

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA HỆ THỐNG THÔNG TIN**



KHÓA LUẬN TỐT NGHIỆP

**XÂY DỰNG HỆ KHUYẾN NGHỊ
DỰA TRÊN NGŨ CẢNH**

Giảng viên hướng dẫn: ThS Đỗ Thị Minh Phụng

Sinh viên thực hiện: Lê Nguyễn Hào Hiệp

MSSV: 08520631

Lớp HTTT03

Khóa 2008 – 2013

Thành phố Hồ Chí Minh, Tháng 02 năm 2013

Lời mở đầu

“We are drowning in data, but starving for knowledge”

John Naisbitt, 1982

(Lời dịch: chúng ta chìm trong dữ liệu nhưng đói tri thức)

Dữ liệu là một thành phần hết sức quan trọng không những trong các lĩnh vực nghiên cứu khoa học mà còn trong đời sống thực tế. Quả thực, ngày nay con người tiếp cận với rất nhiều nguồn dữ liệu khác nhau, đa dạng về số lượng lẫn chất lượng nhưng lượng thông tin, tri thức mà con người có thể nhận thức được rất hạn chế. Vì vậy, sự ra đời của khai phá dữ liệu trong lĩnh vực khoa học máy tính đã góp phần to lớn trong việc giúp đỡ con người tiếp cận được những thông tin, tri thức quý giá tiềm tàng từ những nguồn dữ liệu khổng lồ. Cụ thể hơn trong lĩnh vực thương mại điện tử, e-learning, giải trí trực tuyến... những câu hỏi rất phổ biến đối với một người dùng cá nhân như: làm thế nào có thể chọn mua một món hàng phù hợp, làm sao chọn một khóa học phù hợp hay đơn giản là nghe bài hát gì. Hàng ngàn câu hỏi đặt ra chỉ với một mục đích duy nhất là tiếp cận được những đối tượng (sản phẩm, khóa học, bài hát...) phù hợp nhất với nhu cầu người dùng. Để giải quyết vấn đề đó, hệ khuyến nghị ra đời, tìm ra các thuật toán và áp dụng vào những lĩnh vực cụ thể của đời sống để đưa ra khuyến nghị, đó xem như là những lời gợi ý, lời khuyên cho người dùng. Hệ khuyến nghị là một hướng nghiên cứu trong khai phá dữ liệu đặc biệt được quan tâm của các nhà khoa học từ những đầu năm thập niên 90 và không ngừng được nghiên cứu và phát triển đến hiện nay.

Thông qua quá trình nghiên cứu của các nhà khoa học, nhiều hướng tiếp cận ra đời nhằm giải quyết vấn đề khuyến nghị cho người dùng như lọc nội dung, lọc cộng tác... trong đó tập trung chủ yếu vào các khái niệm: người dùng, sản phẩm, đánh giá, khuyến nghị. Bên cạnh đó, thông qua quá trình nghiên cứu khảo sát, người ta còn nhận thấy yếu tố ngữ cảnh đóng một vai trò hết sức quan trọng trong việc khuyến nghị. Vì thế một khía cạnh mới của hệ khuyến nghị kết hợp với yếu tố ngữ cảnh ra đời với mục đích hoàn thiện việc khuyến nghị và nâng cao chất lượng khuyến nghị cho người dùng. Hệ khuyến nghị dựa trên ngữ cảnh mang một phạm trù rộng lớn, tồn tại rất nhiều vấn đề và thử thách cho các nhà khoa học, nhà nghiên cứu. Vì vậy, tôi lựa chọn đề tài này

thực hiện trong khóa luận tốt nghiệp đại học để nghiên cứu một số vấn đề cơ bản của hệ khuyến nghị dựa trên ngữ cảnh. Mục đích chính của khóa luận là tổng quát hóa các khía cạnh của hệ khuyến nghị dựa trên ngữ cảnh một cách tương đối toàn diện nhất, đưa ra một bức tranh toàn cảnh bao gồm ngữ cảnh, các phương pháp xử lý ngữ cảnh, các kỹ thuật khuyến nghị dựa trên ngữ cảnh, giải thuật khuyến nghị dựa trên lọc cộng tác kết hợp với ngữ cảnh và các phương pháp đánh giá. Qua đó, phân tích được ưu nhược điểm của từng phương pháp, giải thuật để giúp các nhà phát triển phần mềm có thể lựa chọn được hướng tiếp cận phù hợp với từng ứng dụng khác nhau. Bên cạnh đó, khóa luận triển khai xây dựng một dịch vụ khuyến nghị dựa trên ngữ cảnh trên môi trường Internet và triển khai minh họa một số ứng dụng tương tác với dịch vụ. Cụ thể là hệ thống website nghe nhạc dựa trên ngữ cảnh tâm trạng và thời gian với một ứng dụng quy mô nhỏ hơn là ứng dụng khuyến nghị quà tặng dựa trên các yêu cầu tặng quà khác nhau.

