trực cảm
<i>IF_AVS</i>	Chiến lược biểu quyết phê duyệt trên tập mờ trực cảm
<i>IF_FS</i>	Chiến lược công bằng trên tập mờ trực cảm
<i>IF_CRS</i>	Chiến lược Copeland Rule trên tập mờ trực cảm
<i>IF_CIS</i>	Chiến lược đồng thuận sử dụng tích phân Choquet trên tập mờ trực cảm
<i>LMS</i>	Chiến lược ít thiệt thòi nhất

<i>MPS</i>	Chiến lược ưu thế
<i>MS</i>	Chiến lược tích

Danh sách bảng

<i>Bảng 2.1</i> Phép toán tổng hợp đánh giá bằng chiến lược cộng.....	44
<i>Bảng 2.2</i> Phép toán tổng hợp với tích phân Choquet	45
<i>Bảng 2.3</i> Sai số trung bình của mô hình với kích thước trong khuyến nghị	65
<i>Bảng 2.4</i> Sai số trung bình của mô hình với kích thước nhóm.....	65
<i>Bảng 2.5</i> độ đo công bằng thứ nhất theo kích thước khuyến nghị	66
<i>Bảng 2.6</i> Đánh giá trung bình của nhóm theo kích thước khuyến nghị	67
<i>Bảng 2.7</i> Thời gian tính của các mô hình HTVN	72
<i>Bảng 3.1</i> Độ đo tỷ lệ hài lòng của các thuật toán HTVN mờ động.....	101

Danh sách hình vẽ

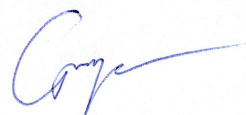
<i>Hình 1.1</i> Hệ tư vấn nhóm theo tiếp cận hợp đánh giá	12
<i>Hình 1.2</i> Pha đồng thuận của Hệ tư vấn nhóm	13
<i>Hình 1.3</i> Một ví dụ về HTVN động.....	27
<i>Hình 2.1</i> Quy trình tổng quát hệ tư vấn nhóm theo tiếp cận hợp khuyến nghị.....	49
<i>Hình 2.2</i> Đánh giá trung bình của nhóm với kích thước nhóm bằng 3	68
<i>Hình 2.3</i> Đánh giá trung bình của nhóm với kích thước nhóm bằng 10.....	69
<i>Hình 2.4</i> Đánh giá trung bình của nhóm với kích thước nhóm bằng 20	69
<i>Hình 2.5</i> Độ công bằng của khuyến nghị cho nhóm có kích thước bằng 3	70
<i>Hình 2.6</i> Độ công bằng của khuyến nghị cho nhóm có kích thước bằng 10	70
<i>Hình 2.7</i> Độ công bằng của khuyến nghị cho nhóm có kích thước bằng 20	71
<i>Hình 3.1</i> Sơ đồ mô hình Hệ tư vấn nhóm động trên tập mờ trực cảm.....	77
<i>Hình 3.2</i> Kết quả học tham số mô hình	98
<i>Hình 3.3</i> Đánh giá trung bình của nhóm theo các HTVN	99
<i>Hình 3.4</i> Đánh giá trung bình của nhóm theo bốn HTVN tốt nhất.....	100
<i>Hình 3.5</i> Độ công bằng của các HTVN động theo tiếp cận mờ trực cảm	102
<i>Hình 3.6</i> Độ công bằng của bốn HTVN động theo tiếp cận mờ trực cảm tốt nhất.....	103
<i>Hình 3.7</i> Độ công bằng GINI của các HTVN động theo tiếp cận mờ trực cảm.....	104
<i>Hình 3.8</i> Độ công bằng GINI của bốn HTVN động theo tiếp cận mờ trực cảm tốt nhất ..	104

CAM KẾT

Tôi là Cù Nguyên Giáp, nghiên cứu sinh chuyên ngành Khoa học máy tính tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam. Tôi xin cam đoan công trình nghiên cứu trong luận án này là công trình nghiên cứu của tôi, với sự hướng dẫn tận tình của tập thể người hướng dẫn PGS. TS. Lê Hoàng Sơn và TS. Nguyễn Như Sơn. Các kết quả trong nghiên cứu này là trung thực và chưa được công bố trong bất kỳ công trình nào khác trước đây. Các dữ liệu thể hiện trong các hình ảnh, bảng biểu được NCS trích xuất từ kết quả trong quá trình nghiên cứu.

Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm trước Hội đồng cũng như Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam về kết quả luận án của mình.

Nghiên cứu sinh



Cù Nguyên Giáp

LỜI CẢM ƠN

Trong quá trình nghiên cứu, để hoàn thiện luận án này NCS đã nhận được sự giúp đỡ, động viên khuyến khích từ Thầy/ Cô, đồng nghiệp, bạn bè và gia đình.

Trước nhất, NCS xin gửi lời cảm ơn sâu sắc nhất tới hai thầy hướng dẫn, PGS.TS Lê Hoàng Sơn, TS. Nguyễn Như Sơn đã dành thời gian quý báu để hướng dẫn và hỗ trợ NCS hoàn thành luận án tiến sĩ về đề tài "Nghiên cứu phát triển Hệ tư vấn nhóm theo tiếp cận mờ trực cảm và tích phân Choquet". NCS muốn bày tỏ lòng biết ơn của mình đến hai Thầy hướng dẫn về những chia sẻ và định hướng cho NCS trong suốt quá trình nghiên cứu. Những chỉ dẫn từ các Thầy đã giúp NCS hoàn thành luận án một cách chuyên nghiệp, đồng thời mở rộng kiến thức và kỹ năng nghiên cứu trong lĩnh vực khoa học máy tính.

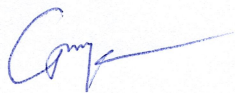
NCS cũng muốn bày tỏ sự biết ơn đến Viện Công Nghệ Thông Tin, Học Viện Khoa Học Công Nghệ thuộc Viện hàn lâm khoa học Việt Nam đã tạo điều kiện tốt nhất trong quá trình nghiên cứu.

NCS xin cảm ơn Khoa HTTTKT & TMDT, Trường Đại học Thương Mại đã tạo điều kiện về mặt công tác để NCS thực hiện luận án tiến sĩ của mình.

NCS xin trân trọng cảm ơn bạn bè, đồng nghiệp và đặc biệt là gia đình đã giành sự động viên to lớn, đưa ra những góp ý mang tính xây dựng để giúp NCS vượt qua những thách thức trong quá trình nghiên cứu.

Một lần nữa, NCS xin gửi lời cảm ơn chân thành nhất đến hai Thầy hướng dẫn, nhà trường, bạn bè đồng nghiệp và gia đình đã giúp đỡ trong quá trình nghiên cứu.

Trân trọng,



Cù Nguyên Giáp

MỞ ĐẦU

1. Tính cấp thiết của nghiên cứu

Hiện nay, vấn đề nghiên cứu và ứng dụng hệ tư vấn (HTV) đang nhận được sự quan tâm mạnh mẽ từ các nhà nghiên cứu và các công ty phát triển ứng dụng. Với nhu cầu sử dụng ngày càng rộng rãi các hệ thống thông tin quản lý, các nền tảng mạng xã hội và thương mại điện tử, chúng ta đang chứng kiến sự bùng nổ dữ liệu chứa thông tin về các đối tượng mà người dùng có thể quan tâm. Khối lượng thông tin, dữ liệu có thể rất lớn, thường sẽ vượt qua khả năng tự phân tích, đánh giá và so sánh bởi người dùng một cách thủ công. Do đó, các công cụ tìm kiếm dựa trên mục tiêu tìm kiếm rõ ràng, được biểu thị qua các từ khóa, các điều kiện .v.v. dần trở nên không đáp ứng đủ cho nhu cầu của người dùng. Dễ thấy, khi sử dụng công cụ tìm kiếm với một bộ từ khóa có thể cho hàng trăm ngàn, hàng triệu kết quả tìm kiếm, và người dùng không thể xem và đánh giá toàn bộ các kết quả này.

Trong bối cảnh đó, các HTV được phát triển như một công cụ lọc thông tin cho người dùng, được xem là giải pháp khả thi và đã được nghiên cứu, phát triển, ứng dụng trong hầu hết các hệ thống thông tin (HTTT) có số lượng người dùng lớn hiện nay [1]. Nhờ vào quá trình tự động phân tích dữ liệu liên quan đến người dùng và các sản phẩm, dịch vụ mà người dùng có thể quan tâm, hệ thống có thể đưa ra các khuyến nghị phù hợp thay vì chỉ hiển thị một lượng lớn kết quả tìm kiếm liên quan đến từ khóa. Một ví dụ điển hình thành công là các hệ tư vấn sản phẩm, dịch vụ được áp dụng trong các sàn thương mại điện tử, hệ tư vấn nội dung trên các nền tảng chia sẻ video, phim, hay khuyến nghị công việc trên các nền tảng giới thiệu việc làm.

Trong giai đoạn đầu, các HTV được phát triển tập trung vào việc đưa ra các lựa chọn phù hợp cho **một người dùng** đơn lẻ, nhưng song hành với quá trình phát triển của các ứng dụng, HTV được mở rộng để đáp ứng việc đưa ra lựa chọn phù hợp cho **một nhóm người dùng**. Các HTV hướng tới giải quyết bài toán đưa ra các khuyến nghị phù hợp cho một nhóm người dùng được gọi là Hệ tư vấn nhóm (HTVN) [2]. Về mặt mô hình hóa, HTVN là một mô hình khái quát của HTV đơn, HTVN sẽ trở thành HTV đơn khi mỗi nhóm chỉ có duy nhất một thành viên.

HTVN được ứng dụng trong các HTTT cung cấp các dịch vụ chung như tư vấn du lịch cho các nhóm du khách, tư vấn dịch vụ giải trí theo nhóm (xem phim, nghe nhạc hay lựa chọn tác phẩm văn học cho nhóm), các gói khám sàng lọc trong y tế, v.v. Trên thực tế, phạm vi ứng dụng của HTVN vẫn còn hạn chế so với hệ thống tư vấn cho người dùng đơn lẻ, nhưng theo thời gian, phạm vi ứng dụng này đang ngày càng được mở rộng.

So sánh với HTV đơn được phát triển trước đó, những thách thức với HTVN có những điểm giống và khác nhất định. Giống với các HTV nói chung, HTVN cần giải quyết bài toán dự đoán sự phù hợp của các sản phẩm, dịch vụ cần được khuyến nghị cho người dùng. Nhưng bên cạnh đó, HTVN còn phải giải quyết bài toán kết hợp sở thích, hay mong đợi của các thành viên trong một nhóm, và đây là điểm mấu chốt để các HTVN có thể đạt được hiệu quả hoạt động cao.

Những năm gần đây, tầm quan trọng của nhánh nghiên cứu về HTVN không ngừng được nâng lên, đánh dấu từ những nghiên cứu và ứng dụng đầu tiên của Masthoff năm 2004 [3], [4] về ứng dụng HTVN vào tư vấn lựa chọn chương trình truyền hình, và các nghiên cứu ứng dụng vào các lĩnh vực như du lịch, dịch vụ giải trí [2], [5]–[8]. HTVN sẽ ngày càng trở nên phổ biến hơn khi nhu cầu ra quyết định chung cho NGD trong các hoạt động cộng tác trở nên phổ biến [9].

Tổng quan nghiên cứu về HTVN cho thấy các nghiên cứu đã được công bố có thể chia thành các nhóm sau: 1) HTVN chủ yếu tập chung vào phát triển phép toán tổng hợp sở thích của các thành viên để tạo khuyến nghị cho nhóm người dùng [4], [10]; 2) HTVN ứng dụng kỹ thuật lọc cộng tác [6], [11]; 3) HTVN có tích hợp mô hình hóa ảnh hưởng xã hội [12]–[14]; 4) HTVN có sử dụng thông tin bối cảnh [15]; 5) HTVN sử dụng tiếp cận tính toán mờ. Các hướng nghiên cứu trên thể hiện mong muốn xử lý các thách thức khác nhau xuất hiện khi phát triển HTVN từ hệ tư vấn đơn. Khi khái quát hóa theo cách tiếp cận, các nghiên cứu về HTVN có thể chia làm hai nhóm tiếp cận chính: (1) Tiếp cận hợp đánh giá và (2) Tiếp cận hợp khuyến nghị. Tổng quan tài liệu cho thấy cách tiếp cận thứ hai chiếm ưu thế rất lớn so với cách tiếp cận thứ nhất [6].

Khi phát triển một mô hình HTVN, có nhiều tiêu chí có thể được sử dụng để đánh giá một HTVN [16], như độ chính xác, sự đa dạng, mức độ bao phủ hay độ công bằng, ... Trong đó, độ công bằng hay sự đồng thuận là một yêu cầu và cũng là tiêu chí đánh giá riêng biệt và rất quan trọng cho HTVN so với HTV đơn [17]. Tổng quan nghiên cứu về HTVN cho thấy trong xu hướng nghiên cứu gần đây, độ công bằng của khuyến nghị được đặc biệt quan tâm và cần tiếp tục được nghiên cứu mở rộng cho phù hợp với ứng dụng thực tiễn [18].

Bên cạnh đó, tổng quan nghiên cứu cũng chỉ ra rằng nghiên cứu về HTVN theo tiếp cận tính toán động và tiếp cận tính toán mờ là rất cần thiết, nhưng vẫn chưa được nghiên cứu nhiều [15], [19]. Tiếp cận động trong HTVN là các phương pháp khai thác thông tin xem xét sự biến động theo thời gian, chẳng hạn như sự thay đổi mức độ hấp dẫn của sản phẩm, dịch vụ. Tiếp cận tính toán mờ trong HTVN là các phương pháp sử dụng lý thuyết mờ để xử lý thông tin khi đưa ra khuyến nghị. Việc kết hợp hai yếu tố "động" và "tính toán mờ" có thể giúp bài toán HTVN phản ánh chính xác sự bất định và tính không chắc chắn trong việc đánh giá của người dùng, sự biến động trong đánh giá của người dùng, và sự thay đổi tính hấp dẫn của sản phẩm theo thời gian. Điều này giúp mô hình HTVN được xây dựng trở nên gần gũi hơn với thực tiễn. Đặc biệt, sự kết hợp giữa HTVN động theo tiếp cận tính toán mờ trực cảm và đảm bảo độ công bằng trong các khuyến nghị cho người dùng nhóm có thể mang lại những khả năng phát triển ứng dụng to lớn trong thực tế.

Dựa trên các công bố về HTVN nói chung, HTVN theo tiếp cận tính toán mờ, và HTVN động hiện nay thì vẫn còn tồn tại một số hạn chế như sau:

- Một số nghiên cứu về tư vấn nhóm đã xem xét đến độ công bằng của khuyến nghị tuy nhiên những đánh giá về độ công bằng và phương pháp giải quyết vấn đề độ công bằng của khuyến nghị đồng thời đảm bảo yếu tố tối đa hóa lợi ích tổng thể của nhóm trong HTVN còn thiếu. Do đó, nghiên cứu về nâng cao độ công bằng khuyến nghị trong HTVN hiện nay là đặc biệt quan trọng và cần tiếp tục nghiên cứu, cải thiện.

- Phát triển HTVN sử dụng tiếp cận tính toán mờ, đặc biệt là với các tập mờ mở rộng, hiện vẫn còn ít được nghiên cứu. Trong khi đó, các tập mờ mở rộng đã chứng minh được ưu điểm trong việc xử lý dữ liệu có tính do dự, không chắc chắn, như đánh giá của người dùng, trong các lĩnh vực khác, bao gồm hệ tư vấn đơn. Do đó, cần tăng cường nghiên cứu về HTVN theo tiếp cận tính toán mờ với các tập mờ mở rộng, trong đó hướng tới tập mờ trực cảm, một loại tập mờ có nhiều ứng dụng trong hệ tư vấn đơn. Đồng thời, cần mở rộng nghiên cứu HTVN theo tiếp cận động, với các mô hình HTVN có khả năng xử lý sự biến động của thông tin theo thời gian. Việc kết hợp xử lý thông tin theo tiếp cận tính toán mờ và động sẽ giúp HTVN phản ánh và xử lý thông tin một cách sát với thực tiễn hơn.

- Nghiên cứu kết hợp HTVN động theo tiếp cận tính toán mờ và hướng tới độ công bằng của khuyến nghị chưa được nghiên cứu.

Tổng hợp các hạn chế, khoảng trống nghiên cứu kể trên, trong luận án này, NCS đề xuất thực hiện “**Nghiên cứu phát triển hệ tư vấn nhóm theo tiếp cận mờ trực cảm và tích phân Choquet**”. Luận án đã phát triển mô hình HTVN sử dụng độ đo mờ để nâng cao độ công bằng của khuyến nghị, áp dụng lý thuyết tập mờ mở rộng để biểu diễn, xử lý tốt hơn thông tin có tính do dự, bất định trong phản hồi và đánh giá của người dùng, và đồng thời xem xét tính động của dữ liệu, cụ thể là ảnh hưởng của thời gian tới đánh giá của người dùng với sản phẩm, dịch vụ trong HTVN.

2. Mục tiêu nghiên cứu của luận án

Mục tiêu chung: Nghiên cứu phát triển mô hình hệ tư vấn nhóm động theo tiếp cận mờ trực cảm đảm bảo sự công bằng trong khuyến nghị.

Mục tiêu cụ thể: Các mục tiêu cụ thể của luận án bao gồm các vấn đề nghiên cứu chính như sau:

- 1) Đề xuất một mô hình HTVN với pha đồng thuận sử dụng tích phân Choquet nhằm cải thiện độ công bằng của khuyến nghị đồng thời đảm bảo mục tiêu tối đa hóa đánh giá tổng hợp của nhóm người dùng với sản phẩm, dịch vụ.
- 2) Đề xuất hệ tư vấn nhóm động dựa trên tiếp cận tính toán mờ trực cảm, trong đó xem xét ảnh hưởng của thời gian tới sự thay đổi đánh giá của các thành

viên trong nhóm, đồng thời hướng tới đảm bảo tính công bằng của khuyến nghị.

Từ các nghiên cứu về các khía cạnh kể trên, một số thuật toán được phát triển cho hệ tư vấn nhóm động theo tiếp cận tính toán mờ trực cảm có khả năng xử lý các dữ liệu gần với thực tế và có tính ứng dụng cao.

3. Đối tượng và Phương pháp nghiên cứu của luận án

Đối tượng nghiên cứu của luận án là mô hình hệ tư vấn nhóm.

Phạm vi nghiên cứu: phạm vi nghiên cứu là các mô hình hệ tư vấn nhóm sử dụng tiếp cận hợp khuyến nghị.

Phương pháp nghiên cứu của luận án bao gồm phương pháp phân tích tổng hợp, phương pháp logic và thực nghiệm so sánh.

Phương pháp phân tích tổng hợp và phương pháp logic:

Được sử dụng để thu thập, tổng hợp và phân tích các nghiên cứu, các đề xuất về phương pháp và các thuật toán đã công bố về vấn đề xây dựng HTVN nói riêng và hệ tư vấn nói chung. Phân tích thực trạng nghiên cứu phát triển các HTVN để xác định nhu cầu, các thách thức cần được giải quyết, đồng thời phân tích ưu điểm, hạn chế của các nghiên cứu đã công bố, để từ đó xác định các vấn đề còn tồn tại, các khoảng trống nghiên cứu trong phạm vi bài toán HTVN.

Trên cơ sở đó, vấn đề nghiên cứu, mục tiêu nghiên cứu của luận án được xác định. Đồng thời, các cơ sở lý thuyết liên quan đến bài toán được nghiên cứu và phân tích sâu để đưa ra phương hướng cải tiến, phát triển các mô hình HTVN mới dựa trên nền tảng kế thừa các nền tảng đã có. Các mô hình, thuật toán cải tiến được đề xuất được chứng minh về mặt lý thuyết bởi các định lý, mệnh đề.

Thực nghiệm và so sánh:

Thực nghiệm được sử dụng để phân tích và đánh giá các mô hình HTVN được đề xuất luận án dựa trên cơ sở nghiên cứu lý thuyết. Trong đó, các mô hình HTVN

được đề xuất sẽ được lập trình và thử nghiệm, từ đó dữ liệu kết quả chạy thực nghiệm được thu thập, tổng hợp. Kết quả thử nghiệm được phân tích, so sánh và đánh giá với các thuật toán khác được công bố trong các nghiên cứu trước đó trên cùng bộ dữ liệu. Nghiên cứu trong luận sử dụng dữ liệu công khai, dữ liệu được công bố và sử dụng bởi các nghiên cứu trước đó. Kết quả thực nghiệm cho phép đánh giá hiệu năng của các mô hình được đề xuất dựa trên nghiên cứu lý thuyết trước đó.

4. Bố cục của luận án

Bố cục của luận án gồm phần mở đầu, ba chương nội dung chính và phần kết luận. Trong đó:

Chương 1 giới thiệu các kiến thức khái quát về hệ tư vấn nhóm và các vấn đề liên quan trong phát triển một HTVN phù hợp. Chương này trình bày tổng quan nghiên cứu về HTVN, và tập trung vào các nghiên cứu có xem xét tới tính công bằng của khuyến nghị, xem xét về tính động trong thông tin của hệ tư vấn và xem xét tiếp cận tính toán mờ để xử lý thông tin do dự, không chắc chắn trong phát triển HTVN. Chương 1 đồng thời giới thiệu cơ sở lý thuyết về tập mờ trực cảm và tích phân Choquet, là các kiến thức nền tảng để phát triển các giải pháp được đề xuất trong luận án. Trên cơ sở đó, luận án phân tích các khoảng trống nghiên cứu và nêu rõ vấn đề và mục tiêu nghiên cứu, cách thức giải quyết và các kết quả đạt được của luận án một cách ngắn gọn.

Chương 2 trình bày nghiên cứu về hệ tư vấn nhóm với tiếp cận hợp khuyến nghị hướng tới cải thiện độ công bằng khuyến nghị dựa trên một độ đo mờ. Dễ thấy, khi xây dựng một HTVN có xem xét đến tính công bằng của khuyến nghị, việc chỉ tìm ra và đưa tới nhóm người dùng các sản phẩm dịch vụ dựa trên sự tối đa hóa tổng lợi ích của các thành viên trong nhóm là không đầy đủ. HTVN trong trường hợp này cần kết hợp giữa hai mục tiêu về tổng lợi ích của các thành viên và tính công bằng giữa các thành viên, do đó chúng ta cần phải giải quyết bài toán tối ưu đa mục tiêu trong HTVN.

Từ yêu cầu trên, Chương 2 tập trung trình bày đề xuất một độ đo mờ thể hiện cách kết hợp hai mục tiêu trên. Nguyên tắc này dựa trên việc ứng dụng tích phân

Choquet trong pha đồng thuận của HTVN. Với hàm dung lượng phi cộng tính được đề xuất, HTVN sẽ đưa ra các khuyến nghị cho nhóm người dùng đảm bảo duy trì lợi ích tổng thể cho các thành viên trong nhóm, nhưng đồng thời giảm thiểu sự bất công bằng giữa các thành viên.

Chương 3 trình bày đề xuất sử dụng lý thuyết mờ trực cảm để nghiên cứu và đề xuất một hệ tư vấn nhóm động. Trong phạm vi luận án, khái niệm động được sử dụng để chỉ các tiếp cận trong HTVN có xem xét đến sự thay đổi thông tin phản ánh đánh giá của người dùng với sản phẩm, dịch vụ và sự hấp dẫn của bản thân sản phẩm dịch vụ đó. Ngoài ra khái niệm động còn phản ánh sự thay đổi về đánh giá của người dùng với một sản phẩm khi người dùng đó sử dụng sản phẩm cùng với những nhóm người dùng khác nhau. Một hệ tư vấn nhóm động dựa trên tập mờ trực cảm được phát triển phải đồng thời với việc duy trì tính công bằng của khuyến nghị cho nhóm. Do đó trong mô hình HTVN động được đề xuất, một phép toán tổng hợp đánh giá sử dụng tích phân Choquet cho tập mờ trực cảm được phát triển và thử nghiệm để tìm được một mô hình HTVN phù hợp với thực tiễn, và có khả năng ứng dụng cao.

Các kết quả nghiên cứu trong Chương 1, Chương 2 và Chương 3 được công bố trong các bài báo khoa học tại các tạp chí quốc tế thuộc danh mục ISI và hội thảo quốc gia, quốc tế. Các công trình này được trình bày trong mục “Danh mục các công trình của tác giả”.

Cuối cùng, phần kết luận trình bày tóm lược các kết quả nghiên cứu đã đạt được, những thách thức đặt ra trong quá trình nghiên cứu, những hạn chế còn tồn tại đồng thời từ đó đưa ra những đề xuất để phát triển, mở rộng nghiên cứu trong tương lai.

Chương 1. TỔNG QUAN VỀ HỆ TƯ VẤN NHÓM

Chương này trình bày ngắn gọn một số vấn đề cơ bản trong nghiên cứu về hệ tư vấn nhóm, các phương pháp phát triển hệ tư vấn nhóm bao gồm tiếp cận tính toán mờ, và kiến thức cơ bản về tập mờ trực cảm và tích phân Choquet. Các kiến thức nền tảng trình bày ở chương này là cơ sở để phát triển các đề xuất của luận án được trình bày chi tiết trong các chương 2 và 3.

1.1. Giới thiệu về Hệ tư vấn nhóm

1.1.1. Hệ tư vấn đơn người dùng

Hệ tư vấn (HTV) là một lĩnh vực nghiên cứu tập trung vào việc xây dựng các hệ thống thông minh có khả năng đưa ra gợi ý cho người dùng [20]–[22]. HTV được phát triển nhằm xác định quan hệ giữa người dùng và sản phẩm, dịch vụ và từ đó xây dựng khuyến nghị bao gồm các sản phẩm được dự đoán là người dùng quan tâm, hoặc được dự đoán là phù hợp với người dùng trong những bối cảnh nhất định [23].

Nghiên cứu về HTV có thể được phân thành nhiều hướng nghiên cứu, dựa trên phương thức khai thác dữ liệu có thể chia các nghiên cứu hai hướng tiếp cận chính [21], [24], bao gồm: Cách tiếp cận truyền thống và Cách tiếp cận mở rộng. Trong đó, cách tiếp cận thứ nhất bao gồm các mô hình lọc tin, chủ yếu là các mô hình: Lọc theo nội dung, Lọc cộng tác và Lọc kết hợp. HTV theo cách tiếp cận truyền thống sử dụng dữ liệu phổ biến là thông tin về người dùng, dữ liệu về đặc trưng sản phẩm và đánh giá của người dùng cho các sản phẩm đó. HTV sử dụng cách tiếp cận mở rộng tập trung vào các hướng chính như: Hệ tư vấn theo ngữ cảnh, Hệ tư vấn dựa trên mạng xã hội, hệ tư vấn dựa trên sở thích cá nhân hoặc cải tiến các phương pháp lọc thông tin kết hợp. Với hướng tiếp cận này, ngoài các thông tin cơ sở, HTV sử dụng thêm các thông tin bổ xung (như dữ liệu phản ánh ngữ cảnh, dữ liệu thể hiện mối liên kết trên mạng xã hội, dữ liệu hành vi .v.v.) để nâng cao chất lượng của các khuyến nghị.

HTV thường được coi là một phương thức lọc dữ liệu chủ động, có vai trò hỗ trợ quá trình ra quyết định, với mục tiêu gợi ý cho người dùng về các sản phẩm, dịch vụ phù hợp với nhu cầu của người dùng theo ngữ cảnh. Nói cách khác, hệ tư vấn giúp

giảm thiểu tình trạng quá tải thông tin bằng cách dự đoán sự quan tâm của người dùng đối với các sản phẩm mới, chưa được đánh giá, đồng thời đưa ra danh sách sản phẩm, dịch vụ mà người dùng có thể sẽ quan tâm [25]. Số lượng sản phẩm được gợi ý thông thường sẽ nhỏ hơn rất nhiều so với số lượng sản phẩm, dịch vụ tồn tại trong các hệ thống.

Ngày nay, HTV đã trở thành yếu tố đặc biệt quan trọng, ảnh hưởng tới thành công và sự phát triển của các hệ thống thông tin, đặc biệt là các sàn giao dịch thương mại điện tử và ngày càng trở nên không thể thiếu trong các HTTT hiện đại [1], [20]. Rõ ràng, với sự tăng cường nhanh chóng về khối lượng dữ liệu, các HTV là một trong các công cụ quan trọng để người dùng nhận được thông tin phù hợp nhất nhằm cải thiện trải nghiệm người dùng, và tăng sự hài lòng của khách hàng.

Hiện nay, khi phát triển HTV các phương pháp lọc thông tin đã được chứng minh khả năng mang lại hiệu quả cao. Trong đó thuật toán Lọc cộng tác đã và đang trở thành một hướng tiếp cận nổi bật, được sử dụng phổ biến trong nghiên cứu và triển khai thực tế.

Các cách tiếp phổ dụng trong hệ tư vấn [20], [26]:

- **Lọc nội dung:** Phương pháp này đề xuất nội dung tương tự với những gì người dùng đã thể hiện sự quan tâm, dựa trên các đặc điểm của chính nội dung đó (ví dụ: văn bản, thể loại, thể, từ khóa). Ví dụ: nếu người dùng thường xuyên xem phim hành động, thuật toán có thể đề xuất các phim khác thuộc thể loại hành động. Phương pháp này có những ưu điểm đó là: Hoạt động tốt ngay cả với người dùng hoặc mục mới (không cần dữ liệu cơ sở người dùng lớn); Dễ giải thích vì các đề xuất dựa trên các tính năng nội dung mà hệ thống có thể dễ dàng xác định.

- **Lọc cộng tác:** Phương pháp “Lọc cộng tác dựa trên người dùng” đề xuất nội dung dựa trên sở thích và hành vi của những người dùng tương tự. Ví dụ: nếu người dùng A thích nội dung X và người dùng B thích nội dung X và Y, và A và B là hai người dùng tương tự nhau thì hệ thống sẽ đề xuất nội dung Y cho người dùng A. Phương pháp “Lọc cộng tác dựa trên tập mục” đề xuất các mục tương tự với các mục

mà người dùng đã tương tác trước đó. Nếu người dùng thích một bộ phim cụ thể, thuật toán sẽ đề xuất các bộ phim khác mà người dùng có sở thích tương tự thích.

- **Tiếp cận lai ghép:** Đây là cách sử dụng đồng thời cả hai phương pháp lọc nội dung và lọc cộng tác một cách đồng thời. Sự kết hợp này cho phép khai thác điểm mạnh của mỗi phương pháp kể trên và giảm thiểu hạn chế của mỗi phương pháp.

Ngoài cách phương pháp truyền thống kể trên, một số kỹ thuật mở rộng khác cũng có thể được áp dụng, ví dụ như cách tiếp cận sử dụng công nghệ học máy hay phương pháp phân tích ma trận nhân tử trong để phát triển HTV. Ứng dụng công nghệ học máy, đặc biệt ứng dụng công nghệ học sâu vào trong các HTV là một xu hướng nghiên cứu nhận được sự quan tâm khá lớn trong những năm gần đây do sự phát triển đột phá của các kỹ thuật học sâu và hạ tầng tính toán [27]–[29]. Trong đó, công nghệ học sâu chủ yếu được sử dụng để tạo thêm các thông tin bổ sung cho các mô hình truyền thống về hệ tư vấn [27], và nó thường được kết hợp với các thuật toán lọc như lọc công tác, hay học nội dung.

Nghiên cứu của luận án này không tập trung vào HTV đơn người dùng, do đó chi tiết kỹ thuật của các hướng tiếp cận kể trên không được trình bày sâu hơn. Kiến thức tổng quát về HTV đơn người dùng là nền tảng để nghiên cứu và phát triển hệ tư vấn nhóm, là đối tượng nghiên cứu chính trong luận án.

1.1.2. Hệ tư vấn nhóm

Ban đầu, các Hệ tư vấn chủ yếu được nghiên cứu phát triển để xây dựng các hệ thống có khả năng sinh ra gợi ý tới từng người dùng độc lập, tuy nhiên hiện nay, chúng đã được mở rộng để cung cấp khuyến nghị phù hợp cho một nhóm người dùng, khi đó HTV được gọi là hệ tư vấn nhóm. HTVN được sử dụng ngày càng phổ biến hơn khi các hệ thống cung cấp dịch vụ dùng chung ngày càng nhiều, chẳng hạn như các hệ thống đưa ra khuyến nghị cho nhóm du khách, hệ thống tư vấn sản phẩm, dịch vụ giải trí dùng chung (nhóm người dùng nhỏ cùng xem phim hay thưởng thức âm nhạc cùng nhau) .v.v. [2], [3], [30], [31]. Với mỗi nhóm người dùng, cảm nhận của từng thành viên với kết quả khuyến nghị tự động là không đồng nhất, một cá nhân có thể hài lòng tuyệt đối với khuyến nghị, nhưng các cá nhân khác có thể không. Thêm

vào đó đánh giá của một thành viên về một sản phẩm, dịch vụ còn chịu tác động từ các thành viên khác, một người có thể có cảm nhận khác nhau về cùng một sản phẩm, dịch vụ khi cùng sử dụng sản phẩm này với các nhóm thành viên khác nhau. Từ đó, có thể thấy khi xây dựng HTVN, ngoài những thách thức mà một HTV đơn người dùng gặp phải, còn cần phải giải quyết thêm các vấn đề khác, riêng có với HTVN, như độ công bằng của khuyến nghị, hay ảnh hưởng của yếu tố tương tác người dùng trong nhóm.

Khái niệm hệ tư vấn nhóm: Hệ tư vấn nhóm có thể được hiểu là hệ thống đưa ra các khuyến nghị về một nhóm sản phẩm, dịch vụ được cho là phù hợp với một nhóm người dùng [4]. HTVN cần kết hợp mức độ yêu thích hay đánh giá của các thành viên để tạo ra một đánh giá của nhóm về sản phẩm, dịch vụ từ đó đưa ra khuyến nghị cho nhóm theo trật tự đánh giá. Bởi vì đánh giá của mỗi thành viên đối với tập sản phẩm, dịch vụ là khác nhau, nên thách thức lớn nhất khi phát triển HTVN là xây dựng phương thức kết hợp các đánh giá này sao cho phù hợp với các thành viên.

Mô hình HTVN đơn giản nhất có thể được biểu diễn như sau:

Cho $U = \{u_1, u_2, \dots, u_n\}$ và $I = \{i_1, i_2, \dots, i_m\}$ là tập hợp người dùng và sản phẩm, dịch vụ; cho $R \subseteq U \times I \rightarrow D$ là tập hợp các đánh giá của người dùng cho các sản phẩm trên miền đánh giá D . Cho nhóm người dùng $g = \{u_1, u_2, \dots, u_l\} | u_i \in U$, khi đó hệ khuyến nghị nhóm được mô hình hóa như sau:

$$HTVN(g, I) = \arg \max_{i \in I} [pref(g, i)] \quad (1.1)$$

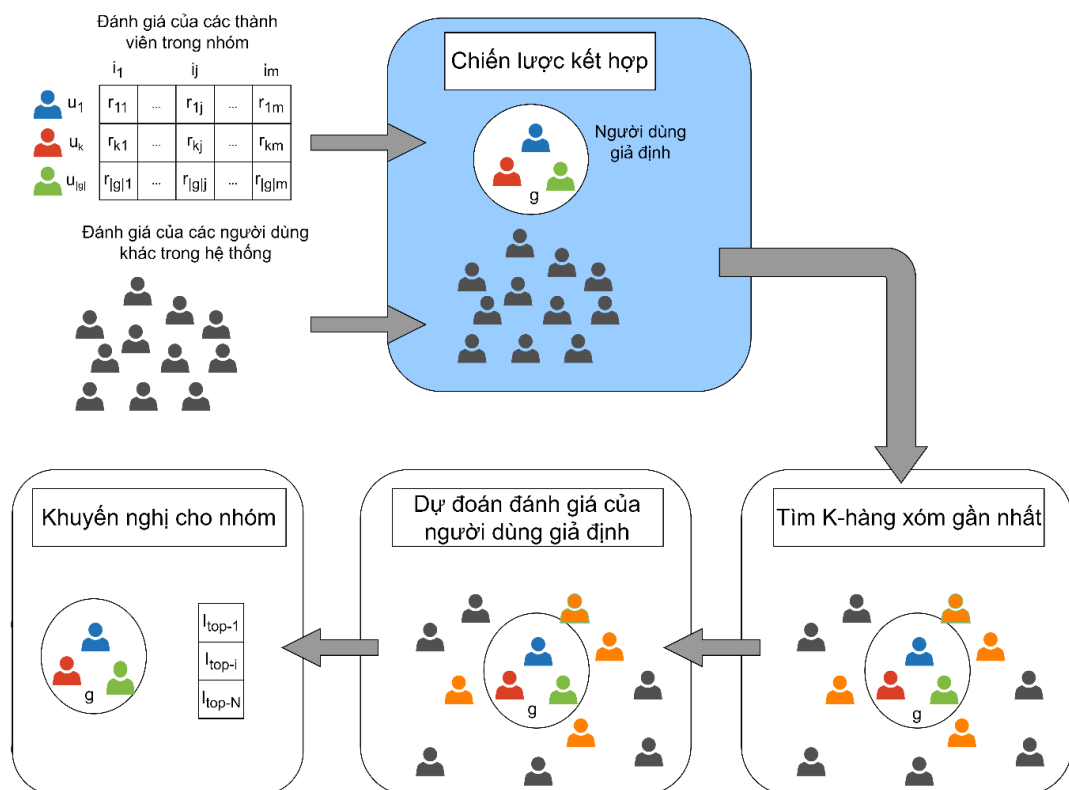
Trong đó $pref(g, i)$ là giá trị dự báo thể hiện mức độ quan tâm hay đánh giá của nhóm người dùng g cho một sản phẩm dịch vụ i .

Trong trường hợp đơn giản đánh giá của nhóm về một sản phẩm, dịch vụ xác định từ đánh giá của các thành viên với sản phẩm đó bằng một phép toán hợp đơn giản (ví dụ như phép tính trung bình). Nhưng nó cũng có thể được ước lượng theo những cách tiếp cận phức tạp hơn phụ thuộc vào ứng dụng.

1.1.3. Các cách tiếp cận xây dựng HTVN

Xét một cách tổng quát, nghiên cứu về HTVN được chia làm hai hướng nghiên cứu chính [2] đó là HTVN theo tiếp cận hợp đánh giá, và HTVN theo tiếp cận hợp khuyến nghị. Cách tiếp cận thứ nhất tạo ra một “người dùng giả định” cho nhóm người dùng từ dữ liệu đánh giá sản phẩm của toàn bộ các thành viên. Người dùng giả định được sử dụng như một người dùng thông thường trong HTV đơn để tạo ra các khuyến nghị phù hợp. Khuyến nghị cho người dùng giả định đó sẽ được coi là khuyến nghị cho nhóm người dùng [14]. Hình 1.1 thể hiện quy trình của HTVN theo tiếp cận này.

Cách tiếp cận này đặt ra yêu cầu về việc xây dựng một hồ sơ người dùng từ tất cả các hồ sơ người dùng trong nhóm. Do đó có thể đặt ra một số yêu cầu riêng về vấn đề tiếp cận thông tin.