Do thời gian nghiên cứu giới hạn, nội dung đề tài còn nhiều hạn chế và không tránh khỏi những thiếu sót, kính mong quý Thầy Cô, bạn bè đóng góp ý kiến để đề tài được hoàn thiện hơn, mang nhiều ý nghĩa thực tiễn hơn.

Thành phố Hồ Chí Minh, ngày 26 tháng 02 năm 2013

Sinh viên thực hiện

Lê Nguyễn Hào Hiệp

Lời cảm ơn

Trước tiên, tôi xin cảm ơn tất cả các giảng viên khoa Hệ thống Thông Tin – Đại học Công Nghệ Thông Tin đã trực tiếp giảng dạy và cung cấp cho tôi nhiều kiến thức nền tảng quý giá trong suốt thời gian học tập tại môi trường đại học. Đặc biệt, tôi xin chân thành cảm ơn sâu sắc đến ThS Đỗ Thị Minh Phụng đã trực tiếp hướng dẫn, hỗ trợ giúp đỡ tôi trong suốt quá trình thực hiện khóa luận.

Tiếp theo, tôi xin chân thành cảm ơn tác giả Umberto Panniello¹ đã cung cấp tài liệu tham khảo với những giải đáp bổ ích về thuật toán, tác giả Hideki Asoh² và Andrej Kösir³ đã cung cấp dữ liệu kiểm thử Food và Comoda trong quá trình thực hiện đánh giá kiểm thử. Tôi cũng không quên sự giúp đỡ của hai cựu sinh viên⁴ đã cung cấp luận văn tốt nghiệp mà họ đã thực hiện trước đó để tôi tham khảo cách trình bày cũng như những vấn đề về nội dung liên quan. Xin chân thành cảm ơn tất cả những bạn bè đã giúp đỡ tôi rất nhiều trong quá trình thu thập dữ liệu kiểm thử, chính họ, những người mà tôi có thể không biết tên cũng là những nhân tố hết sức quan trọng để tôi có thể hoàn thành luận văn này.

Cuối cùng, tôi xin bày tỏ lòng biết ơn sâu sắc đến gia đình, bạn bè, người thân đã khích lệ, động viên tôi suốt thời gian thực hiện đề tài này.

Thành phố Hồ Chí Minh, ngày 26 tháng 02 năm 2013

Sinh viên thực hiện

Lê Nguyễn Hào Hiệp

¹ Umberto Panniello, u.panniello@poliba.it, Politecnico di Bari (Italy), Viale Japigia 182.

² Hideki Asoh, h.asoh@aist.go.jp, Information Technology RI, AIST, 1-1-1, Umezono, Tsukuba, Ibaraki 305-8568 Japan.

³ Andrej Kösir, andrej.kosir@fe.uni-lj.si, Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, 1000 Ljubljana, Slovenia.

⁴ Cựu sinh viên Đại học Công nghệ thông tin: Huỳnh Tùng và Tào Quang Hùng – Sinh viên khoa Hệ thống thông tin, khóa 2007 – 2012.

Lời cam đoan

Tôi, Lê Nguyễn Hào Hiệp xác nhận nội dung trình bày trong báo cáo này dựa trên những tổng hợp lý thuyết và kiến thức tích lũy của bản thân. Các số liệu, những kết luận nghiên cứu được trình bày trong luận văn này là trung thực và chưa từng được công bố dưới bất cứ hình thức nào. Mọi thông tin trích dẫn đều được chú thích và liệt kê rõ ràng thành các tài liệu tham khảo.

Tôi xác nhận báo cáo khóa luận tốt nghiệp này là công trình nghiên cứu của tôi dưới sự hướng dẫn của ThS Đỗ Thị Minh Phụng và sự giúp đỡ của những cá nhân khác đã được ghi nhận.

Thành phố Hồ Chí Minh, ngày 26 tháng 02 năm 2013

Lê Nguyễn Hào Hiệp

Lời nhận xét của giảng viên hướng dẫn

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Thành phố Hồ Chí Minh, ngày 26 tháng 02 năm 2013

Giảng viên hướng dẫn

ThS Đỗ Thị Minh Phụng

Lời nhận xét của giảng viên phản biện

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Thành phố Hồ Chí Minh, ngày tháng 03 năm 2013

Giảng viên phản biện

Mục lục

Danh sách bảng	x
Danh sách hình ảnh	xi
Danh mục từ viết tắt	xii
Danh mục thuật ngữ Anh – Việt	xiii
1 Tổng quan đề tài	1
1.1 Đặt vấn đề	1
1.2 Mục tiêu	4
1.3 Ý nghĩa của đề tài	4
1.4 Đối tượng, phạm vi và phương pháp nghiên cứu	5
1.5 Nội dung thực hiện.....	6
1.6 Kết quả dự kiến	7
1.7 Bố cục báo cáo	7
2 Hệ khuyến nghị dựa trên ngữ cảnh	9
2.1 Hệ khuyến nghị	9
2.1.1 Chức năng.....	10
2.1.2 Vấn đề của hệ khuyến nghị	11
2.1.3 Kỹ thuật tiếp cận.....	11
2.2 Ngữ cảnh.....	14
2.2.1 Định nghĩa	14
2.2.2 Thu thập ngữ cảnh	15
2.2.3 Biểu diễn ngữ cảnh trong hệ khuyến nghị.....	16
2.2.4 Lựa chọn ngữ cảnh trong hệ khuyến nghị.....	18
2.3 Hệ khuyến nghị dựa trên ngữ cảnh	19
2.4 Các phương pháp tiếp cận trong hệ khuyến nghị dựa trên ngữ cảnh	20
2.4.1 Tiền xử lý ngữ cảnh (Contextual Pre-Filtering).....	23
2.4.2 Hậu xử lý ngữ cảnh (Contextual Post-Filtering)	24

2.4.3	Mô hình hóa ngữ cảnh (Contextual Modeling)	25
2.4.4	Kết hợp các kỹ thuật.....	26
3	Kỹ thuật lọc cộng tác	27
3.1	Kỹ thuật lọc cộng tác	28
3.1.1	Lọc cộng tác	28
3.1.2	Kết hợp ngữ cảnh trong lọc cộng tác	29
3.2	Vấn đề trong lọc cộng tác	31
3.2.1	Dữ liệu thừa.....	31
3.2.2	Khả năng đáp ứng.....	31
3.2.3	Đồng nghĩa	32
3.2.4	Vấn đề khác	32
3.3	Đánh giá kỹ thuật lọc cộng tác.....	32
3.3.1	Đánh giá dự đoán	33
3.3.2	Đánh giá khuyến nghị.....	33
3.4	Giải thuật lọc cộng tác trong khuyến nghị truyền thống	37
3.4.1	User-Based	37
3.4.2	Baseline Predictor.....	39
3.4.3	Matrix Factorization	40
3.5	Giải thuật lọc cộng tác trong hệ khuyến nghị dựa trên ngữ cảnh	42
3.5.1	Item Splitting.....	42
3.5.2	Weigh POF và Filter POF	43
3.5.3	Contextual Neighbors.....	45
3.5.4	Context-Aware Matrix Factorization	47
3.6	Kết luận chung	49
4	Đánh giá	53
4.1	Những vấn đề trong kiểm thử đánh giá hệ khuyến nghị.....	53
4.2	Các phương pháp phân chia dữ liệu.....	54
4.3	Dữ liệu đánh giá.....	55
4.4	Phương pháp phân chia dữ liệu đánh giá.....	57

4.5	Kết quả đánh giá thực tế	58
4.5.1	So sánh khuyến nghị 2D và Pre-Filtering	60
4.5.2	So sánh khuyến nghị 2D và Post-Filtering	62
4.5.3	So sánh khuyến nghị 2D và Contextual Modeling.....	63
4.5.4	So sánh Pre-Filtering, Post-Filtering và Contextual Modeling	64
4.5.5	So sánh giải thuật xử lý và không xử lý ngữ cảnh liên quan	67
4.6	Kết luận chung	70
5	Ứng dụng minh họa	72
5.1	Dịch vụ khuyến nghị dựa trên ngữ cảnh.....	73
5.1.1	Kiến trúc của dịch vụ.....	73
5.1.2	Thành phần của dịch vụ.....	74
5.1.3	Hoạt động của dịch vụ.....	75
5.2	Các ứng dụng tương tác với dịch vụ khuyến nghị dựa trên ngữ cảnh	77
5.2.1	Hệ thống nghe nhạc HMusic	77
5.2.2	Ứng dụng khuyến nghị quà tặng HGift.....	81
5.3	Kết luận chung	83
6	Kết luận và Hướng phát triển	84
6.1	Kết quả	84
6.2	Khó khăn và hạn chế.....	85
6.3	Hướng phát triển	86
6.4	Tổng kết	87
	Tài liệu tham khảo	88
	Phụ lục A. Giá trị tham số trong các giải thuật	92