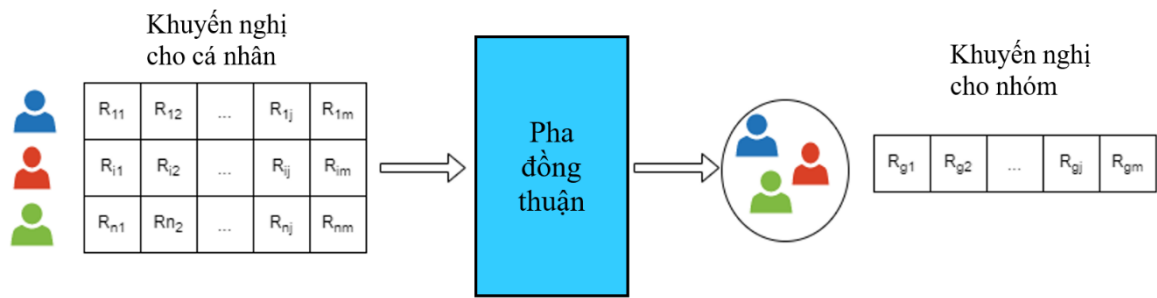
Hình 1.1 Hệ tư vấn nhóm theo tiếp cận hợp đánh giá

Cách tiếp cận thứ hai để xây dựng HTVN là tiếp cận hợp khuyến nghị, trong đó dựa trên pha đồng thuận để tạo ra khuyến nghị cho nhóm người dùng, trong đó sử dụng

dụng đánh giá của các thành viên về các sản phẩm [11]. Quy trình tổng quát của các tiếp cận gồm hai pha, pha sinh khuyến nghị và pha đồng thuận.

Pha sinh khuyến nghị sẽ sinh ra tập danh sách các sản phẩm, dịch vụ được lựa chọn để khuyến cho từng thành viên trong nhóm dựa trên dữ liệu của cá nhân thành viên đó. Quá trình này tương đương với một HTV đơn người dùng.

Tại pha đồng thuận, một danh sách bao gồm các sản phẩm, dịch vụ được lựa chọn để khuyến nghị tới nhóm người dùng từ danh sách ưu tiên của từng thành viên đã được khuyến nghị trong pha thứ nhất. Giai đoạn hợp khuyến nghị này được gọi là pha "đồng thuận" của HTVN. Sơ đồ quy trình xử lý trong pha đồng thuận được trình bày trong hình 1.2 sau đây.



Hình 1.2 Pha đồng thuận của Hệ tư vấn nhóm

Xây dựng khuyến nghị cho nhóm bằng cách tổng hợp các khuyến nghị cho từng thành viên cho phép phát triển nhiều kỹ thuật để cải thiện chất lượng của HTVN, đáp ứng các mục tiêu khác với hệ tư vấn truyền thống như tính công bằng. Ngoài ra, cách thiết kế này cho phép các hệ thống thông tin triển khai đồng thời tư vấn cho nhóm và tư vấn đơn người dùng trên cùng một cơ chế hệ thống đồng nhất. Tổng quan nghiên cứu cũng cho thấy HTVN theo tiếp cận hợp khuyến nghị đang được ứng dụng rộng rãi, và có ưu thế hơn [2].

1.1.4. HTVN sử dụng cách tiếp cận hợp khuyến nghị

Hệ tư vấn nhóm theo tiếp cận hợp khuyến nghị sử dụng pha đồng thuận là hướng tiếp cận được ứng dụng rộng rãi nhất hiện nay [2], [32]. Có nhiều phép toán tổng hợp đánh giá của thành viên khác nhau được đề xuất cho pha đồng thuận của HTVN. Mỗi

phép toán cho thấy một chiến lược khác nhau được thiết kế để tạo ra một giá trị chung đại diện cho đánh giá của nhóm cho mỗi sản phẩm, dịch vụ. Các chiến lược hướng tới đạt được giá trị tổng lợi ích là tối đa, ví dụ như chiến lược cộng, chiến lược trung bình, hay các chiến lược xem xét đến các thành viên đặc biệt trong nhóm ví dụ như “chiến lược ít thiệt thòi nhất” và “chiến lược ưu thế nhất”. Một số chiến lược khác mô tả cách thức con người tạo ra sự đồng thuận trong thực tế. Ví dụ như cơ chế biểu quyết trong thực tế được thể hiện trong chiến lược biểu quyết phê duyệt. Các chiến lược khác như chiến lược luật Copeland hay chiến lược chỉ số Borda và “chiến lược công bằng” thể hiện các cách thức xử lý khác mà con người đã áp dụng trong thực tế. Các chiến lược kể trên đang là các chiến lược chính được sử dụng trong các nghiên cứu đã được công bố về HTVN.

Cụ thể hơn, giả sử ta có nhóm người dùng g , với mỗi thành viên $u | u \in g$ ta có đánh giá của u với sản phẩm i là r_{ui} . Từ đó, giá trị tổng hợp đại diện cho đánh giá của nhóm g với sản phẩm i , ký hiệu là $pref(g, i)$, được xác định theo các chiến lược có cách tính cụ thể như sau:

Chiến lược cộng (AUS) [3]: Chiến lược tính tổng đánh giá của các thành viên để tạo thành đánh giá chung của nhóm. Công thức như sau:

$$pref(g, i) = \sum_{u \in g} r_{ui} \quad (1.2)$$

Chiến lược tích (MS) [3]: Chiến lược này tính tích đánh giá của các thành viên để tạo thành đánh giá chung của nhóm. Công thức như sau:

$$pref(g, i) = \prod_{u \in g} r_{ui} \quad (1.3)$$

Chiến lược ít thiệt thòi nhất (LMS) [3]: Chiến lược này dùng đánh giá thấp nhất của thành viên nhóm làm đánh giá chung của nhóm. Công thức như sau:

$$pref(g, i) = \min_{u \in g} (r_{ui}) \quad (1.4)$$

Chiến lược ưu thế (MPS) [3]: Chiến lược này dùng đánh giá cao nhất của thành viên nhóm làm đánh giá chung của nhóm. Công thức như sau:

$$pref(g, i) = \max_{u \in g} (r_{ui}) \quad (1.5)$$

Chiến lược biểu quyết phê duyệt (AVS) [3]: Chiến lược này lấy tổng số thành viên bỏ phiếu đồng ý với một sản phẩm được lựa chọn trong danh sách khuyến nghị làm đánh giá của nhóm. Công thức như sau:

$$pref(g, i) = \sum_{u \in g} v_{ui} \quad (1.6)$$

Trong đó, $v_{ui} = 1$ khi người dùng u bỏ phiếu đồng ý với sản phẩm i và ngược lại $v_{ui} = 0$ khi người dùng này bỏ phiếu không đồng ý.

Chiến lược luật Copeland (CRS) [3]: Chiến lược này mô phỏng cách xây dựng chỉ số Copeland, trong đó đánh giá của nhóm cho một sản phẩm, dịch vụ được tính theo công thức.

$$pref(g, i) = \sum_{j \in I} c_{i,j} \quad (1.7)$$

Trong đó $c_{i,j} = 1$ khi sản phẩm i “đánh bại” j . Khái niệm đánh bại được xác định khi số lượng thành viên đánh giá sản phẩm i cao hơn sản phẩm j nhiều hơn so với số thành viên có đánh giá ngược lại. $c_{i,j} = -1$ nếu j đánh bại i và trong trường hợp còn lại $c_{i,j} = 0$.

Chiến lược chỉ số Borda (BCS) [3]: Sử dụng một chỉ số mới dựa trên phương pháp chỉ số Borda làm chỉ số thay thế. Cụ thể, đánh giá của nhóm cho một đối tượng là tổng điểm do thành viên nhóm đánh giá cho sản phẩm đó, và chỉ số này sẽ được dùng thay cho các đánh giá ban đầu. Chỉ số Borda được xác định dựa trên thứ tự xếp hạng giảm dần của các đánh giá của thành viên cho một sản phẩm. Chỉ số Borda có nhiều biến thể, và người phát triển các HTVN có thể thay đổi tùy thuộc vào bài toán cụ thể.

Chiến lược công bằng (FS) [3]: Chiến lược này cho phép lần lượt từng thành viên trong nhóm chọn một sản phẩm, dịch vụ người đó ưu tiên để đưa vào kết quả khuyến nghị chung. Sau đó, chu trình này tiếp tục được lặp lại để đảm bảo danh sách

khuyến nghị đủ số lượng theo yêu cầu. Một sản phẩm đã được lựa chọn vào danh sách khuyến nghị sẽ không được lựa chọn lại bởi những người dùng khác. Chiến lược công bằng thể hiện thông qua quyền ưu tiên lựa chọn lần lượt được chuyển giao cho các thành viên trong nhóm.

Đánh giá độ phức tạp HTVN:

Dựa trên công thức trình bày ở trên, pha đồng thuận được xây dựng theo các chiến lược khác nhau sẽ có độ phức tạp tính toán khác nhau. Từ đánh giá của nhóm cho tất cả các sản phẩm HTVN lựa chọn top- N các sản phẩm có đánh giá cao nhất để đưa vào khuyến nghị. Chúng ta có thể xác định độ phức tạp tính toán của HTVN sử dụng các chiến lược khác nhau như sau.

Để tạo đánh giá của nhóm g đối với sản phẩm i , các chiến lược như chiến lược cộng, chiến lược tích, chiến lược ít thiệt thòi nhất, chiến lược ưu thế và chiến lược biểu quyết phê duyệt đều tổng hợp đánh giá của các thành viên bằng một phép toán số học đơn giản. Do đó, độ phức tạp của của pha đồng thuận sử dụng các chiến lược trên là như nhau và bằng $O(|g|)$. Để lựa chọn N sản phẩm tốt nhất, ta lần lượt lựa chọn từng sản phẩm có đánh giá cao nhất từ tập I sản phẩm, vậy cần tính giá trị đánh giá của nhóm cho tất cả I sản phẩm. Dễ thấy để chọn ra N sản phẩm tốt nhất từ tập I sản phẩm, HTVN sử dụng các chiến lược đồng thuận kể trên có độ phức tạp tính toán là $O(|I| \cdot |g| \cdot N)$.

Bên cạnh đó, chiến lược luật Copeland và chiến lược chỉ số Borda trước hết cần so sánh đánh giá của thành viên nhóm để sinh ra chỉ số mới thay cho giá trị đánh giá ban đầu. Do đó các chiến lược này có độ phức tạp tính toán cao hơn. Độ phức tạp cho thuật toán thực hiện việc chuyển đổi chỉ số là $O(|g| \cdot |I|)$. Sau khi có chỉ số thay thế, việc tổng hợp các chỉ số này dựa trên phép toán số học cơ bản, từ đó ta HTVN sử dụng hai chiến lược này sẽ có độ phức tạp tính toán là $O(|I|^2 \cdot |g| \cdot N)$.

Độ phức tạp trong pha đồng thuận phản ánh sự khác biệt về độ phức tạp tính toán của thuật toán hệ khuyến nghị nhóm so với hệ khuyến nghị đơn người dùng. Các thuật toán thể hiện các chiến thuật đồng thuận có độ phức tạp tính toán thấp có thể được cài đặt và thực hiện đơn giản nhưng các chiến lược này lại không xét đến các mục tiêu khác hay các vấn đề phát sinh trong HTVN như ảnh hưởng của nhóm tới thành viên, hay độ công bằng và các yếu tố khác.

1.1.5. Các phương pháp đánh giá Hệ tư vấn nhóm.

Hiệu quả của các HTVN được đánh giá bằng hai phương thức chính là đánh giá "trực tuyến" và đánh giá "ngoại tuyến" [16]. Phương thức đánh giá trực tuyến dựa trên việc xây dựng hệ thống áp dụng HTVN trên một ứng dụng thực tế, trong đó quá trình đánh giá sẽ được tính trực tiếp từ phản hồi trên ứng dụng của người dùng. Tuy nhiên, phương pháp này rất ít được ứng dụng do chi phí phát triển ứng dụng lớn, thời gian và nguồn lực cần để thu thập phản hồi kéo dài.

Cách thức tốt nhất hiện nay trong việc đánh giá HTV nói chung và HTVN nói riêng là phương pháp đánh giá ngoại tuyến. Cụ thể, phương pháp này dùng một tập dữ liệu có sẵn, trong đó có dữ liệu cần thiết để phát triển HTVN. Từ đó, dữ liệu được chia làm hai tập dữ liệu riêng biệt: tập dữ liệu huấn luyện và tập dữ liệu kiểm thử. Tập dữ liệu thứ nhất được sử dụng để phát triển các mô hình HTVN, sau đó các mô hình này được dùng để tạo ra các khuyến nghị cho nhóm người dùng trong tập dữ liệu kiểm thử. Phân tích kết quả các khuyến nghị được đưa ra từ mô hình HTVN và dữ liệu thực tế đã thu thập sẽ cho thấy hiệu quả của các mô hình này.

Với cả hai cách tiếp cận kể trên, để đo lường sự khác biệt giữa kết quả đưa ra bởi HTVN và kết quả thực cần có một số thang đo để phản ánh hiệu năng của các mô hình. Các độ đo này được thiết kế để phản ánh mức độ chính xác của mô hình trong dự báo đánh giá, dự đoán phân lớp, mức độ đồng thuận hay tính đa dạng của khuyến nghị. Đây là các thước đo phổ biến nhất được ứng dụng để phân tích hiệu năng của các HTVN [16], cụ thể như sau:

- Độ đo phân lớp [16]: là các độ đo đánh giá hiệu quả HTVN sử dụng giá trị “độ nhạy” và “độ đặc hiệu” cho khía cạnh đánh giá khả năng phân lớp chính xác. Khi ứng dụng trong đánh giá HTVN, độ nhạy được hiểu là tỷ lệ giữa tổng số sản phẩm được khuyến nghị được coi là “có liên quan” so với tổng số sản phẩm trong danh sách đề xuất. Độ đặc hiệu được hiểu là tỷ lệ giữa số lượng sản phẩm được coi là “có liên quan” trong danh sách đề xuất với tổng số lượng sản phẩm được coi là “có liên quan” trong dữ liệu. Một cách tổng quát, các độ đo này cho thấy tỷ lệ sản phẩm được xác định hợp lý (có liên quan) với nhóm người dùng trong danh sách khuyến nghị. Trong đó, sai số trong các pha dự báo sẽ không được quan tâm.
- Độ đo sai số [16]: Trong HTVN, để xác định sai số khi dự báo đánh giá của nhóm hay của người dùng, sai số tuyệt đối trung bình - MAE và sai số toàn phương trung bình - RMSE là hai thước đo phổ biến được sử dụng trong các công bố trước đây. MAE và RMSE đều cho thấy sai số của các mô hình khi dự báo giá trị đánh giá của người dùng, nhóm người dùng.
- Độ chính xác trong phân hạng – ĐCXPH [16]: Độ đo này thể hiện sự liên quan của các sản phẩm, dịch vụ được khuyến nghị với kết quả thực đồng thời xét đến thứ tự của sản phẩm, dịch vụ trong danh mục khuyến nghị đó. Tổng quan nghiên cứu cho thấy phương pháp giá trị “lợi nhuận tích lũy triết khấu” (DCG) là phương pháp phổ biến nhất để đánh giá ĐCXPH.
- Độ bao phủ và tính ngẫu nhiên [16]: tính chất này của các HTVN có thể được xem xét và phát triển theo các quan điểm khác nhau. Một phương thức phổ biến đó là sử dụng "độ bao phủ nhóm" (GC) thể hiện tỷ lệ số lượng nhóm nhận được tư vấn (ít nhất là nhận được một kết quả tư vấn) trên tổng số lượng nhóm cần tư vấn. Một độ đo khác đó là "độ bao phủ sản phẩm" (CC) thể hiện tỷ lệ tổng số các sản phẩm, dịch vụ được sử dụng để tạo khuyến nghị ít nhất một lần trên tổng số sản phẩm, dịch vụ có trong cơ sở dữ liệu.

- Tính đồng thuận và độ công bằng [16]: Đây là nhóm các độ đo thể hiện tỷ lệ số thành viên nhóm hài lòng với kết quả khuyến nghị, hay sự chênh lệch giữa đánh giá của các thành viên về sản phẩm .v.v. Thực tế hiện nay, khi phát triển các mô hình HTVN trong các hệ thống thông tin, tính công bằng của khuyến nghị ngày càng có vai trò quan trọng để đảm bảo sự hài lòng của người dùng. Nội dung này sẽ được nghiên cứu kỹ hơn trong luận án do mức độ quan trọng của nó, và các độ đo được sử dụng khi đánh giá tính công bằng của khuyến nghị của HTVN được trình bày trong mục 1.2.1.

1.2. Tổng quan nghiên cứu về HTVN

HTVN đầu tiên được giới thiệu trong những nghiên cứu nổi bật của Mathoff và cộng sự [4], và sau đó đến những năm gần đây HTVN mới thực sự trở thành một nhánh nghiên cứu nổi bật. Các giai đoạn phát triển của HTVN có thể tóm tắt như sau.

Giai đoạn đầu (cuối thập niên 1990- đầu thập niên 2000), HTVN được phát triển như một phần mở rộng của HTV truyền thống. Trong đó, các nghiên cứu tập trung chủ yếu vào các phép toán cho phép tổng hợp một cách phù hợp nhất các đánh giá cá nhân để tạo ra các khuyến nghị phù hợp cho nhóm [4], [10]. Giai đoạn cuối thập niên 2000 chứng kiến sự tích hợp các kỹ thuật lọc cộng tác vào Hệ thống tư vấn nhóm [6], [11]. Kỹ thuật lọc cộng tác cho phép phân tích sự tương đồng giữa các người dùng hay sản phẩm dựa trên dữ liệu lịch sử, từ đó xây dựng mô hình dự đoán đánh giá của nhóm người dùng với sản phẩm và tạo tư vấn. Cách tiếp cận này nâng cao độ chính xác của các khuyến nghị đưa ra bởi các HTVN.

Những năm 2010 chứng kiến các HTVN có tích hợp mô hình hóa ảnh hưởng xã hội [12]–[14]. Dữ liệu mô hình hóa ảnh hưởng xã hội thường được thu thập từ sự liên kết giữa các hệ thống với các nền tảng truyền thông và mạng xã hội. Các hệ thống này xem xét các kết nối trong mạng xã hội, sự tương tác qua lại giữa người dùng trong và ngoài nhóm để điều chỉnh các đề xuất. Trong giai đoạn này, các HTVN theo bối cảnh cũng được nghiên cứu và phát triển như một trong các trọng tâm của HTVN [15]. Các hệ thống này tích hợp các yếu tố theo ngữ cảnh như thời gian, địa điểm hay

loại thiết bị người dùng sử dụng để truy cập vào quá trình đánh giá, phân tích và đưa ra các khuyến nghị cho các nhóm. Sử dụng thông tin về bối cảnh có ưu điểm là có thể nâng cao trải nghiệm của người dùng bằng cách liên tục điều chỉnh các khuyến nghị cho phù hợp với sự thay đổi đa dạng của bối cảnh.

Nửa cuối thập niên 2010 chứng kiến sự phát triển của các HTVN tập trung vào việc nâng cao sự đa dạng và tính công bằng trong đề xuất [17], [18], [33]. Các nhà nghiên cứu tập trung vào các kỹ thuật để giảm thiểu sự bất bình đẳng, đảm bảo sự đại diện và tính công bằng trong các đề xuất cho một nhóm. Bên cạnh đó, giai đoạn này cũng chứng kiến các nghiên cứu về HTVN theo tiếp cận động [15], [34] và các nghiên cứu về HTVN theo tiếp cận tính toán mờ [19], [35].

Những năm gần đây, phương pháp tiếp cận kết hợp trở thành xu hướng nghiên cứu nổi bật [8], [36] bên cạnh các tiếp cận về HTVN động, HTVN mờ [37], [38]. HTVN hiện tại thường áp dụng các phương pháp kết hợp, kết hợp lọc cộng tác, lọc nội dung, kết hợp sử dụng thông tin về bối cảnh, thời gian và các kỹ thuật khác như tính toán mờ để tối ưu hóa độ chính xác, tính đa dạng và sự hài lòng của người dùng. Các mô hình kết hợp tận dụng điểm mạnh của các thuật toán khác nhau để khắc phục các hạn chế và đưa ra đề xuất mạnh mẽ hơn cho các nhóm người dùng đa dạng.

Tổng quan chi tiết hơn các nghiên cứu về HTVN hướng tới nâng cao độ công bằng, HTVN động và HTVN theo tiếp cận tính toán mờ được trình bày trong các phần dưới đây. Từ đó vấn đề nghiên cứu, mục tiêu nghiên cứu, và giải pháp đề xuất trong luận án sẽ được trình bày ngắn gọn.

1.2.1. Tổng quan nghiên cứu về độ công bằng trong HTVN

Nửa cuối thập niên 2010 các nghiên cứu về HTVN đã tập trung nhiều hơn vào đề độ công bằng trong các đề xuất cho một nhóm [17], [18], [33]. Thực tế với HTVN, nâng cao tính đa dạng và công bằng là rất quan trọng để nâng cao sự hài lòng và tin cậy của người dùng với các HTVN. Khi độ công bằng giữa các thành viên được cải thiện, sự thỏa mãn nói chung của nhóm sẽ tăng và điều này sẽ làm tăng mức độ ứng

dụng thực tiễn của các HTVN [39]. Một số tiếp cận nổi bật trong việc nâng cao tính công bằng của khuyến nghị trong HTVN như sau:

Nghiên cứu [18], các tác giả đưa ra hai khái niệm về độ công bằng đó là “tỷ lệ hài lòng mức m ”, và “độ ghen tị mức m ”. Từ hai khái niệm này, các tác giả xây dựng một hàm thể hiện đánh giá của một nhóm người dùng với một nhóm sản phẩm được khuyến nghị, và HTVN được xây dựng dựa trên thuật toán tham lam để tìm lần lượt các sản phẩm được đánh giá cao nhất để đưa vào tập khuyến nghị cho nhóm.

Trong nghiên cứu [33] đưa ra khái niệm độ công bằng trong tư vấn dịch vụ y tế thông qua khái niệm tỷ lệ người hài lòng trong nhóm. Giá trị độ đo đánh giá của nhóm với khuyến nghị là tích của độ công bằng và độ liên quan của dịch vụ với nhóm. Từ đó các tác giả xây dựng hàm tối ưu trong HTVN là $D^* = \arg \max_{|D|=z} (value(g, D))$, trong đó số lượng sản phẩm khuyến nghị là z , g là nhóm người dùng, và D là tập hợp sản phẩm được khuyến nghị. Trong nghiên cứu [40] đề cập đến khái niệm công bằng trong hệ khuyến nghị, tuy không trực tiếp xây dựng HTVN, nhưng cách tiếp cận này có thể áp dụng trong HTVN. Trong đó các tác giả tiếp cận dựa trên nguyên tắc xây dựng một độ đo để tìm những người hàng xóm phù hợp nhất để đưa ra dự đoán đánh giá của một người dùng. Từ đó, tập sản phẩm khuyến nghị cho người dùng sẽ công bằng hơn.

Trong nghiên cứu [17], các tác giả sử dụng giải bài toán tối ưu đa mục tiêu theo tiếp cận tối ưu Pareto. Trong đó mục tiêu tìm kiếm của HTVN là một giải pháp kết hợp giữa độ đo công bằng (social welfare) và tính hữu ích của sản phẩm. Hàm thể hiện giá trị đánh giá của nhóm người dùng với sản phẩm là: $\lambda.SW(g, i) + (1 - \lambda)F(g, i)$. Trong đó, g là nhóm người dùng, i là sản phẩm và λ là hệ số thể hiện sự ưu tiên giữa hai độ đo.

Độ đo tính công bằng trong HTVN

Trong nghiên cứu về hệ tư vấn nói chung, có nhiều cách nhìn khác nhau về độ công bằng [40], tuy nhiên trong nghiên cứu về HTVN, các nghiên cứu về tính công

bằng chủ yếu tập trung vào sự khác biệt giữa mức độ thỏa mãn hoặc đánh giá của các thành viên trong nhóm đối với các sản phẩm, dịch vụ được khuyến nghị. Một số độ đo đã được phát triển và công bố để ước lượng và so sánh hiệu quả giữa các chiến lược tiếp cận cho pha đồng thuận của HTVN. Một số nghiên cứu điển hình đã trình bày các định nghĩa và độ đo liên quan đến khái niệm công bằng trong HTVN, như sau.

Các nghiên cứu trước đây như [16], [17] các nhà nghiên cứu định nghĩa khái niệm “công bằng” trong HTVN là sự mất cân bằng giữa sự hài lòng của các cá nhân trong nhóm. Từ đó các độ đo về “độ công bằng trong khuyến nghị nhóm” đã được các nhà nghiên cứu đề xuất. Giả sử $pref(u, i)$ thể hiện tính hữu ích của sản phẩm được đề xuất i với người dùng u , một số độ đo tính công bằng được các tác giả đề xuất như sau:

+ Độ đo công bằng “*Least misery*”:

$$fairness_{LM}(g, i) = \min\{pref(u, i), \forall u \in g\} \quad (1.8)$$

Độ đo công bằng này dựa trên đánh giá thấp nhất của một người dùng trong nhóm về một khuyến nghị của HTVN để làm độ đo tính công bằng.

+ Độ đo công bằng theo phương sai:

$$fairness_{var}(g, i) = 1 - \text{Var}(pref(u, i), \forall u \in g) \quad (1.9)$$

Độ đo công bằng này dựa trên sự khác biệt giữa đánh giá của các người dùng trong nhóm về một khuyến nghị của HTVN để làm độ đo tính công bằng.

+ Độ đo công bằng “*Jain*”:

$$fairness_{Jain}(g, i) = \frac{\left(\sum_{u \in g} pref(u, i)\right)^2}{|g| * \sum_{u \in g} pref(u, i)^2} \quad (1.10)$$

Độ đo công bằng này tạo ưu thế cho các khuyến nghị mà đánh giá của các người dùng trong nhóm về khuyến nghị đó là tương đồng nhau.

+ Độ đo công bằng theo tỷ lệ Min-Max:

$$fairness_{\min-\max}(g, i) = \frac{\min\{pref(u, i), \forall u \in g\}}{\max\{pref(u, i), \forall u \in g\}} \quad (1.11)$$

Độ đo công bằng tỷ lệ giữa đánh giá thấp nhất và cao nhất trong nhóm người dùng.

Bên cạnh đó, nghiên cứu [18] đề xuất bổ xung hai độ đo tính công bằng mới. Trong tiếp cận đầu tiên, các tác giả đánh giá độ công bằng dựa trên tỷ lệ của số lượng người dùng trong nhóm hài lòng với ít nhất m sản phẩm, dịch vụ được khuyến nghị và kích thước nhóm. Cụ thể, khái niệm *m-proportional* của một “gói” các sản phẩm, dịch vụ thể hiện gói sản phẩm, dịch vụ đó có ít nhất m sản phẩm, dịch vụ được đánh giá là hài lòng bởi một người dùng. Từ đó, ta có độ đo công bằng được giới thiệu như sau:

$$fairness_{m-proportional}(U, I) = \frac{|U_I|}{|U|} \quad (1.12)$$

Trong đó U_I tập các người dùng nhận I là một *m-proportional* của nó.

Độ đo bổ xung thứ hai dựa trên khái niệm “không ghen tị cấp m ” (*m-envy-freeness*). Giá trị *m-envy-freeness* của một gói các sản phẩm, dịch vụ được khuyến nghị thể hiện rằng tập sản phẩm, dịch vụ đó chứa ít nhất m sản phẩm, dịch vụ được thích bởi người dùng u cũng được thích bởi người dùng khác. Trên cơ sở đó, ta có độ đo công bằng được giới thiệu như sau :

$$fairness_{m-envy-freeness}(U, I) = \frac{|U_I|}{|U|} \quad (1.13)$$

Trong đó U_I là tập người dùng nhận I là một *m-envy-freeness* của nó.

- Trong nghiên cứu [16], một độ đo công bằng khác được các tác giả đề xuất. Cụ thể, công thức tính độ khuyến nghị của một sản phẩm, dịch vụ i cho một nhóm người dùng g được thể hiện như sau:

$$fairness(g, i) = \frac{\left| \bigcup_{u \in g} r_{ui} \geq \theta \right|}{|g|} \quad (1.14)$$

Trong đó r_{ui} là đánh giá của người dùng u trong nhóm cho sản phẩm, dịch vụ i và θ là một ngưỡng cho thấy người dùng u hài lòng với khuyến nghị sản phẩm i . Từ đó công thức tính độ công bằng của một tập sản phẩm, dịch vụ I được khuyến nghị cho một nhóm người dùng g được tính bằng cách tính tổng độ công bằng của từng sản phẩm, dịch vụ đơn lẻ.

Tổng quan nghiên cứu về HTVN cho thấy vấn đề độ đo tính công bằng đã được nghiên cứu và công bố nhưng còn rời rạc, một số độ đo chỉ được sử dụng một lần hoặc trong rất ít các nghiên cứu mà không có tính kế thừa. Cách tiếp cận chủ yếu là định nghĩa một độ đo về độ công bằng của khuyến nghị, sau đó phát triển một độ đo đại diện cho đánh giá của nhóm người dùng với sản phẩm là sự kết hợp có trọng số giữa độ đo công bằng và độ đo tính hữu ích của sản phẩm. HTVN sẽ tìm tập sản phẩm tối ưu theo độ đo đánh giá của nhóm. Cách tiếp cận trên tuy đơn giản và dễ thực hiện, nhưng vẫn có những điểm yếu nhất định do cách kết hợp giữa độ đo công bằng và độ đo tính hữu ích sản phẩm là cố định, như nhau đối với mọi người dùng và mọi nhóm người dùng. Trên thực tế mỗi nhóm người dùng có ưu tiên khác nhau giữa hai mục tiêu trên.

Ngoài ra, tương tác giữa các thành viên trong nhóm và ảnh hưởng của nó tới đánh giá của từng thành viên tới sản phẩm, dịch vụ trong khuyến nghị chưa được các nghiên cứu trước đây xem xét khi phát triển HTVN hướng tới cải thiện độ công bằng của khuyến nghị. Do đó để cải tiến và nâng cao hiệu quả của HTVN trong khía cạnh nâng cao độ công bằng, nghiên cứu trong luận án sẽ phân tích và xem xét các phương pháp mới, đồng thời xem xét cả ảnh hưởng của tương tác giữa các thành viên nhóm.

Nghiên cứu [41] cho thấy ứng dụng tích phân Choquet với hàm dung lượng phi cộng tính có thể tìm kiếm giải pháp tốt hơn thỏa mãn cân bằng giữa hai mục tiêu tính công bằng khuyến nghị và tối ưu hóa lợi ích. Nhưng nghiên cứu này giới hạn trong phạm vi bài toán Hệ hỗ trợ ra quyết định đa tiêu chí, đồng thời gặp phải giới hạn xử lý rất lớn.

Trong luận án này, NCS sử dụng tiếp cận này trong HTVN có thể tạo các khuyến nghị cân bằng hơn giữa độ công bằng và tính hữu ích. Đồng thời có thể xem xét đến ảnh hưởng giữa các thành viên trong nhóm với nhau khi cùng sử dụng một sản phẩm, dịch vụ. Tuy nhiên các nghiên cứu trước đây cũng chỉ ra việc xây dựng hàm dung lượng phi cộng tính tối ưu là một bài toán thuộc lớp NP-Khó. Do đó, việc tìm kiếm hàm dung lượng phù hợp là một thách thức cần giải quyết. Giải quyết vấn đề này là một đóng góp nghiên cứu quan trọng trong luận án, và nội dung chi tiết được trình bày trong Chương 2.

1.2.2. Tổng quan hệ tư vấn nhóm động

Nghiên cứu sự biến động của thông tin trong HTV:

Trong hệ tư vấn nói chung, các phương pháp khai thác thông tin xem xét đến sự biến động theo thời gian có thể được chia thành bốn nhóm [42]. Mỗi nhóm thể hiện một quan điểm khác nhau khi khai thác thông tin động theo bối cảnh và theo thời gian.

- Tiếp cận gần đúng: Do sở thích của người dùng hiện tại có thể được thể hiện tốt hơn bởi đánh giá hiện tại so với đánh giá cũ, nên các phương pháp xử lý trong hướng “tiếp cận gần đúng” khai thác các ảnh hưởng của thời gian tới đánh giá của người dùng bằng cách “chiết khấu” theo thời gian các tác động của đánh giá trong quá khứ đối với hành vi trong tương lai của người dùng trong khi vẫn duy trì xu hướng chính, có hữu trong dữ liệu [43], [44].
- Cách tiếp cận dựa trên cơ sở gom nhóm: Các phương pháp trong cách tiếp cận dựa trên cơ sở gom nhóm có nhiều dạng khác nhau, nhưng tất cả đều nhóm dữ liệu theo khoảng thời gian mà dữ liệu được thu thập [45]–[47]. Từ đó, khi phát triển một mô hình tư vấn cả dữ liệu huấn luyện và kiểm thử trong phương pháp tiếp cận này đều sinh từ cùng dữ liệu trong một khoảng thời gian và chúng thường được tạo bằng cách phân chia ngẫu nhiên tất cả dữ liệu trong khoảng thời gian này. Do đó, những dự đoán về sở thích của người dùng đối với các

mục, sản phẩm chưa đánh giá thực sự là những dự đoán phụ thuộc vào thời gian về sở thích của người dùng.

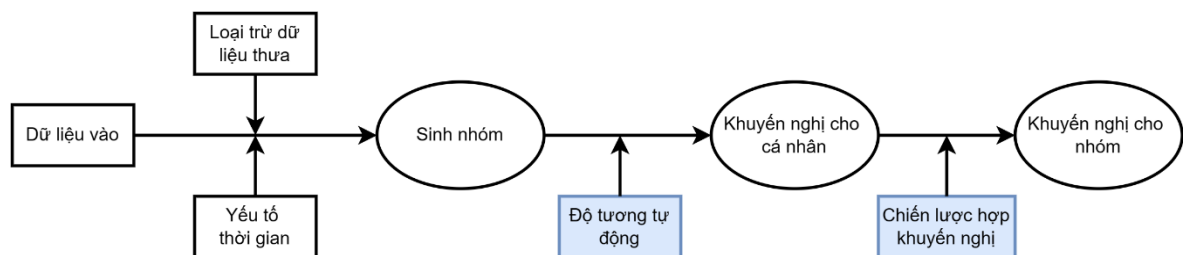
- Phương pháp cập nhật trực tuyến: Các phương pháp trong cách tiếp cận này có mục đích chung là mô hình hóa cách sử dụng hệ thống tư vấn trong thực tế. Do đó, tại thời điểm khởi đầu chỉ thông tin trong quá khứ được sử dụng để dự đoán những điều chưa biết trong tương lai. Sau đó, mô hình cơ bản phải được cập nhật để phản ánh phản hồi mới đến trong hệ thống. Do đó, các cơ chế cập nhật trực tuyến được phát triển để cập nhật các tham số mô hình một cách linh hoạt [48], [49]. Tuy nhiên, các phương pháp của tiếp cận này thường nhấn mạnh vào khả năng mở rộng của giai đoạn cập nhật và bỏ qua việc mô hình hóa các tính chất động của thông tin liên quan tới thị hiếu người dùng và mức độ hấp dẫn của sản phẩm, dịch vụ.
- Cách tiếp cận dựa trên tính động: Cách tiếp cận này dựa trên mô hình hóa rõ ràng các biến động theo thời gian trong phản hồi để theo dõi xu hướng thay đổi của các yếu tố như sở thích của người dùng và mức độ hấp dẫn của sản phẩm, dịch vụ. Trong một số phương pháp, mô hình không gian trạng thái ngẫu nhiên hoặc mô hình hồi quy với yếu tố thời gian thường được áp dụng để theo dõi xu hướng sở thích của người dùng [50], [51]. Tuy nhiên, sự hấp dẫn của sản phẩm trong cách tiếp cận này được giả định là tĩnh trong mô hình không gian trạng thái. Bên cạnh đó, giả thiết các phân phối trong hệ thống là phân phối Gaussian có thể đơn giản hóa quá mức các tình huống thực tế trong thực tế và có thể dẫn đến hiệu suất dự đoán thấp.

Nghiên cứu sự biến động của thông tin trong HTVN:

Những phương pháp xử lý dữ liệu động được áp dụng cho HTV nói chung đều có thể được dùng cho HTVN, từ đó phát triển các HTVN động phù hợp với thực tiễn. Tuy vậy, tổng quan nghiên cứu cho thấy HTVN là một vấn đề được đặt ra và phát triển muộn hơn so với HTV đơn người dùng, và các nghiên cứu về HTVN trước đây thường chỉ tập trung vào giải quyết bài toán kết hợp các đánh giá của thành viên để

tạo ra đánh giá chung. Số lượng nghiên cứu và công bố về HTVN có sử dụng tiếp cận động còn tương đối hạn chế. Một số nghiên cứu điển hình có thể chỉ ra dưới đây:

Nghiên cứu của Jinpeng Chen và cộng sự [52] là một nghiên cứu điển hình đã chỉ ra vấn đề cần xem xét sự thay đổi của đánh giá của người dùng với sản phẩm theo thời gian, và sử dụng điều này để phát triển một mô hình HTVN. Trong đó, các nhà nghiên cứu đặt giả thiết rằng sự khác biệt trong đánh giá của những người dùng chịu ảnh hưởng bởi khoảng chênh lệch về mặt thời gian mà những đánh giá đó được đưa ra. Sự khác biệt về mặt thời gian đánh giá càng lớn thì sự khác biệt giữa những người dùng đó càng lớn. Với giả thiết này, các tác giả đã đưa ra mô hình đánh giá sự tương tự động (dynamic similarity) giữa những người dùng và từ đó xây dựng một HTVN động. Các bước chính trong mô hình được đề xuất thể hiện trong sơ đồ dưới đây (Hình 1.3).



Hình 1.3 Một ví dụ về HTVN động

Một nghiên cứu khác về HTVN động đó là nghiên cứu của Huang về pha đồng thuận của HTVN có xem xét mối quan hệ thứ bậc của các sản phẩm theo thời gian [53]. Nghiên cứu xem xét ảnh hưởng của thời điểm xuất hiện đối với thứ tự các sản phẩm trong khuyến nghị đối với người dùng với các sản phẩm. Ví dụ một sản phẩm A được khuyến nghị đứng trước sản phẩm B. Nghiên cứu này đề xuất một mô hình mới để khai phá các quy luật đồng thuận theo thời gian, từ đó sử dụng các quy luật này để tạo các chuỗi sản phẩm có thứ tự đạt được sự đồng thuận tối đa.

Xem xét tính động của thông tin theo bối cảnh, nghiên cứu của Jia và cộng sự đề xuất một HTVN động trong đó xem xét mức độ tích cực của thành viên ảnh hưởng

tới nhóm trong đánh giá sản phẩm, dịch vụ như thế nào [34]. Nghiên cứu đề xuất phương thức ước lượng sự thay đổi của đánh giá mỗi thành viên theo thời gian như sau:

$$b_{ui}(t) = \mu + b_u(t) + b_i(t) \quad (1.15)$$

Trong đó, $b_{ui}(t)$ thể hiện đánh giá của thành viên u với sản phẩm i vào thời điểm t , và μ thể hiện đánh giá trung bình, $b_u(t)$, $b_i(t)$ thể hiện sự ảnh hưởng của thời gian tới đánh giá của mỗi thành viên và sự hấp dẫn của sản phẩm. Chi tiết ước lượng các sự biến đổi này được trình bày trong tài liệu gốc của nghiên cứu, và không được giới thiệu lại tại đây [34]. Khi đó, trong pha đồng thuận ảnh hưởng của mức độ “tích cực” của thành viên sẽ được xem xét và sử dụng.

Tổng quát, mức độ tích cực của người dùng được đánh giá bằng công thức:

$$A_{uG}(t_{rec}) = \frac{|R_u(t)|}{\sum_{u \in G} |R_u(t)|}, \quad t \in [t_{rec} - T, t_{rec}] \quad (1.16)$$

Trong đó, mức độ tích cực được xem xét trong một cửa sổ trượt theo thời gian với kích thước T , và độ tích cực được sử dụng như trọng số trong phép toán hợp đánh giá các thành viên trong nhóm để tạo thành đánh giá chung của nhóm.

Có thể thấy nghiên cứu về HTVN theo tiếp cận động còn rất hạn chế, mặc dù đây là một tiếp cận quan trọng để bài toán gần với thực tiễn hơn. Trong luận án này, NCS nghiên cứu đề xuất sử dụng tiếp cận phù hợp để đo lường sự ảnh hưởng của yếu tố thời gian lên đánh giá của người dùng tới sản phẩm, và sự suy hao tính hấp dẫn của sản phẩm theo thời gian. Chương 3 của luận án sẽ trình bày chi tiết đề xuất và kết quả của hướng nghiên cứu này.

1.2.3. Hệ tư vấn nhóm theo tiếp cận tính toán mờ

Bên cạnh tiếp cận thông tin động, sử dụng tiếp cận tính toán mờ trong xây dựng hệ tư vấn là một chiến lược tiếp cận được theo đuổi bởi nhiều nhà nghiên cứu. Tiếp cận này có nhiều ưu điểm như là có thể biểu diễn và xử lý thông tin không chắc chắn được thể hiện trong đánh giá của người dùng với sản phẩm. Nghiên cứu ứng

dụng trực tiếp tính toán mờ trong HTVN có phần hạn chế hơn so với ứng dụng tính toán mờ trong HTV đơn người dùng. Một số nghiên cứu tiêu biểu như:

Nghiên cứu [54], các tác giả sử dụng các tiếp cận tính toán mờ trên tập mờ loại I để cải thiện việc biểu diễn và tính toán đánh giá của người dùng. Từ đó tác giả sử dụng tiếp cận này trong mô hình lọc cộng tác để xây dựng một HTVN. Trong nghiên cứu này, các tác giả cũng nghiên cứu ảnh hưởng tương tác qua lại giữa các thành viên trong nhóm. Các tác giả thực nghiệm mô hình đề xuất với dữ liệu công khai và cho kết quả đáng khích lệ.

Trong một công bố khác, tác giả Jorge Castro và cộng sự [19] ứng dụng tiếp cận tính toán mờ để khử nhiễu trong bài toán HTVN từ đó nâng cao chất lượng của mô hình tư vấn. Trong nghiên cứu này, các tác giả xây dựng mô hình biểu diễn xu hướng trong đánh giá của người dùng và sản phẩm. Khi các đánh giá không đi theo xu hướng, nó sẽ được xem xét và xử lý như một yếu tố nhiễu.

Trong nghiên cứu khác của Jorge Castro và cộng sự [35], các tác giả sử dụng tiếp cận tính toán mờ để xử lý vấn đề mất mát thông tin trong phép toán hợp khi xây dựng đánh giá của nhóm người dùng từ đánh giá của các thành viên. Trong đó tập mờ do dự được sử dụng như một tập nền. Nghiên cứu hướng đến xây dựng khuyến nghị có tính đa dạng, và kết quả nghiên cứu chỉ ra rằng trong mô hình được đề xuất tính đa dạng của khuyến nghị tỷ lệ thuận với kích thước nhóm.

Có thể thấy nghiên cứu phát triển HTVN theo tiếp cận động và phát triển HTVN theo tiếp cận tính toán mờ đều có những ưu điểm nổi bật, và có thể hỗ trợ lẫn nhau. Từ đó xây dựng được một mô hình HTVN tốt hơn.

Bên cạnh đó, có thể thấy nghiên cứu về hệ tư vấn nhóm động theo tiếp cận tính toán mờ còn mờ, chưa được nghiên cứu có hệ thống. Cụ thể, cho tới hiện tại mới có rất ít tập mờ mở rộng được ứng dụng trong HTVN, và với các tập mờ mở rộng đã được ứng dụng thì các phép toán được dùng trong HTVN vẫn là các phép toán cơ bản. Trong khi đó, kết hợp hai yếu tố “động” và “tính toán mờ” có thể giúp bài toán HTVN biểu diễn đúng đặc trưng về sự bất định và tính không chắc chắn khi đưa ra

các đánh giá của người dùng, và sự biến động trong đánh giá của người dùng, sự thay đổi về tính hấp dẫn của sản phẩm theo thời gian. Đặc biệt là sự kết hợp giữa HTVN động theo tiếp cận tính toán mờ trực cảm và mục tiêu tăng cường tính công bằng của kết quả tư vấn đối với các nhóm người dùng có thể mang lại tiềm năng ứng dụng lớn trong thực tế và chưa được nghiên cứu. Do đó, trong chương tiếp theo của chuyên đề này, một mô hình HTVNĐ cho tập mờ trực cảm được đề xuất, mô hình này cũng sử dụng phép toán tổng hợp đánh giá của các thành viên trong nhóm dựa trên tích phân Choquet cho tập mờ trực cảm trong pha đồng thuận để cải thiện tính công bằng của khuyến nghị.

1.3. Giới thiệu về lý thuyết tập mờ trực cảm

1.3.1. Khái quát về Tập mờ trực cảm

Trong các tập mờ và tập mờ mở rộng được nghiên cứu, tập mờ trực cảm có những ưu thế nhất định trong biểu diễn và xây dựng các hệ tư vấn [55]. Với những ưu điểm của mình, tập mờ trực cảm là nền tảng để phát triển hệ tư vấn nhóm mờ động trong luận án. Trong nội dung dưới đây, kiến thức tổng quan về tập mờ trực cảm được trình bày một cách ngắn gọn.

Định nghĩa về tập mờ trực cảm được giới thiệu bởi Atanasov [56], [57].

Định nghĩa 1.1: Cho không gian nền X , một tập mờ trực cảm A trên X được cho bởi:

$$A = \{(x, \mu_A(x), \nu_A(x)) \mid x \in X\} \quad (1.17)$$

Thỏa mãn

$$\mu_A : X \rightarrow [0,1], \nu_A : X \rightarrow [0,1]$$

Trong đó, $\mu_A(x) \in [0,1]$ thể hiện độ thuộc của một phần tử x với A , và $\nu_A(x) \in [0,1]$ thể hiện độ không thuộc của x với A , và điều kiện $0 \leq \mu_A(x) + \nu_A(x) \leq 1$ thỏa mãn $\forall x \in X$.

Định nghĩa 1.2: Số mờ trực cảm khái quát $A = \{(x, \mu_A(x), \nu_A(x)) \mid x \in X\}$, được xác định bởi:

$$\mu_A(x) = \begin{cases} f_\mu^L(x) & , a \leq x \leq b \\ u & , b \leq x \leq c \\ f_\mu^R(x) & , c \leq x \leq d \\ 0 & , otherwise \end{cases} \quad (1.18)$$

$$\nu_A(x) = \begin{cases} g_\nu^L(x) & , a_1 \leq x \leq b \\ v & , b \leq x \leq c \\ g_\nu^R(x) & , c \leq x \leq d_1 \\ 1 & , otherwise \end{cases} \quad (1.19)$$

Trong đó $f_\mu^L(x)$, $g_\nu^R(x)$ là các hàm đơn điệu tăng, và $f_\mu^R(x)$, $g_\nu^L(x)$ là các hàm đơn điệu giảm, giá trị u, v là hai tham số, và a, b, c, d, a_1, b_1 là các ngưỡng.

Các phép toán cơ bản trên tập mờ trực cảm, bao gồm các phép toán tập hợp như phép hợp, phép giao, các phép toán logic như so sánh bằng, thứ bậc, phép toán đại số như phép cộng, nhân, lũy thừa được định nghĩa cụ thể như sau [56], [58]:

Các phép toán tập hợp	
$A \cup B = \{ \langle x, \max(\mu_A(x), \mu_B(x)), \min(\nu_A(x), \nu_B(x)) \rangle \mid x \in X \}$	(1)
$A \cap B = \{ \langle x, \min(\mu_A(x), \mu_B(x)), \max(\nu_A(x), \nu_B(x)) \rangle \mid x \in X \}$	(2)
$A^c = \{ \langle x, \nu_A(x), 1 - \mu_A(x) \rangle \mid x \in X \}$	(3)
$A \subseteq B$ if $\mu_A(x) \leq \mu_B(x) \ \& \ \nu_A(x) \geq \nu_B(x)$ $A \supseteq B$ if $\mu_A(x) \geq \mu_B(x) \ \& \ \nu_A(x) \leq \nu_B(x)$	(4)
$A = B$ if $\mu_A(x) = \mu_B(x) \ \& \ \nu_A(x) = \nu_B(x) \ \forall x \in X$	(5)
$A \leq B$ if $\mu_A(x) \leq \mu_B(x) \ \& \ \nu_A(x) \geq \nu_B(x) \ \forall x \in X$	(6)

Các phép toán số học	
$A \oplus B = \langle \mu_A(x) + \mu_B(x) - \mu_A(x) \cdot \mu_B(x), \nu_A(x) \cdot \nu_B(x) \rangle$	(7)
$\lambda A = \langle 1 - (1 - \mu_A(x))^\lambda, \nu_A(x)^\lambda \rangle, \lambda > 0$	(8)
$A \otimes B = \langle \mu_A(x) \cdot \mu_B(x), \nu_A(x) + \nu_B(x) - \nu_A(x) \nu_B(x) \rangle$	(9)
$A^\lambda = \langle \mu_A(x)^\lambda, 1 - (1 - \nu_A(x))^\lambda \rangle, \lambda > 0$	(10)
$A \odot B = \begin{cases} \left\langle \frac{\mu_A - \mu_B}{1 - \mu_B}, \frac{\nu_A}{\mu_B} \right\rangle, & \text{if } 0 \leq \frac{\nu_A}{\mu_B} \leq \frac{1 - \mu_A}{1 - \mu_B} \leq 1 \\ \langle 0, 1 \rangle, & \text{otherwise} \end{cases}$	(11)

Các phép toán đại số trong tập mờ trực cảm ở trên là nền tảng để phát triển các thuật toán xử lý với dữ liệu mờ trực cảm.

1.3.2. Khoảng cách và độ tương tự cho tập mờ trực cảm

Khoảng cách và độ tương tự cho tập mờ trực cảm:

Khoảng cách giữa các tập mờ: Các phép tính khoảng cách giữa các tập mờ trực cảm là một trong các phép toán quan trọng nhất để thực hiện xử lý dữ liệu trong các thuật toán sử dụng logic mờ trực cảm. Dưới đây là một số phép tính khoảng cách được các nhà nghiên cứu phát triển cho các tập mờ trực cảm. Các phép tính này là sự mở rộng từ các phép tính khoảng cách cho các tập mờ truyền thống.

Khoảng cách Hamming, và khoảng cách Euclidean là các phép toán tính khoảng cách giữa các tập mờ trực cảm được giới thiệu bởi Szmidt và Kacprzyk [59]:

Khoảng cách Hamming:

Cho A, B là hai tập mờ trực cảm trên tập nền $X = \{x_1, \dots, x_n\}$

$$d_{IFS}(A, B) = \sum_{i=1}^n (|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)| + |\pi_A(x_i) - \pi_B(x_i)|) \quad (1.20)$$

Khoảng cách Euclidean:

$$e_{IFS}(A, B) = \sqrt{\sum_{i=1}^n \left((\mu_A(x_i) - \mu_B(x_i))^2 + (\nu_A(x_i) - \nu_B(x_i))^2 + (\pi_A(x_i) - \pi_B(x_i))^2 \right)} \quad (1.21)$$

Tuy nhiên, Grzegorzewski [60] không thỏa mãn với các công thức đề xuất bởi Szmidt và Kacprzyk. Dựa trên độ đo Hausdorff, Grzegorzewski đưa ra các công thức khác về khoảng cách giữa các tập mờ trực cảm như.

Khoảng cách Hamming chuẩn hóa:

Cho A, B là các tập mờ trực cảm trên tập nền $X = \{x_1, \dots, x_n\}$

$$d_G(A, B) = \frac{1}{n} \sum_{i=1}^n \max\{|\mu_A(x_i) - \mu_B(x_i)|, |\nu_A(x_i) - \nu_B(x_i)|\} \quad (1.22)$$

Khoảng cách Euclidean chuẩn hóa:

$$e_G(A, B) = \sqrt{\frac{1}{n} \sum_{i=1}^n \max\{(\mu_A(x_i) - \mu_B(x_i))^2, (\nu_A(x_i) - \nu_B(x_i))^2\}} \quad (1.23)$$

Độ tương tự giữa các tập mờ trực cảm:

Bên cạnh các phép tính khoảng cách, độ tương tự giữa các tập mờ trực cảm cũng là một khái niệm quan trọng được sử dụng khi tính toán và so sánh trong các thuật toán sử dụng số mờ trực cảm. Jun Ye (2011) giới thiệu các độ đo tương tự giữa các tập mờ trực cảm [61], các công thức tính cụ thể được trình bày dưới đây.

Độ tương tự Cosine (Cosine similarity):

Độ tương tự Cosine của 2 tập IFSs [61]:

$$\text{sim}_{\cosine}(A, B) = \frac{1}{n} \sum_{i=1}^n \frac{\mu_A(x_i) \cdot \mu_B(x_i) + \nu_A(x_i) \cdot \nu_B(x_i)}{\sqrt{\mu_A^2(x_i) + \mu_B^2(x_i)} + \sqrt{\nu_A^2(x_i) + \nu_B^2(x_i)}} \quad (1.24)$$

Độ tương tự Cosine có trọng số (Weighted Cosine similarity) của hai IFSs [61]

$$wsim_{\cosine}(A, B) = \sum_{i=1}^n w_i \frac{\mu_A(x_i) \cdot \mu_B(x_i) + \nu_A(x_i) \cdot \nu_B(x_i)}{\sqrt{\mu_A^2(x_i) + \mu_B^2(x_i)} + \sqrt{\nu_A^2(x_i) + \nu_B^2(x_i)}} \quad (1.25)$$

Độ tương quan: độ đo tương quan giữa các tập mờ trực cảm.

Định nghĩa: độ tương quan $r_{A-IFS}(A, B)$ giữa hai $A-IFSs$, A và B trên X , được xác định bằng công thức:

$$r_{A-IFS}(A, B) = \frac{1}{3} (r_1(A, B) + r_2(A, B) + r_3(A, B)) \quad (1.26)$$

$$r_1(A, B) = \frac{\sum_{i=1}^n (\mu_A(x_i) - \bar{\mu}_A) \cdot (\mu_B(x_i) - \bar{\mu}_B)}{\sqrt{\sum_{i=1}^n (\mu_A(x_i) - \bar{\mu}_A)^2} \cdot \sqrt{\sum_{i=1}^n (\mu_B(x_i) - \bar{\mu}_B)^2}} \quad (1.27)$$

$$r_2(A, B) = \frac{\sum_{i=1}^n (\nu_A(x_i) - \bar{\nu}_A) \cdot (\nu_B(x_i) - \bar{\nu}_B)}{\sqrt{\sum_{i=1}^n (\nu_A(x_i) - \bar{\nu}_A)^2} \cdot \sqrt{\sum_{i=1}^n (\nu_B(x_i) - \bar{\nu}_B)^2}} \quad (1.28)$$

$$r_3(A, B) = \frac{\sum_{i=1}^n (\pi_A(x_i) - \bar{\pi}_A) \cdot (\pi_B(x_i) - \bar{\pi}_B)}{\sqrt{\sum_{i=1}^n (\pi_A(x_i) - \bar{\pi}_A)^2} \cdot \sqrt{\sum_{i=1}^n (\pi_B(x_i) - \bar{\pi}_B)^2}} \quad (1.29)$$

1.3.3. Phép toán trung bình với các số mờ trực cảm

Bên cạnh độ đo khoảng cách và độ tương quan, các phép toán trên tập các số mờ trực cảm là một phần quan trọng để phát triển các phép toán cho HTVN. Đặc biệt là cách xây dựng các chiến lược hợp khuyến nghị cho HTVN động theo tiếp cận mờ trực cảm. Dưới đây là các phép toán hợp và trung bình quan trọng.

Cho $A = \{(\mu_{a_i}, \nu_{a_i}); i = 1, 2, \dots, n\}$ là tập các số mờ trực cảm. Ta có:

Trung bình cộng có trọng số:

$$IFAW(A, W) = \bigoplus_{i=1}^n a_i w_i = \left\langle 1 - \prod_{i=1}^n (1 - \mu_{a_i})^{w_i}, \prod_{i=1}^n \nu_{a_i}^{w_i} \right\rangle \quad (1.30)$$

Trung bình cộng Bonfferroni (Intuitionistic Bonfferroni Mean) [62]

$$IFB^{p,q}(A) = \left(\frac{1}{n(n-1)} \left(\bigoplus_{\substack{i,j=1 \\ i \neq j}}^n a_i^p \otimes a_j^q \right) \right)^{\frac{1}{p+q}} \quad (1.31)$$

$$IFB^{p,q}(A) = \left\langle \left(1 - \prod_{\substack{i,j=1 \\ i \neq j}}^n (1 - \mu_{a_i}^p \mu_{a_j}^q)^{\frac{1}{n(n-1)}} \right)^{\frac{1}{p+q}}, 1 - \left(1 - \prod_{\substack{i,j=1 \\ i \neq j}}^n (1 - (1 - \nu_{a_i})^p (1 - \nu_{a_j})^q)^{\frac{1}{n(n-1)}} \right)^{\frac{1}{p+q}} \right\rangle$$

Trong đó $p, q \geq 0$.

1.4. Giới thiệu về tích phân Choquet

Để phát triển được phép toán tổng hợp khuyến nghị cho pha đồng thuận của HTVN dựa trên tích phân Choquet, khái niệm và cách thức tính tích phân Choquet, khái niệm về hàm dung lượng và độ đo mờ sẽ được giới thiệu lại một cách ngắn gọn trong mục này.

1.4.1. Định nghĩa hàm dung lượng

Nghiên cứu về tích phân Choquet và các độ đo mờ được nhiều tác giả quan tâm [41], [63]. Trong nghiên cứu [41] tác giả T. Lust giới thiệu và ứng dụng của độ đo mờ cho bài toán hỗ trợ ra quyết định đa tiêu chí trong đó hướng tới lựa chọn kết hợp được hai mục tiêu, tối đa tổng giá trị đánh giá theo từng tiêu chí và đồng thời tối thiểu hóa sự khác biệt trong đánh giá giữa từng tiêu chí. Nghiên cứu [41] chỉ ra sự tồn tại về mặt lý thuyết một lựa chọn thỏa mãn đồng thời hai yêu cầu trên, tuy nhiên nghiên cứu chưa đưa ra giải thuật phù hợp trong ứng dụng.

Trong nghiên cứu thuộc luận án này, NCS kế thừa và mở rộng nghiên cứu [41] cho bài toán tổng hợp khuyến nghị trong pha đồng thuận của HTVN. Do đó, các định

nghĩa về tích phân Choquet rời rạc và hàm dung lượng được đề cập như cơ sở lý thuyết để phát triển mô hình pha đồng thuận mới.

Trong nghiên cứu [41] để xây dựng độ đo mờ trên một tập các điều kiện, một hàm dung lượng cần được xác định trên tập các điều kiện đó. Hàm dung lượng được định nghĩa như sau:

Định nghĩa 1.1 [41]: cho tập M là một tập các điều kiện thỏa mãn $M \neq \emptyset$, hàm dung lượng là một hàm $\xi: 2^M \rightarrow [0,1]$ thỏa mãn:

$$\left| \begin{array}{l} \xi(\emptyset) = 0, \xi(M) = 1 \\ \forall A, B \subset 2^M \text{ \& } A \subseteq B \Rightarrow \xi(A) \leq \xi(B) \end{array} \right. \quad (1.32)$$

Trong đó 2^M là tất cả các tập hợp được tạo thành từ các điều kiện trong M . Các ràng buộc $\xi(\emptyset) = 0, \xi(M) = 1$ là điều kiện giới hạn của hàm dung lượng và ràng buộc

$\forall A, B \subset 2^M \text{ \& } A \subseteq B \Rightarrow \xi(A) \leq \xi(B)$ thể hiện tính đơn điệu của hàm.

Hàm dung lượng có tính chất “cộng tính” nếu thỏa mãn điều kiện $\forall A, B \subset 2^M, A \cap B = \emptyset \Rightarrow \xi(A \cup B) = \xi(A) + \xi(B)$, ngược lại hàm dung lượng được gọi là hàm “phi cộng tính”.

1.4.2. Tích phân Choquet

Với các hàm dung lượng được định nghĩa trong phần 1.4.1, ta có định nghĩa tích phân Choquet rời rạc cho một véc tơ thể hiện đánh giá theo $m = |M|$ tiêu chí như sau.

Định nghĩa 1.2 [41]: Cho một vector $r \in R^m$, $r = \{r_1, \dots, r_m\}$ thể hiện đánh giá theo tập tiêu chí, ta có tích phân Choquet của r theo hàm dung lượng ξ được định nghĩa như sau:

$$CQ_\xi(r) = \sum_{i=1}^m (r_i^\uparrow - r_{i-1}^\uparrow) \xi(Y_i^\uparrow) \quad (1.33)$$

Trong đó $r^\uparrow = \{r_1^\uparrow, \dots, r_m^\uparrow\}$ là một hoán vị theo thứ tự tăng dần của r , trong đó $0 = r_0^\uparrow \leq r_1^\uparrow \leq \dots \leq r_m^\uparrow$, và tập hợp $Y_i^\uparrow$ là tập con các điều kiện trong M tương ứng với hoán

vị trên, trong đó $Y_i^\uparrow = \{c_j \in M, r_j \geq r_i^\uparrow\} = \{c_{i^\uparrow}, c_{(i+1)^\uparrow}, \dots, c_{m^\uparrow}\}$ với $i \leq m$ và $Y_{m+1}^\uparrow = \emptyset$; $c_{i^\uparrow}$ là tiêu chí trong M tương ứng với hoán vị $r^\uparrow$.

Ví dụ:

Cho một véc tơ $r = \{r_1, r_2, r_3\}$, trong đó mỗi thành phần r_i thể hiện giá trị đánh giá theo một điều kiện. Các thành phần này có thể đại diện cho giá trị được chấm cho các tiêu chí hoặc điều kiện khác nhau như giá cả, chất lượng, dịch vụ v.v.

Giả sử cho các giá trị $r_1 = 0,7$, $r_2 = 0,9$ và $r_3 = 0,8$; và một độ đo mờ được xây dựng trên hàm dung lượng ξ được xác định trên tập ba điều kiện kể trên như sau: $\xi(\{1\}) = 0,3$, $\xi(\{2\}) = 0,4$, $\xi(\{3\}) = 0,3$ lần lượt thể hiện giá trị hàm dung lượng cho mình điều kiện thứ nhất, thứ hai, thứ 3 đứng độc lập. Và $\xi(\{1,2\}) = 0,7$, $\xi(\{1,3\}) = 0,6$, $\xi(\{2,3\}) = 0,7$, $\xi(\{1,2,3\}) = 1$ lần lượt thể hiện giá trị hàm dung lượng cho các tổ hợp kết hợp giữa các điều kiện với nhau.

Để tích tích phân Choquet cho véc tơ kể trên, ta cần sắp xếp thứ tự các thành phần theo thứ tự tăng dần: $r_1 = 0,7$, $r_3 = 0,8$ và $r_2 = 0,9$.

Khi đó:

$$CQ_\xi(r) = (0,7 - 0) * \xi(\{1,3,2\}) + (0,8 - 0,7) * \xi(\{3,2\}) + (0,9 - 0,8) * \xi(\{2\})$$

$$CQ_\xi(r) = 0,7 * 1 + 0,1 * 0,7 + 0,1 * 0,4 = 0,81$$

Khi ứng dụng trong HTVN, giá trị đánh giá của nhóm g được tổng hợp từ đánh giá của các thành viên bằng tích phân Choquet. Khi đó, véc tơ $r \in \mathbb{R}^{|g|}$, là tập giá trị thể hiện đánh giá của từng thành viên với một sản phẩm, dịch vụ i nào đó, $r = \{r_{u_i}, \dots, r_{u_{|g|}}\}$. Hàm dung lượng được định nghĩa trên tập người dùng của g . Và cụ thể chỉ cần xác định các giá trị hàm dung lượng một số tập con người dùng trong g .

Với pha đồng thuận sử dụng một phép toán tổng hợp các đánh giá dựa trên tích phân Choquet, nếu hàm dung lượng được sử dụng là một hàm phi cộng tính thì phép tổng hợp này thể hiện một độ đo mờ phản ánh mục tiêu của phép toán [41], [64].

1.5. Kết luận

Chương 1 đã giới thiệu khái quát một số kiến thức nền tảng về HTVN dựa trên cách thức mở rộng hệ tư vấn đơn người dùng. Dễ thấy rằng khuyến nghị cho một nhóm người dùng bao quát tiếp cận khuyến nghị đơn người dùng. Hơn nữa cách tiếp cận này dễ dàng thể hiện sự kế thừa các thành tựu (ứng dụng tiếp cận tính toán mờ, tiếp cận động) trong các nghiên cứu của lĩnh vực HTV nói chung vào nghiên cứu trong HTVN nói riêng.

Chương 1 cũng trình bày tổng quan nghiên cứu về HTVN một cách ngắn gọn có hệ thống, bao gồm các nghiên cứu về các chiến lược tiếp cận, phương thức đánh giá, các nghiên cứu về độ công bằng của khuyến nghị, các nghiên cứu sử dụng thông tin dạng tĩnh, thông tin dạng động và các nghiên cứu sử dụng tiếp cận tính toán mờ. Các nội dung này được trình bày và phân tích theo các ưu nhược điểm của mỗi phương pháp, mỗi đề xuất, trên cơ sở đó đưa ra các vấn đề cần được nghiên cứu và là mục tiêu của luận án.

Cụ thể luận án nghiên cứu về HTVN và đề xuất các thuật toán mới, bao gồm:

- 1) Phát triển thuật toán HTVN sử dụng độ đo mờ nhằm cải thiện độ công bằng của khuyến nghị. Trong đó pha đồng thuận sử dụng tích phân Choquet và hàm dung lượng phi cộng tính phù hợp.
- 2) Phát triển thuật toán HTVN động dựa trên tiếp cận tính toán mờ trực cảm, trong đó xem xét ảnh hưởng của thời gian và bối cảnh nhóm tới sự thay đổi các thông tin trong hệ tư vấn đồng thời hướng tới tính công bằng của khuyến nghị.

Chương 1 cũng trình bày một số cơ sở lý thuyết về tập mờ trực cảm và các phép toán trên tập mờ trực cảm, lý thuyết về tích phân Choquet rời rạc. Lý thuyết về

tích phân Choquet và tập mờ trực cảm là nền tảng để NCS phát triển các mô hình HTVN được trình bày trong các chương tiếp theo.

Tổng quan nghiên cứu về hệ tư vấn, hệ tư vấn nhóm được công bố một phần trong một bài hội thảo quốc tế, công trình thứ 2 trong danh mục các công trình của tác giả. Bên cạnh đó, nội dung và kết quả nghiên cứu chi tiết của các mô hình đề xuất sẽ được trình trong Chương 2 và Chương 3 của luận án.

Chương 2. TĂNG CƯỜNG ĐỘ CÔNG BẰNG KHUYẾN NGHỊ CỦA HỆ TƯ VẤN NHÓM VỚI ĐỘ ĐO MỜ

2.1. Mở đầu

Trong hệ tư vấn nhóm theo tiếp cận hợp khuyến nghị, pha đồng thuận của mô hình cần sử dụng một phương pháp nào đó để tổng hợp khuyến nghị của từng thành viên trong nhóm thành khuyến nghị chung cho nhóm. Có nhiều chiến lược khác nhau để thực hiện pha đồng thuận [3], [11], [65], và những chiến lược này thể hiện các mục tiêu khác nhau. Nhưng dễ thấy, các chiến lược cơ sở trong các nghiên cứu đã công bố đều không đề cập tới vấn đề độ công bằng giữa các thành viên trong nhóm, kể cả chiến lược có tên “chiến lược công bằng”. Chiến lược này chỉ sử dụng tiếp cận trong đó mỗi thành viên trong nhóm lần lượt được ưu tiên đưa ra lựa chọn của mình. Trong khi đó, như đã phân tích tổng quan nghiên cứu trong Chương 1, vấn đề đảm bảo tính công bằng giữa các thành viên trong nhóm là một yếu tố ngày càng quan trọng để đảm bảo tính hiệu quả của HTVN.

Tuy nhiên, chúng ta cũng không thể sử dụng tiếp cận trong đó chỉ chú ý tới tính công bằng. Nguyên nhân là nếu chỉ tập trung vào tính công bằng có thể dẫn đến đưa ra khuyến nghị mà mọi thành viên nhóm đều “ghét” như nhau. Khi đó sự công bằng đạt tối đa, nhưng lợi ích cho nhóm người dùng hay tổng đánh giá của các thành viên lại rất thấp. Vậy một HTVN quan tâm tới tính công bằng trong khuyến nghị cần đảm bảo lựa chọn các sản phẩm để đưa vào khuyến nghị, với hai tiêu chí là tối đa hóa tổng đánh giá của các thành viên cho sản phẩm, và tối thiểu hóa sự khác biệt giữa đánh giá của các thành viên. Bài toán này trở thành bài toán tối ưu đa mục tiêu.

Có thể thấy, cùng với sự phát triển của các mô hình HTVN với chiến lược hợp khuyến nghị, vấn đề nâng cao tính công bằng đã trở thành một trong các trọng tâm nghiên cứu. Tuy nhiên, để giải quyết vấn đề này, thách thức lớn đầu tiên gặp phải đó là khái niệm công bằng có tính chất xã hội, nói cách khác nó được hiểu khác nhau với các cộng đồng khác nhau và rất khó đo lường [66]. Trong các nghiên cứu gần đây, một số định nghĩa định độ công bằng của khuyến nghị trong HTVN được đưa ra và các công thức được đề xuất để đo lường và tính toán [16]–[18]. Trong đó, một số

nghiên cứu coi độ công bằng của khuyến nghị trong HTVN là tỷ lệ số thành viên hài lòng với các sản phẩm, dịch vụ trong khuyến nghị so với kích thước nhóm [16]. Các nghiên cứu khác sử dụng độ lệch giữa độ hài lòng của các thành viên trong nhóm là độ đo công bằng [17]. Trong các nghiên cứu này, độ đo công bằng cho từng sản phẩm, dịch vụ trong danh sách khuyến nghị là độc lập. Trong nghiên cứu khác, độ công bằng của khuyến nghị của HTVN xem xét danh sách sản phẩm, dịch vụ trong khuyến nghị như là một “gói” duy nhất thay vì là tập hợp rời rạc các sản phẩm dịch vụ [18]. Tổng quan nghiên cứu cho thấy có nhiều cách tiếp cận về độ đo công bằng khác nhau, và khó có được một khái niệm duy nhất được đồng thuận để sử dụng đại trà. Thay vào đó, dựa vào từng miền ứng dụng, khái niệm và thiết kế độ đo về độ công bằng sẽ được lựa chọn và áp dụng phù hợp với mục tiêu ứng dụng.

Với các nghiên cứu đã công bố, từ các độ đo công bằng được đề xuất các tác giả đã phát triển các phép toán kết hợp giữa độ công bằng và tính hữu ích của sản phẩm với nhóm người dùng, nhằm tạo ra một độ đo thể hiện đánh giá của nhóm đối với sản phẩm. Từ đó, HTVN có thể tìm kiếm các sản phẩm có giá trị độ đo cao nhất. Cách tiếp cận này có ưu điểm là độ phức tạp tính toán không cao và việc kết hợp các mục tiêu trong HTVN là rõ ràng. Tuy nhiên, các nghiên cứu này cũng gặp phải một số hạn chế nhất định như sau.

Thứ nhất, việc xây dựng hàm kết hợp giữa hai mục tiêu độ công bằng của khuyến nghị và tính hữu ích của sản phẩm đối với một nhóm được xây dựng cố định cho mọi người dùng và mọi nhóm. Trong thực tế, mỗi người dùng và nhóm người dùng có thể có mức độ ưu tiên khác nhau giữa hai mục tiêu này.

Thứ hai, một thách thức khác gặp phải khi xây dựng giải thuật tối ưu về độ công bằng trong HTVN dựa trên pha “đồng thuận” đó là cảm nhận hay đánh giá của mỗi thành viên khi sử dụng một sản phẩm dịch vụ sẽ chịu sự tác động từ các thành viên khác trong nhóm [3], [67], [68]. Do đó, khi tính toán sự chênh lệch giữa đánh giá của các thành viên trong nhóm cần ước lượng được ảnh hưởng của tương tác nhóm. Hiện nay, các tiếp cận phổ biến cho phép toán tổng hợp khuyến nghị trong HTVN không

giải quyết được vấn đề này. Cụ thể là, khi xử lý các thuật toán đề xuất thường giả thiết rằng đánh giá của một người dùng với sản phẩm là không đổi khi người dùng đó ở trong các nhóm khác nhau [2], [17]. Từ đó có thể rút ra kết luận rằng các chiến lược đồng thuận truyền thống còn có nhiều hạn chế.

Ngoài ra, mặc dù tổng quan nghiên cứu cho thấy có những nghiên cứu về ảnh hưởng của tương tác trong nhóm [13], [69] sử dụng dữ liệu thu thập trên mạng xã hội, và nghiên cứu [10], [32] sử dụng dữ liệu nội bộ, nhưng những nghiên cứu này lại không hướng tới bài toán độ công bằng.

Trong luận án này, độ công bằng của khuyến nghị trong HTVN được xác định trên độ lệch của đánh giá giữa các thành viên trong nhóm với tập sản phẩm được khuyến nghị. Từ ưu điểm, cũng như hạn chế của các nghiên cứu đã được công bố, trong Chương 2 của luận án sẽ trình bày đề xuất về một phương pháp tiếp cận linh hoạt hơn để giải quyết vấn đề nâng cao độ công bằng khuyến nghị trong HTVN. Cụ thể, mục tiêu nghiên cứu là phát triển mô hình mới cho pha đồng thuận để đảm bảo cân bằng hai mục tiêu, tối đa tổng đánh giá của nhóm và đảm bảo tối ưu mức độ công bằng. Trong đó, thay thế cho các chiến lược đồng thuận đã có trước đó, luận án đề xuất sử dụng tích phân Choquet để tạo đề xuất cho nhóm. Phép toán đồng thuận sử dụng tích phân Choquet mở rộng phạm vi tìm kiếm giải pháp so với chiến lược đồng thuận sử dụng phép toán tổng có trọng số. Do đó pha đồng thuận với tích phân Choquet có thể đưa ra khuyến nghị cân bằng hơn so với các chiến lược trước đó nhờ xây dựng một độ đo mờ phù hợp [41], [70].

Khi ứng dụng tích phân Choquet trong pha đồng thuận để nâng cao tính công bằng của khuyến nghị, nhiệm vụ xây dựng một hàm dung lượng phù hợp với mục tiêu cụ thể của mỗi bài toán là rất quan trọng. Cho tới hiện nay chưa có một thuật toán nào khả thi cả về mặt lý thuyết và thực hành (đảm bảo hiệu năng) để giải quyết được trọn vẹn bài toán này [71]. Trong các tiếp cận nổi bật, một số thuật toán chính xác và thuật toán gần đúng đã được giới thiệu cho vấn đề này trong bài toán hỗ trợ ra quyết định đa tiêu chí [41], [71]–[73]. Tuy nhiên, các tiếp cận trên không phù hợp với bài toán xây dựng phép toán tổng hợp đánh giá người dùng sử dụng tích phân

Choquet cho HTVN do thời gian tính quá lớn, trong khi yêu cầu tạo ra tư vấn cần thực hiện nhiều và thường xuyên.

Luận án này sử dụng tiếp cận khác, đó là xây dựng hàm dung lượng dựa trên một phương thức ước lượng mức độ hoạt động của các thành viên trong nhóm và các nhóm con của nó. Độ đo mờ này đồng thời phần nào thể hiện sự tương tác giữa các thành viên trong nhóm đo hàm dung lượng được xây dựng theo mức độ tích cực của mỗi người dùng trong hệ thống. Trong đó giả thiết rằng người dùng tích cực thường sẽ có ảnh hưởng tích cực với những người cùng tham gia.

Các kết quả trong chương này được công bố trong một bài báo thuộc danh mục SCIE, là công trình số 1 được liệt kê trong phần “*Danh mục các công trình của tác giả*”.

2.2. Đề xuất hệ tư vấn nhóm nâng cao tính công bằng với độ đo mờ

2.2.1. Phép toán tổng hợp có trọng số và phép toán tổng hợp dựa trên tích phân Choquet

Chiến lược cộng là một phương pháp phổ biến trong HTVN với pha đồng thuận nhằm tối đa hóa lợi ích của nhóm. Tuy nhiên, hạn chế của phương pháp này là không xem xét đến sự khác biệt giữa đánh giá của các thành viên trong nhóm. Do đó, một người dùng đánh giá rất cao một sản phẩm sẽ ảnh hưởng lớn tới kết quả tư vấn trong cả nhóm, và kết quả là các lựa chọn từ HTVN có thể rất mất công bằng.

Trong các hệ thống sử dụng HTVN, việc nâng cao tính công bằng trong khuyến nghị nhóm sẽ góp phần tăng cường mức độ thỏa mãn của người dùng đối với hệ thống. Các nghiên cứu được công bố đã chỉ ra rằng sử dụng toán tử tích phân Choquet để xây dựng phép toán tổng hợp các đánh giá có thể giải quyết bài toán này [41], [74]. Sau đây là ví dụ cho thấy tiềm năng để cải thiện HTVN của cách tiếp cận này.

Ví dụ minh họa: Bảng 2.1 thể hiện dữ liệu về một tập hợp gồm ba sản phẩm, dịch vụ được đánh giá bởi ba người dùng khác nhau. Thang đo đánh giá thuộc khoảng [1,10], trong đó giá trị lớn hơn thể hiện người dùng yêu thích sản phẩm tương ứng hơn sản phẩm được đánh giá bởi giá trị nhỏ. Trong pha đồng thuận, để lựa chọn sản

phẩm phù hợp với nhóm người dùng này, chúng ta cần tổng hợp đánh giá của các thành viên cho một sản phẩm theo một chiến lược nào đó. Dễ thấy, khi sử dụng chiến lược cộng ta có được kết quả như bảng 2.1 dưới đây.