Danh sách bảng

Bảng 3.1:	Ma trận đánh giá	28
Bảng 3.2:	Ma trận đánh giá mở rộng với ngữ cảnh.....	30
Bảng 3.3:	Các tham số liên quan đến chỉ số Precision và Recall.....	34
Bảng 3.4:	Kết quả giải thuật MF với số lần lặp thay đổi	41
Bảng 3.5:	Kết quả so sánh giải thuật MF, Weigh MF và Filter MF trên dữ liệu Comoda	45
Bảng 3.6:	Kết quả so sánh giải thuật UB, EPF-UB, CN trên dữ liệu Comoda	47
Bảng 3.7:	Phân tích ưu nhược điểm của kỹ thuật dựa trên bộ nhớ và dựa trên mô hình	50
Bảng 3.8:	Phân tích ưu nhược điểm của một số giải thuật CACF tiêu biểu	51
Bảng 4.1:	Bảng tổng hợp dữ liệu kiểm thử	57
Bảng 4.2:	Số lượng sản phẩm khuyến nghị trên các dữ liệu kiểm thử khác nhau	58
Bảng 4.3:	Bảng tổng hợp các giải thuật kiểm thử	59
Bảng 4.4:	Bảng phân nhóm so sánh các giải thuật và dữ liệu liên quan	59
Bảng 4.5:	Kết quả kiểm thử giải thuật khuyến nghị 2D và Pre-Filtering	60
Bảng 4.6:	Kết quả kiểm thử giải thuật khuyến nghị 2D và Post-Filtering	62
Bảng 4.7:	Kết quả kiểm thử giải thuật khuyến nghị 2D và Contextual Modeling.....	63
Bảng 4.8:	Kết quả kiểm thử giải thuật khuyến nghị Pre, Post-Filtering và Contextual Modeling dựa trên bộ nhớ	65
Bảng 4.9:	Kết quả kiểm thử giải thuật khuyến nghị Pre, Post-Filtering và Contextual Modeling dựa trên mô hình.....	66
Bảng 4.10:	Kết quả so sánh giải thuật ISMF, CAMF khi xử lý ngữ cảnh liên quan trên dữ liệu Comoda, HMusic và HGift.....	68
Bảng 5.1:	Ngữ cảnh trong hệ thống nghe nhạc HMusic	78
Bảng 5.2:	Bảng so sánh giải thuật MF, ISMF và CAMF trong ứng dụng HMusic ...	81
Bảng 5.3:	Ngữ cảnh trong ứng dụng HGift.....	81
Bảng 5.4:	Bảng so sánh giải thuật MF, ISMF và CAMF trong ứng dụng HGift.....	83
Bảng A.1:	Bảng giá trị tham số sử dụng trong các thuật toán trên các dữ liệu.....	92

Danh sách hình ảnh

Hình 2.1:	Không gian 3 chiều User×Item×Time	17
Hình 2.2:	Biểu diễn thông tin ngữ cảnh tổng quát bằng cấu trúc phân cấp	18
Hình 2.3:	Quy trình khuyến nghị 2D	22
Hình 2.4:	Các kỹ thuật khuyến nghị dựa trên ngữ cảnh	23
Hình 2.5:	Khuyến nghị dựa trên ngữ cảnh với kỹ thuật kết hợp	26
Hình 3.1:	Đồ thị Precision – Recall (PR).....	37
Hình 3.2:	Phân chia sản phẩm.....	42
Hình 5.1:	Kiến trúc tổng quát của RESTful Web Service	73
Hình 5.2:	Định dạng JSON các thể hiện trong dịch vụ khuyến nghị.....	74
Hình 5.3:	Sơ đồ trình tự tương tác giữa ứng dụng client và dịch vụ khuyến nghị	77
Hình 5.4:	Thu thập ngữ cảnh và đánh giá trong hệ thống HMusic.....	79
Hình 5.5:	Khuyến nghị cho người dùng trên website hmusic.us	80
Hình 5.6:	Thu thập ngữ cảnh và đánh giá trong ứng dụng HGift.....	82

Danh mục từ viết tắt

CACF	Context-Aware Collaborative Filtering
CAMF	Context-Aware Matrix Factorization
CARS	Context-Aware Recommender Systems
CBF	Content Based Filtering
CF	Collaborative Filtering
CM	Contextual Modeling
CN	Contextual Neighbors
EPF	Exact Pre Filtering
EPF-MF	Exact Pre-Filtering Matrix Factorization
EPF-UB	Exact Pre-Filtering User Based
IS	Item Splitting
ISMF	Item Splitting with Matrix Factorization
ISUB	Item Splitting with User Based
KB	Knowledge Base
MF	Matrix Factorization
POF	Post Filtering
RS	Recommender Systems
SVD	Singular Vector Decomposition
UB	User-Based

Danh mục thuật ngữ Anh – Việt

Bias	Tính thành kiến – thiên vị
Collaborative Filtering	Lọc cộng tác
Community Based	Dựa trên cộng đồng
Context-Aware Collaborative Filtering	Lọc cộng tác ngữ cảnh
Context-Aware Recommender Systems	Hệ khuyến nghị dựa trên ngữ cảnh
Context-Based Filtering	Lọc nội dung
Contextual Generalization	Tổng quát hóa ngữ cảnh
Contextual Modeling	Mô hình hóa ngữ cảnh
Contextual Neighbors	Ngữ cảnh láng giềng
Contextual Post-Filtering	Hậu xử lý ngữ cảnh
Contextual Pre-Filtering	Tiền xử lý ngữ cảnh
Exact Pre-Filtering	Tiền lọc ngữ cảnh chính xác
Hybird Systems	Hệ thống lai
Information Retrieval	Tìm kiếm thông tin
Interactional View	Cách nhìn tương tác
Item Splitting	Phân chia sản phẩm
Knowlegde Based	Dựa trên tri thức
Machine Learning	Máy học
Matrix Factorization	Phân rã ma trận
Memory Based	Dựa trên bộ nhớ
Model Based	Dựa trên mô hình
Pairwised t-test	Kiểm định t theo cặp
Recommender Systems	Hệ khuyến nghị
Relational Property	Thuộc tính quan hệ
Relevant context	Ngữ cảnh liên quan
Representation	Thể hiện
Representational View	Cách nhìn đại diện
Resource	Tài nguyên
Stochastic Gradient Descent	Phương pháp giảm đồi
Confidence Level	Mức tin cậy