Pha đồng thuận với chiến lược cộng:

Bảng 2.1 Phép toán tổng hợp đánh giá bằng chiến lược cộng

Sản phẩm, dịch vụ			
Người dùng	i_1	i_2	i_3
u_1	10	4	3
u_2	1	9	8
u_3	10	5	2
Đánh giá của nhóm	21	18	13

Xếp hạng các sản phẩm, dịch vụ được đánh giá của nhóm là: $i_1 > i_2 > i_3$

Khi đó HTVN đưa ra khuyến nghị theo thứ tự xếp hạng là $i_1 > i_2 > i_3$.

Pha đồng thuận sử dụng tích phân Choquet:

Khi xem xét sử dụng phép toán tổng hợp đánh giá dựa trên tích phân Choquet cho pha đồng thuận sẽ cho thấy sự khác biệt. Theo các nghiên cứu [41], [63], [74] phép toán tổng hợp đánh giá bằng tích phân Choquet sẽ cho kết quả cân bằng hơn, và hàm dung lượng phù hợp sẽ nằm trong lân cận của hàm dung lượng với các giá trị được sinh ngẫu nhiên theo phân bố chuẩn.

Giả sử ta có ξ là hàm dung lượng được xác định như sau:

$$\xi(\{u_1\}) = \xi(\{u_2\}) = \xi(\{u_3\}) = 0.2; \quad \xi(\{u_1, u_2\}) = \xi(\{u_2, u_3\}) = \xi(\{u_1, u_3\}) = 0.4; \quad \text{và} \\ \xi(\{u_1, u_2, u_3\}) = 1.$$

Khi ứng dụng tích phân Choquet, đánh giá của một nhóm người dùng với một sản phẩm I_i , ký hiệu là $\text{pref}_{I_i}^{CQ}$, được tính theo công thức 1.33. Cụ thể như sau:

$$pref_{I_1}^{CQ} = (1-0)\xi(\{u_1, u_2, u_3\}) + (10-1)\xi(\{u_1, u_3\}) + (10-10)\xi(\{u_3\}) = 1 + 9*0.4 = 4.6$$

$$pref_{I_2}^{CQ} = (4-0)\xi(\{u_1, u_2, u_3\}) + (5-4)\xi(\{u_2, u_3\}) + (9-5)\xi(\{u_2\}) = 4 + 1*0.4 + 4*0.2 = 5.2$$

$$pref_{I_3}^{CQ} = (2-0)\xi(\{u_1, u_2, u_3\}) + (3-2)\xi(\{u_1, u_2\}) + (8-3)\xi(\{u_2\}) = 2 + 1*0.4 + 5*0.2 = 3.4$$

Để thấy kết quả đánh giá của nhóm cho các sản phẩm được trình bày trong bảng 2.2.

Bảng 2.2 Phép toán tổng hợp với tích phân Choquet

Sản phẩm, dịch vụ			
Người dùng	i₁	i₂	i₃
u ₁	10	4	3
u ₂	1	9	8
u ₃	10	5	2
Đánh giá của nhóm	4.6	5.2	3.2

Xếp hạng các sản phẩm, dịch vụ được đánh giá của nhóm là: $i_2 > i_1 > i_3$

Khi đó HTVN đưa ra khuyến nghị theo thứ tự xếp hạng là $i_2 > i_1 > i_3$, khác với tiếp cận của chiến lược cộng.

Trong cách tiếp cận thứ hai, i_2 là sản phẩm, dịch vụ có sự chênh lệch ít nhất giữa đánh giá của các thành viên và là sự lựa chọn công bằng nhất. Đồng thời, tổng đánh giá của các thành viên của nhóm cho sản phẩm này cũng thuộc nhóm cao nhất. Ta thấy, i_2 là sản phẩm, dịch vụ được khuyến nghị đầu tiên với tiếp cận sử dụng tích phân Choquet, trong khi i_2 không phải là lựa chọn đầu tiên với chiến lược cộng.

Ví dụ trên tuy cho thấy một tiềm năng của pha đồng thuận sử dụng phép toán tổng hợp đánh giá của các thành viên dựa trên tích phân Choquet nhằm cải thiện độ công bằng của khuyến nghị, nhưng để có thể phát triển và ứng dụng được phép toán này cần có nghiên cứu đầy đủ hơn về mặt lý thuyết, phát triển giải thuật cho hàm dung lượng phù hợp và thực nghiệm. Các nội dung sau đây trong Chương 2 sẽ trình bày đầy đủ về vấn đề này.

2.2.2. Vấn đề tăng cường sự công bằng của hệ tư vấn nhóm dựa trên độ đo mờ.

Dựa trên tiềm năng của phép toán tổng hợp đánh giá người dùng sử dụng tích phân Choquet cho pha đồng thuận nhằm nâng cao tính công bằng khuyến nghị ở trên. Luận án này đề xuất mô hình mới cho HTVN nhằm cải thiện độ công bằng của khuyến nghị, đồng thời bảo toàn mục tiêu tối đa hóa tổng đánh giá chung của nhóm người dùng. Phép toán tổng hợp đánh giá dựa trên tích phân Choquet cho phép tìm kiếm và lựa chọn sản phẩm phù hợp để tạo ra một khuyến nghị cân bằng hơn cho nhóm [41], [70]. Thật vậy, khi ứng dụng tích phân Choquet và sử dụng hàm dung lượng phi cộng tính, phép toán tổng hợp đánh giá bằng tích phân Choquet sẽ thể hiện một độ đo mờ cho phép tìm ra lựa chọn cân bằng nhất giữa việc tối đa hóa tổng đánh giá của các thành viên, và tối thiểu hóa độ lệch giữa các đánh giá đó. Trong trường hợp thiết kế một hàm dung lượng có tính chất cộng tính, thì phép toán tổng hợp đánh giá bằng tích phân Choquet sẽ suy biến thành phép toán hợp có trọng số [41].

Rõ ràng, để xây dựng phép toán tổng hợp đánh giá bằng tích phân Choquet phù hợp, việc xây dựng một độ đo mờ dựa trên một hàm dung lượng phi cộng tính xác định trên dữ liệu thành viên trong nhóm là quan trọng nhất. Hàm dung lượng thể hiện một thước đo mờ cho hàm tối ưu mục tiêu, nhưng hiện nay không có thuật toán nào được đề xuất có thể giải quyết bài toán này một cách triệt để [71]. Theo nghiên cứu [71] xây dựng một độ đo mờ dựa trên việc tạo ra một hàm dung lượng phi cộng tính bằng phép sinh ngẫu nhiên theo phân phối đồng nhất có thể được ứng dụng cho vấn đề này. Tuy nhiên, theo định nghĩa của hàm dung lượng, số lượng tiêu chí sẽ quyết định tổng số giá trị phải tính toán, và khi số tiêu chí tăng cao thì số lượng giá trị cần tính sẽ quá lớn và không phù hợp để triển khai ứng dụng (giới hạn chấp nhận được được chỉ ra trong nghiên cứu này là ứng dụng cho bài toán hỗ trợ ra quyết định có tối đa 7 tiêu chí).

Một số phương pháp nổi bật khác đã được đề cập cho bài toán này [75], trong đó vấn đề cần giải quyết được biến đổi thành bài toán sinh các giá trị của hàm dung lượng một cách tự động theo phân bố chuẩn trên biểu đồ Hasse. Đây là bài toán thuộc

lớp NP-khó, nên các giải thuật tìm kiếm vét cạn để xác định lời giải chính xác sẽ không phù hợp trong thực tiễn vì độ phức tạp tính toán quá lớn. Trong lớp bài toán này, các thuật toán xấp xỉ được một số nghiên cứu đề cập tới như các nghiên cứu [71]–[73]. Chiến lược chung của các nghiên cứu này đó là giảm độ phức tạp tính toán cho các phép toán có độ phức tạp lớn nhất, và chấp nhận một kết quả xấp xỉ đúng. Một số thuật toán điển hình theo tiếp cận này được trình bày trong các nghiên cứu [72], [73], [76], tuy nhiên ngay cả các mô hình này vẫn không thể ứng dụng khi số lượng tiêu chí tăng cao [41].

Có thể thấy, các nghiên cứu ở trên đã chỉ ra về lý thuyết có thể sử dụng cơ chế sinh giá trị ngẫu nhiên theo phân bố xác suất để tạo ra các giá trị của độ đo mờ phù hợp. Tuy nhiên tiếp cận này chỉ là một tiếp cận phù hợp về mặt lý thuyết nhưng lại không khả thi trong thực tế do độ phức tạp tính toán lớn. Để sử dụng tích phân Choquet trong phép toán tổng hợp đánh giá ở pha đồng thuận, dữ liệu về sản phẩm, dịch vụ và kích thước nhóm người dùng thường lớn hơn giới hạn có thể sử dụng các tiếp cận đề cập trong nghiên cứu [71]. Bên cạnh đó là các tiếp cận xấp xỉ đã được đề xuất nhằm phát triển độ mờ phù hợp [71]–[73], nhưng các mô hình được đề xuất trong hướng tiếp cận này vẫn chưa phù hợp cho HTVN vì độ phức tạp tính toán vẫn còn quá lớn. Mà thực tế cho thấy, một HTVN cần đưa ra khuyến nghị trong một khoảng thời gian rất ngắn (thông thường thì giới hạn này là vài giây).

Trong HTVN, việc áp dụng phép toán tổng hợp đánh giá của người dùng trong nhóm dựa trên tích phân Choquet yêu cầu xác định giá trị hàm dung lượng cho từng nhóm người dùng. Một cách tiếp cận khả thi về mặt thời gian tính toán là đề xuất một thuật toán ước lượng trực tiếp các giá trị cần thiết của hàm dung lượng có thể xấp xỉ với phép toán sinh ngẫu nhiên theo phân bố xác suất, đồng thời đảm bảo thời gian tính nhỏ, phù hợp yêu cầu của HTVN. Luận án sẽ sử dụng cách tiếp cận này, cụ thể luận án sẽ xây dựng một số hàm dung lượng từ dữ liệu phản ánh hoạt động của người dùng trong hệ thống, đồng thời phân nào phản ánh tương tác giữa người dùng. Hàm dung lượng cũng thỏa mãn tính chất của độ đo mờ với mục tiêu cải thiện độ công bằng của khuyến nghị trong HTVN.

Trong các ứng dụng của HTVN, thông thường thông tin bổ xung bên ngoài về người dùng là không có sẵn để đánh giá ảnh hưởng giữa các thành viên trong nhóm, do đó, lịch sử tương tác giữa người dùng- sản phẩm là thông tin chính dùng để ước lượng tương tác giữa các thành viên. Cách tiếp cận này không tối ưu về mặt thông tin nhưng lại mang lại lợi ích lớn trong triển khai, và dễ ứng dụng. Từ đó, giá trị của các hàm dung lượng được xây dựng dựa trên một trong các giả thiết sau: (1) Giả thiết thứ nhất là tương tác của thành viên trong nhóm là như nhau, (2) Giả thiết thứ hai giả định rằng thành viên tích cực hơn có ảnh hưởng lớn hơn và (3) Giả thiết thứ ba là các thành viên có độ tương đồng càng cao thì ảnh hưởng tới nhau càng lớn.

Các mô hình HTVN sử dụng phép toán tổng hợp đánh giá bằng tích phân Choquet với các hàm dung lượng được đề xuất theo tiếp cận này không tìm kiếm sản phẩm có tổng đánh giá lớn nhất mà tìm kiếm một giải pháp tối ưu cân bằng giữa hai mục tiêu giảm thiểu sự khác biệt giữa các thành viên và tối đa tổng đánh giá của nhóm. Hiệu năng của các mô hình đề xuất được chứng minh bằng lý thuyết và thực nghiệm, các kết quả này được trình bày ở phần tiếp theo.

2.3. Mô hình Hệ tư vấn nhóm với pha đồng thuận sử dụng tích phân Choquet

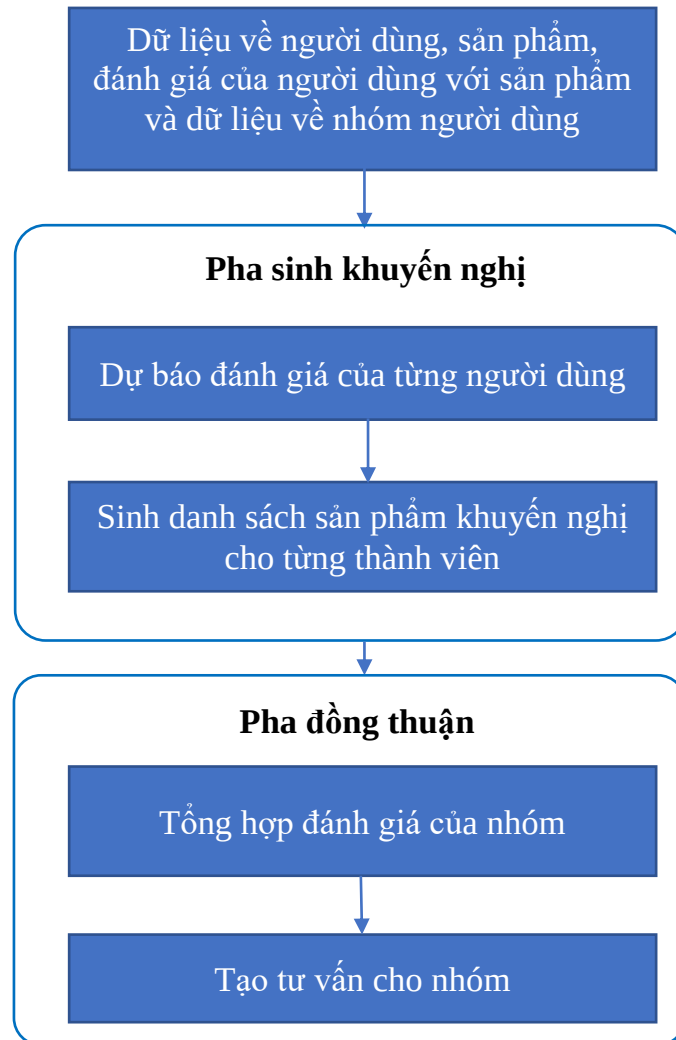
2.3.1. Đề xuất mô hình Hệ tư vấn nhóm với pha đồng thuận dựa trên tích phân Choquet

a) Đề xuất mô hình HTVN dựa trên tích phân Choquet

Trong luận án này, NCS đề xuất mô hình HTVN với hai pha, pha thứ nhất là pha sinh khuyến nghị sẽ dự báo đánh giá của người dùng với các sản phẩm, dịch vụ chưa được đánh giá và pha thứ hai thực hiện cơ chế đồng thuận để tổng hợp đánh giá của các thành viên nhóm. Cụ thể:

- Pha sinh khuyến nghị: Sử dụng phương pháp lọc cộng tác theo người dùng để dự đoán đánh giá cá nhân của những người dùng cho từng sản phẩm, dịch vụ và từ đó tạo ra khuyến nghị cho từng thành viên.
- Pha đồng thuận: Sử dụng phép toán tổng hợp đánh giá của các thành viên trong nhóm dựa trên tích phân Choquet để tính đánh giá của nhóm cho sản phẩm,

dịch vụ dựa trên đánh giá của từng thành viên. Dựa trên kết quả đánh giá của nhóm sẽ đưa ra khuyến nghị cho nhóm theo nguyên tắc lựa chọn N sản phẩm, sản phẩm được nhóm đánh giá cao nhất (top- N).



Hình 2.1 Sơ đồ quy trình hệ tư vấn nhóm theo tiếp cận hợp khuyến nghị

Các phép toán của hai pha trong HTVN đề xuất được trình bày chi tiết dưới đây.

Pha sinh khuyến nghị

Để dự báo đánh giá của một người dùng cho các sản phẩm, dịch vụ chưa được người dùng đó đánh giá, phương pháp lọc cộng tác được lựa chọn sử dụng. Phương pháp lọc cộng tác theo người dùng cơ bản có hai bước : (1) xác định một tập hợp k người hàng xóm gần nhất với người dùng đang xét; (2) sinh ra giá trị dự báo về đánh

giá của người dùng đang xét với một sản phẩm dựa trên đánh giá của các hàng xóm với cùng sản phẩm đó [77]. Cụ thể như sau:

Bước 1: Xác định tập k hàng xóm gần nhất của một người dùng theo độ đo tính tương tự giữa các người dùng. Độ đo tương tự Cosine thường được các nghiên cứu đã công bố sử dụng và cho hiệu quả cao. Trong luận án này NCS cũng lựa chọn sử dụng độ đo này, cụ thể công thức như dưới đây.

$$sim_{cosin}(a, b) = \frac{\sum_{i \in I} r_{ai} r_{bi}}{\sqrt{\sum_{i \in I} r_{ai}^2} \sqrt{\sum_{i \in I} r_{bi}^2}} \quad (2.1)$$

Trong đó, I là tập các sản phẩm, dịch vụ được cả hai người dùng a, b cùng đánh giá.

Bước 2: Sau khi xác định tập hàng xóm, đánh giá của một người dùng a cho sản phẩm j chưa có đánh giá trước đó sẽ được dự đoán. Giả thiết rằng, đánh giá của một người dùng với một sản phẩm phụ thuộc vào bản thân người dùng đó và có sự tương đồng với đánh giá của các hàng xóm. Đánh giá của một người dùng cho sản phẩm được tính bởi công thức sau.

$$r_{aj} = \bar{r}_a + \alpha \sum_{u \in \hat{U}} sim_{cosin}(a, u) (r_{uj} - \bar{r}_u) \quad (2.2)$$

Trong đó $\hat{U}$ là tập các người dùng được lựa chọn từ bước 1 của người dùng a , bao gồm k người dùng có độ tương tự cao nhất với a . α là hệ số hiệu chỉnh và được

tính bởi công thức $\alpha = \frac{1}{\sum_{u \in \hat{U}} sim_{cosin}(a, u)}$. Khi mở rộng nghiên cứu, hệ số hiệu chỉnh có

thể được thay đổi để thể hiện sự ưu tiên giữa ảnh hưởng của người dùng trong tập hàng xóm tới dự đoán và đánh giá trung bình của người dùng với các sản phẩm khác tới kết quả dự báo.

Sau khi dự đoán đánh giá của mỗi người dùng cho các sản phẩm mà người dùng đó chưa đánh giá, việc chọn ra n sản phẩm được đánh giá cao nhất cho cá nhân đó

được thực hiện bởi một thuật toán sắp xếp rất đơn giản. Thuật toán này có thể được cài đặt bằng một thuật toán sắp xếp và lọc như thuật toán sắp xếp chọn.

Pha đồng thuận

Pha thứ hai của mô hình là pha đồng thuận. Pha này sử dụng phép toán tổng hợp đánh giá của các thành viên trong nhóm với một sản phẩm được tính bằng tích phân Choquet, chi tiết công thức tính được trình bày tại nội dung 1.4. Trong đó hàm dung lượng được sử dụng là một trong hai hàm dung lượng được xây dựng như sau.

b) Các hàm dung lượng được đề xuất

Cho nhóm người dùng $\mathcal{G}$, để xác định n sản phẩm, dịch vụ phù hợp nhất (top- n) trong tập hợp sản phẩm I theo phép toán tổng hợp đánh giá của các thành viên trong nhóm dựa trên tích phân Choquet, một hàm dung lượng phi cộng tính trên tập $|\mathcal{G}|$ người dùng cần được định nghĩa. Cụ thể là xác định một hàm dung lượng tối ưu. Việc xác định một hàm dung lượng tối ưu là bài toán NP-khó [41], tuy nhiên trong HTVN, pha đồng thuận chỉ yêu cầu tính giá trị của hàm dung lượng trên một phần không gian đầy đủ. Cụ thể là ước lượng giá trị hàm dung lượng trên $\mathcal{G}$ tương ứng với tập hợp các đánh giá của các thành viên được sắp xếp theo thứ tự tăng dần. Điều đó có nghĩa là chỉ cần xác định $|\mathcal{G}|$ giá trị của hàm dung lượng.

Dưới đây, NCS đề xuất một hàm dung lượng thứ nhất dùng thông tin về đánh giá đã có của người dùng trong hệ thống, và thể hiện mức độ tích cực trong tương tác của người dùng đó, kế thừa nghiên cứu [78]. Hàm dung lượng được đề xuất tuân theo giả thiết cơ bản đó là người dùng càng tích cực thì giá trị của hàm dung lượng càng lớn và ảnh hưởng của người đó tới các thành viên khác trong nhóm càng cao. Ở đây, sự tích cực thể hiện thông qua hai yếu tố, thứ nhất là số lượng sản phẩm mà người dùng đã đánh giá (người dùng đánh giá càng nhiều sản phẩm thì người dùng đó tham gia vào hệ thống khuyến nghị càng tích cực), và thứ hai là mức độ đánh giá (người dùng có xu hướng đánh cao các sản phẩm thì thể hiện tính tích cực cao hơn). Hàm dung lượng được đề xuất cũng đảm bảo khi hai người dùng có mức độ tích cực như

sau thì các giá trị đo của hàm dung lượng với hai người dùng đó cũng bằng nhau. Cụ thể công thức tính trong hàm dung lượng được đề xuất như sau.

Hàm dung lượng thứ nhất:

Giả sử $g = \{u_1, \dots, u_n\}$ là một nhóm người dùng. Hàm dung lượng được xác định trên tập người dùng này như sau:

- 1) Giá trị hàm dung lượng với tập rỗng: $\xi(\emptyset) = 0$
- 2) Với mỗi người dùng trong nhóm:

$$\xi(u_i) = \begin{cases} \frac{\omega(u_i)}{\sum_{u_i \in g} \omega(u_i)}, & \text{if } \sum_{u_i \in g} \omega(u_i) < 1 \\ \omega(u_i), & \text{otherwise} \end{cases} \quad (2.3)$$

Trong đó:

$$\omega(u_i) = \frac{\text{count}(r_{u_i j} \geq 0)}{|I|}, j = 1, \dots, |I| \quad (2.4)$$

- 3) và giá trị hàm dung lượng của một tập con $A \subset g$ là:

$$\xi(A) = \sum_{u_i \in A} \xi(u_i) + \sigma(A) \quad (2.5)$$

Nếu $\xi(A) > 1$ nó được gán bằng 1, và $\xi(g) = 1$.

Trong tiếp cận của hàm dung lượng thứ nhất ta có $\sigma(A)$ được xác định bởi công thức sau.

$$\sigma(A) = \sum_{u_i \in A} \omega'(u_i) \quad (2.6)$$

Trong đó:

$$\omega'(u_i) = \frac{\text{count}(r_{u_i j} > \text{aveRate}) - \text{count}(r_{u_i j} < \text{aveRate})}{|I|} \quad (2.7)$$

Trong đó, $aveRate$ là đánh giá trung bình của tất cả người dùng với các sản phẩm trong I , và I bao gồm tất cả các sản phẩm đã được người dùng u_i đánh giá. $\omega'(u_i)$ thể hiện mức độ tích cực của một người dùng so với những người dùng khác trong đánh giá các sản phẩm.

Dễ thấy hàm dung lượng được định nghĩa như trên thỏa mãn điều kiện của hàm dung lượng theo định nghĩa trong phần 1.4.1.

Chứng minh:

- Trong định nghĩa kể trên, điều kiện về giới hạn của hàm dung lượng là $\xi(\emptyset) = 0$ và $\xi(g) = \xi(u_1, \dots, u_n) = 1$ được quy định trực tiếp trong định nghĩa về hàm dung lượng cho nhóm người dùng.
- Bên cạnh đó, tính chất đơn điệu tăng của hàm dung lượng $\forall A, B \subset g \text{ \& } A \subseteq B \Rightarrow \xi(A) \leq \xi(B)$ cũng được thỏa mãn.

Thật vậy:

+ Dễ thấy, từ công thức (2.3) và (2.4) ta có $\xi(u_i) \geq 0$.

+ Trong trường hợp $\forall A, B \subset g, A \subseteq B$, và $B = A \cup u_x, u_x \subset g$, từ công thức (2.5) và (2.6) ta có thể thấy $\xi(B) = \xi(A) + \xi(u_x) + \omega'(u_x)$; trong đó theo công thức (2.3) và (2.4) ta có:

$$\xi(u_x) + \omega'(u_x) \geq \omega(u_x) + \omega'(u_x) = \frac{\text{count}(r_{u_xj} \geq 0)}{|I|} + \frac{\text{count}(r_{u_xj} > aveRate) - \text{count}(r_{u_xj} < aveRate)}{|I|} \geq 0$$

Vì $\xi(u_x) + \omega'(u_x) \geq 0$, do đó $\xi(B) \geq \xi(A)$.

Từ đó, trong trường hợp tổng quát $\forall A, B \subset g, A \subseteq B$ ta có thể giả thiết $B = A \cup \{u_1, \dots, u_k\}, u_1, \dots, u_k \subset g$. Tương tự như trên ta có thể chứng minh $\xi(B) \geq \xi(A \cup \{u_1, \dots, u_{k-1}\}) \geq \dots \geq \xi(A \cup \{u_1\}) \geq \xi(A)$. Vậy $\xi(B) \geq \xi(A)$.

Bên cạnh đó, hàm dung lượng thứ nhất có tính chất như sau.

Tính chất 2.1: Hàm dung lượng được định nghĩa ở trên thỏa mãn tính chất cộng tính khi $\sigma(A) = 0$ và $\sum_{u_j \in g} \omega(u_j) \leq 1$.

Chứng minh:

Với một nhóm g , giá trị $\omega(u_i) = \frac{\text{count}(r_{u_{ij}} \geq 0)}{|I|} \in [0,1]$, và $\sum_{u_j \in g} \omega(u_j)$ là không đổi. Trong hệ thống thực tế, số lượng sản phẩm, dịch vụ đã được đánh giá của một người dùng thường rất nhỏ nên giá trị $\omega(u_i)$ thường khá nhỏ.

Khi $\sigma(A) = 0$ và $\sum_{u_j \in g} \omega(u_j) \leq 1$, với A, B thỏa mãn $A \cap B = \emptyset$ ta có:

$$\begin{aligned} \xi(A \cup B) &= \sum_{u_i \in A \cup B} \xi(u_i) = \frac{\sum_{u_i \in A \cup B} \omega(u_i)}{\sum_{u_j \in g} \omega(u_j)} = \frac{\sum_{u_i \in A} \omega(u_i) + \sum_{u_i \in B} \omega(u_i)}{\sum_{u_j \in g} \omega(u_j)} \\ &= \xi(A) + \xi(B) \end{aligned}$$

Do đó hàm dung lượng được định nghĩa ở trên có tính chất cộng tính.

Định lý 2.1: Trong trường hợp hàm dung lượng được định nghĩa ở trên có tính chất cộng tính thì phép toán tổng hợp đánh giá của các thành viên trong nhóm dựa trên tích phân Choquet trở thành một phép toán tổng hợp có trọng số.

Chứng minh: Không mất tính tổng quát, giả sử rằng vector $r \in \mathbb{R}^m$, $r = \{r_{u_{1i}}, \dots, r_{u_{mi}}\}$ thỏa mãn $r_{u_{1i}} \leq \dots \leq r_{u_{mi}}$. Khi đó, tích phân Choquet của vector $r \in \mathbb{R}^m$ dựa trên hàm dung lượng ξ được tính bằng công thức sau:

$$\text{pref}_i^{\text{CQ}} = \sum_{j=1}^m (r_{u_{ji}} - r_{u_{j-1}i}) \xi(u_1, \dots, u_j)$$

Trong đó $r_{u_{0i}} = 0$.

Vì hàm dung lượng ξ có tính chất cộng tính nên $\forall A, B \subset g, A \cap B = \emptyset$, ta có

$$\xi(A, B) = \xi(A) + \xi(B). \text{ Do đó, } \xi(\{u_j, u_{j+1}, \dots, u_m\}) = \xi(u_j) + \xi(\{u_{j+1}, \dots, u_m\}) = \sum_{l=j}^m \xi(u_l).$$

Từ đó ta có:

$$\text{pref}_i^{\text{CQ}} = \sum_{j=1}^m (r_{u_{ji}} - r_{u_{j-1}i}) \sum_{l=j}^m \xi(u_l)$$

$\Rightarrow$

$$pref_i^{CQ} = \sum_{j=1}^m r_{u_{ji}} \cdot \xi(u_j)$$

Ta thấy, trong trường hợp này $\xi(u_i)$ chính là trọng số của một phép toán tổng hợp đánh giá.

Từ công thức (2.6) và (2.7) ta thấy điều kiện đề $\delta(A) = 0, \forall A \subset g$ rất khó xảy ra. Khi điều kiện của tính chất 2.1 không được thỏa mãn, hàm dung lượng được đề xuất là một hàm dung lượng có tính chất phi cộng tính. Và phép toán tổng hợp các đánh giá của người dùng trong nhóm dựa trên tích phân Choquet và hàm dung lượng ở trên sẽ thể hiện một độ đo mờ.

Hàm dung lượng thứ hai:

Trong HTVN, sự tương tác giữa các thành viên trong nhóm sẽ tác động đến mức độ thỏa mãn của từng cá nhân với một sản phẩm, dịch vụ. Với cùng một loại hoạt động, một người có thể thấy thích thú nhiều hơn nếu được thực hiện với những người cùng tính cách, sở thích và ít thích thú hơn khi phải cùng thực hiện với người có sự khác biệt lớn. Do hàm dung lượng thứ nhất được đề xuất ở trên chưa tính đến khía cạnh này, một hàm dung lượng thứ hai, mở rộng từ hàm dung lượng thứ nhất, đã được bổ sung một phép đo thể hiện sự tương đồng giữa các thành viên trong nhóm sẽ giải quyết được vấn đề này.

Tuy nhiên, do yêu cầu tính toán của phương pháp này lớn hơn nhiều so với hàm dung lượng thứ nhất nên để đảm bảo nghiên cứu và đề xuất ứng dụng thực tế, NCS sử dụng cả hai đề xuất về hàm dung lượng trong luận án. Khi đánh giá về hiệu quả của các hàm dung lượng được đề xuất, cả hai yếu tố về độ phức tạp tính toán và hiệu quả trong việc cải thiện độ công bằng cần được xét xét đồng thời.

Đề xuất hàm dung lượng thứ hai tương tự như hàm dung lượng thứ nhất nhưng giá trị $\sigma(A)$ trong đó được thay bằng giá trị $\sigma'(A)$. Giá trị này được xác định như sau.

$$\sigma'(A) = \frac{1}{dens(A)} * \sigma(A) \quad (2.8)$$

Hàm mật độ thể hiện mức độ đồng nhất của nhóm, được tính dựa trên khoảng cách của mọi người dùng với nhau. Trong luận án, độ đo khoảng cách Mahalanobis được sử dụng để ước lượng hàm mật độ, cụ thể như sau:

$$dens(A) = Average(dis(u_i)), \forall u_i \in A \quad (2.9)$$

Độ đo mờ dựa trên hàm dung lượng thứ hai trình bày ở trên cho thấy các nhóm có độ đồng nhất cao (thì giá trị $dens(A)$ sẽ nhỏ) sẽ có ảnh hưởng tích cực đến sự hài lòng của các thành viên trong đó.

Tương tự hàm dung lượng thứ nhất, dễ thấy hàm dung lượng được xác định theo cách mở rộng như trên thỏa mãn điều kiện của hàm dung lượng theo định nghĩa trong phần 1.4.1.

c) Cơ chế đồng thuận sử dụng tích phân Choquet

Với mỗi sản phẩm i trong tập sản phẩm I , đánh giá của một nhóm g được ước lượng bởi phép toán tổng hợp các đánh giá của thành viên bằng tích phân Choquet như sau.

Không mất tính tổng quát, giả sử thành viên nhóm g đánh giá một sản phẩm i được biểu diễn bởi véc tơ r và được sắp xếp theo thứ tự tăng dần của đánh giá:

$$r = \{r_{u_1 i}, \dots, r_{u_{|g|} i}\}$$

Ta có đánh giá tổng hợp của nhóm cho sản phẩm i dùng tích phân Choquet được tính bởi công thức sau:

$$pref^{CQ}(g, i) = \sum_{k=1}^{|g|} (r_{u_k i} - r_{u_{k-1} i}) \xi(\Upsilon_k) \quad (2.10)$$

Trong đó, $r_{u_0 i}$ là đánh giá giả định được gán bằng 0, thỏa mãn:

$$0 = r_{u_0 i} \leq r_{u_1 i} \leq \dots \leq r_{u_{|g|} i},$$

và tập con người dùng trong nhóm $\Upsilon_k = \{u_k, \dots, u_{|g|}\}$ với $n \leq |g|$.

Với mỗi nhóm, đánh giá cho tất cả sản phẩm trong I được xác định, từ đó chọn lựa ra các sản phẩm tốt nhất cho nhóm, $top-N$ sản phẩm, được thực hiện một cách đơn giản bởi một thuật toán sắp xếp và lựa chọn. Trong luận án này, thuật toán sắp xếp chọn được lựa chọn để cài đặt. Thuật toán sắp xếp chọn có tốc độ tính toán chậm hơn so với thuật toán sắp xếp nhanh, hay thuật toán sắp xếp vun đống. Thuật toán sắp xếp chọn phù hợp với mục tiêu của luận án là phát triển và thực nghiệm các thuật toán mới cho pha đồng thuận. Khi triển khai thực tế, người dùng có thể cân nhắc chuyển sang sử dụng thuật toán sắp xếp nhanh, hay thuật toán sắp xếp vun đống.

2.3.2. Thuật toán đề xuất

Như đã trình bày ở trên, mô hình HTVN được đề xuất có hai pha. Trong pha đầu tiên, đánh giá của tất cả thành viên cho các sản phẩm chưa có đánh giá được dự đoán bằng mô hình lọc cộng tác theo người dùng. Các phép toán phục vụ pha này đã trình bày trong phần (a) mục 2.3.1. Trong đề xuất này, phần quan trọng nhất là pha đồng thuận sử dụng chiến lược tổng hợp được đề xuất phần (b) mục 2.3.2 nói trên. Trong phần này, giả mã của thuật toán đồng thuận trong HTVN sử dụng tích phân Choquet với các hàm dung lượng đã được định nghĩa ở trên được trình bày chi tiết.

Thuật toán cho pha đồng thuận sử dụng tích phân Choquet

Đầu vào: Cho tập người dùng U với tập sản phẩm I . Cho g là nhóm người dùng, tập I' là các sản phẩm ứng viên, R là ma trận đánh giá của tập người dùng U với tập sản phẩm I .

Đầu ra: khuyến nghị N sản phẩm, dịch vụ phù hợp nhất cho nhóm g , trong đó $N \leq |I'|$

Thuật toán:

1. $kich_thuoc_nhom = |g|$
 2. $Trong_so = [None] * kich_thuoc_nhom$
 3. $Trong_so_hc = [None] * kich_thuoc_nhom$
 4. **for** i **in** $range(kich_thuoc_nhom)$:
 5. xác định $Trong_so[i]$ theo công thức (2.4)
 6. xác định $Trong_so_hc[i]$ theo công thức (2.7)
- # xác định giá trị hàm dung lượng

```

7. def get_capacity(Permutation, Index):
8.     gia_tri_ham_dl=0
9.     for i in range (Chi_so, len(Permutation)):
10.         gia_tri_ham_dl += Trong_so[i]
11.         if (len(Permutation)-Chi_so > 1):
12.             for j in range (Chi_so, len(Permutation)):
13.                 gia_tri_ham_dl += Trong_so_hc[j]
14.                 if lựa chọn hiệu chỉnh:
15.                     gia_tri_ham_dl*= 1/dens(g) # ct (2.8)
16. return gia_tri_ham_dl
    # tính đánh giá của g với từng mục thuộc tập ứng viên
17. so_luong_ung_vien=|I'|
18. danh_gia_nhom=[None]* so_luong_ung_vien
19. for i in range (so_luong_ung_vien):
20.     mang_nd_sap_xep = get_permutation(g,I',R)
        # Hợp đánh giá các thành viên dự trên tích phân
        #Choquet.
21.     danh_gia_nhom_1_sp=0
22.     danh_gia=0
23.     for j in range(number_user):
24.         thanh_vien=user_permuation[j]
25.         khoang_cach_DG = R[thanh_vien,i] - danh_gia
26.         danh_gia_nhom_1_sp += khoang_cach_DG
            *get_capacity(mang_nd_sap_xep,j)
27.         previous_rate = R[thanh_vien,i]
28.     danh_gia_nhom[i]= danh_gia_nhom_1_sp
        # trả về khuyến nghị cho nhóm và kết thúc
29. Trả về top-N sản phẩm có giá trị đánh giá cao nhất.

```

Lưu ý rằng, trong đoạn mã giả ở trên, hàm **get_permutation** nhận đầu vào nhóm người dùng, sản phẩm đang xét và ma trận đánh giá từ đó trả về một mạng chứa định danh của thành viên trong nhóm được sắp xếp theo thứ tự tăng dần của đánh giá với

sản phẩm đang xét. Thuật toán này được cài đặt bằng cách sử dụng thuật toán sắp xếp chọn để xác định và trả về một hoán vị thể hiện thứ tự của thành viên nhóm trong danh sách được sắp xếp theo thứ tự tăng dần của đánh giá.

Do số lượng thành viên nhóm thường không quá lớn nên không cần thiết phải cài đặt thuật toán sắp xếp phức tạp hơn như sắp xếp nhanh hay sắp xếp vun đống. Thuật toán sắp xếp chọn sử dụng một vòng lặp với mục tiêu tại mỗi lần lặp sẽ xác định được phần tử nhỏ nhất của phần chưa sắp xếp và bổ sung vào phần đã sắp xếp. Trong đó, với mỗi lần lặp lại cần thực hiện một phép duyệt mảng để tìm phần tử nhỏ nhất. Thuật toán này rất phổ biến và tương đối đơn giản, do đó NCS không mô tả chi tiết ở đây.

2.3.3. Độ phức tạp thuật toán pha đồng thuận

Để so sánh độ phức tạp tính toán của mô hình đề xuất mới cho pha đồng thuận và các chiến lược đồng thuận được trình bày trong mục 1.1.4, chúng ta cần đánh giá độ phức tạp của thuật toán dựa trên khái niệm độ phức tạp tính toán O lớn.

Trước hết, ta thấy quá trình tính toán các giá trị của hàm dung lượng cho một tập người dùng A yêu cầu phải tính các giá trị $\omega(u)$ và $\omega'(u)$. Độ phức tạp của thuật toán tính $\omega(u)$ và $\omega'(u)$ phụ thuộc vào số lượng các sản phẩm và dễ thấy độ phức tạp của quá trình này là $O(|I|)$. Từ đó để xác định giá trị hàm dung lượng cho một tập người dùng A độ phức tạp của hàm `get_capacity` (từ câu lệnh 7-16) là $O(|A|.|I|)$.

Để ước lượng đánh giá của nhóm g với sản phẩm i theo mô hình đề xuất sử dụng tích phân Choquet (các câu lệnh từ 20-27), ta cần tính $m = |g|$ giá trị hàm dung lượng, độ phức tạp tính toán của quá trình này là $O(|g|.|I|.|g|)$ (lưu ý rằng $|A| \leq |g|$). Pha đồng thuận cần tính giá trị đánh giá của nhóm cho tất cả các sản phẩm, dịch vụ thuộc nhóm ứng viên nên sẽ có độ phức tạp tính toán là $O(|I'|.|g|^2. |I|)$. Và để hoàn thiện cần xác định ra N sản phẩm theo thứ tự đánh giá cao nhất toàn bộ pha đồng thuận sẽ tương ứng với độ phức tạp tính toán là $O(|I'|.|g|^2. |I|.N)$.