TỔNG QUAN ĐỀ TÀI

Để mở đầu cho khóa luận tốt nghiệp “Xây dựng hệ khuyến nghị dựa trên ngữ cảnh”, nội dung chương sẽ giới thiệu tổng quát về đề tài. Nội dung trình bày của chương bao gồm: khái quát vấn đề chính của đề tài, mục tiêu cùng với phương pháp nghiên cứu và nội dung thực hiện. Bên cạnh đó, phạm vi của đề tài và các kết quả dự kiến của đề tài cũng sẽ được trình bày chi tiết trong chương này.

1.1 Đặt vấn đề

Sự ra đời và bùng nổ của Internet đã kết nối thế giới với hàng ngàn kho dữ liệu khổng lồ bên cạnh sự phát triển của những dịch vụ trực tuyến như e-learning, giải trí trực tuyến và đặc biệt là thương mại điện tử (e-commerce) với một số hệ thương mại nổi tiếng như amazon, itunes... Một số hệ thương mại thường cung cấp cho nhiều đối tượng người dùng số lượng lớn sản phẩm đa chủng loại, điều này gây ra trở ngại lớn cho người dùng trong việc tìm ra những đối tượng sản phẩm liên quan đến nhu cầu của mình, để giải quyết vấn đề này, người dùng cần được **khuyến nghị**. Có rất nhiều định nghĩa khuyến nghị như: khuyến nghị đưa ra lời khuyên, lời gợi ý, khuyến nghị là đề nghị một đối tượng người nhận sẽ thích hoặc đưa ra một vấn đề người nhận mong chờ. Điều này đặt ra nhu cầu cao trong việc khuyến nghị hiệu quả cho người dùng. Với những lý do đó, **hệ khuyến nghị** ra đời, đó là những ứng dụng thông minh có thể *tự nhận biết* và *đề nghị* những nội dung, thông tin, sản phẩm... phù hợp nhất với nhu cầu và sở thích của người dùng. Trong những hệ khuyến nghị truyền thống này, khái niệm người dùng, sản phẩm và đánh giá là những thông tin tham gia vào quá trình xử lý đề

đưa ra khuyến nghị. Tuy nhiên, ngoài những thành công và hiệu quả mà hệ khuyến nghị mang lại (xem 2.1.1), những khó khăn và thách thức cũng luôn tồn tại song song (xem 2.1.2).

Hệ khuyến nghị truyền thống thường khai thác thông tin lịch sử người dùng trước đó để đưa ra những dự đoán trong tương lai với giả định hành vi của người dùng không thay đổi nhanh chóng trong thời gian ngắn. Giả định này chỉ hoàn toàn đúng khi hành vi thực của người dùng tương đối ổn định, tuy nhiên trên thực tế, những hành vi của người dùng còn chịu tác động của nhiều điều kiện và nhân tố khác nhau. Ví dụ đơn giản như việc lựa chọn địa điểm du lịch của người dùng còn phụ thuộc vào yếu tố thời tiết, bạn đồng hành... Phần lớn hệ khuyến nghị truyền thống gặp phải vấn đề trong việc đáp ứng những trường hợp này, khi đó chất lượng của khuyến nghị trở nên giảm sút và kém tin cậy [1]. Những yếu tố tác động đến việc khuyến nghị đó được gọi là *ngữ cảnh*, đây là một đa khái niệm trong nhiều lĩnh vực khác nhau. Cụ thể, trong lĩnh vực khoa học tính toán, Dey [2] đã định nghĩa “*ngữ cảnh là bất kỳ thông tin nào được sử dụng để mô tả những yếu tố hoàn cảnh của một thực thể. Một thực thể là một người, một địa điểm hoặc một đối tượng liên quan đến tương tác giữa người dùng và ứng dụng*” (xem chương 2).

Việc mở rộng khuyến nghị kết hợp với thông tin ngữ cảnh đã làm thay đổi cơ bản hiệu suất của các kỹ thuật giải quyết khuyến nghị truyền thống trước đó và đặt ra yêu cầu tìm ra các kỹ thuật giải quyết. Adomavicius và cộng sự [3] đã mô hình hóa các kỹ thuật khuyến nghị kết hợp ngữ cảnh với ba kỹ thuật chính là: tiền xử lý ngữ cảnh (contextual pre-filtering), hậu xử lý ngữ cảnh (contextual post-filtering) và mô hình hóa ngữ cảnh (contextual modeling). Kỹ thuật tiền xử lý ngữ cảnh dùng thông tin ngữ cảnh để điều chỉnh dữ liệu, kỹ thuật hậu xử lý ngữ cảnh dùng thông tin ngữ cảnh để điều chỉnh sản phẩm khuyến nghị trong khi đó kỹ thuật mô hình hóa ngữ cảnh sử dụng trực tiếp thông tin ngữ cảnh trong quá trình xử lý dự đoán đánh giá. Các kỹ thuật này đều có nhiều thuật toán khác nhau và đều có những ưu nhược điểm riêng.