Như đã trình bày trong mục 1.2.2, với các chiến lược đồng thuận đơn giản, để xác định N sản phẩm theo thứ tự đánh giá cao nhất từ tập I sản phẩm, độ phức tạp tính toán của pha đồng thuận trong HTVN là $O(|I| \cdot |g| \cdot N)$. Bên cạnh đó, các chiến lược như chiến lược luật Copeland và chiến lược chỉ số Borda cần so sánh các sản phẩm để tính toán các giá trị chỉ số mới thay cho đánh giá ban đầu của người dùng. Độ phức tạp tính toán của phép toán tổng hợp đánh giá của các thành viên trong các thuật toán này sẽ lớn hơn và bằng $O(|g| \cdot |I|)$. Khi đó, độ phức tạp tính toán của pha đồng thuận theo các chiến lược này là $O(|I|^2 \cdot |g| \cdot N)$.

Ta thấy, vì kích thước nhóm $|g|$ thường rất nhỏ so với số lượng sản phẩm, dịch vụ $|I|$, chiến lược đồng thuận sử dụng toán tử tổng hợp đánh giá của thành viên trong nhóm người dùng với tích phân Choquet có độ phức tạp tính toán rất gần với các chiến lược như chiến lược luật Copeland và chiến lược chỉ số Borda. Do đó chiến lược này hoàn toàn có tính khả dụng. Ngoài ra, khi sử dụng hàm dung lượng thứ hai, cần ước lượng giá trị thể hiện độ đồng nhất của nhóm, theo mật độ $dens(A)$, và thuật toán này làm tăng khá lớn độ phức tạp tính toán.

2.4. Kết quả thực nghiệm và bàn luận

2.4.1. Mục tiêu thực nghiệm

Để thực hiện đánh giá thuật toán HTVN với “pha đồng thuận” sử dụng tích phân Choquet được đề xuất, luận án thực nghiệm và so sánh thuật toán đề xuất với thuật toán đi theo các chiến lược đồng thuận thường được sử dụng trong các nghiên cứu trước đó. Về độ phức tạp tính toán, việc so sánh được thực hiện trên nền tảng so sánh lý thuyết, dựa trên đánh giá độ phức tạp thời gian O lớn được trình bày trong mục 2.3.3 ở trên. Do đó, mục tiêu so sánh trong thực nghiệm tập trung vào độ phù hợp của các thuật toán xét theo khía cạnh tối đa hóa tổng đánh giá của nhóm người dùng và giảm sự chênh lệch các thành viên trong nhóm (tăng độ công bằng của khuyến nghị). Hai mục tiêu này được gắn liền với nhau để đảm bảo không có một đánh giá thiên lệch về một xu thế duy nhất.

2.4.2. Dữ liệu thực nghiệm

Dựa trên tổng quan nghiên cứu ta thấy rằng không có dữ liệu công khai nào có sẵn dữ liệu về nhóm người dùng và đánh giá chung của cả nhóm. Do đó, khi muốn áp dụng HTVN, các nghiên cứu trước đây đều sử dụng dữ liệu công khai dùng cho HTV đơn người dùng, và sau đó dùng một cơ chế sinh tự động để tạo dữ liệu về các nhóm dựa trên giả thiết nào đó về đặc điểm của nhóm người dùng.

Do đó, để so sánh mô hình HTVN được đề xuất (mô hình sử dụng tích phân Choquet) và các chiến lược đồng thuận khác trong mục 1.4, dữ liệu MovieLens-1M được dùng để làm dữ liệu thực nghiệm. Đặc trưng của dữ liệu được thể hiện trong bảng sau.

Table 1: Dữ liệu MovieLens-1M

Tập dữ liệu	Số lượng người dùng	Số lượng phim	Số lượng đánh giá
MovieLens 1M	610	9742	100836

Dữ liệu MovieLens-1M là bộ dữ liệu phổ biến để đánh giá các nghiên cứu về HTVN [2], [79], nhưng dữ liệu không chứa thông tin nào phản ánh nhóm người dùng sử dụng sản phẩm cùng nhau. Do đó, để áp dụng cho vấn đề đánh giá HTVN cần có phương pháp tạo nhóm người dùng [11], [17], [32]. Có hai cách tiếp cận để lựa chọn người dùng vào trong thành viên một nhóm [11]. Cách thứ nhất sử dụng cơ chế sinh ngẫu nhiên nhằm lựa chọn các thành viên, với giả thiết rằng không có một ràng buộc nào về việc hai người dùng có thể tham gia cùng một nhóm; và cách thức thứ hai đặt thêm các ràng buộc về sự tương quan của các thành viên trong nhóm [11]. Để thực hiện kiểm nghiệm HTVN đề xuất, các nhóm người dùng được sinh ra theo giả thiết rằng không có ràng buộc nào giữa các thành viên khi tạo lập nhóm. Do đó, trong luận án, với nhóm người dùng g_i một phương thức lấy mẫu ngẫu nhiên sẽ lựa chọn từng người dùng (không lặp) để làm thành viên của $|g_i|$. Một bộ dữ liệu duy nhất về các nhóm người dùng sinh ra và bộ dữ liệu này được sử dụng đồng nhất trong quá trình kiểm thử các tiếp cận của chiến lược đồng thuận trong HTVN. Trong tương lai, với

một bài toán ứng dụng chuyên biệt, việc thử nghiệm với các bộ dữ liệu được sinh ra từ chiến lược tạo nhóm người dùng có áp dụng điều kiện ràng buộc giữa các thành viên sẽ được thực hiện.

2.4.3. Phương thức đánh giá

Cần nhấn mạnh rằng có nhiều phương thức tổng hợp đáng giá có thể được áp dụng trong phát triển HTVN theo tiếp cận hợp khuyến nghị, và mỗi phương thức tổng hợp đáng giá này được xây dựng dựa trên các giả thiết khác nhau. Đã có nhiều độ đo được phát triển để phân tích hiệu quả của các HTVN dựa trên các phương thức đồng thuận này. Để so sánh hiệu quả của HTVN đề xuất với các chiến lược đồng thuận phổ biến khác được trình bày trong phần 1.1.4 ở trên với cách tiếp cận đề xuất trong luận án, các độ đo được sử dụng bao gồm độ đo về sai số, độ đo về tính công bằng và tổng đánh giá của nhóm. Chi tiết các độ đo được trình bày dưới đây.

Độ đo sai số:

Sai số dự báo của HTVN được đánh giá bằng độ đo sai số tuyệt đối – MAE. Sai số này được ước lượng như sau.

$$MAE(g, I) = \frac{\sum_{i \in I} |r_{g,i} - \hat{r}_{g,i}|}{|I|} \quad (2.11)$$

Với I tập hợp sản phẩm đề xuất cho nhóm g , và $\hat{r}_{g,i}$, $r_{g,i}$ tương ứng là giá trị do mô hình dự báo và giá trị thực đánh giá của nhóm. Trong đó :

$$r_{g,i} = \frac{\sum_{u \in g} r_{ui}}{|g|} \quad (2.12)$$

Giá trị đánh giá của nhóm:

Nhắc lại, mục tiêu của HTVN vẫn bao gồm việc lựa chọn sản phẩm có tổng đánh giá của nhóm là cao nhất để đưa vào khuyến nghị, bên cạnh yếu tố độ công bằng. Trong luận án một độ đo thể hiện mức độ hài lòng với các sản phẩm trong danh sách khuyến nghị cho nhóm được tính như sau:

$$group_pref(g, I) = \frac{\sum_{i \in I} r_{g,i}}{|I|} \quad (2.13)$$

Với $r_{g,i} = \frac{\sum_{u \in g} r_{ui}}{|g|}$ là giá trị trung bình đánh giá của tất cả thành viên cho sản phẩm

tương ứng.

Các độ đo công bằng:

Đánh giá độ công bằng của khuyến nghị từ mô hình HTVN, luận án lựa chọn hai độ đo được sử dụng phổ biến trong các nghiên cứu được công bố trước đây, và đã được trình bày trong mục 1.2.1.

- Độ đo thứ nhất thể hiện trong công thức (1.14) như sau.

$$fairness(g, i) = \frac{\left| \bigcup_{u \in g} : r_{ui} \geq \theta \right|}{|g|}$$

Với θ là ngưỡng thể hiện một thành viên hài lòng với sản phẩm. Giá trị độ đo càng cao thì thuật toán được xem là càng có hiệu năng cao.

- Độ đo thứ hai theo công thức (1.9) như sau:

$$fairness_{var}(g, i) = 1 - Var\{r_{ui}, \forall u \in g\}$$

Độ đo công bằng này cho thấy giá trị độ công bằng càng cao thì thuật toán tương ứng càng tốt. Giá trị tốt nhất đạt được khi $Var\{r_{ui}, \forall u \in g\} = 0$ và khi đó $fairness_{var}(g, i) = 1$.

2.4.4. Kết quả và bàn luận

Kết quả thực nghiệm với HTVN theo tiếp cận hợp khuyến nghị đề xuất, sử dụng tích phân Choquet trong phát triển phép toán tổng hợp các đánh giá của các thành viên trong nhóm trong pha đồng thuận, sẽ được trình bày và phân tích tại nội dung này. Trong đó HTVN được đề xuất được so sánh với các thuật toán đồng thuận phổ biến khác dựa trên thực nghiệm với cùng một bộ dữ liệu.

Đề xuất HTVN sử dụng toán tử Choquet được cài đặt bằng ngôn ngữ lập trình Python, các tính năng căn bản được cài đặt để có thể thực nghiệm và đánh giá các chiến lược khác nhau trong pha đồng thuận, đồng thời có thể phát triển tiếp trong các giai đoạn nghiên cứu mở rộng, hoặc triển khai tích hợp trong các hệ thống thực tiễn.

Kết quả thực nghiệm được thực hiện với dữ liệu được mô tả ở mục 2.4.2, và thông tin về các nhóm được sinh ra từ một thuật toán tạo nhóm ngẫu nhiên. Cơ chế này dựa trên giả thiết rằng không có ràng buộc nào về việc hình thành một nhóm người dùng từ các người dùng cá nhân trong hệ thống. Để đánh giá thuật toán, dữ liệu về các nhóm được sinh ra thỏa mãn: các nhóm có kích thước $group_size = \{3, 5, 10, 15, 20\}$, và 30 nhóm được sinh ra tương ứng với mỗi kích thước nhóm, tổng cộng có 150 nhóm. Khi thống kê kết quả, hiệu năng của một thuật toán theo kích thước nhóm sẽ là giá trị trung bình từ kết quả đo được với 30 nhóm có kích thước tương ứng. Kích thước nhóm được thiết kế tăng dần để thấy được ảnh hưởng của kích thước nhóm tới hiệu năng của thuật toán.

Trong pha khuyến nghị, trước pha đồng thuận, thuật toán lọc cộng tác theo người dùng được sử dụng. Trong mô hình lọc cộng tác theo người dùng, độ đo tương đồng cosine được sử dụng để đánh giá tương quan giữa hai người dùng, và giá trị ngưỡng k trong thuật toán K-hàng xóm được xác định bằng 10% số lượng người dùng trong dữ liệu.

Các HTVN đề xuất, sử dụng tích phân Choquet trong pha đồng thuận được, tương ứng với hai cách thức xây dựng hàm dung lượng được mô tả trong phần 2.3.2 và tương ứng được ký hiệu là CIS_CF1 và CIS_CF2 (trong đó CIS là cách viết tắt của cụm từ Choquet integral based strategy, và CF viết tắt cho từ Capacity function).

Sau khi tiến hành thực nghiệm, kết quả của hai mô hình được đề xuất được so sánh với các HTVN sử dụng các chiến lược phổ đồng thuận biến khác. Kết quả được thể hiện trong các so sánh dưới đây.

Kết quả thực nghiệm

Như đã nêu trong các phần trước, độ chính xác trong quá trình dự báo đánh giá của người dùng với một sản phẩm chủ yếu phụ thuộc vào mô hình lọc cộng tác theo người dùng. Pha đồng thuận với các chiến lược khác nhau không tác động nhiều tới sai số này. Vì vậy, NCS sử dụng độ đo sai số dự báo để cung cấp thông tin về khả năng ứng dụng thực tiễn của các thuật toán, mục tiêu so sánh độ chính xác của HTVN khi sử dụng các chiến lược đồng thuận không phải mục tiêu chính.

Bảng 2.3 cho thấy sai số dự đoán trung bình, sử dụng độ đo MAE, của các mô hình HTVN sử dụng các chiến lược đồng thuận khác nhau.

Bảng 2.3 Sai số trung bình của mô hình với kích thước trong khuyến nghị

Top-N (N)	AUS	MS	LMS	MPS	AVS	CRS	CIS_CF1	CIS_CF2
1	0.226	0.254	0.464	0.240	0.409	0.224	0.336	0.319
2	0.219	0.252	0.423	0.247	0.377	0.217	0.264	0.251
3	0.201	0.217	0.411	0.236	0.410	0.197	0.251	0.248
4	0.185	0.207	0.394	0.236	0.382	0.184	0.243	0.244
5	0.173	0.208	0.385	0.225	0.356	0.175	0.257	0.258
6	0.171	0.207	0.378	0.204	0.374	0.171	0.265	0.268
7	0.171	0.201	0.363	0.190	0.370	0.173	0.257	0.262
8	0.163	0.200	0.359	0.187	0.366	0.163	0.274	0.278
9	0.172	0.203	0.353	0.187	0.352	0.172	0.277	0.281
10	0.178	0.202	0.338	0.203	0.352	0.178	0.282	0.285

Bảng 2.4 Sai số trung bình của mô hình với kích thước nhóm

KT nhóm	AUS	MS	LMS	MPS	AVS	CRS	CIS_CF1	CIS_CF2
3	0.095	0.081	0.092	0.071	0.080	0.095	0.182	0.182
5	0.173	0.197	0.386	0.242	0.380	0.171	0.220	0.220

10	0.197	0.222	0.424	0.272	0.421	0.197	0.259	0.259
15	0.224	0.249	0.469	0.255	0.490	0.224	0.290	0.300
20	0.240	0.327	0.564	0.238	0.503	0.240	0.402	0.385

Bảng 2.3 và 2.4 cho thấy sai số tuyệt đối (MAE) cao hơn khi kích thước khuyến nghị nhỏ (đưa ra ít sản phẩm) và giảm dần khi kích thước khuyến nghị tăng. Đồng thời, sai số này cũng có xu hướng tăng dần lên khi kích thước của nhóm tăng. Điều này đúng với tất cả các mô hình cho pha đồng thuận, bao gồm HTVN đề xuất sử dụng tích phân Choquet.

Tính công bằng của khuyến nghị:

Xét các độ đo tính công bằng, với độ đo công bằng đầu tiên phản ánh tỷ lệ thành viên hài lòng với khuyến nghị và kích thước nhóm. Trong luận án, một thành viên được coi là hài lòng với một sản phẩm nếu thành viên đó đánh giá sản phẩm, dịch vụ tương ứng bằng hoặc cao hơn ngưỡng θ , được thiết lập trước. Trong luận án, ngưỡng $\theta = 3.5$ được lựa chọn dựa trên thang đo đánh giá của người dùng của bộ dữ liệu được quy định là $[1-5]$, trong đó 1 thể hiện người dùng không thích sản phẩm và 5 thể hiện người dùng rất thích sản phẩm.

Bảng 2.5 cho thấy “chiến lược biểu quyết phê duyệt” (**AVS**) và “chiến lược ít thiệt thòi nhất” (**LMS**) là hai HTVN tốt nhất từ hướng nhìn thuần về độ công bằng theo độ đo thứ nhất. Đồng thời, HTVN với pha đồng thuận sử dụng “chiến lược ưu thế” (**MPS**) có kết quả tồi nhất.

Bảng 2.5 độ đo công bằng thứ nhất theo kích thước khuyến nghị

Top-N (N)	AUS	MS	LMS	MPS	AVS	CRS	CIS_CF1	CIS_CF2
1	0.701	0.760	0.820	0.651	0.851	0.701	0.694	0.675
2	0.721	0.757	0.831	0.642	0.846	0.721	0.687	0.672
3	0.725	0.747	0.820	0.640	0.852	0.725	0.681	0.676

4	0.718	0.752	0.813	0.636	0.834	0.718	0.676	0.675
5	0.716	0.753	0.800	0.627	0.831	0.716	0.681	0.680
6	0.715	0.742	0.797	0.624	0.827	0.715	0.680	0.680
7	0.715	0.742	0.787	0.622	0.819	0.715	0.680	0.682
8	0.706	0.740	0.782	0.618	0.811	0.706	0.686	0.688
9	0.701	0.734	0.779	0.615	0.807	0.701	0.689	0.690
10	0.702	0.730	0.775	0.618	0.807	0.702	0.690	0.690

Tuy nhiên, ta thấy rằng độ đo công bằng này không cho biết được sự khác nhau trong đánh giá (mức độ yêu thích) của các thành viên trong nhóm.

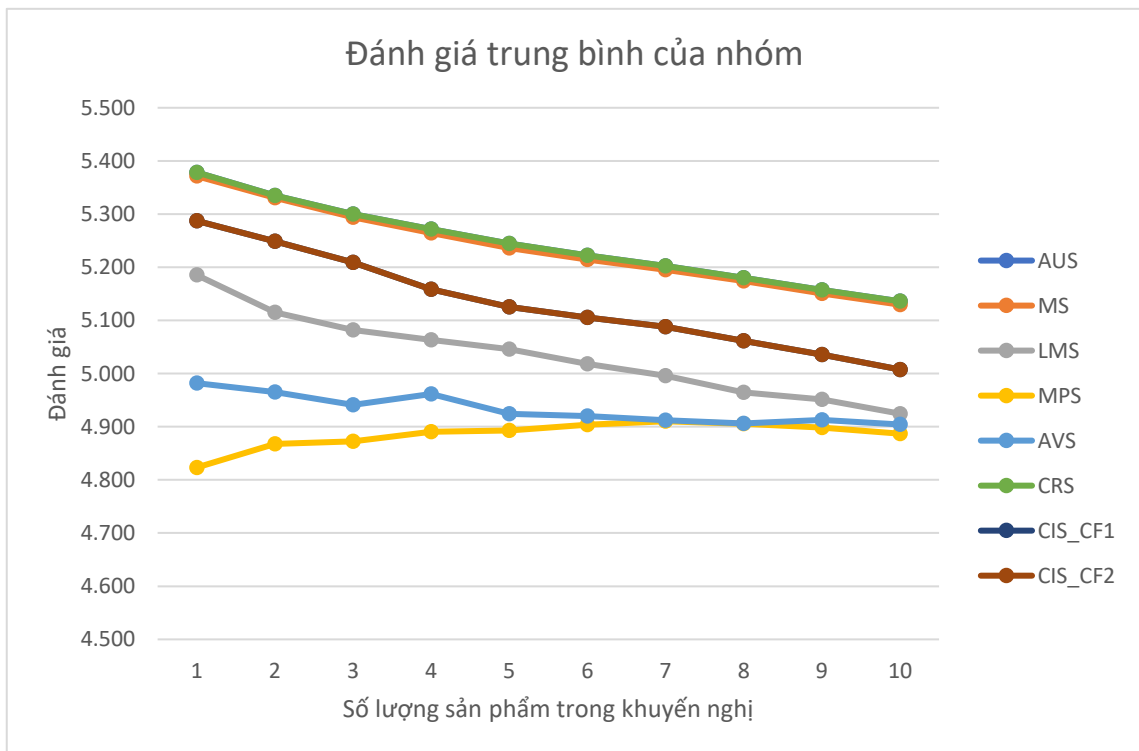
Bảng 2.6 Đánh giá trung bình của nhóm theo kích thước khuyến nghị

Top-N (N)	AUS	MS	LMS	MPS	AVS	CRS	CIS_CF1	CIS_CF2
1	4.717	4.694	4.373	4.384	4.412	4.717	4.556	4.560
2	4.689	4.670	4.350	4.404	4.383	4.689	4.536	4.537
3	4.669	4.658	4.339	4.387	4.357	4.669	4.514	4.516
4	4.652	4.645	4.337	4.396	4.350	4.652	4.503	4.505
5	4.637	4.628	4.331	4.397	4.349	4.637	4.481	4.483
6	4.625	4.615	4.327	4.401	4.342	4.625	4.465	4.467
7	4.614	4.605	4.313	4.402	4.341	4.614	4.451	4.453
8	4.600	4.594	4.301	4.398	4.339	4.600	4.442	4.444
9	4.586	4.578	4.292	4.390	4.335	4.586	4.433	4.433
10	4.572	4.564	4.279	4.378	4.331	4.572	4.420	4.420

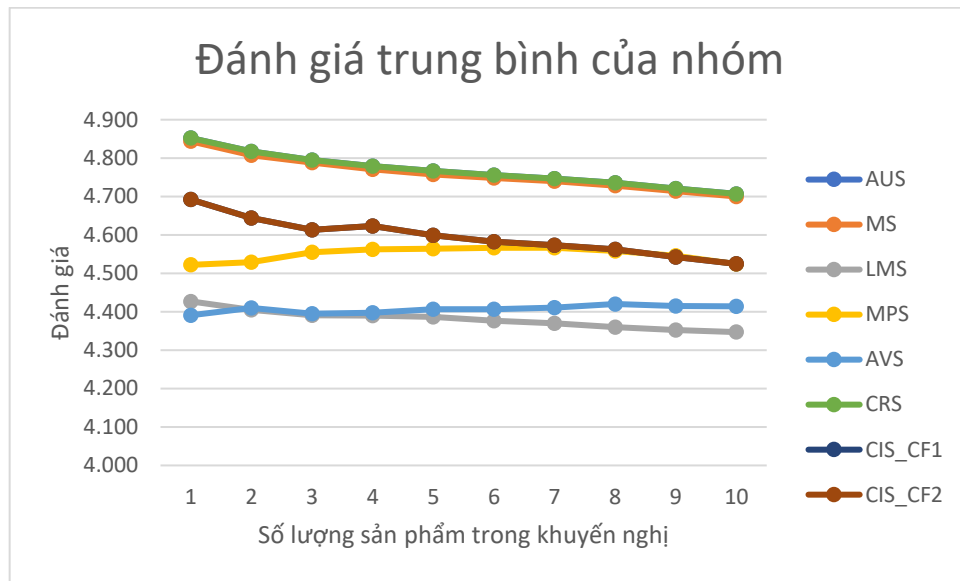
Từ bảng 2.6 ta thấy hiệu năng của các HTVN sử dụng chiến lược đồng thuận khác nhau về đánh giá trung bình của nhóm. Rõ ràng chiến lược tổng (AUS) sẽ luôn cho hiệu năng cao nhất khi xét đến yếu tố này, bởi vì chiến lược này thuần túy xác định sản phẩm tốt nhất dựa trên tổng đánh giá của các thành viên trong nhóm. HTVN sử dụng chiến lược luật Copeland (CRS) có kết quả tốt tương đương với chiến lược AUS. Hai thuật toán đề xuất cho HTVN với pha đồng thuận sử dụng hai phép toán tổng hợp đánh giá của thành viên nhóm dựa trên tích phân Choquet được đề xuất với hàm dung lượng thứ nhất (CIS_CF1) và hàm dung lượng thứ hai (CIS_CF2) có hiệu

năng như nhau và rất gần với HTVN sử dụng chiến lược AUS và các mô hình đề xuất này tốt hơn HTVN sử dụng các chiến lược còn lại. Bên cạnh đó, với mọi chiến lược sử dụng cho pha đồng thuận, khi số lượng sản phẩm, dịch vụ trong một khuyến nghị tăng lên, độ đo mức độ thỏa mãn, thể hiện qua đánh giá trung bình của nhóm với toàn bộ sản phẩm trong khuyến nghị, có xu hướng giảm xuống.

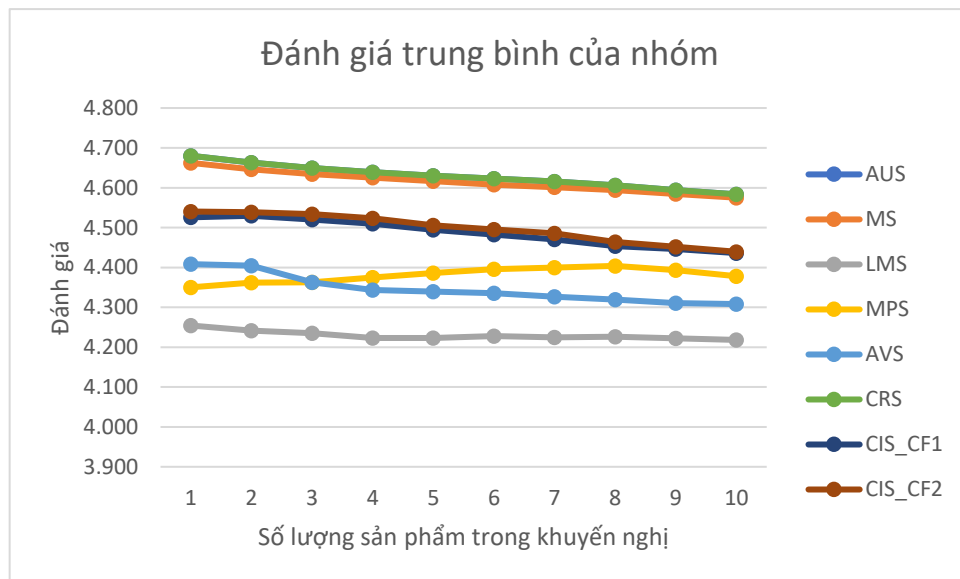
Khi xem xét đến ảnh hưởng của kích thước nhóm, các hình 2.2, 2.3, 2.4 dưới đây cho thấy hiệu quả của các HTVN theo ba kích thước nhóm khác nhau. Kết quả cho thấy ảnh hưởng của tham số số lượng các sản phẩm, dịch vụ trong khuyến nghị tới sự thay đổi về đánh giá trung bình của nhóm. Khi số kích thước khuyến nghị tăng lên, sự chênh lệch về hiệu năng giữa các chiến lược đồng thuận có xu hướng giảm.



Hình 2.2 Đánh giá trung bình của nhóm với kích thước nhóm bằng 3

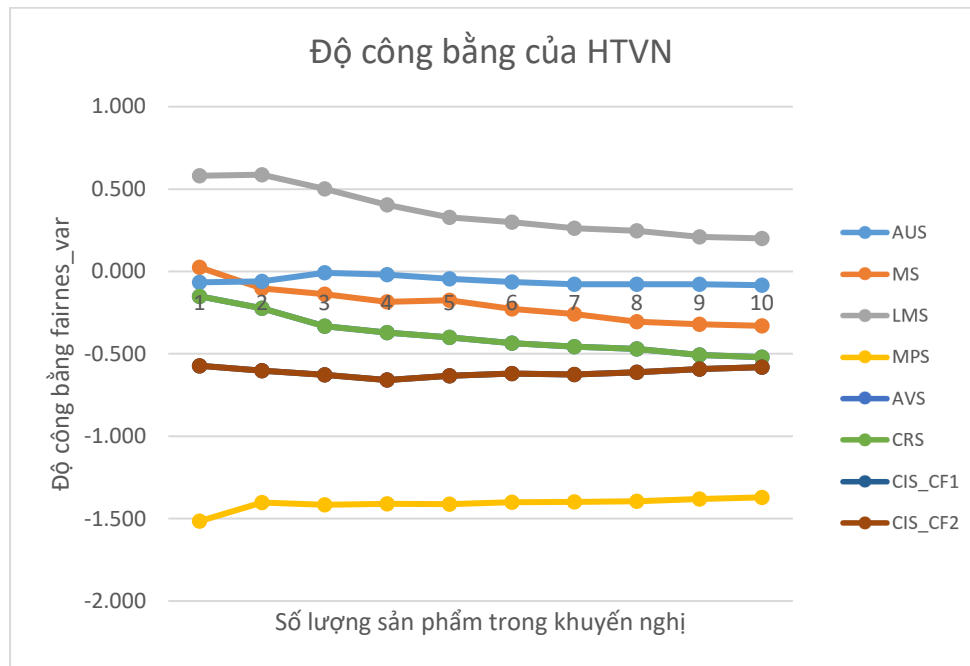


Hình 2.3 Đánh giá trung bình của nhóm với kích thước nhóm bằng 10

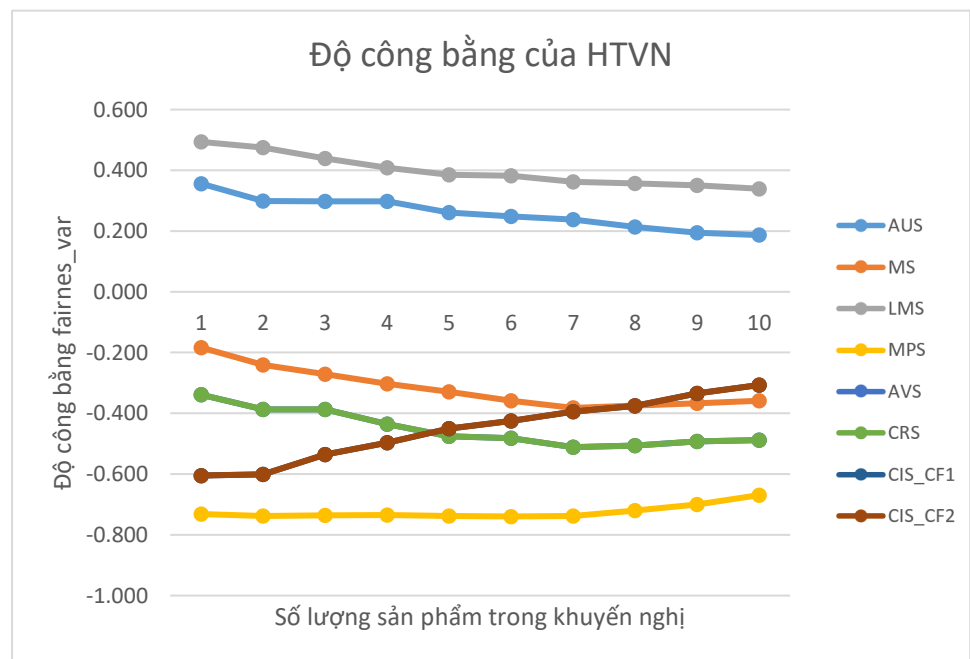


Hình 2.4 Đánh giá trung bình của nhóm với kích thước nhóm bằng 20

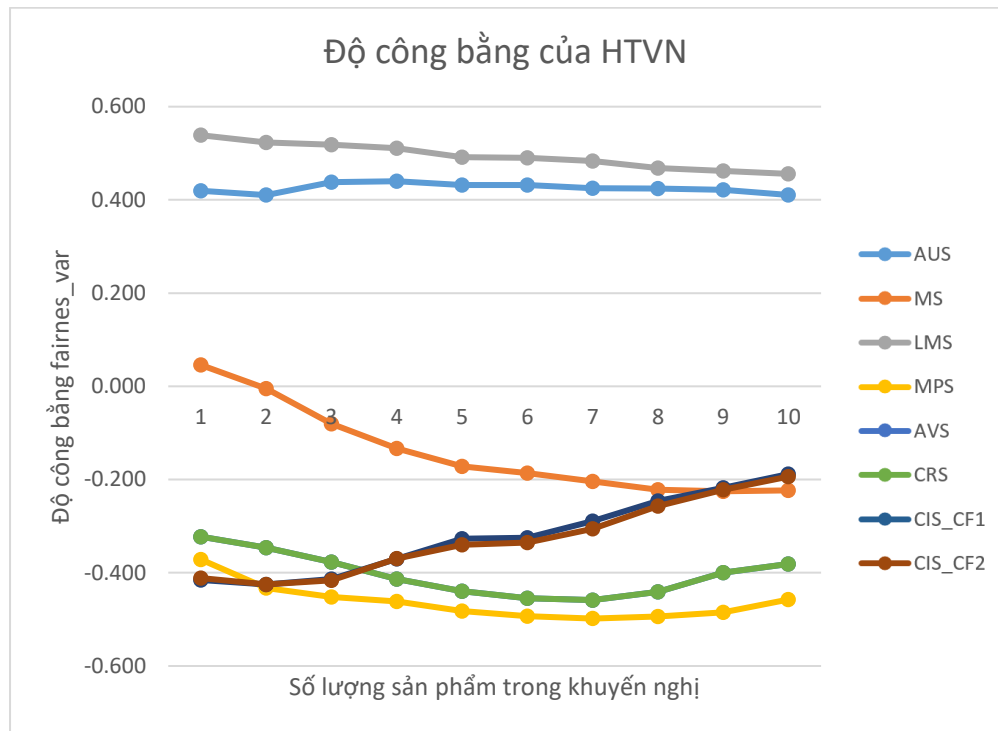
Khi xét đến hiệu năng của các mô hình HTVN theo độ đo công bằng thứ hai, độ đo cho thấy sự chênh lệch giữa đánh giá của các thành viên với sản phẩm được khuyến nghị, hình 2.5, 2.6, 2.7 dưới đây thể hiện hiệu quả của các mô hình theo kích thước nhóm.



Hình 2.5 Độ công bằng của khuyến nghị cho nhóm có kích thước bằng 3



Hình 2.6 Độ công bằng của khuyến nghị cho nhóm có kích thước bằng 10



Hình 2.7 Độ công bằng của khuyến nghị cho nhóm có kích thước bằng 20

Từ hình 2.5, 2.6 và 2.7 chúng ta có thể thấy rằng các thuật toán đồng thuận sử dụng Chiến lược ít thiệt thòi nhất (LMS) và Chiến lược biểu quyết phê duyệt (AVS) có độ đo công bằng cao nhất, nhưng đồng thời đây là hai chiến lược đưa ra khuyến nghị bao gồm các sản phẩm có mức độ đánh giá trung bình thấp nhất. Thuật toán đồng thuận sử dụng Chiến lược ưu thế (MPS) cho thấy hiệu quả về độ công bằng thấp nhất, và đồng thuật toán này sinh ra khuyến nghị có đánh giá trung bình của nhóm cũng kém nhất. Ở chiều ngược lại, các thuật toán đồng thuận sử dụng Chiến lược tổng (AUS) và chiến lược luật Copeland (CRS) đưa ra khuyến nghị có độ đo về đánh giá trung bình của nhóm là tốt nhất, nhưng các thuật toán này lại có độ đo công bằng thấp nhất.

Trong khi đó, mô hình HTVN được đề xuất với pha đồng thuận sử dụng hai phép toán tổng hợp đánh giá của các thành viên nhóm dựa trên tích phân Choquet với hai hàm dung lượng, hàm dung lượng thứ nhất (CIS_CF1) và hàm dung lượng thứ hai (CIS_CF2) không tốt hơn các HTVN sử dụng chiến lược đồng thuận khác ở bất kỳ khía cạnh đơn lẻ nào (xét theo từng độ đo riêng rẽ) trong hai khía cạnh: đánh giá

trung bình của nhóm và độ công bằng của khuyến nghị. Nhưng, dễ thấy hai HTVN đề xuất có kết quả thực nghiệm tốt nhất khi cân bằng được giữa cả hai mục tiêu. HTVN sử dụng chiến lược tích (MS) có kết quả gần với hai mô hình đề xuất xét theo cách nhìn này, nhưng hiệu quả của mô hình này kém đi khi tăng kích thước nhóm hoặc sản phẩm khuyến nghị.

Do đó, khi phát triển HTVN với mục tiêu cân bằng cả hai khía cạnh độ công bằng và lợi ích tổng thể của các thành viên trong nhóm thì pha đồng thuận sử dụng tích phân Choquet có ưu thế hơn các chiến lược khác. Hơn nữa độ đo mờ với hàm hàm dung lượng được xây dựng từ dữ liệu hành vi của người dùng trong nhóm cho phép giải bài toán tối ưu hai mục tiêu ở trên một cách mềm dẻo hơn (không cần định nghĩa một phép toán kết hợp hai mục tiêu đồng nhất cho mọi nhóm người dùng). Bên cạnh đó, khi phân tích cách thức xây dựng hàm dung lượng, chúng ta có thể thấy hàm này phần nào đã xử lý được mối tương tác giữa thành viên trong nhóm. Thành viên tích cực (thực hiện nhiều đánh giá, và đánh giá có tính tích cực) sẽ có ưu thế và ảnh hưởng tích cực đến các người dùng khác trong nhóm.

Về mặt độ phức tạp tính toán, mô hình đề xuất sẽ phức tạp hơn so với các chiến lược đồng thuận thông thường. Thực nghiệm trên môi trường máy tính có cấu hình như sau Intel(R) Core(TM) i7-7Y75 CPU 1.60 GHz, RAM 16.0 GB, Hệ điều hành Window 10 – 64 bit, thời gian tính trung bình để tạo khuyến nghị cho một nhóm người dùng tính theo đơn vị giây thể hiện trong bảng 2.7. Lưu ý rằng thời gian này bao gồm cả thời gian của pha sinh khuyến nghị, dùng để dự báo các đánh giá còn thiếu của người dùng.

Bảng 2.7 Thời gian tính của các mô hình HTVN

Kích thước nhóm	AUS	MS	LMS	MPS	AVS	CRS	CIS_CF1	CIS_CF2
3	2.159	2.159	2.160	2.159	2.160	2.160	3.929	3.901
5	3.356	3.356	3.356	3.356	3.356	3.356	4.436	4.838
10	6.990	6.989	6.990	6.990	6.990	6.994	9.247	11.248
15	9.776	9.777	9.782	9.782	9.782	9.784	11.779	15.432
20	11.287	11.289	11.295	11.295	11.295	11.296	13.768	18.770

2.5. Kết luận chương 2

Chương 2 trình bày kết quả nghiên cứu về hướng tiếp xây dựng HTVN nâng cao độ công bằng khuyến nghị với độ đo mờ. Nghiên cứu cơ bản đã thực hiện được mục tiêu đặt ra, đó là nghiên cứu xây dựng và phát triển HTVN ứng dụng toán tử Choquet nhằm tăng sự công bằng của khuyến nghị đồng thời đảm bảo giữ được tổng đánh giá của người dùng trong nhóm ở mức cao.

Cụ thể, mô hình HTVN đề xuất dựa trên tiếp cận hợp khuyến nghị, trong đó pha đồng thuận dùng toán tử Choquet được sử dụng để phát triển một phép toán tổng hợp các đánh giá của người dùng trong nhóm thành đánh giá chung của nhóm. Ứng dụng các hàm dung lượng được đề xuất từ hoạt động của người dùng trong hệ thống, pha đồng thuận trong HTVN đã xử lý tương tác giữa các thành viên trong nhóm, và biểu diễn được một độ đo mờ phù hợp. Từ đó khuyến nghị được sinh ra bởi HTVN cân bằng được giữa hai mục tiêu tối ưu tổng lợi ích của người dùng trong nhóm và đảm bảo sự công bằng giữa mọi người dùng trong nhóm. Đồng thời thuật toán đề xuất giải pháp cho vấn đề tìm hàm dung lượng phù hợp bằng các thuật toán gần đúng có độ phức tạp tính toán thấp, phù hợp với ứng dụng thực tiễn của HTVN.

Mô hình đề xuất đã được cài đặt bằng ngôn ngữ lập trình Python. Mặc dù phiên bản hiện tại của thuật toán đề xuất trong luận án được cài đặt theo hướng hỗ trợ việc so sánh với các phương pháp khuyến nghị khác của HTVN. Phiên bản này có thể được tùy chỉnh một cách dễ dàng để tích hợp vào một ứng dụng thực tế của HTVN. Khi sử dụng với dữ liệu có định dạng khác định dạng mặc định, người dùng cần bổ sung thêm quá trình tiền xử lý dữ liệu để có dữ liệu đầu vào phù hợp. Bên cạnh đó, khi triển khai mô hình đề xuất để giảm hơn nữa thời gian tính có thể chuyển đổi thuật toán sang mô hình thuật toán song song để đạt được hiệu suất tối ưu.

Kết quả nghiên cứu trong Chương 2 được công bố trong một bài báo trên tạp chí thuộc danh mục SCIE, là công trình số 1 trong danh mục công trình công bố của tác giả.

Chương 3. HỆ TƯ VẤN NHÓM ĐỘNG THEO TIẾP CẬN TÍNH TOÁN MỜ TRỰC CẢM ĐẢM BẢO TÍNH CÔNG BẰNG

3.1. Mở đầu

Như đã phân tích trong chương 1, các nghiên cứu được công bố gần đây cho thấy sự cần thiết của HTVN và tiềm năng ứng dụng trong các nền tảng, hệ thống thông tin khác nhau [2], [6], [12], [34], [52]. Khi ứng dụng HTVN, sự kết hợp các đặc điểm của sản phẩm, dịch vụ với sở thích của người dùng một cách sát với thực tiễn là điều cần thiết cho sự thành công của hệ thống. Trong đó xử lý dữ liệu theo tiếp cận động và xử lý liệu có chứa thông tin về tính do dự không nhất quán bằng tiếp cận tính toán mờ là tiếp cận quan trọng để có được một mô hình HTVN gần với thực tế.

Dễ thấy, cảm nhận của người dùng đối với sản phẩm, dịch vụ sẽ thay đổi theo thời gian và do đó đánh giá của người về một sản phẩm là một giá trị biến đổi theo thời gian. Bên cạnh đó, mức độ phổ biến, hấp dẫn của sản phẩm cũng luôn thay đổi, ví dụ như khi một sản phẩm mới xuất hiện sẽ ảnh hưởng tới các sản phẩm cùng loại trước đó. Bản chất thay đổi của sở thích người dùng và sự thay đổi về mức độ hấp dẫn của sản phẩm, dịch vụ đặc biệt nổi bật trong trường hợp người dùng lặp đi lặp lại việc tiêu thụ một số sản phẩm thuộc các danh mục nhất định, chẳng hạn như phim, tin tức và âm nhạc [50], [80]. Chính vì vậy, nghiên cứu và xây dựng mô hình HTVN trong đó xem xét sự biến động của các yếu tố như sở thích của người dùng, sự thay đổi về mức độ hấp dẫn của sản phẩm, dịch vụ sẽ cho phép đưa ra các khuyến nghị sát với thực tiễn hơn.

Bên cạnh đó, khi thu thập thông tin về các đánh giá của người dùng đối với các sản phẩm, dữ liệu này thường chứa các thông tin thể hiện tính do dự hoặc không nhất quán. Ví dụ, một người dùng có thể đánh giá ở hai mức thang đo khác nhau cho các đối tượng tương tự, hay đánh giá về cùng một đối tượng với các mức thang đo khác nhau trong những thời điểm khác nhau. Hơn thế nữa, hiện nay để thu thập đánh giá thể hiện sở thích của người dùng các hệ thống thông tin hầu hết sẽ sử dụng cơ chế xếp hạng, ví dụ hệ thống lựa chọn số sao gắn với sản phẩm, dịch vụ. Cách thức này rất dễ triển khai, tuy nhiên điểm yếu của cơ chế đó là nó luôn chứa thông tin không

chắc chắn. Người dùng thường chắc chắn hơn khi lựa chọn các xếp hạng ở các cực của thang đo, nhưng họ thường do dự hơn về lựa chọn các xếp hạng ở các cấp độ trung gian của thang đo. Do đó, xây dựng một phương pháp tốt và thực tế hơn để xử lý dữ liệu không chắc chắn, hay sự do dự của người dùng là cần thiết.

Để xử lý vấn đề này, một trong các tiếp cận nổi bật là sử dụng lý thuyết mờ để biểu diễn và xử lý thông tin. Lý thuyết về tập mờ và logic mờ đã được Zadeh giới thiệu vào năm 1965 [81] trong đó tập mờ được đặc trưng bởi một hàm thuộc có thể nhận bất kì giá trị nào trong khoảng đóng $[0,1]$. Từ lý thuyết về tập mờ ban đầu, có nhiều lý thuyết về các tập mờ mở rộng được các nhà nghiên cứu giới thiệu và ứng dụng. Giữa các lý thuyết về tập mờ mở rộng đã được đề xuất [82]–[87], tập mờ trực cảm và tập Neutrosophic có thể biểu diễn thông tin không chắc chắn, do dự một cách phù hợp [55], [85], [86], [88]. Trong luận án này, tập mờ trực cảm là trọng tâm nghiên cứu để xây dựng một biểu diễn phù hợp cho thông tin về đánh giá của người dùng với các sản phẩm, dịch vụ [89]–[91]. Tập mờ trực cảm được đề xuất bởi Atanasov [56], [57], là một bộ bao gồm ba yếu tố thành phần: độ thuộc, độ không thuộc và độ do dự và ràng buộc giữa các thành phần này. Tập mờ trực cảm với các thành phần của nó rất phù hợp để biểu diễn thông tin về đánh giá của người dùng về một sản phẩm, trong đó độ do dự thể hiện tính không chắc chắn của người dùng về đánh giá với một sản phẩm. Bên cạnh đó, so với một số tập mờ mở rộng khác, tập mờ trực cảm được nghiên cứu và phát triển trên cơ sở toán học vững chắc với nhiều phép toán phù hợp cho việc phát triển HTVN.

Dựa trên hai yếu tố quan trọng kể trên để có thể xây dựng HTVN gần với thực tiễn hơn, trong Chương 3 này NCS trình bày nghiên cứu về “Hệ tư vấn nhóm động theo tiếp cận tính toán mờ đảm bảo tính công bằng”. Mục tiêu chính cần đạt được như sau:

- Nghiên cứu HTVN động trên Tập mờ trực cảm để đồng thời xử lý thông tin về sự do dự, không chắc chắn trong HTVN và sử dụng tiếp cận động để xem xét tác động của yếu tố thời gian khi dự đoán đánh giá của một người dùng.

- Đề xuất mô hình HTVN động trên tập mờ trực cảm trong đó phép toán tổng hợp đánh giá của các thành viên sử dụng tích phân Choquet làm cơ sở để đảm bảo tính công bằng của kết quả khuyến nghị.

Bên cạnh đó, khi phát triển các mô hình HTVN, các siêu tham số sẽ có ảnh hưởng chi phối tới kết quả cuối cùng của hệ thống. Bên cạnh tiếp cận thử sai cho việc tìm kiếm các siêu tham số của mô hình, các nhà khoa học cũng đã phát triển một số phương pháp tìm kiếm siêu tham số tối ưu hơn. Trong luận án nghiên cứu và ứng dụng phương pháp học siêu tham số theo tiếp cận “tối ưu siêu tham số Bayesian” để tìm được các bộ siêu tham số phù hợp nhất cho mô hình thuật toán được phát triển.

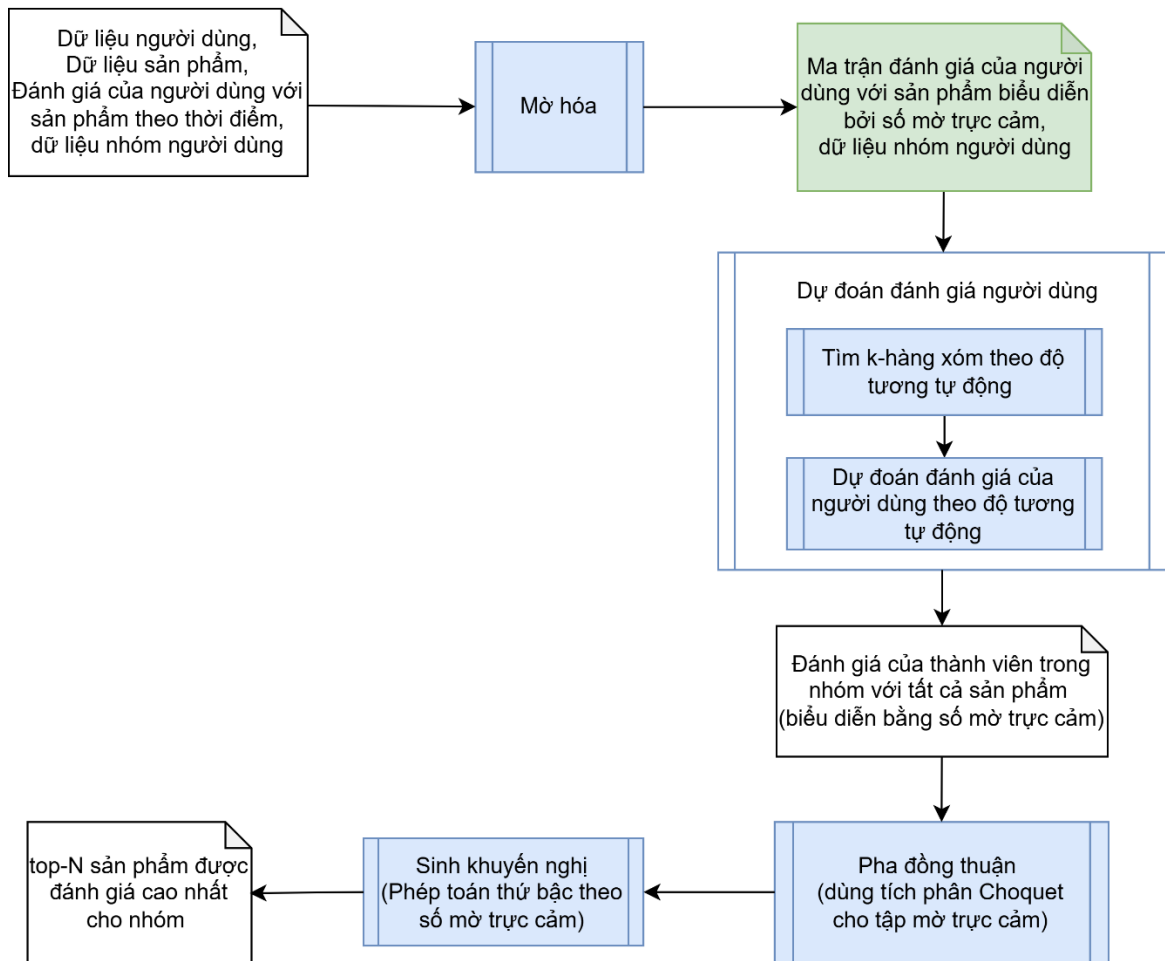
Kết quả nghiên cứu trong chương 3 được công bố trong một bài báo tạp chí thuộc danh mục SCIE và một bài hội thảo là công trình số 3 và 4 liệt kê trong phần “*Danh mục các công trình của tác giả*”.

3.2. Đề xuất mô hình Hệ tư vấn nhóm động trên tập mờ trực cảm

3.2.1. Đề xuất mô hình Hệ tư vấn nhóm động trên tập mờ trực cảm

a) Mô hình khái quát

Dựa trên tiếp cận xây dựng HTVN theo tiếp cận hợp khuyến nghị, mô hình hệ tư vấn nhóm động (HTVNĐ) trên tập mờ trực cảm được NCS đề xuất có các bước xử lý khái quát được biểu diễn theo sơ đồ trong hình 3.1. Trong đó những cải tiến quan trọng nhất nằm ở giai đoạn dự đoán đánh giá của người dùng trên nền tảng số mờ trực cảm và xử lý thông tin động và giai đoạn đồng thuận sử dụng tích phân Choquet cho tập mờ trực cảm.



Hình 3.1 Sơ đồ quy trình Hệ tư vấn nhóm động trên tập mờ trực cảm đề xuất

Các bước chính bao gồm:

- **Bước 1:** Mờ hóa ma trận dữ liệu đánh giá sản phẩm của người dùng.
- **Bước 2:** Tìm K hàng xóm gần nhất dựa trên độ tương tự động, *độ tương tự được tính với ảnh hưởng của sự khác nhau giữa các thời điểm đánh giá của người dùng.*
- **Bước 3:** Dự đoán đánh giá của người dùng trong nhóm dự trên *độ đo tương tự động.*
- **Bước 4:** Sử dụng phép toán tổng hợp đánh giá của các thành viên trong nhóm để ước lượng đánh giá của nhóm, *có xem xét tới sự hiệu chỉnh theo ảnh hưởng nhóm.*
- **Bước 5:** Đưa ra khuyến nghị dựa trên thuật toán khuyến nghị top-N

Để thấy, HTVN mờ động được đề xuất có thể xử lý thông tin đánh giá của người dùng với sản phẩm tốt hơn nhờ biểu diễn các đánh giá này bằng số mờ trực cảm. Tuy nhiên, khi sử dụng số mờ trực cảm để biểu diễn thông tin, mô hình HTVN đề xuất cũng yêu cầu các phép toán phức tạp hơn. Cụ thể là các phép toán liên quan đến xử lý dữ liệu động của HTVN như độ tương tự động, ảnh hưởng của thời điểm dự báo tới kết quả dự báo đánh giá. Việc phát triển các phép toán này là đóng góp có ý nghĩa quan trọng của nghiên cứu trong luận án.

b) Các công thức được sử dụng trong mô hình

Các công thức được sử dụng trong các bước của mô hình bao gồm: quá trình mờ hóa sử dụng hàm mờ hóa dạng Gauss; ảnh hưởng của thời gian được biểu diễn bởi dạng hàm phản ánh độ suy hao theo thời gian, và ảnh hưởng của nhóm tới các thành viên phụ thuộc vào kích thước nhóm. Chi tiết các công thức được đề xuất để phù hợp với từng bước trong thuật toán như sau.

Bước 1: Mờ hóa các đánh giá

Phương thức mờ hóa các đánh giá của người dùng được thể hiện trong công thức dưới đây.

$$(t_{ui}, r_{ui}) \rightarrow (t_{ui}, < \mu_{ui}, \nu_{ui}, \pi_{ui} >) \quad (3.1)$$

Trong đó t_{ui} là thời điểm người dùng u đánh giá sản phẩm i , $\pi_{ui} = 1 - (\mu_{ui} + \nu_{ui})$, và μ_{ui} , ν_{ui} được ước lượng bởi hàm dạng Gauss, được giới thiệu trong [62] như sau:

$$\mu_{ui}(r_{ui}) = \begin{cases} 1 & , r_{ui} = uth \\ a_1 \cdot \exp\left(-\frac{(r_{ui} - b_1)^2}{2c_1^2}\right) & , uth < r_{ui} < lth \\ 0 & , r_{ui} = lth \end{cases} \quad (3.2)$$

$$\nu_{ui}(r_{ui}) = \begin{cases} 0 & , r_{ui} = uth \\ a_2 \cdot \exp\left(-\frac{(r_{ui} - b_2)^2}{2c_2^2}\right) & , uth < r_{ui} < lth \\ 1 & , r_{ui} = lth \end{cases} \quad (3.3)$$

Trong đó uth là ngưỡng trên và lth là ngưỡng dưới. Các ngưỡng này và các tham số a_1, b_1, c_1 và a_2, b_2, c_2 được thiết lập dựa trên đặc điểm của dữ liệu.

Phương thức mờ hóa trên được xây dựng theo hướng đảm bảo các giá trị độ thuộc sẽ phản ánh mức độ yêu thích của người dùng với sản phẩm, độ không thuộc phản ánh mức độ không thích của người dùng và độ do dự thể hiện phần không chắc chắn trong đánh giá của người dùng.

Trong các phần tiếp theo, đánh giá của một người dùng u cho sản phẩm i thể hiện bằng số mờ trực cảm được ký hiệu là $a_{ui} = \langle \mu_{ui}, \nu_{ui}, \pi_{ui} \rangle$

Bước 2: Xác định độ tương tự động

Độ tương tự động khái quát: Giả thiết rằng mức độ tương tự của người dùng không chỉ phụ thuộc vào đánh giá cho sản phẩm của người dùng mà còn phụ thuộc vào thời điểm đánh giá. Điều đó có nghĩa là khoảng cách giữa thời gian hai người dùng đưa ra đánh giá về một sản phẩm, dịch vụ càng lớn thì những người dùng này càng ít giống nhau hơn. Trong nghiên cứu này, thời gian của đánh giá được chuyển đổi thành “độ dốc” dựa trên thời gian có gốc là thời điểm đánh giá của người dùng đầu tiên xuất hiện.

Định nghĩa 3.1: Cho t_{ui}, t_{vi} là thời điểm đánh giá của hai người dùng u, v đánh giá cùng một sản phẩm i . Ảnh hưởng của yếu tố thời gian lên độ tương tự của người dùng được đo bởi công thức:

$$f(\Delta t_i) = \frac{1}{1 + \lambda \Delta t} = \frac{1}{1 + \lambda |t_{ui} - t_{vi}|} \quad (3.4)$$

Từ đó, công thức tính độ tương tự động giữa hai người dùng u, v dựa trên đánh giá cho một sản phẩm i được thể hiện như sau:

$$Dsim(u, v) = Sim(u, v) \cdot f(\Delta t_i) \quad (3.5)$$

Công thức (3.5) và (3.4) cho thấy thời điểm đánh giá của hai người dùng càng xa nhau thì độ tương tự giữa hai người dùng sẽ giảm. Nếu thời điểm đánh giá của hai người dùng là cùng một thời điểm thì $t_{ui} - t_{vi} = 0$ thì $f(\Delta t) = \frac{1}{1 + \lambda |t_{ui} - t_{vi}|} = 1$, và do đó

$Dsim(u, v) = Sim(u, v)$, tức là độ tương tự của hai người dùng không chịu sự tác động của sự sai khác về thời gian đánh giá. Ngược lại, khi $t_{ui} - t_{vi} \neq 0$ thì $0 < f(\Delta t) = \frac{1}{1 + \lambda |t_{ui} - t_{vi}|} < 1$, do đó $Dsim(u, v) < Sim(u, v)$. Vậy so với cách tiếp cận dữ liệu tĩnh trong HTVN, cách tiếp cận đề xuất xem xét đến ảnh hưởng của thời gian làm thay đổi mức độ tương tự giữa các người dùng.

Mở rộng hơn, khi sử dụng độ tương tự động được phát triển từ độ tương tự Cosine cho số mờ trực cảm, chúng ta có độ tương tự động giữa hai người dùng u, v đã đánh giá chung tập hợp gồm n sản phẩm, dịch vụ sẽ được tính như sau.

Độ tương tự động phát triển từ độ tương tự Cosine:

$$DCos sim(u, v) = \frac{1}{n} \sum_{i=1}^n \frac{\mu_{ui} \cdot \mu_{vi} + \nu_{ui} \cdot \nu_{vi}}{\sqrt{\mu_{ui}^2 + \mu_{vi}^2} + \sqrt{\nu_{ui}^2 + \nu_{vi}^2}} f(\Delta t_i) \quad (3.6)$$

Bước 3: Dự đoán đánh giá của người dùng có xem xét tới ảnh hưởng của thời gian

Dựa trên độ tương tự động và tiếp cận của thuật toán k-hàng xóm gần nhất, trong HTVN sử dụng tiếp cận lọc cộng tác theo người dùng, ta cần phải đưa ra ước lượng về đánh giá của một người dùng cho một sản phẩm từ những người dùng lân cận của người dùng đó. Có nhiều cách xây dựng phép toán dự báo này, trong đó ta có thể sử dụng phép toán trung bình có trọng số cho số mờ trực cảm, hoặc phép toán trung bình Bonfferroni.

Nhắc lại, công thức tính giá trị trung bình có trọng số của tập mờ trực cảm cho một véc tơ $A = \{a_1, a_2, \dots, a_n\}$ được tính như sau:

$$IFAW(A, W) = \bigoplus_{i=1}^n a_i w_i = \left\langle 1 - \prod_{i=1}^n (1 - \mu_{a_i})^{w_i}, \prod_{i=1}^n \nu_{a_i}^{w_i} \right\rangle$$

So với cách tiếp cận này, phép toán trung bình Bonfferroni cho các số mờ trực cảm mềm dẻo hơn nhờ các tham số p, q có thể được hiệu chỉnh cho phù hợp với đặc trưng bài toán ứng dụng:

$$IFB^{p,q}(A) = \left\langle \left(1 - \prod_{\substack{i,j=1 \\ i \neq j}}^n \left(1 - \mu_{a_i}^p \mu_{a_j}^q \right)^{\frac{1}{n(n-1)}} \right)^{\frac{1}{p+q}}, 1 - \left(1 - \prod_{\substack{i,j=1 \\ i \neq j}}^n \left(1 - (1 - \nu_{a_i})^p (1 - \nu_{a_j})^q \right)^{\frac{1}{n(n-1)}} \right)^{\frac{1}{p+q}} \right\rangle$$

Trong luận án này, phép tính trung bình Bonfferroni đã được sử dụng để dự báo đánh giá của một người dùng cho một sản phẩm từ k-hàng xóm gần nhất, và các đánh giá đã biết của chính người dùng đó. Tiếp cận này được sử dụng phổ biến trong lọc cộng tác theo người dùng bởi vì nó phản ánh được đặc điểm của chính người dùng và ảnh hưởng của mỗi người dùng trong k-hàng xóm đến người dùng đó. Công thức ước lượng cụ thể như sau.

Giả sử để ước lượng đánh giá a_{ui} của một người dùng u cho sản phẩm i (biểu diễn trên số mờ trực cảm) ta có thể sử dụng một trong hai công thức sau.

$$a_{ui} = \bigoplus_{j=1}^n \frac{1}{n} a_{uj} \oplus IFAW(\tilde{A}_v, W_{DCosim(v,u)}) \quad (3.7)$$

$$a_{ui} = IFB^{p,q} a_u \oplus IFAW(\tilde{A}_v, W_{DCosim(v,u)}) \quad (3.8)$$

Trong đó:

$$IFAW(\tilde{A}_v, W_{DCosim(v,u)}) = \bigoplus_{v \in \hat{U}} \tilde{a}_{vi} Dcosim(v,u) \quad (1.9)$$

$$\tilde{a}_{vi} = a_{vi} - IFB^{p,q} a_v \quad (3.10)$$

Trong đó $\hat{U}$ là tập hàng xóm của người dùng u .

Trong các công thức trên, $IFB^{p,q} a_u$ thể hiện đánh giá trung bình của một người dùng u với tất cả sản phẩm họ đã đánh giá. $IFAW(\tilde{A}_v, W_{DCosim(v,u)})$ thể hiện ảnh hưởng của các hàng xóm tới dự báo đánh giá của người dùng u sử dụng độ tương tự động. $\tilde{a}_{vi} = a_{vi} - IFB^{p,q} a_v$ thể hiện sự khác biệt giữa đánh giá của mỗi hàng xóm với sản phẩm i và mức độ đánh giá trung bình của người hàng xóm đó với các sản phẩm đã đánh giá của họ.

Bước 4: Phép toán tổng hợp đánh giá của các thành viên trong nhóm.

Sau khi ước lượng được đánh giá của một người dùng với một sản phẩm từ đánh giá của các hàng xóm lân cận, đánh giá của một người dùng khi đặt trong các nhóm khác nhau sẽ được hiệu chỉnh theo đặc trưng của nhóm.

Hiệu chỉnh đánh giá của người dùng dựa trên ảnh hưởng của nhóm, cụ thể là kích thước nhóm:

Giả định là sự thỏa mãn của người dùng bị ảnh hưởng bởi quy mô nhóm và sự tương đồng giữa các thành viên trong nhóm. Giả thiết rằng người dùng sẽ cảm thấy ít hài lòng hơn khi tham gia một nhóm đông thành viên và các thành viên đó có sự khác biệt lớn với mình. Trong đó, sự thỏa mãn của người dùng được điều chỉnh theo công thức sau.

$$a_{ui}^g = eff_u^g . a_{ui} = \left(\frac{1}{|g|-1} \sum_{v \in g, v \neq u} sim(v, u) \right) a_{ui} \quad (3.11)$$

Trong đó a_{ui}^g là đánh giá hiệu chỉnh của người dùng u cho sản phẩm i khi người dùng này thuộc nhóm người dùng g .

Trong trường hợp $eff_u^g = 1$ đánh giá thể hiện sự yêu thích của thành viên không bị ảnh hưởng bởi nhóm, nếu $eff_u^g > 1$ đánh giá thể hiện sự yêu thích của thành viên sẽ tăng từ ảnh hưởng của nhóm và ngược lại.

Ngoài các công thức được trình bày ở trên, trong bước 4 sử dụng phép toán tổng hợp đánh giá của các thành viên trong nhóm dựa trên tích phân Choquet để ước lượng đánh giá của nhóm với sản phẩm. Các công thức sử dụng trong bước 4 này sẽ được trình bày ở phần tiếp theo.

Bước 5: Đưa ra khuyến nghị dựa trên thuật toán khuyến nghị top-N

Tại bước 5, các sản phẩm được sắp xếp theo thứ tự giảm dần dựa trên đánh giá của nhóm và tạo khuyến nghị theo thuật toán top-N. Trong đó phương thức so sánh các số mờ trực cảm được sử dụng là công thức sử dụng trong cách tính phép toán

tổng hợp đánh giá của các thành viên trong nhóm dựa trên tích phân Choquet nên cũng được trình bày lồng ghép trong đó.

3.2.2. Phép toán đồng thuận cho HTVN động trên tập mờ trực cảm

Pha đồng thuận trong HTVN mờ trực cảm được xây dựng với phép toán tích phân Choquet nhằm đảm bảo tính công bằng của khuyến nghị cho các người dùng trong nhóm. Tương tự với phép toán được phát triển trong Chương 2, phép toán tổng hợp đánh giá của các thành viên với tích phân Choquet đòi hỏi xây dựng một độ đo mờ dựa trên hàm dung lượng phi cộng tính, và trong chương này các hàm dung lượng được phát triển trong Chương 2 đều có thể được sử dụng lại. Tuy nhiên, do độ phức tạp tính toán của hệ tư vấn động theo tiếp cận mờ trực cảm cao hơn nhiều so với HTVN trên số rõ, do đó để đảm bảo tính khả thi trong tính toán độ đo mờ được áp dụng trong chương 3 sẽ đơn giản hơn so với phép toán trong Chương 2. Cụ thể trong tiếp cận ở Chương 3 chỉ sử dụng hàm dung lượng thứ nhất được trình bày trong Chương 2.

Với một hàm dung lượng đã xác định, định nghĩa của phép tích phân Choquet trên tập mờ trực cảm rời rạc được định nghĩa trong [92] như sau:

Định nghĩa 3.2 [92]: Cho $A = \{(\mu_{a_i}, \nu_{a_i}); i = 1, 2, \dots, n\}$ là tập các số mờ trực cảm thể hiện đánh giá trên tập các điều kiện X , và ξ là một hàm dung lượng trên tập X . Phép tính cho tích phân Choquet rời rạc của A được thể hiện trong công thức dưới đây:

$$\begin{aligned} IFC_{\xi}(A) &= IFC_{\xi}(\{\tilde{a}_1, \dots, \tilde{a}_n\}) = \tilde{a}_1 (\xi(A_1^{\uparrow}) - \xi(A_2^{\uparrow})) \oplus \tilde{a}_2 (\xi(A_2^{\uparrow}) - \xi(A_3^{\uparrow})) \oplus \dots \oplus \tilde{a}_n (\xi(A_n^{\uparrow}) - \xi(A_{n+1}^{\uparrow})) \\ &= \sum_{i=1}^n \oplus \tilde{a}_i (\xi(A_i^{\uparrow}) - \xi(A_{i+1}^{\uparrow})) \\ &= \left(1 - \prod_{i=1}^n (1 - \mu_{u_i})^{\xi(A_i^{\uparrow}) - \xi(A_{i+1}^{\uparrow})}, \prod_{i=1}^n \nu_{u_i}^{\xi(A_i^{\uparrow}) - \xi(A_{i+1}^{\uparrow})} \right) \end{aligned} \quad (3.12)$$

Trong đó $\{\tilde{a}_1, \dots, \tilde{a}_n\}$ là một sắp xếp tăng dần của các thành phần trong A , và tương ứng tập các điều kiện $A_i^{\uparrow} = (i, \dots, n)$, $A_{n+1}^{\uparrow} = \emptyset$.

Định nghĩa về thứ bậc với tập mờ trực cảm:

Hàm điểm số (Score function) và hàm độ chính xác (accuracy function) của các số mờ trực cảm:

Cho $a_1 = (\mu_{a_1}, \nu_{a_1})$, $b_1 = (\mu_{b_1}, \nu_{b_1})$ là hai số mờ trực cảm.

Cho định nghĩa về hàm điểm số và hàm độ chính xác của một số IF như sau:

$$S(a_1) = \mu_{a_1} - \nu_{a_1}, H(a_1) = \mu_{a_1} + \nu_{a_1}$$

Thứ bậc:

- Nếu $S(a_1) < S(b_1)$ thì a_1 là nhỏ hơn b_1 , và ghi là $a_1 \prec b_1$
- Nếu $S(a_1) = S(b_1)$ thì
 - Nếu $H(a_1) < H(b_1)$ thì a_1 là nhỏ hơn b_1 , ghi là $a_1 \prec b_1$
 - Nếu $H(a_1) = H(b_1)$ thì $a_1 = b_1$

Dựa trên công thức tính tích phân Choquet với tập mờ trực cảm, phép toán đồng thuận trong HTVN động trên tập mờ trực cảm được phát triển như sau:

Hàm dung lượng

Dựa theo nghiên cứu được thực hiện trong Chương 2, để nâng cao độ công bằng khuyến nghị cho nhóm người dùng, một hàm dung lượng thể hiện độ đo mờ được xây dựng để các khuyến nghị cho nhóm người dùng sinh ra từ HTVN động theo tiếp cận mờ trực cảm cân bằng giữa mục tiêu tối đa hóa tổng đánh giá của nhóm và sự công bằng giữa các thành viên trong nhóm.

Trong Chương 3 này, luận án sử dụng hàm dung lượng tương tự, nhưng rút gọn hơn so với các hàm dung lượng được đề xuất trong Chương 2. Cụ thể, trong Chương 3 chỉ sử dụng hàm dung lượng thứ nhất được đề xuất trong Chương 2.

Với hàm dung lượng được định nghĩa ở Chương 2, phép toán tổng hợp đánh giá của thành viên trong nhóm dựa trên tích phân Choquet được dùng để xác định đánh giá của một nhóm người dùng cho sản phẩm trong HTVN mờ động được tính như sau.

Định nghĩa 3.3: Cho nhóm người dùng $g = \{u_1, u_2, \dots, u_n\}$, và véc tơ

$R_g = \{(\mu_{u_i}, \nu_{u_i}); i = 1, 2, \dots, n\}$ là một tập các số mờ trực cảm thể hiện đánh giá của các

thành viên trong nhóm g cho một sản phẩm. Không mất tính tổng quát, ta giả sử R_g có thứ tự tăng dần. Cho ξ là một hàm dung lượng trên tập người dùng g , ta có phép toán tổng hợp đánh giá của các thành viên sử dụng tích phân Choquet cho R_g như sau:

$$IF_CA_{\xi}(R_g) = \left(1 - \prod_{i=1}^n (1 - \mu_{u_i})^{\xi(A_i^{\uparrow}) - \xi(A_{i+1}^{\uparrow})}, \prod_{i=1}^n \nu_{u_i}^{\xi(A_i^{\uparrow}) - \xi(A_{i+1}^{\uparrow})} \right) \quad (3.13)$$

Trong đó $A_i^{\uparrow}$ tập hợp các thành viên trong nhóm tương ứng, và $A_i^{\uparrow} = \{u_i, u_{i+1}, \dots, u_n\}$ là các tập con của g , $A_{n+1}^{\uparrow} = \phi$.

3.2.3. Các chiến lược tổng hợp đánh giá khác cho pha đồng thuận

Để đánh giá hiệu quả của mô hình HTVN động theo tiếp cận mờ trực cảm sử dụng tích phân Choquet, và do các nghiên cứu về HTVN trên số mờ trực cảm là chưa có trước đây, luận án đề xuất cách chuyển đổi các chiến lược đồng thuận phổ biến cho số rõ thành chiến lược đồng thuận áp dụng trên tập mờ trực cảm. Đây là một cách mở rộng tự nhiên của các phép toán tổng hợp đánh giá của các thành viên nhóm trong pha đồng thuận thường dùng và phản ánh đúng triết lý như tiếp cận áp dụng trong tập số thực. Các chiến lược phổ biến bao gồm Chiến lược trung bình cộng, Chiến lược ít thiệt thòi nhất, Chiến lược ưu thế, Chiến lược bỏ phiếu phê duyệt, Chiến lược luật Copeland được mở rộng để áp dụng trên các giá trị IF. Sau đây là mô tả ngắn gọn về các chiến lược tổng hợp được sử dụng.

Chiến lược trung bình “cộng” trên IFS (IF_AUS): Trong chiến lược này, phép toán trung bình số học có trọng số cho số mờ trực cảm được áp dụng để tính đánh giá của nhóm:

$$g_pref_i^g = IFAW(R_{gi}, W_g) = \bigoplus_{u=1}^{|g|} a_{ui} w_u = \left\langle 1 - \prod_{u=1}^{|g|} (1 - \mu_{ui})^{w_u}, \prod_{u=1}^{|g|} \nu_{ui}^{w_u} \right\rangle \quad (3.14)$$

Trong đó $a_{ui} = \langle \mu_{ui}, \nu_{ui}, \nu_{ui} \rangle$ là đánh giá của người dùng u trong nhóm g cho sản phẩm i thể hiện bằng số mờ trực cảm.

Chiến lược trung bình cộng cho thấy sản phẩm có đánh giá trung bình cao hơn từ các thành viên cao hơn sẽ được ưu tiên lựa chọn.

Chiến lược “ít thiệt thòi nhất” trên IFS (IF_LMS): Giá trị đánh giá nhỏ nhất trong các thành viên được sử dụng để làm giá trị đại diện đánh giá cho nhóm.

$$g_pref_i^g = \min_{u \in g}(a_{ui}) = \min_{u \in g}(\langle \mu_{ui}, \nu_{ui}, \nu_{ui} \rangle) \quad (3.15)$$

Trong chiến lược này, cần lưu ý rằng phép toán so sánh giữa hai số mờ trực cảm được sử dụng, và phép so sánh được trình bày trong mục 3.2.2 ở trên.

Chiến lược ưu thế trên IFS (IF_MPS): Giá trị đánh giá lớn nhất của các thành viên trong nhóm để làm đại diện đánh giá cho nhóm.

$$g_pref_i^g = \max_{u \in g}(a_{ui}) = \max_{u \in g}(\langle \mu_{ui}, \nu_{ui}, \nu_{ui} \rangle) \quad (3.16)$$

Tiếp cận trong chiến lược này ngược lại với chiến lược ít thiệt thòi nhất.

Chiến lược biểu quyết phê duyệt trên IFS (IF_AVS): Tổng hợp biểu quyết của các thành viên được tính để làm giá trị biểu diễn cho đánh giá của nhóm với một sản phẩm.

$$g_pref_i^g = \sum_{u \in g} p_{ui} \quad (3.17)$$

Trong đó $p_{ui} = 1$ nếu người dùng u “biểu quyết” cho sản phẩm i trong tập sản phẩm ứng cử viên I' và $p_{ui} = 0$ khi sản phẩm này không được biểu quyết.

Chiến lược luật Copeland trên IFS (IF_CRS): Dựa trên tư tưởng luật Copeland, giá trị thể hiện đánh giá của một nhóm về sản phẩm được thay thế bằng tổng hợp kết quả so sánh sản phẩm đó với các sản phẩm khác trong nhóm sản phẩm ứng viên. Cụ thể:

$$g_pref_i^g = \sum_{j \in I'} CI_{ij} \quad (3.18)$$

Trong đó $CI_{ij} = 1$ nếu sản phẩm i “đánh bại” j trong danh sách ứng cử viên I' . Khái niệm “đánh bại” ở đây được hiểu là số thành viên trong nhóm đánh giá i cao hơn j là lớn hơn số thành viên đánh giá ngược lại. $CI_{ij} = -1$ nếu j “đánh bại” i và còn lại thì $CI_{ij} = 0$.

3.3. Giả mã của giải thuật

3.3.1. Giả mã của thuật toán Hệ tư vấn nhóm động trên tập mờ trực cảm

Giả mã của thuật toán hệ tư vấn nhóm động trên tập mờ trực cảm tiếp cận đảm bảo độ công bằng của khuyến nghị theo đề xuất ở trên được trình bày sau đây:

Giả mã của thuật toán Hệ tư vấn nhóm động trên tập mờ trực cảm – IF_CIS

Đầu vào: Nhóm người dùng g , tập người dùng U , tập sản phẩm I , ma trận đánh giá R^t thể hiện đánh giá của người dùng U với các sản phẩm I theo thời điểm.

Đầu ra: Khuyến nghị tập N sản phẩm phù hợp nhất cho g , $N \leq |I'|$

1. Mờ hóa ma trận đánh giá $R_{IF}^t \leftarrow R^t$ theo công thức (3.1)
2. $kich_thuoc_nhom = |g|$
3. Gán tập ứng viên $I' = \emptyset$
4. for i in range ($kich_thuoc_nhom$):
5. Xác định tập hàng xóm N_{u_i} của u_i bằng độ đo (3.6)
6. foreach sản phẩm, dịch vụ chưa đánh giá $i \in I$ của người dùng u_i
7. dự đoán $IF_R_{u_i}^t$ bằng phép toán (3.8)
8. Lựa chọn N_{u_i} sản phẩm, dịch vụ phù hợp cho người dùng u_i và thêm các sản phẩm, dịch vụ đó vào I' nếu nó chưa tồn tại.
9. trong_so=[None]* $kich_thuoc_nhom$
10. trong_so_hc=[None]* $kich_thuoc_nhom$
11. for i in range ($kich_thuoc_nhom$):
12. tính trong_so[i] theo công thức (2.4)
13. tính trong_so_hc[i] theo công thức (2.7)

```

# hàm dung lượng
14. def get_capacity(Permutation, Index):
15.     gia_tri_ham_dl=0
16.     for i in range (Index, len(Permutation)):
17.         gia_tri_ham_dl += trong_so[i]
18.         if (len(Permutation)-Index>1):
19.             for j in range (Index, len(Permutation)):
20.                 gia_tri_ham_dl += cap_ trong_so_hc[j]
21.     return gia_tri_ham_dl
        # tính đánh giá của nhóm cho một sản phẩm,dịch vụ
22.     so_luong_ung_vien= $|I'|$ 
23.     danh_gia_nhom=[None]*so_luong_ung_vien
24.     for i in range (so_luong_ung_vien):
25.         user_permutation = get_permutation( $g, I'_i, R$ )
            # Phép toán tổng hợp cho tập mờ trực cảm dựa trên
            #tích phân Choquet.
26.         Tính so_luong_ung_vien[i] bằng công thức (3.13)
            # Đưa ra khuyến nghị
27.     Từ véc tơ lưu đánh giá của nhóm danh_gia_nhom lựa chọn
        và trả về top-N sản phẩm phù hợp cao nhất.