Trong khuyến nghị truyền thống, có rất nhiều hướng tiếp cận được đề xuất và áp dụng như lọc cộng tác, lọc nội dung, dựa trên tri thức... Tuy nhiên, lọc cộng tác được xem như hướng tiếp cận hiệu quả và thành công nhất trong hệ khuyến nghị hiện

nay. Hướng tiếp cận này nhanh chóng được áp dụng rộng rãi trong các hệ thương mại điện tử [4] và được chia thành ba nhóm: dựa trên bộ nhớ (memory-based), dựa trên mô hình (model-based) và lai (hybrid) (xem chương 3) và hướng tiếp cận này cũng được áp dụng trong các kỹ thuật tiền xử lý ngữ cảnh, hậu xử lý ngữ cảnh và mô hình hóa ngữ cảnh trong các hệ khuyến nghị kết hợp với yếu tố ngữ cảnh. Cụ thể, Panniello và cộng sự [5], [6] đã đưa các giải thuật khác nhau (EPF, Weigh PoF, Filter PoF và Contextual Neighbors) và so sánh các giải thuật này để đưa ra cái nhìn tổng quát cũng như mức độ hiệu quả của các giải thuật này, kết quả cho thấy các kỹ thuật khuyến nghị kết hợp với ngữ cảnh hiệu quả hơn (hoặc rất ít) so với giải thuật khuyến nghị truyền thống. Các giải thuật được Panniello và cộng sự sử dụng [5], [6] đều là các kỹ thuật lọc cộng tác dựa trên bộ nhớ và dữ liệu được dùng trong quá trình kiểm thử là dữ liệu giao dịch (*transaction-based*) với giá trị đánh giá là tần số lần mua hàng của các sản phẩm. Bên cạnh đó, Baltrunas và cộng sự [1], [7], [8] đã đề xuất hai giải thuật Item Splitting và phân rã ma trận tương ứng với kỹ thuật Pre-Filtering và Contextual Modeling sử dụng hướng tiếp cận dựa trên mô hình (model-based) (xem chương 3).

Hệ khuyến nghị dựa trên ngữ cảnh đã đặt ra nhiều khó khăn và thách thức cho các nhà khoa học trong việc tìm ra phương pháp, giải thuật để cải thiện hiệu quả khuyến nghị, đặc biệt là khó khăn với dữ liệu kiểm thử. Phần lớn dữ liệu đánh giá kiểm thử rất hạn chế về số lượng cũng như chất lượng và chưa có bộ dữ liệu ngữ cảnh “tiêu chuẩn” để phục vụ cho việc đánh giá và so sánh giữa các thuật toán với nhau. Một số bộ dữ liệu thỏa mãn yêu cầu chất lượng và số lượng đều là những bộ dữ liệu chỉ được sử dụng trong mục đích nghiên cứu nội bộ của các tổ chức, cơ quan trong và ngoài nước.

Mục đích chính của hệ khuyến nghị là việc đưa ra cho người dùng nội dung mà họ quan tâm một cách nhanh nhất và hiệu quả nhất. Với lý do đó, khóa luận cũng tiến hành thực hiện xây dựng các hệ khuyến nghị. Cụ thể trong khóa luận này sẽ xây dựng một dịch vụ khuyến nghị dựa trên ngữ cảnh là một dịch vụ web (web service) với chức năng thực thi các thuật toán khuyến nghị và đưa ra kết quả cho các ứng dụng sử dụng dịch vụ. Bên cạnh đó, để minh họa sự tương tác web service này, khóa luận cũng xây dựng hai ứng dụng minh họa là hệ thống website nghe nhạc với ngữ cảnh là tâm trạng

và thời gian và một ứng dụng khác đó là ứng dụng khuyến nghị quà tặng với các thông tin ngữ cảnh là đối tượng tặng quà, giới tính, độ tuổi người nhận và lý do tặng quà. Hai ứng dụng này sẽ sử dụng dịch vụ khuyến nghị dựa trên ngữ cảnh để đưa ra các khuyến nghị cho người dùng tại mỗi ứng dụng mà không cần phải xây dựng thuật toán khuyến nghị, các ứng dụng này sẽ được minh họa với các giải thuật khác nhau sẽ được trình bày chi tiết trong chương 5.