```

Giả mã của giải thuật IF_CIS được trình bày dưới dạng giả mã python. Giả mã này có chuyển đổi nhanh chóng sang một chương trình thực nghiệm lập trình bằng ngôn ngữ lập trình python.

3.3.2. Độ phức tạp của thuật toán

Về độ phức tạp tính toán, có thể thấy rằng thuật toán IF_CIS đề xuất có ba quy trình tuần tự chiếm ưu thế: 1) mờ hóa ma trận đánh giá của người dùng; 2) dự đoán

sở thích của thành viên nhóm và 3) pha đồng thuận sử dụng toán tử tổng hợp đánh giá của thành viên nhóm dựa trên tích phân Choquet.

Trong phần thứ nhất, khối lượng công việc của quá trình mờ hóa phụ thuộc vào số lượng người dùng và số lượng sản phẩm dịch vụ. Bên cạnh đó, độ phức tạp để mờ hóa một đánh giá là hằng số. Độ phức tạp của quá trình này được biểu thị bằng công thức sau:

$$\begin{aligned} T_{fuzzification}(|U|, |I|) &= O(|U| \cdot |I|) O(fuzzilize(r_{ui})) \\ &= O(c \cdot |U| \cdot |I|) \end{aligned} \quad (3.19)$$

Trong phần thứ hai, với một nhóm người dùng g để xác định đánh giá của mỗi thành viên cho các sản phẩm sẽ cần xác định k -hàng xóm gần nhất của thành viên đó theo độ tương tự động thể hiện ở công thức (3.6), từ đó sẽ tính được đánh giá của thành viên cho mỗi sản phẩm theo công thức . Độ phức tạp tính toán của phần này thể hiện bởi công thức:

$$\begin{aligned} T_{prediction}(|g|, |U|, |I|) &= O(|g|) \cdot (O(|U| \cdot |I| \cdot c) + O(k) \cdot O(|I|^2 \cdot c)) \\ &= O(|g| \cdot k \cdot |I|^2 \cdot c) \end{aligned} \quad (3.20)$$

Trong đó, $O(|U| \cdot |I| \cdot c)$ là độ phức tạp để tìm các hàng xóm của một người dùng và $O(k) \cdot O(|I|^2 \cdot c)$ là độ phức tạp tính toán để ước lượng giá trị đánh giá của người dùng với các sản phẩm chưa được đánh giá. Và thông thường số người dùng sẽ rất nhỏ so với số lượng sản phẩm.

Trong phần thứ ba, với mỗi sản phẩm trong tập ứng viên I' , độ phức tạp tính toán của pha đồng thuận phụ thuộc vào chiến lược đồng thuận. Với toán tử tổng hợp đánh giá thành viên nhóm sử dụng tích phân Choquet cho tập các số mờ trực cảm, chúng ta cần tính hàm dung lượng cho các tập con $A_j = (1, \dots, |g|)$. Giá trị độ đo mờ được tính bằng công thức (3.12) và (2.5), độ phức tạp tính toán là $O(|g|^2 \cdot |I| \cdot c)$. Do đó, độ phức tạp của pha đồng thuận được đề xuất sẽ là $T_{aggregation}(|g|, |I'|) = O(|g|^2 \cdot |I| \cdot |g|)$.

Với các chiến lược phổ biến bao gồm Chiến lược trung bình cộng, Chiến lược ít thiệt thòi nhất, Chiến lược ưu thế, Chiến lược bỏ phiếu phê duyệt, Chiến lược luật

Copeland được mở rộng để áp dụng trên các giá trị IF, độ phức tạp có thể ước lượng được. Trong đó các chiến lược thực hiện phép toán tổng hợp đại số trên véc tơ các số mờ trực cảm nói chung đều có độ phức tạp $O_{aggregation}(|g|.c)$, ngoại trừ chiến lược luật Copeland có độ phức tạp $O_{aggregation}(|I'|.c)$.

Trong các công thức trên $O(c)$ thể hiện độ phức tạp hằng số.

Độ phức tạp của toàn bộ thuật toán IF_CIS bằng hợp của độ phức tạp ba công việc trên:

$$T(|U|, |I|, |g|) = T_{fuzzilization}(|U|, |I|) + T_{prediction}(|U|, |I|, |g|) + T_{aggregation}(|I'|, |g|) \quad (3.21)$$

3.3.3. Phương pháp học tham số của thuật toán

Trong thuật toán IF_CIS một số tham số sẽ ảnh hưởng tới hiệu quả việc dự đoán đánh giá của một người dùng với các sản phẩm, dịch vụ. Các tham số này bao gồm khoảng thời gian được sử dụng để xem xét sự biến động về sở thích người dùng và sự hấp dẫn của sản phẩm trong công thức (3.8); hệ số λ thể hiện ảnh hưởng của thời gian tới sự thay đổi trong đánh giá; và kích thước tập hàng xóm được sử dụng trong thuật toán K-hàng xóm gần nhất.

Tổng quan nghiên cứu cho thấy, với bài toán tìm kiếm tham số phù hợp cho một mô hình tuy đã có các nghiên cứu được công bố nhưng không có một phương pháp nào hoàn toàn tối ưu [93]. Một số hướng nghiên cứu điển hình trong bài toán tìm kiếm tham số mô hình như giải thuật di truyền, tìm kiếm lưới, tối ưu hóa Bayes là các phương pháp được ứng dụng phổ biến [94]. Trong những năm gần đây, phương pháp tối ưu hóa Bayes (BO) được dùng ngày càng phổ biến như một kỹ thuật tối ưu hóa toàn cục cho các bài toán tối ưu siêu tham số (HPO) [95]–[97]. Do đó, trong nghiên cứu này, để tìm được bộ tham số tối ưu cho giai đoạn dự đoán đánh giá của HTNV, một quy trình tìm tham số tối ưu được thiết kế dựa trên phương pháp tối ưu hóa Bayes.

Thuật toán tối ưu hóa Bayes bao gồm quá trình lặp thể hiện chiến lược tìm kiếm một bộ tham số $\tilde{p}$ của mô hình sao cho giá trị của hàm mất mát là tối thiểu:

$\tilde{p} = \arg \min_{p \in P} f_{loss}(p)$. Trong đó p thể hiện không gian tìm kiếm các tham số của mô hình.

Trong bài toán IF_CIS, NCS đề xuất hàm mất mát được thiết kế dựa trên khoảng cách giữa đánh giá của cá nhân với một sản phẩm được dự đoán và đánh giá thực tế của cá nhân đó, và đều biểu diễn trên số mờ trực cảm. Chi tiết hơn, hàm mất mát sẽ thể hiện sai lệch trong dự báo đánh giá của các thành viên trong một tập các nhóm người dùng với một tập các sản phẩm, dịch vụ.

Hàm mất mát cho phương pháp tìm tham số của thuật toán IF_CIS:

Cho tập các nhóm người dùng G và tập các sản phẩm, dịch vụ I đã được đánh giá, ta có hàm mất mát được tính như sau.

$$\begin{aligned} f_{loss}(p, G, I) &= \frac{1}{|G|} \sum_{g \in G} \frac{1}{|g|} \sum_{u \in g} \frac{1}{|I|} \sum_{i \in I} e_{IFS}(r_{ui}, \hat{r}_{ui}) \\ &= \frac{1}{|G|} \sum_{g \in G} \frac{1}{|g|} \sum_{u \in g} \frac{1}{|I|} \sum_{i \in I} \sqrt{\frac{1}{2} \left((\mu_{r_{ui}} - \mu_{\hat{r}_{ui}})^2 + (\nu_{r_{ui}} - \nu_{\hat{r}_{ui}})^2 + (\pi_{r_{ui}} - \pi_{\hat{r}_{ui}})^2 \right)} \end{aligned} \quad (3.22)$$

Trong đó r_{ui} là đánh giá thực của người dùng u với sản phẩm i và $\hat{r}_{ui}$ là giá trị đánh giá dự báo.

Hàm mất mát kể trên thể hiện sự chênh lệch giữa giá trị về sở thích cá nhân được dự báo và các đánh giá thật, trong đó sự chênh lệch được tính bằng độ đo khoảng cách Euclide cho các số mờ trực cảm. Trong đó, khác với các nghiên cứu khác thường sử dụng một hàm đánh giá điểm thông qua độ thuộc, độ không thuộc mà bỏ qua độ do dự. Nhưng trong bài toán HTVN, thành phần độ do dự có ảnh hưởng nhất định đến quyết định của người dùng, do đó công thức (3.22) đã thể hiện được điều này.

Bên cạnh hàm mất mát, không gian tìm kiếm của bộ các tham số là một phần quan trọng của thuật toán tối ưu Bayes. Không tìm kiếm các bộ tham số được xác định từ quá trình nghiên cứu các tài liệu liên quan và phân tích từ bộ dữ liệu thực tiễn. Trong đó với những lĩnh vực có tính chất đặc biệt, đặc thù cần tham khảo ý kiến chuyên gia lĩnh vực đó để có thể thiết kế không gian các tham số cho tối ưu.

3.4. Thực nghiệm thuật toán

3.4.1. Mục tiêu thử nghiệm

Để thực hiện đánh giá thuật toán hệ tư vấn nhóm mờ động trên tập nền mờ trực cảm, luận án thực nghiệm và so sánh thuật toán đề xuất với thuật toán đi theo các chiến lược đồng thuận thường được sử dụng trong các nghiên cứu trước đó. Đầu ra của các thuật toán HTVN mờ này đều là đánh giá về mức độ của các sản phẩm và dịch vụ với một nhóm thể hiện trên số mờ trực cảm và một xếp hạng khuyến nghị dựa trên thứ bậc của các giá trị đánh giá đó. Về độ phức tạp tính toán, việc so sánh được thực hiện trên nền tảng so sánh lý thuyết dựa trên đánh giá về độ phức tạp thời gian O lớn trình bày ở mục 3.3.2. Do đó, mục tiêu so sánh trong thực nghiệm tập trung vào độ phù hợp của các thuật toán xét theo khía cạnh tối đa hóa lợi ích tổng thể của nhóm và độ công bằng giữa các thành viên trong nhóm. Hai mục tiêu này được gắn liền với nhau để đảm bảo không có một đánh giá quá thiên lệch về một xu thế duy nhất.

3.4.2. Dữ liệu thực nghiệm

Dữ liệu thực nghiệm của thuật toán hệ khuyến nghị nhóm động theo tiếp cận mờ trực cảm được thực hiện trên cùng bộ dữ liệu MovieLens-1M đã được trình bày trong Chương 2. Dữ liệu nhóm được sinh ra bởi thuật toán tạo nhóm người dùng, thuật toán này được trình bày trong mục 2.4.2 và không được trình bày lại ở đây.

3.4.3. Phương thức đánh giá

Như đã đề cập ở trên, để đánh giá được sự phù hợp của các thuật toán ta cần tính được đánh giá của một nhóm với sản phẩm, dịch vụ được khuyến nghị, và xác định giá trị độ đo độ công bằng. Nhưng do các đánh giá của nhóm về các sản phẩm dịch vụ được thể hiện trên giá trị đo là một số mờ trực cảm nên cần phát triển phép toán để đo mức độ thỏa mãn và độ công bằng trên tập số mờ trực cảm này. Các phép toán được xây dựng như sau.

a) Hàm hiệu quả

Hàm hiệu quả là phép toán cho phép ước lượng đánh giá của một người dùng, nhóm người dùng với các sản phẩm, dịch vụ thể hiện trên số mờ trực cảm thành một giá trị độ đo thể hiện bằng số thực $[0,1]$. Về mặt nguyên tắc, đây thực tế là một quá trình giải mờ. Quá trình giải mờ này đảm bảo rằng, những đánh giá với độ yêu thích cao hơn phải có giá trị của hàm hiệu quả là lớn hơn. Trong quá trình mờ hóa, thành phần độ thuộc sẽ phản ánh mức độ yêu thích, thành phần độ không thuộc sẽ phản ánh mức độ không thích và thành phần còn lại phản ánh độ do dự của người dùng khi đánh giá về một sản phẩm. Do đó, hàm hiệu quả sẽ phụ thuộc nhiều nhất vào giá trị độ thuộc.

Trong luận án này, hàm hiệu quả sẽ sử dụng quá trình giải mờ như sau.

Giả sử đánh giá của một người dùng hay nhóm người dùng với một sản phẩm, dịch vụ được thể hiện bằng một số mờ trực cảm $r_{ui} = \{\mu_{r_{ui}}, \pi_{r_{ui}}, v_{r_{ui}}\}$, quá trình ước lượng hàm hiệu quả là một quá trình giải mờ theo “trọng tâm” [98], [99]. Công thức thể hiện hàm hiệu quả theo phép toán giải mờ theo trọng tâm như sau [100]:

$$utility_score(r_{ui}) = \frac{\sum_{i=1}^N w_i \mu_{r_{ui}}}{\sum_{i=1}^N w_i} \quad (3.23)$$

Trong đó I là một tập gồm N sản phẩm được khuyến nghị cho người dùng u , r_{ui} là tập hợp đánh giá của người dùng u với từng sản phẩm trong I , và w_i thể hiện mức độ ưu tiên cho từng sản phẩm, dịch vụ. Khi coi các sản phẩm vụ như nhau thì trọng số đối với các sản phẩm là như nhau và $w_i = 1, \forall i \in (1, N)$.

Có thể thấy, công thức giải mờ như trên chỉ sử dụng giá trị độ thuộc, hay giá trị biểu diễn độ yêu thích để đánh giá hàm hiệu quả. Xét một cách đầy đủ, hàm hiệu quả cần ước lượng ảnh hưởng của sự không yêu thích và mức độ do dự của một người dùng với sản phẩm khi so sánh. Do đó, trước khi thực hiện giải mờ, một biến đổi được thực hiện để chuyển đổi giá trị $r_{ui} = \{\mu_{r_{ui}}, \pi_{r_{ui}}, v_{r_{ui}}\}$ thành một giá trị được chuẩn

hóa $r_{ui}^* = \{\mu_{r_{ui}^*}, v_{r_{ui}^*}\}$. Phương thức chuẩn hóa được thực hiện bằng một trong hai cách [100] như sau:

(*)

$$r_{ui}^* = \langle \mu_{r_{ui}} + \alpha \pi_{r_{ui}}, v_{r_{ui}} + (1 - \alpha) \pi_{r_{ui}} \rangle \quad (3.24)$$

(**)

$$r_{ui}^* = \left\langle \mu_{r_{ui}} + \frac{\mu_{r_{ui}}}{\mu_{r_{ui}} + v_{r_{ui}}} \pi_{r_{ui}}, v_{r_{ui}} + \frac{v_{r_{ui}}}{\mu_{r_{ui}} + v_{r_{ui}}} \pi_{r_{ui}} \right\rangle \quad (3.25)$$

Trong đó độ do dự của số mờ trực cảm được chuẩn hóa là

$$\pi_{r_{ui}^*}(x) = 1 - \mu_{r_{ui}^*}(x) - v_{r_{ui}^*}(x).$$

b) Các độ đo hiệu quả của thuật toán

Độ đo sai số

Độ đo đầu tiên và cũng là độ đo thường được sử dụng trong hệ khuyến nghị đó là độ đo sai số của khuyến nghị. Độ đo này thể hiện độ lệch giữa đánh giá của nhóm người dùng được dự báo cho một sản phẩm và đánh giá thực của nhóm người dùng với sản phẩm đó. Đánh giá thực của một nhóm người dùng cho một sản phẩm được ước lượng bởi giá trị trung bình đánh giá của tất cả các thành viên trong nhóm. Độ lệch của dự báo so với giá trị thật được mở rộng từ độ đo sai số toàn phương trung bình (RMSE). Công thức tính sai số như sau:

$$Error(G, I) = \frac{1}{|G|} \sum_{g \in G} \frac{1}{|g|} \sum_{u \in g} \frac{1}{|I'|} \sum_{i \in I'} e_{IFS}(r_{ui}, \hat{r}_{ui}) \quad (3.26)$$

Trong đó G là tập hợp các nhóm người dùng, I là tập hợp các sản phẩm, dịch vụ được khuyến nghị cho tất cả các nhóm, I' là các sản phẩm dịch vụ được khuyến nghị cho một nhóm nào đó mà các sản phẩm đó có giá trị đánh giá thực. r_{ui} là giá trị đánh giá dự báo của người dùng u đối với sản phẩm i , còn $\hat{r}_{ui}$ là đánh giá thực tương ứng. Độ lệch giữa hai đánh giá được tính theo công thức tính khoảng cách Euclidean giữa hai số mờ trực cảm $e_{IFS}(r_{ui}, \hat{r}_{ui})$. Độ đo khoảng cách đã phản ánh sai số toàn phương trung bình giữa ba thành phần độ thuộc, độ không thuộc và độ do dự của một

số mờ trực cảm, nên trong công thức trên chúng ta không thực hiện phép toán bình phương khoảng cách hai số mờ trực cảm. Khi cần thiết, chúng ta có thể mở rộng công thức tính sai số ở trên.

Lưu ý rằng, công thức tính sai số ở trên tương tự như công thức xây dựng hàm mất mát cho phương thức học tham số theo Bayse. Tuy nhiên, hàm huấn luyện tìm tham số tối ưu chỉ thực hiện trong pha dự báo mà không bao gồm pha đồng thuận. Độ đo sai số được thực hiện sau khi toàn bộ quá trình tư vấn nhóm kết thúc, tức là sau khi pha đồng thuận đã kết thúc.

Độ đo đánh giá trung bình của nhóm

Nhắc lại là mục tiêu của HTVN động theo tiếp cận mờ trực cảm (IF_CIS) được thiết kế để đạt được sự cân bằng giữa mục tiêu tối đa hóa tổng đánh giá của nhóm và tối thiểu hóa sự khác biệt giữa các thành viên. Bên cạnh đó, đánh giá của các thành viên đều thể hiện bởi một số mờ trực cảm nên giá trị thể hiện sự đánh giá trung bình của nhóm với khuyến nghị sẽ được ước lượng qua đánh giá trung bình của nhóm với tất cả các sản phẩm trong khuyến nghị I được tính bởi công thức sau.

$$g - pref(g, I) = \frac{1}{|I|} \sum_{i \in I} \left(\frac{\sum_{u \in g} utility_score(r_{ui})}{|g|} \right) \quad (3.27)$$

Trong đó r_{ui} là đánh giá của một người dùng u trong nhóm g với một sản phẩm i . Công thức trên thể hiện đánh giá trung bình của nhóm với mỗi sản phẩm là độc lập với nhau.

Các độ đo về độ công bằng

Khái niệm công bằng có thể được tùy biến khác nhau trong HTVN, do các nghiên cứu tập trung vào khía cạnh nào của bài toán. Trong luận án này, một số độ đo công bằng được công bố sẽ được sử dụng lại để đánh giá hiệu năng của thuật toán. Ngoài ra, một độ đo độ công bằng mới được đề xuất dựa trên việc phát triển khái niệm tính bình đẳng (equity) thường được sử dụng trong xã hội đó là tính bình đẳng theo độ đo GINI. Độ đo mới được đưa vào nhằm mục tiêu đánh giá HTVN mờ sát với thực tế ứng dụng hơn.

- **Độ đo tỷ lệ hài lòng:**

Thiết lập độ đo công bằng đầu tiên được sử dụng đó là tỷ lệ thành viên hài lòng với khuyến nghị trên tổng số người của nhóm. Độ đo này cho thấy tỷ lệ này càng lớn thì số người hài lòng càng cao.

$$fairness(g, i) = \frac{\left| \bigcup_{u \in g} r_{ui} \succ r_i^{th} \right|}{|g|} \quad (3.28)$$

Trong đó r_i^{th} là đánh giá trung bình của mỗi sản phẩm được tính bằng công thức sau.

$$r_i^{th} = IFAW(R_i, W) = \bigoplus_{u=1}^n r_{ui} w_u = \left\langle 1 - \prod_{u=1}^n (1 - \mu_{r_{ui}})^{w_u}, \prod_{u=1}^n \nu_{r_{ui}}^{w_u} \right\rangle \quad (3.29)$$

Với R_i là tập hợp các đánh giá đã biết với một sản phẩm i .

Độ đo tỷ lệ hài lòng của nhóm người dùng với một nhóm sản phẩm được tính bằng trung bình cộng độ đo tỷ lệ hài lòng cho từng sản phẩm trong nhóm sản phẩm đó. Độ đo tỷ lệ hài lòng có giá trị thuộc đoạn $[0, 1]$, trong đó giá trị độ đo càng lớn thể hiện mức độ công bằng càng cao.

- **Độ đo công bằng:**

Mặc dù độ đo tỷ lệ hài lòng là một độ đo được đề cập nhiều trong các nghiên cứu về tính công bằng của khuyến nghị, dễ thấy rằng độ đo trên không cho thấy sự khác biệt hay bất bình đẳng giữa các thành viên trong nhóm. Do đó, độ đo thứ hai được sử dụng để đánh giá sự khác biệt trong đánh giá của từng thành viên nhóm với sản phẩm được khuyến nghị. Độ đo này được gọi là độ đo công bằng, và được tính như sau.

$$equity(g, I) = \frac{1}{|I|} \sum_{i \in I} \left\{ \frac{1}{|g|^2} \sum_{l=1}^{|g|} \sum_{k=1}^{|g|} dis(r_{li}, r_{ki}) \right\} \quad (3.30)$$

Trong đó sự khác biệt giữa đánh giá của hai người dùng l, k cho một sản phẩm i được tính bằng khoảng cách giữa hai giá trị mờ trực cảm được trình bày trong mục 1.3.2.

- **Độ đo công bằng GINI:**

Độ đo công bằng thứ ba là độ đo được đề xuất mới dựa trên khái niệm về chỉ số GINI thể hiện mức độ bất bình đẳng giữa một nhóm người dùng. Chỉ số bất bình đẳng là một chỉ số được sử dụng rất rộng rãi trong nghiên cứu về kinh tế xã hội, cho phép ước lượng mức độ bất bình đẳng của một cộng đồng. Trong HTVN mờ động, độ đo công bằng GINI được đề xuất và tính như sau:

$$Gini_equity(g, I) = \frac{\sum_{l=1}^{|g|} \sum_{k=1}^{|g|} |utility_score(r_{li}) - utility_score(r_{ki})|}{2|g| \sum_{u=1}^{|g|} utility_score(r_{ui})} \quad (3.31)$$

Với độ đo công bằng GINI, giá trị càng nhỏ thì sự bất bình đẳng là càng nhỏ và như vậy khuyến nghị đề xuất bởi HTVN là càng tốt.

3.4.4. Kết quả và bàn luận

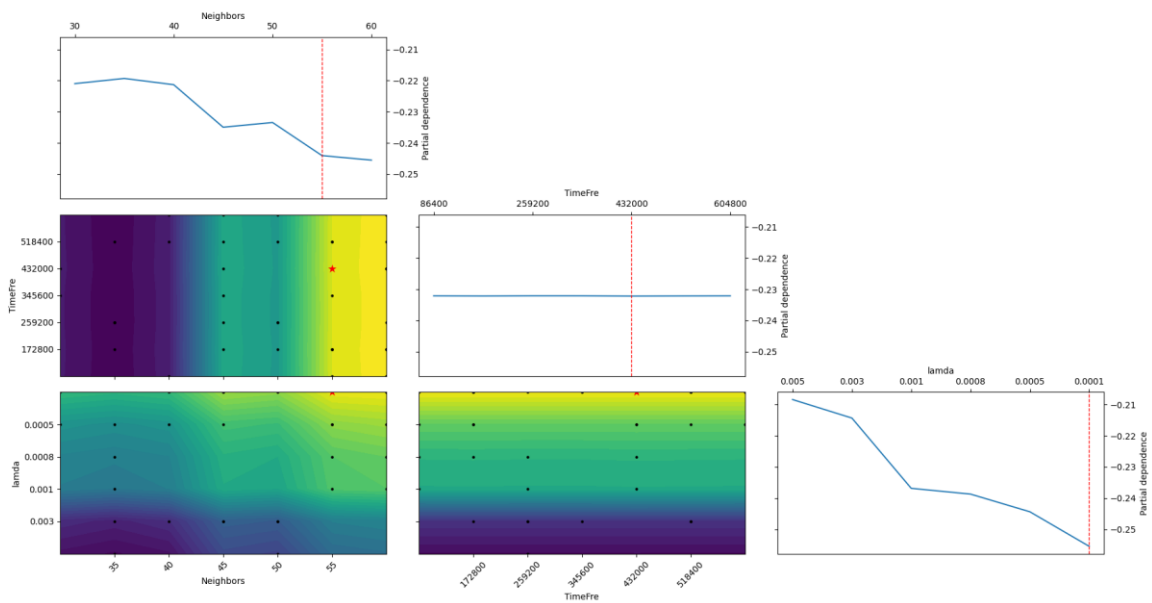
Trong phần kết quả, luận án sẽ trình bày phân kết quả huấn luyện mô hình học tham số được sử dụng bởi phương pháp học tối ưu Bayes. Tham số đưa ra bởi thuật toán này sẽ được dùng để thiết lập cho các bước dự báo đánh giá trong pha đầu tiên của HTVN động theo tiếp cận mờ trực cảm. Có nghĩa là các mô hình thực nghiệm sẽ sử dụng cùng một hệ thống tham số như nhau cho cùng một hệ thống dữ liệu đầu vào như sau.

a) Kết quả phương pháp học tham số Bayes

Khi ứng dụng phương pháp tìm tham số bằng thuật toán tối ưu Bayes phần quan trọng đầu tiên là thiết kế không gian tìm kiếm cho tham số. Với các bài toán có không gian tìm kiếm tham số là không gian vô hạn, việc thiết kế không gian tìm kiếm phù hợp sẽ giúp quá trình hội tụ của thuật toán đạt được trong khoảng thời gian chấp nhận được.

Dựa trên phân tích đặc điểm của bộ dữ liệu, bộ dữ liệu về phim ảnh (Movielens), sau khi nghiên cứu các tài liệu liên quan đến bộ dữ liệu này, không gian tìm kiếm cho các tham số được thiết lập như sau: tham số về số lượng k-hàng xóm là một số nguyên thuộc đoạn $[20, 60]$, tương đương với $[2\%-10\%]$ người dùng trong bộ dữ liệu; khoảng phân chia thời gian được tính theo đơn vị ngày và thuộc đoạn $[1, 7]$ (lưu ý rằng một ngày tương đương với 86,400 đơn vị đo thời gian trong tập dữ liệu) giới hạn này được thiết kế để do ảnh hưởng trong phạm vi tuần; tham số về ảnh hưởng của độ dốc thời gian trong bộ dữ liệu λ là một số thực thuộc đoạn $[0.0001-0.005]$. Không gian tìm kiếm được xác định thông qua nghiên cứu các nghiên cứu trước đó, đồng thời sử dụng một số lần chạy thực nghiệm để xác định.

Kết quả của thuật toán học được thể hiện trong hình 3.2 dưới đây.



Hình 3.2 Kết quả học tham số mô hình

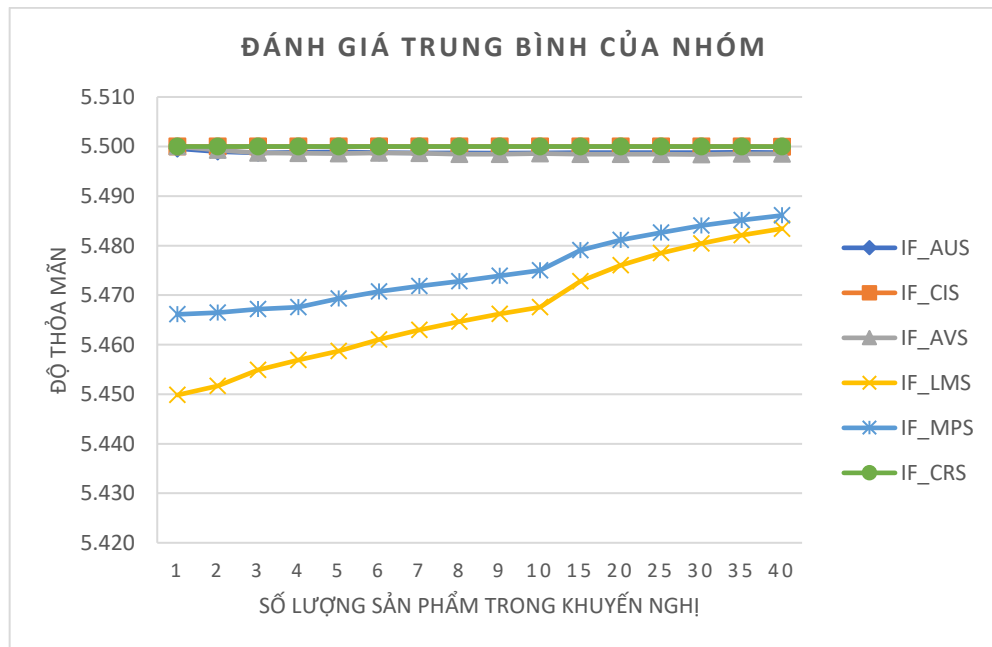
Hình 3.2 ở trên thể hiện hiệu quả của quá trình học tham số theo thuật toán Bayes, trong đó các tham số tối ưu tìm được trong không gian tham số thiết kế là $\{k\text{-neighbors}=55; \text{time-slot} = 5 \text{ day}; \lambda = 0.0001\}$. Bộ tham số này là bộ tham số dùng để thực nghiệm các thuật toán đồng thuận trong HTVN (IF_CIS). Chi tiết các kết quả thực nghiệm được trình bày trong phần tiếp theo của luận án.

b) So sánh thuật toán

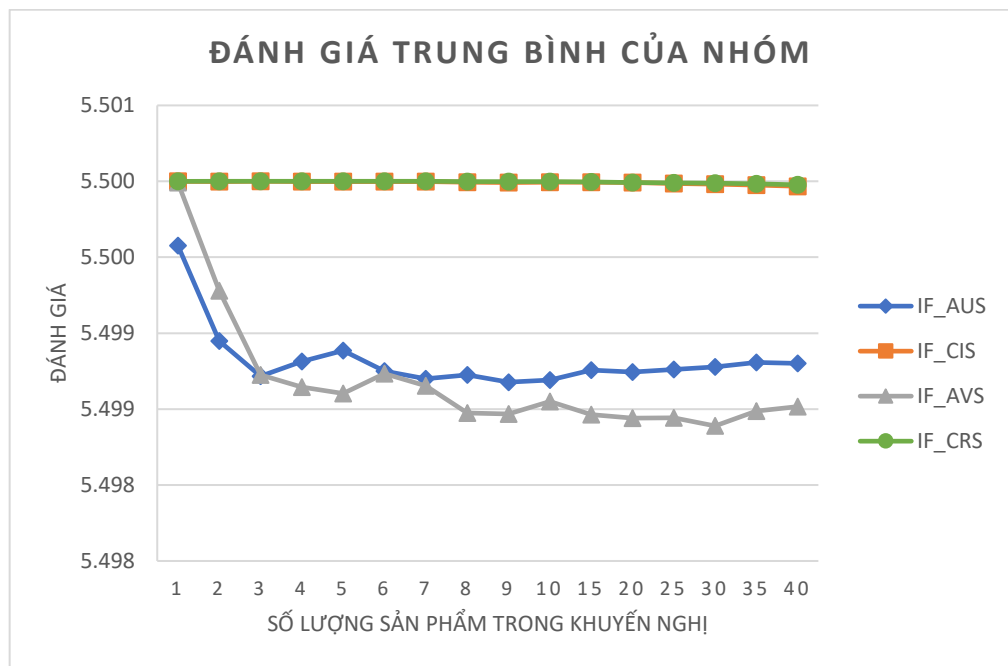
Trong phần này của luận án, hiệu quả của thuật toán đồng thuận sử dụng phép toán tổng hợp đánh giá của các thành viên trong nhóm dựa trên thích phân Choquet cho HTVN mờ động (IF_CIS) sẽ được trình bày và so sánh với các chiến lược đồng thuận khác cho HTVN trên tập mờ trực cảm.

Các mô hình được thực nghiệm trên bộ dữ liệu công khai Movielens 1M, và các nhóm được người dùng được tạo dựng với các kích thước {3, 5, 10, 15, 20}, và có 30 được sinh ra tương ứng với mỗi kích thước nhóm để thực nghiệm. Cách thức sinh dữ liệu nhóm được trình bày trong mục 2.4.2. Kích thước các nhóm thực nghiệm được lựa chọn như trên dựa trên nghiên cứu các công bố trước đó trong cùng lĩnh vực, và với mục tiêu đánh giá ảnh hưởng của kích thước nhóm với hiệu năng của các mô hình HTVN được thực nghiệm. Hiệu năng của các mô hình HTVN động trên tập mờ trực cảm được trình bày khái quát trong các biểu đồ sau, trong đó kết quả thống kê được lấy là giá trị trung bình của tất cả các nhóm tương ứng với tất cả kích thước nhóm.

Về độ đo đánh giá trung bình theo nhóm:



Hình 3.3 Đánh giá trung bình của nhóm theo các HTVN



Hình 3.4 Đánh giá trung bình của nhóm theo bốn HTVN tốt nhất

Hai biểu đồ 3.3 và 3.4 ở trên thể hiện sự khác nhau giữa sáu mô hình HTVN động theo tiếp cận mờ trực cảm sử dụng các chiến lược đồng thuận khác nhau về độ đo đánh giá trung bình của nhóm. Các số liệu thể hiện đánh giá trung bình của nhóm so với kích thước khuyến nghị, lưu ý rằng giá trị này càng cao thì mô hình HTVN càng tốt hơn. Hình 3.3 cho thấy các HTVNĐ theo tiếp cận mờ trực cảm sử dụng thuật toán đồng thuận IF_LMS và IF_MPS có hiệu quả kém hơn nhiều so với bốn HTVN mờ động còn lại. Hình 3.4 cho thấy sự khác biệt nhỏ giữa nhóm bốn HTVN tốt hơn, trong đó thực nghiệm chứng tỏ rằng các HTVNĐ theo tiếp cận mờ trực cảm sử dụng pha đồng thuận IF_CIS và IF_CRS có hiệu suất tốt hơn hai tiếp cận còn lại. Điều này cho thấy mô hình HTVN động trên tập mờ trực cảm, sử dụng tích phân Choquet để xây dựng phép toán tổng hợp đánh giá của thành viên nhóm có hiệu quả tương đương với thuật toán tốt nhất về mặt đánh giá trung bình nhóm.

Về độ công bằng của khuyến nghị

- Xét theo độ đo tỷ lệ hài lòng:

Trên khía cạnh độ công bằng, độ đo đầu tiên được sử dụng đó là tỷ lệ thành viên trong nhóm hài lòng (có đánh giá cao hơn đánh giá trung bình của sản phẩm) được thể hiện trong bảng sau:

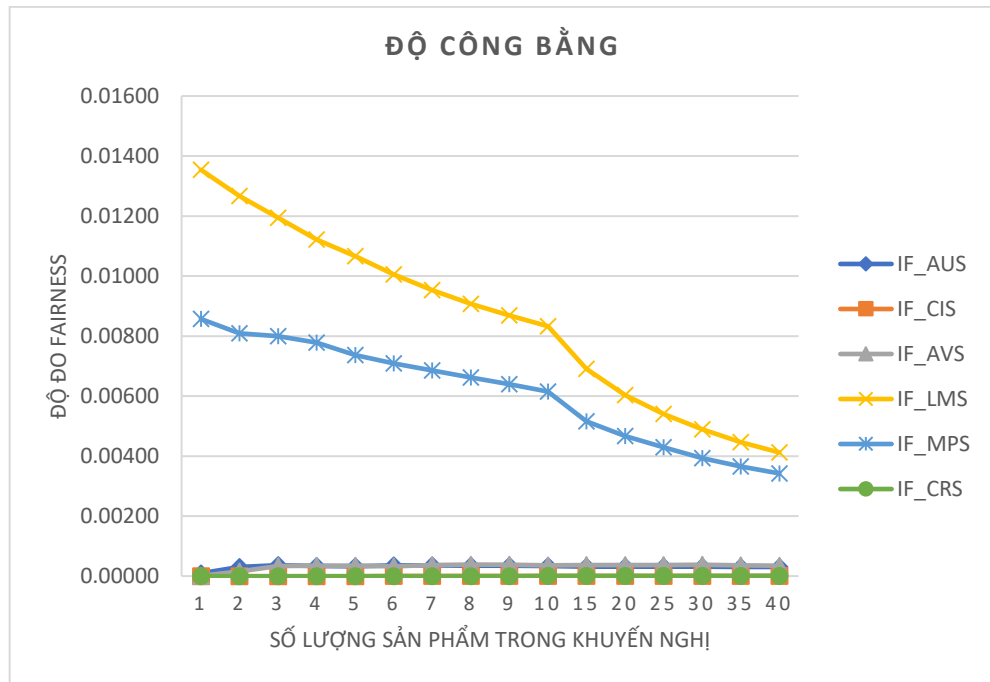
Bảng 3.1 Độ đo tỷ lệ hài lòng của các thuật toán HTVN mờ động

Top-N	IF_AUS	IF_CIS	IF_AVS	IF_LMS	IF_MPS	IF_CRS
1	1.00000	1.00000	1.00000	0.96433	0.97633	1.00000
2	0.99978	1.00000	0.99978	0.96756	0.97839	1.00000
3	0.99970	1.00000	0.99985	0.96948	0.97759	1.00000
4	0.99978	1.00000	0.99978	0.97194	0.97861	1.00000
5	0.99982	1.00000	0.99982	0.97364	0.98024	1.00000
6	0.99978	1.00000	0.99985	0.97574	0.98169	1.00000
7	0.99976	1.00000	0.99987	0.97738	0.98262	1.00000
8	0.99979	1.00000	0.99983	0.97924	0.98390	1.00000
9	0.99977	1.00000	0.99985	0.98084	0.98495	1.00000
10	0.99979	1.00000	0.99987	0.98218	0.98598	1.00000
15	0.99983	1.00000	0.99974	0.98736	0.98964	1.00000
20	0.99985	1.00000	0.99978	0.98974	0.99124	1.00000
25	0.99984	1.00000	0.99978	0.99134	0.99245	1.00000
30	0.99986	1.00000	0.99979	0.99262	0.99349	1.00000
35	0.99987	1.00000	0.99982	0.99363	0.99424	1.00000
40	0.99987	1.00000	0.99984	0.99435	0.99485	1.00000

Dễ thấy, hai HTVN có pha đồng thuận sử dụng chiến lược IF_CIS và IF_CRS có hiệu năng tốt nhất xét theo chỉ số “tỷ lệ hài lòng”. Tuy vậy, hầu hết các thuật toán đều cho kết quả rất cao xét theo độ đo này. Nguyên nhân chủ yếu ở đây là ngưỡng lựa chọn cho khái niệm về độ hài lòng là tương đối thấp. Do đó hầu hết các người dùng đều được xếp vào nhóm hài lòng với khuyến nghị.