1.2 Mục tiêu

Khóa luận được thực hiện với ba mục tiêu chính:

Thứ nhất, tìm hiểu về hệ khuyến nghị và yếu tố ngữ cảnh, các kỹ thuật tiếp cận hệ khuyến nghị và các kỹ thuật xử lý khuyến nghị với ngữ cảnh, đặc biệt là kỹ thuật lọc cộng tác (collaborative filtering) với một số thuật toán tiêu biểu. Bên cạnh đó, khóa luận cũng tìm hiểu phương pháp đánh giá kiểm thử trong hệ khuyến nghị bao gồm các chỉ số đánh giá, cách tính các chỉ số và một số phương pháp phân chia dữ liệu.

Thứ hai, thực hiện xây dựng cài đặt các thuật toán và thử nghiệm trên một số bộ dữ liệu khác nhau kết hợp các số chỉ số đánh giá để đưa ra so sánh giữa các thuật toán.

Thứ ba, xây dựng dịch vụ khuyến nghị dựa trên ngữ cảnh (context-aware recommender service) cung cấp chức năng khuyến nghị cho các ứng dụng tương tác với dịch vụ thông qua một số thuật toán tiêu biểu. Khóa luận cũng thực hiện xây dựng một số ứng dụng tương tác với web service với thuật toán khuyến nghị khác nhau.

1.3 Ý nghĩa của đề tài

Ý nghĩa thực tế: Xây dựng dịch vụ khuyến nghị dựa trên ngữ cảnh là một dịch vụ web dùng chung cho các ứng dụng khuyến nghị, dịch vụ giúp tiết kiệm thời gian xây dựng ứng dụng khuyến nghị. Mặt khác, các ứng dụng minh họa tương tác với dịch vụ như khuyến nghị thông qua hệ thống “Nghe nhạc dựa trên tâm trạng và thời gian” giúp người dùng có thể tìm thấy bài hát phù hợp với sở thích cũng như tâm trạng trong kho nhạc số khổng lồ, hay ứng dụng khuyến nghị quà tặng với các thông tin ngữ cảnh là đối tượng tặng quà, độ tuổi, giới tính người nhận và lý do tặng quà. Những ứng dụng này mang đến cho người sử dụng khả năng tiếp cận với nội dung phù hợp nhất một

cách nhanh chóng và hiệu quả, nói cách khác những ứng dụng này làm giảm sự quá tải thông tin cho người sử dụng. Nội dung của đề tài còn cho thấy tầm quan trọng và lợi ích to lớn của hệ khuyến nghị nói chung và hệ khuyến nghị dựa trên ngữ cảnh nói riêng trong đời sống con người.

Ý nghĩa nghiên cứu khoa học: Đề tài khuyến nghị dựa trên ngữ cảnh là đề tài khá mới mẻ và được rất nhiều các nhà khoa học quan tâm hiện nay. Việc khái quát hóa các vấn đề về khuyến nghị, ngữ cảnh, các kỹ thuật giải quyết, thuật toán, kiểm thử một cách chi tiết đã cho thấy một bức tranh toàn cảnh về hệ khuyến nghị dựa trên ngữ cảnh thông qua các công trình nghiên cứu liên quan, một số vấn đề đã được thực hiện cũng như một số vấn đề chưa được chú trọng tập trung. Hệ khuyến nghị thừa kế và áp dụng nhiều kiến thức từ nhiều lĩnh vực khác nhau như xác suất thống kê hoặc khai phá dữ liệu, điều này cho thấy hệ khuyến nghị và các môn khoa học cơ bản (toán, thống kê) có một mối liên hệ chặt chẽ với nhau.

1.4 Đối tượng, phạm vi và phương pháp nghiên cứu

Đối tượng nghiên cứu

- Hệ khuyến nghị, chức năng, vấn đề tồn tại và các kỹ thuật giải quyết.
- Ngữ cảnh, phương pháp thu thập ngữ cảnh, kết hợp ngữ cảnh trong hệ khuyến nghị và các kỹ thuật xử lý khuyến nghị dựa trên ngữ cảnh.
- Phương pháp đánh giá hệ khuyến nghị.

Phạm vi nghiên cứu

- Tập trung vào hệ khuyến nghị dựa trên ngữ cảnh và kỹ thuật lọc cộng tác (collaborative filtering).
- Tìm hiểu kỹ thuật lọc cộng tác và kỹ thuật lọc cộng tác kết hợp với yếu tố ngữ cảnh. Các thuật toán tương ứng cho kỹ thuật lọc cộng tác bao gồm User-Based, phân tích ma trận (matrix factorization). Các thuật toán kết hợp ngữ cảnh: kỹ thuật pre-filtering bao gồm giải thuật *EPF*, *Item Splitting* và *Generalization Context*, kỹ thuật post-filtering bao gồm giải thuật *Weigh PoF*, *Filter PoF* và kỹ thuật contextual modeling bao gồm giải thuật *Contextual Neighbors* và *Context-Aware Matrix Factorization*.

- Thực hiện đánh giá hệ khuyến nghị trên các chỉ số *MAE*, *RMSE*, *Precision*, *Recall* thông qua phương pháp kiểm thử chéo (m-fold cross validation).