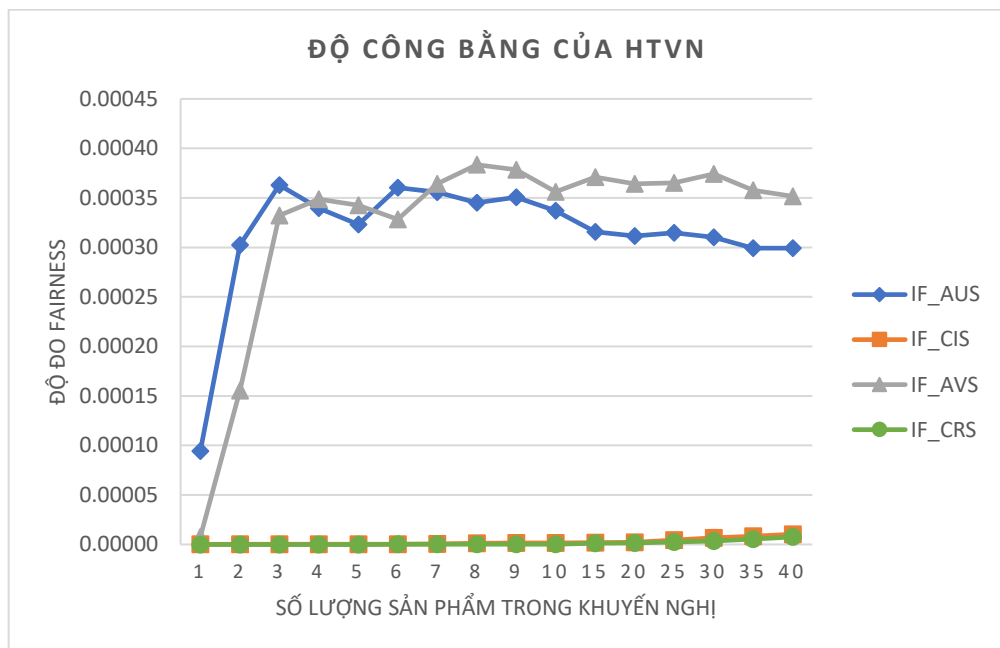
- **Xét theo độ đo độ công bằng:**

Độ đo thứ hai về độ công bằng của sáu mô hình HTVN động theo tiếp cận mờ trực cảm được thể hiện trong hình dưới đây.



Hình 3.5 Độ công bằng của các HTVN động theo tiếp cận mờ trực cảm

Trong hình 3.5 ở trên, dễ thấy hai HTVNĐ theo tiếp cận mờ trực cảm sử dụng chiến lược ít thiệt thòi nhất (IF_LMS) và chiến lược hài lòng nhất (IF_MPS) lần lượt là những thuật toán có độ công bằng kém nhất trong các thuật toán thử nghiệm. Bốn thuật toán ưu thế hơn được phân tích kỹ hơn bên dưới.

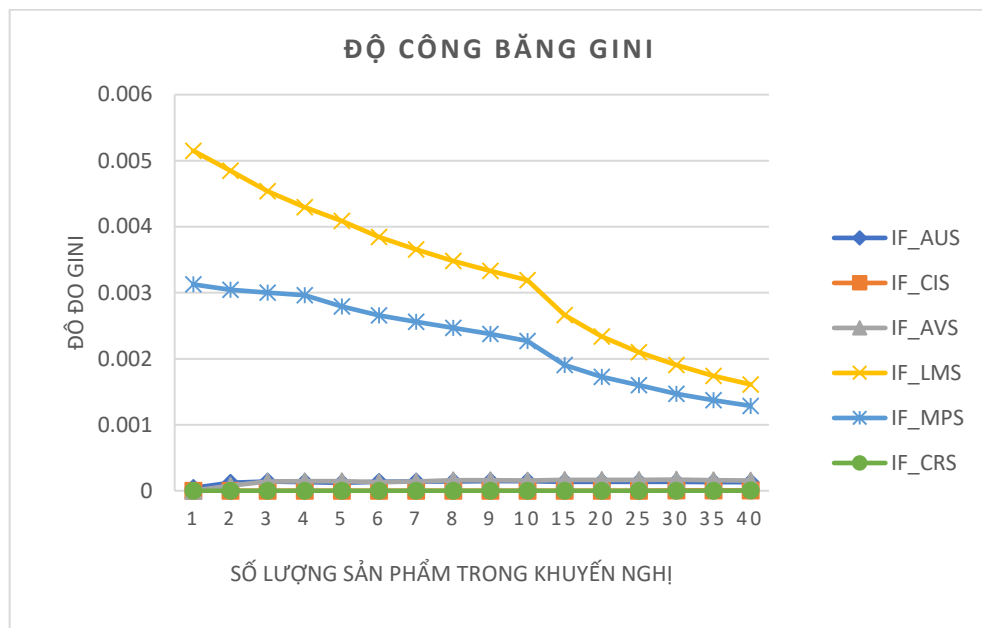


Hình 3.6 Độ công bằng của bốn HTVN động theo tiếp cận mờ trực cảm tốt nhất

Hình 3.6 trên cho thấy hiệu quả của bốn mô hình có ưu thế hơn về khía cạnh độ đo công bằng thứ hai. Trong bốn thuật toán này IF_AUS, IF_CIS, IF_AVS và IF_CRS, dễ thấy thuật toán IF_CIS được đề xuất và thuật toán IF_CRS cho kết quả tốt hơn. Từ đó ta thấy thuật toán HTVN động theo tiếp cận mờ trực cảm sử dụng toán tử tổng hợp đánh giá thành viên nhóm với tích phân Choquet (IF_CIS) cho kết quả thuộc nhóm tốt nhất.

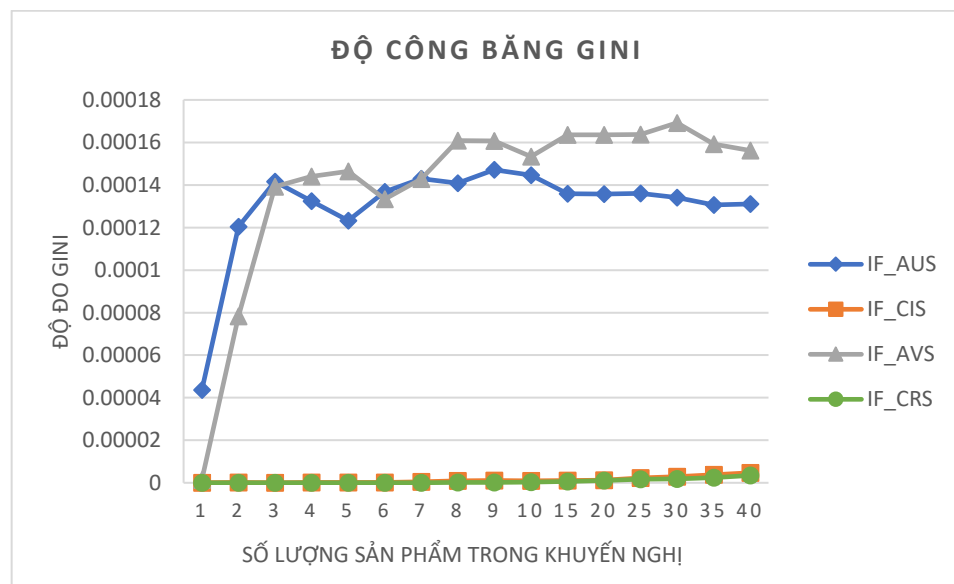
- **Xét theo độ đo công bằng Gini:**

Độ công bằng xét theo độ đo thứ ba (độ đo công bằng Gini) của sáu mô hình HTVN động theo tiếp cận mờ trực cảm được thể hiện trong hình dưới đây.



Hình 3.7 Độ công bằng GINI của các HTVN động theo tiếp cận mờ trực cảm

Để thấy hai hệ tư vấn nhóm mờ động sử dụng chiến lược ít thiệt thòi nhất (IF_LMS) và chiến lược hài lòng nhất (IF_MPS) lần lượt là những thuật toán có độ bất bình đẳng cao nhất trong các thuật toán thử nghiệm, nghĩa là hai thuật toán trên có hiệu năng kém nhất. Bốn thuật toán ưu thế hơn được phân tích kỹ hơn bên dưới.



Hình 3.8 Độ công bằng GINI của bốn HTVN động theo tiếp cận mờ trực cảm tốt nhất

Hình 3.8 ở trên thể hiện hiệu quả của bốn thuật toán có ưu thế hơn về khía cạnh độ đo công bằng thứ ba: độ bất bình đẳng. Trong bốn thuật toán này IF_AUS, IF_CIS, IF_AVS và IF_CRS, dễ thấy thuật toán IF_CIS được đề xuất và thuật toán IF_CRS cho kết quả tốt hơn. Từ đó ta thấy thuật toán HTVN mờ động sử dụng phép toán tổng hợp đánh giá của các thành viên trong nhóm với tích phân Choquet (IF_CIS) cho kết quả thuộc nhóm tốt nhất xét theo độ đo công bằng GINI.

Tóm lược ngắn gọn, xét theo tất cả các độ đo về khía cạnh độ công khuyến nghị, bằng thực nghiệm cho thấy HTVN động theo tiếp cận mờ trực cảm sử dụng tích phân Choquet được đề xuất (IF_CIS), và mô hình sử dụng chiến lược luật Copeland (IF_CAS) cho kết quả thực nghiệm tốt hơn các mô hình HTVN sử dụng chiến lược đồng thuận khác. Hai mô hình này đều tạo ra các khuyến nghị có tính cân bằng giữa hai mục tiêu tối đa hóa đánh giá trung bình của nhóm và đảm bảo sự công bằng giữa người dùng trong nhóm.

Tuy nhiên, giữa hai chiến lược này, chiến lược luật Copeland là một tiếp cận cứng và không thể biến đổi. Chiến lược đồng thuận này cũng không cho phép xem xét sự tương tác giữa các thành viên trong nhóm. Ngược lại, khi sử dụng độ đo mờ dựa trên hàm dung lượng phi cộng tính và cách thức xây dựng phép toán tổng hợp đánh giá của các thành viên trong nhóm dựa trên tích phân Choquet trong pha đồng thuận có thể mở rộng, đồng thời cho phép xử lý hiệu ứng tương tác giữa các thành viên nhóm dựa trên việc phát triển các hàm dung lượng phi cộng tính phù hợp. Do đó tiếp cận được đề xuất với tích phân Choquet sẽ có khả năng phát triển lớn hơn để ứng dụng trong thực tiễn.

Bên cạnh ưu điểm kể trên, mô hình HTVN được đề xuất trong chương 3 của luận án còn là một mô hình xử lý thông tin trên tập nền mờ trực cảm và xem xét ảnh hưởng của yếu tố thời gian tới sự thay đổi đánh giá của người dùng và sự hấp dẫn của sản phẩm. Dựa trên khung thuật toán được đề xuất, các phép toán mới để ước lượng độ tương tự động có thể được bổ xung hay tùy chỉnh để phù hợp với từng bài toán cụ thể. Dù độ phức tạp tính toán sẽ tăng như phân tích trong mục 3.3.2, dễ thấy cách tiếp

cận này phù hợp với các bài toán thực tế hơn là cách tiếp cận thông tin tĩnh trong các nghiên cứu khác về HTVN.

3.5. Kết luận chương 3

Nghiên cứu trong Chương 3 cơ bản đã thực hiện được mục tiêu đặt ra. Đó là nghiên cứu xây dựng và phát triển hệ tư vấn nhóm động dựa trên tập mờ trực cảm và tích phân Choquet.

Mô hình đề xuất biểu diễn thông tin về đánh giá của người dùng trên tập nền là tập mờ trực cảm, đồng thời trong các bước xử lý thông tin theo mô hình lọc cộng tác ảnh hưởng của yếu tố thời gian được xem xét. Cụ thể, thông tin động ở đây được xử lý theo giả thiết rằng sở thích của của người dùng, sự hấp dẫn của sản phẩm đều là thông tin biến đổi, suy hao theo thời gian. Độ đo “độ tương tự động” và phép toán dùng để dự báo đánh giá của người dùng đối với sản phẩm được đề xuất mới, phù hợp với dữ liệu được biểu diễn trên tập mờ trực cảm. Bên cạnh đó, trong pha đồng thuận của mô hình, toán tử Choquet được sử dụng để phát triển một phép toán tổng hợp các đánh giá của thành viên trong nhóm.

Ứng dụng toán tử Choquet trên tập mờ trực cảm và hàm dung lượng được đề xuất trong pha đồng thuận của HTVN động trên tập mờ trực cảm đã giải quyết được đồng thời các vấn đề như: xử lý thông tin về độ do dự trong đánh giá của người dùng, ảnh hưởng của yếu tố thời gian tới sự thay đổi của đánh giá của người dùng, và độ công bằng của khuyến nghị.

Mô hình đề xuất đã được cài đặt bằng ngôn ngữ lập trình Python và có thể được tùy chỉnh một cách dễ dàng để tích hợp vào một ứng dụng thực tế. Tuy vậy, khi áp dụng mô hình đề xuất với các bài toán khác trong thực tế, người dùng cần thiết kế và thử nghiệm lại các tham số của mô hình học tối ưu tham số. Ngoài ra, việc ứng dụng tiếp cận mờ trực cảm và tích phân Choquet sẽ làm tăng thời gian tính của HTVN so với tiếp cận truyền thống. Nhưng thuật toán có ưu điểm là có thể phù hợp với các hệ thống tính toán song song, đó đó khi gặp phải vấn đề về hiệu năng tính toán (đặc biệt với dữ liệu lớn) người dùng có thể cân nhắc chuyển đổi sang thuật toán song song cho mô hình được đề xuất này.

Các kết quả nghiên cứu ở Chương 3 được công bố trong công trình 3 và 4, phần *“Danh mục công trình của tác giả”*.

KẾT LUẬN

1) Những kết quả chính của luận án

Luận án nghiên cứu về Hệ tư vấn nhóm, một hệ tư vấn khái quát của hệ tư vấn đơn người dùng truyền thống, theo tiếp cận tính toán mờ trực cảm và tích phân Choquet. Trong đó các thách thức của HTVN như độ công bằng của khuyến nghị, thông tin động, độ do dự và tính không chắc chắn trong dữ liệu đã được nghiên cứu và đề xuất các hướng giải quyết. Để nâng cao tính công bằng của khuyến nghị, luận án đề xuất ứng dụng tích phân Choquet làm nền tảng để xây dựng thuật toán tổng hợp đánh giá trong pha đồng thuận của HTVN. Về xử lý thông tin động, trong phạm vi của luận án, tính động được xem xét từ khía cạnh sự thay đổi của sở thích người dùng và độ hấp dẫn của sản phẩm theo thời gian. Để xử lý vấn đề thông tin do dự và không chắc chắn, luận án sử dụng tập mờ trực cảm và các phép toán tương ứng để biểu diễn và xử lý thông tin trong HTVN. Sử dụng đồng bộ ba yếu tố kể trên, luận án đã đề xuất được những giải pháp phù hợp nhằm giải quyết một số thách thức lớn gặp phải khi phát triển HTVN.

Dựa trên nghiên cứu lý thuyết và thực nghiệm, kết quả chính đạt được của luận án bao gồm:

1. Đề xuất thuật toán cho pha đồng thuận sử dụng tích phân Choquet áp dụng xây dựng hệ tư vấn nhóm nhằm tăng cường sự công bằng của các khuyến nghị và đảm bảo giữ được tổng thể lợi ích cho nhóm người dùng.

2. Đề xuất hệ tư vấn nhóm theo mô hình động áp dụng tính toán mờ trực cảm và tích phân Choquet trong pha đồng thuận đáp ứng giải quyết những bài toán có đánh giá của người dùng thay đổi theo thời gian.

Kết quả của luận án có thể được ứng dụng trong thực tế nhằm phát triển các hệ thống thông tin thông minh, cung cấp khả năng lọc thông tin phù hợp và khuyến nghị cho nhóm người dùng.

2) Hướng phát triển của luận án

Bên cạnh những đóng góp đã đạt được, nghiên cứu hiện tại còn đối mặt với một số thách thức chưa được giải quyết. Trong đó giả thiết về ảnh hưởng của thời gian được phát triển cho thuật toán HTVN động trong luận án đó là mô hình đường dốc thời gian. Mô hình này phù hợp với các sản phẩm dịch vụ có xu thế lỗi thời dần theo thời gian. Trên thực tế có những dữ liệu có đặc điểm khác về điều này. Do đó hướng phát triển tiếp theo của luận án sẽ tập trung vào hai nhóm vấn đề chính đó là nghiên cứu sâu hơn về ảnh hưởng của thời gian tới đánh giá của người dùng và tương tác giữa các thành viên trong nhóm. Bên cạnh đó, việc phát triển các thiết kế mới cho mô hình học tham số tối ưu cần tiếp tục nghiên cứu.

Cụ thể hướng phát triển tiếp theo của luận án như sau:

1. Tiếp tục nghiên cứu, đề xuất các thuật toán học tham số mới để tìm kiếm các tham số thể hiện ảnh hưởng của thời gian tới sự thay đổi sở thích người dùng và độ hấp dẫn sản phẩm được tối ưu hơn.

2. Tiếp tục nghiên cứu, đề xuất các thuật toán cho phép xử lý tốt hơn ảnh hưởng tương tác giữa các thành viên trong nhóm theo hướng sử dụng thêm thông tin từ mạng xã hội.

Khi được triển khai, các hướng nghiên cứu này sẽ tiếp tục mở rộng kết quả của luận án để mô hình HTVN đề xuất có thể phù hợp với nhiều bài toán thực tế hơn, và có khả năng ứng dụng cao hơn nữa.

DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ

1	Cu Nguyen Giap , Nguyen Nhu Son*, Nguyen Long Giang, Hoang Thi Minh Chau, Tran Manh Tuan, Le Hoang Son, “A New Approach for Fairness Increment of Consensus-driven Group Recommender Systems Based on Choquet Integral”, <i>International Journal of Data Warehousing and Mining (IJDWM)</i> , 2022, 18(1), 1-22, (SCIE, IF: 1,2).
2	Cu Nguyen Giap , Le Thi Huyen Dieu, Luong Thi Hong Lan, Tran Thi Ngan, Tran Manh Tuan, “Utilize Deep learning to increase the performance of a Book recommender system using the Item-based Collaborative Filtering”. Hội thảo quốc tế <i>RICE</i> , 2022, 109-113.
3	Nguyễn Như Sơn, Cù Nguyên Giáp , Lê Hoàng Sơn , Nguyễn Long Giang, Dương Thị Thanh Loan, Trần Mạnh Tuấn, Dương Thị Thu Huyền, “ <i>Kỹ thuật tư vấn nhóm dựa trên tập mờ trực cảm và ứng dụng</i> ”, Hội thảo Quốc gia lần thứ XXV "Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông" (2022), Hà Nội, Việt Nam, 2022, 81-86.
4	Nguyen Nhu Son, Cu Nguyen Giap* , Le Hoang Son*, Nguyen Long Giang, Tran Manh Tuan, Vassilis C. Gerogiannis, Dimitrios Tzimos, “A Dynamic Fuzzy Group Recommender System based on Intuitionistic Fuzzy Choquet Integral Aggregation”, <i>Soft Computing</i> , 2024, 1-14, (SCIE, IF: 4,1).

TÀI LIỆU THAM KHẢO

- [1] A. De Biasio, N. Navarin, and D. Jannach, “Economic recommender systems – a systematic review,” *Electron. Commer. Res. Appl.*, vol. 63, no. October 2023, 2024, doi: 10.1016/j.elerap.2023.101352.
- [2] S. Dara, C. R. Chowdary, and C. Kumar, “A survey on group recommender systems,” *J. Intell. Inf. Syst.*, vol. 54, no. 2, pp. 271–295, 2020, doi: 10.1007/s10844-018-0542-3.
- [3] J. Masthoff, “Group recommender systems: Combining individual models,” in *Recommender systems handbook*, Springer, 2011, pp. 677–702.
- [4] J. Masthoff, “Group modeling: Selecting a sequence of television items to suit a group of viewers,” in *Personalized digital television*, Springer, 2004, pp. 93–141.
- [5] A. Delic, J. Neidhardt, T. N. Nguyen, and F. Ricci, “An observational user study for group recommender systems in the tourism domain,” *Inf. Technol. & Tour.*, vol. 19, pp. 87–116, 2018.
- [6] R. Katarya, “A systematic review of group recommender systems techniques,” *Proc. Int. Conf. Intell. Sustain. Syst. ICISS 2017*, no. Iciss, pp. 425–428, 2018, doi: 10.1109/ISS1.2017.8389444.
- [7] W. Wang, G. Zhang, and J. Lu, “Hierarchy visualization for group recommender systems,” *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 49, no. 6, pp. 1152–1163, 2017.
- [8] A. Valera, A. L. Murciego, and M. N. Moreno-García, “Group Recommender Systems in the Music Domain: A Systematic Literature Review,” *New Trends Disruptive Technol. Tech Ethics Artif. Intell. DITTET Collect. 1*, pp. 296–307, 2022.
- [9] P. Alves, D. Gomes, C. Rodrigues, J. Carneiro, P. Novais, and G. Marreiros, “Groupplanner: a group recommender system for tourism with multi-agent microservices,” in *International Conference on Practical Applications of*

Agents and Multi-Agent Systems, 2022, pp. 454–460.

- [10] Y.-L. Chen, L.-C. Cheng, and C.-N. Chuang, “A group recommendation system with consideration of interactions among group members,” *Expert Syst. Appl.*, vol. 34, no. 3, pp. 2082–2090, 2008.
- [11] L. Baltrunas, T. Makcinskas, and F. Ricci, “Group recommendations with rank aggregation and collaborative filtering,” in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 119–126.
- [12] A. Delic, J. Masthoff, J. Neidhardt, and H. Werthner, “How to use social relationships in group recommenders: Empirical evidence,” *UMAP 2018 - Proc. 26th Conf. User Model. Adapt. Pers.*, pp. 121–129, 2018, doi: 10.1145/3209219.3209226.
- [13] I. A. Christensen and S. Schiaffino, “Social influence in group recommender systems,” *Online Inf. Rev.*, 2014.
- [14] J. K. Kim, H. K. Kim, H. Y. Oh, and Y. U. Ryu, “A group recommendation system for online communities,” *Int. J. Inf. Manage.*, vol. 30, no. 3, pp. 212–219, 2010.
- [15] J. Chen, Y. Liu, and D. Li, “Dynamic group recommendation with modified collaborative filtering and temporal factor,” *Int. Arab J. Inf. Technol.*, vol. 13, no. 2, pp. 294–301, 2016.
- [16] A. Felfernig, L. Boratto, M. Stettinger, and M. Tkalčič, “Evaluating group recommender systems,” in *Group recommender systems*, Springer, 2018, pp. 59–71.
- [17] L. Xiao, Z. Min, Z. Yongfeng, G. Zhaoquan, L. Yiqun, and M. Shaoping, “Fairness-aware group recommendation with pareto-efficiency,” in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, 2017, pp. 107–115.
- [18] D. Serbos, S. Qi, N. Mamoulis, E. Pitoura, and P. Tsaparas, “Fairness in package-to-group recommendations,” in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 371–379.

- [19] J. Castro, R. Yera, and L. Martínez, “A fuzzy approach for natural noise management in group recommender systems,” *Expert Syst. Appl.*, vol. 94, pp. 237–249, 2018, doi: 10.1016/j.eswa.2017.10.060.
- [20] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, “Recommender systems survey,” *Knowledge-based Syst.*, vol. 46, pp. 109–132, 2013.
- [21] P. M.-V. Sindhvani and P. Melville, “Recommender systems,” *Commun. ACM*, pp. 1–21, 2010.
- [22] C. C. Aggarwal and others, *Recommender systems*, vol. 1. Springer, 2016.
- [23] Y. Deldjoo *et al.*, “A review of modern fashion recommender systems,” *ACM Comput. Surv.*, vol. 56, no. 4, pp. 1–37, 2023.
- [24] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, “Recommender system application developments: a survey,” *Decis. Support Syst.*, vol. 74, pp. 12–32, 2015.
- [25] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005.
- [26] D. Das, L. Sahoo, and S. Datta, “A survey on recommendation system,” *Int. J. Comput. Appl.*, vol. 160, no. 7, 2017.
- [27] M. Almaghrabi and G. Chetty, “A deep learning based collaborative neural network framework for recommender system,” in *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*, 2018, pp. 121–127.
- [28] Z. Batmaz, A. Yurekli, A. Bilge, and C. Kaleli, “A review on deep learning for recommender systems: challenges and remedies,” *Artif. Intell. Rev.*, vol. 52, pp. 1–37, 2019.
- [29] S. Zhang, L. Yao, A. Sun, and Y. Tay, “Deep learning based recommender system: A survey and new perspectives,” *ACM Comput. Surv.*, vol. 52, no. 1, pp. 1–38, 2019.

- [30] A. Jameson and B. Smyth, "Recommendation to groups," in *The adaptive web*, Springer, 2007, pp. 596–627.
- [31] J. Masthoff, "Group modeling: Selecting a sequence of television items to suit a group of viewers," *User Model. User-Adapted Interact.*, vol. 14, no. 1, pp. 37–85, 2004, doi: 10.1023/B:USER.0000010138.79319.f0.
- [32] J. Castro, F. J. Quesada, I. Palomares, and L. Martinez, "A consensus-driven group recommender system," *Int. J. Intell. Syst.*, vol. 30, no. 8, pp. 887–906, 2015.
- [33] M. Stratigi, H. Kondylakis, and K. Stefanidis, "Fairness in group recommendations in the health domain," *Proc. - Int. Conf. Data Eng.*, pp. 1481–1488, 2017, doi: 10.1109/ICDE.2017.217.
- [34] J. Jia, Y. Yao, Z. Lei, and P. Liu, "Dynamic Group Recommendation Algorithm Based on Member Activity Level," *Sci. Program.*, vol. 2021, 2021.
- [35] J. Castro, M. J. Barranco, R. M. Rodriguez, and L. Martinez, "Group recommendations based on hesitant fuzzy sets," *Int. J. Intell. Syst.*, vol. 33, no. 10, pp. 2058–2077, 2018.
- [36] T. De Pessemier, J. Dhondt, and L. Martens, "Hybrid group recommendations for a travel service," *Multimed. Tools Appl.*, vol. 76, no. 2, pp. 2787–2811, 2017, doi: 10.1007/s11042-016-3265-x.
- [37] R. Yera, A. A. Alzahrani, and L. Martinez, "A fuzzy content-based group recommender system with dynamic selection of the aggregation functions," *Int. J. Approx. Reason.*, vol. 150, pp. 273–296, 2022.
- [38] P. Ji and X. Ma, "A fuzzy intelligent group recommender method in sparse-data environments based on multi-agent negotiation," *Expert Syst. Appl.*, vol. 213, p. 119294, 2023.
- [39] N. N. Htun, E. Lecluse, and K. Verbert, "Perception of fairness in group music recommender systems," in *26th International Conference on Intelligent User Interfaces*, 2021, pp. 302–306.

- [40] R. Burke, “Multisided fairness for recommendation,” *arXiv Prepr. arXiv1707.00093*, 2017.
- [41] T. Lust, “Choquet integral versus weighted sum in multicriteria decision contexts,” in *International Conference on Algorithmic Decision Theory*, 2015, pp. 288–304.
- [42] C. Luo, X. Cai, and N. Chowdhury, “Self-training temporal dynamic collaborative filtering,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2014, pp. 461–472.
- [43] Y. Ding and X. Li, “Time weight collaborative filtering,” in *Proceedings of the 14th ACM international conference on Information and knowledge management*, 2005, pp. 485–492.
- [44] J. Wang, B. Sarwar, and N. Sundaresan, “Utilizing related products for post-purchase recommendation in e-commerce,” in *Proceedings of the fifth ACM conference on Recommender systems*, 2011, pp. 329–332.
- [45] A. Karatzoglou, “Collaborative temporal order modeling,” in *Proceedings of the fifth ACM conference on Recommender systems*, 2011, pp. 313–316.
- [46] J. J. McAuley and J. Leskovec, “From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews,” in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 897–908.
- [47] Y. Koren, “Collaborative filtering with temporal dynamics,” in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 447–456.
- [48] G. Ling, H. Yang, I. King, and M. R. Lyu, “Online learning for collaborative filtering,” in *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 2012, pp. 1–8.
- [49] M. Aharon *et al.*, “Off-set: one-pass factorization of feature sets for online recommendation in persistent cold start settings,” in *Proceedings of the 7th ACM Conference on Recommender Systems*, 2013, pp. 375–378.
- [50] Z. Lu, D. Agarwal, and I. S. Dhillon, “A spatio-temporal approach to

- collaborative filtering,” in *Proceedings of the third ACM conference on Recommender systems*, 2009, pp. 13–20.
- [51] S. Nowakowski and A. Boyer, “Automatic tracking and control for web recommendation New approaches for web recommendation,” *Int. J. Adv. Intell. Syst.*, 2013.
- [52] J. Chen, Y. Liu, and D. Li, “Dynamic group recommendation with modified collaborative filtering and temporal factor.,” *Int. Arab J. Inf. Technol.*, vol. 13, no. 2, pp. 294–301, 2016.
- [53] T. C.-K. Huang, “Recommendations of closed consensus temporal patterns by group decision making,” *Knowledge-Based Syst.*, vol. 54, pp. 318–328, 2013.
- [54] N. Capuano, F. Chiclana, E. Herrera-Viedma, H. Fujita, and V. Loia, “Fuzzy group decision making for influence-aware recommendations,” *Comput. Human Behav.*, vol. 101, pp. 371–379, 2019.
- [55] L. H. Son and N. T. Thong, “Intuitionistic fuzzy recommender systems: An effective tool for medical diagnosis,” *Knowledge-Based Syst.*, vol. 74, pp. 133–150, 2015, doi: 10.1016/j.knosys.2014.11.012.
- [56] K. T. Atanassov, “Intuitionistic fuzzy sets,” *International Journal Bioautomation*, vol. 20, pp. S1–S6, 2016, doi: 10.1007/978-3-7908-1870-3_1.
- [57] K. T. Atanassov, “Intuitionistic fuzzy sets,” in *Intuitionistic fuzzy sets*, Springer, 1999, pp. 1–137.
- [58] Z. Xu and X. Cai, *Intuitionistic fuzzy information aggregation: Theory and applications*, vol. 9783642295, no. January. 2012.
- [59] E. Szmidt and J. Kacprzyk, “Distances between intuitionistic fuzzy sets,” *Fuzzy Sets and Systems*, vol. 114, no. 3, pp. 505–518, 2000, doi: 10.1016/S0165-0114(98)00244-9.
- [60] P. Grzegorzewski, “Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on the Hausdorff metric,” *Fuzzy sets Syst.*, vol. 148, no. 2, pp. 319–328, 2004.
- [61] J. Ye, “Cosine similarity measures for intuitionistic fuzzy sets and their

- applications,” *Math. Comput. Model.*, vol. 53, no. 1–2, pp. 91–97, 2011.
- [62] M. Xia, Z. Xu, and B. Zhu, “Generalized intuitionistic fuzzy Bonferroni means,” *Int. J. Intell. Syst.*, vol. 27, no. 1, pp. 23–47, 2012, doi: 10.1002/int.20515.
- [63] M. Grabisch, M. Roubens, and others, “Application of the Choquet integral in multicriteria decision making,” *Fuzzy Meas. Integr. Appl.*, pp. 348–374, 2000.
- [64] C. Labreuche, “Construction of a Choquet integral and the value functions without any commensurateness assumption in multi-criteria decision making,” in *Proceedings of the 7th Conference of the European Society for Fuzzy Logic and Technology, EUSFLAT 2011 and French Days on Fuzzy Logic and Applications, LFA 2011*, 2011, vol. 1, no. 1, doi: 10.2991/eusflat.2011.130.
- [65] I. Cantador and P. Castells, “Group recommender systems: new perspectives in the social web,” in *Recommender systems for the social web*, Springer, 2012, pp. 139–157.
- [66] J. M. Peiró, V. Martínez-Tur, and C. Moliner, “Perceived Fairness,” in *Encyclopedia of Quality of Life and Well-Being Research*, A. C. Michalos, Ed. Dordrecht: Springer Netherlands, 2014, pp. 4693–4696.
- [67] H. Wang, X. Zhang, and G. Chen, “Mining a complete set of both positive and negative association rules from large databases,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2008, pp. 777–784, doi: 10.1007/978-3-540-68125-0_75.
- [68] C. N. Wang and V. T. Phan, “Enhancing the accurate of grey prediction for GDP growth rate in Vietnam,” *Proc. - 2014 Int. Symp. Comput. Consum. Control. IS3C 2014*, pp. 1137–1139, 2014, doi: 10.1109/IS3C.2014.295.
- [69] L. Quijano-Sanchez, J. A. Recio-Garcia, B. Diaz-Agudo, and G. Jimenez-Diaz, “Social factors in group recommender systems,” *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 1, pp. 1–30, 2013.
- [70] J.-L. Marichal, “An axiomatic approach of the discrete Choquet integral as a

- tool to aggregate interacting criteria,” *IEEE Trans. fuzzy Syst.*, vol. 8, no. 6, pp. 800–807, 2000.
- [71] E. F. Combarro, I. DiAz, and P. Miranda, “On random generation of fuzzy measures,” *Fuzzy Sets Syst.*, vol. 228, pp. 64–77, 2013.
 - [72] D. Lerche and P. B. Sørensen, “Evaluation of the ranking probabilities for partial orders based on random linear extensions,” *Chemosphere*, vol. 53, no. 8, pp. 981–992, 2003.
 - [73] R. Bubley and M. Dyer, “Faster random generation of linear extensions,” *Discrete Math.*, vol. 201, no. 1–3, pp. 81–88, 1999.
 - [74] L. Galand, P. Perny, and O. Spanjaard, “Choquet-based optimisation in multiobjective shortest path and spanning tree problems,” *Eur. J. Oper. Res.*, vol. 204, no. 2, pp. 303–315, 2010.
 - [75] K. De Loof, H. De Meyer, and B. De Baets, “Exploiting the lattice of ideals representation of a poset,” *Fundam. Informaticae*, vol. 71, no. 2, 3, pp. 309–321, 2006.
 - [76] A. Karzanov and L. Khachiyan, “On the conductance of order Markov chains,” *Order*, vol. 8, no. 1, pp. 7–15, 1991.
 - [77] M. D. Ekstrand, J. T. Riedl, J. A. Konstan, and others, “Collaborative filtering recommender systems,” *Found. Trends®in Human--Computer Interact.*, vol. 4, no. 2, pp. 81–173, 2011.
 - [78] H. X. Huynh, G. N. Cu, T. M. Huynh, H. H. Luong, and others, “Recommender Systems Based on Resonance Relationship of Criteria With Choquet Operation,” *Int. J. Data Warehous. Min.*, vol. 16, no. 4, pp. 44–62, 2020.
 - [79] A. Pujahari and V. Padmanabhan, “Group recommender systems: Combining user-user and item-item collaborative filtering techniques,” in *2015 International conference on information technology (ICIT)*, 2015, pp. 148–152.
 - [80] L. Xiong, X. Chen, T.-K. Huang, J. Schneider, and J. G. Carbonell,

- “Temporal collaborative filtering with bayesian probabilistic tensor factorization,” in *Proceedings of the 2010 SIAM international conference on data mining*, 2010, pp. 211–222.
- [81] L. A. Zadeh, “Fuzzy sets,” in *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh*, World Scientific, 1996, pp. 394–432.
- [82] M. Palanikumar, N. Kausar, S. F. Ahmed, S. A. Edalatpanah, E. Ozbilge, and A. Bulut, “New applications of various distance techniques to multi-criteria decision-making challenges for ranking vague sets,” *AIMS Math.*, vol. 8, no. 5, pp. 11397–11424, 2023.
- [83] S. El-Morsy, “Stock portfolio optimization using pythagorean fuzzy numbers,” *J Oper Strateg Anal*, vol. 1, no. 1, pp. 8–13, 2023.
- [84] F. Smarandache, “Soft set product extended to hypersoft set and indetermsoft set cartesian product extended to indetermhypersoft set,” *J. fuzzy Ext. Appl.*, vol. 3, no. 4, pp. 313–316, 2022.
- [85] M. Ihsan, M. Saeed, and A. U. Rahman, “Optimizing hard disk selection via a fuzzy parameterized single-valued neutrosophic soft set approach,” *J. Oper. Strateg Anal*, vol. 1, no. 2, pp. 62–69, 2023.
- [86] S. C. Oner, B. Oztaysi, and M. Oner, “Social platform based interval valued intuitionistic fuzzy location recommendation system,” *J. Intell. Fuzzy Syst.*, vol. 38, no. 1, pp. 1027–1042, 2020.
- [87] P. Qiu, A. Sorourkhah, N. Kausar, T. Cagin, and S. A. Edalatpanah, “Simplifying the Complexity in the Problem of Choosing the Best Private-Sector Partner,” *Systems*, vol. 11, no. 2, 2023, doi: 10.3390/systems11020080.
- [88] R. Parvathi and C. Radhamani, “Fuzzification functions for temporal intuitionistic fuzzy sets,” *Notes Intuitionistic Fuzzy Sets*, vol. 25, no. 3, pp. 1–12, 2019, doi: 10.7546/nifs.2019.25.3.1-12.
- [89] F. Xiao, “A Distance Measure for Intuitionistic Fuzzy Sets and Its Application to Pattern Classification Problems,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 6, pp. 3980–3992, 2021,

doi: 10.1109/TSMC.2019.2958635.

- [90] R. Verma, “On intuitionistic fuzzy order- α divergence and entropy measures with MABAC method for multiple attribute group decision-making,” *J. Intell. Fuzzy Syst.*, vol. 40, no. 1, pp. 1191–1217, 2021.
- [91] M. Krawczak and G. Szkatuła, “On matching of intuitionistic fuzzy sets,” *Inf. Sci. (Ny)*, vol. 517, pp. 254–274, 2020.
- [92] C. Tan and X. Chen, “Intuitionistic fuzzy Choquet integral operator for multi-criteria decision making,” *Expert Syst. Appl.*, vol. 37, no. 1, pp. 149–157, 2010.
- [93] D. H. Wolpert and W. G. Macready, “No free lunch theorems for optimization,” *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, 1997.
- [94] H. Alibrahim and S. A. Ludwig, “Hyperparameter optimization: Comparing genetic algorithm against grid search and bayesian optimization,” in *2021 IEEE Congress on Evolutionary Computation (CEC)*, 2021, pp. 1551–1559.
- [95] F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019.
- [96] F. Hutter, H. H. Hoos, and K. Leyton-Brown, “Sequential model-based optimization for general algorithm configuration,” in *Learning and Intelligent Optimization: 5th International Conference, LION 5, Rome, Italy, January 17-21, 2011. Selected Papers 5*, 2011, pp. 507–523.
- [97] H. Cho, Y. Kim, E. Lee, D. Choi, Y. Lee, and W. Rhee, “Basic enhancement strategies when using Bayesian optimization for hyperparameter tuning of deep neural networks,” *IEEE access*, vol. 8, pp. 52588–52608, 2020.
- [98] P. Angelov, “Crispification: defuzzification of intuitionistic fuzzy sets,” *BUSEFAL*, vol. 64, pp. 51–55, 1995.
- [99] C. Radhamani, “Crispification of temporal intuitionistic fuzzy sets,” in *AIP Conference Proceedings*, 2020, vol. 2277, no. 1, p. 90014.
- [100] D. P. Filev and R. R. Yager, “A generalized defuzzification method via BAD distributions,” *Int. J. Intell. Syst.*, vol. 6, no. 7, pp. 687–697, 1991.