Phạm vi công nghệ

- Đối với dịch vụ khuyến nghị dựa trên ngữ cảnh, ngôn ngữ lập trình sử dụng là java với kiến trúc RESTful web service dựa trên framework Jersey⁵ (JAX-RS) và hệ quản trị dữ liệu phi quan hệ (NoSQL) MongoDB⁶.
- Đối với ứng dụng minh họa cài đặt thuật toán và kiểm thử, ngôn ngữ lập trình là java và hệ quản trị MySQL. Việc so sánh thuật toán được thực hiện trên năm bộ dữ liệu (Food, HMusic, Comoda, Movielens Semi-Synthetic, HGift) với các thông tin ngữ cảnh khác nhau và số lượng đánh giá khác nhau.
- Đối với ứng dụng minh họa các ứng dụng khuyến nghị dựa trên ngữ cảnh, ngôn ngữ lập trình được sử dụng là PHP, hệ quản trị MySQL. Dữ liệu thu thập thực tế trong thời gian ngắn với một lượng người dùng giới hạn.

Phương pháp nghiên cứu

- Phương pháp khảo sát, phân tích và ứng dụng các đối tượng nghiên cứu để xây dựng ứng dụng minh họa.

1.5 Nội dung thực hiện

- Khảo sát nghiên cứu các vấn đề về hệ khuyến nghị, ngữ cảnh, các phương pháp, giải thuật giải quyết và so sánh ưu nhược điểm của từng phương pháp.
- Thu thập dữ liệu thực tế những đánh giá từ người dùng thực với các sản phẩm khác nhau và ngữ cảnh khác nhau.
- Xây dựng dịch vụ khuyến nghị dựa trên ngữ cảnh dưới dạng web service dùng chung trong môi trường Internet cho nhiều ứng dụng khác nhau.
- Xây dựng hai ứng dụng khuyến nghị tương tác với dịch vụ khuyến nghị.
- Kiểm thử, đánh giá các thuật toán, đưa ra so sánh và báo cáo kết quả.

⁵ <http://jersey.java.net/> , truy cập lần cuối 22/02/2013

⁶ <http://www.mongodb.org/> , truy cập lần cuối 22/02/2013

1.6 Kết quả dự kiến

Các kết quả dự kiến của khóa luận như sau:

- Xây dựng ứng dụng thu thập dữ liệu đánh giá của người dùng về quà tặng và âm nhạc với số lượng ngữ cảnh, sản phẩm và người dùng khác nhau.
- Xây dựng các thuật toán khuyến nghị dựa trên ngữ cảnh và kiểm thử để đưa ra những nhận xét và so sánh các thuật toán.
- Triển khai xây dựng dịch vụ khuyến nghị dựa trên ngữ cảnh dưới dạng web service dùng chung cho nhiều ứng dụng khuyến nghị khác nhau.
- Triển khai hệ khuyến nghị nghe nhạc dựa trên tâm trạng và thời gian đưa ra những khuyến nghị nghe nhạc cho người dùng dựa vào ngữ cảnh được nhận biết thông qua tương tác của người dùng với hệ thống.
- Triển khai hệ khuyến nghị quà tặng với các thông tin ngữ cảnh là đối tượng tặng quà, độ tuổi, giới tính người nhận và lý do tặng quà.
- Hoàn thành báo cáo khóa luận tốt nghiệp đại học.

1.7 Bố cục báo cáo

Báo cáo được chia thành sáu chương, phần còn lại được bố cục như sau:

Chương 2 – Hệ khuyến nghị dựa trên ngữ cảnh

Nội dung chương sẽ khái quát hóa những vấn đề về hệ khuyến nghị bao gồm định nghĩa hệ khuyến nghị, các chức năng, vấn đề và kỹ thuật tiếp cận của hệ khuyến nghị. Bên cạnh đó cũng trình bày chi tiết về ngữ cảnh, phương pháp thu thập ngữ cảnh, kết hợp ngữ cảnh trong hệ khuyến nghị cùng với các kỹ thuật xử lý khuyến nghị dựa trên ngữ cảnh trong hệ khuyến nghị.

Chương 3 – Kỹ thuật lọc cộng tác

Hiện nay, có rất nhiều phương pháp tiếp cận giải quyết trong hệ khuyến nghị, nhưng kỹ thuật lọc cộng tác là một phương pháp được nghiên cứu và triển khai phổ biến nhất hiện nay vì tính hiệu quả của nó. Vì vậy, các vấn đề về kỹ thuật lọc cộng tác, kết hợp ngữ cảnh trong lọc cộng tác, những thách thức, thuật toán và các tiêu chí đánh giá trong lọc cộng tác sẽ được trình bày chi tiết trong chương này.

Chương 4 – Đánh giá

Để kiểm chứng tính hiệu quả của các thuật toán đã được trình bày trong chương 3, các thuật toán phải được kiểm thử thực tế thông qua các bộ dữ liệu khác nhau và dựa trên các chỉ số đánh giá khác nhau. Nội dung của chương sẽ trình bày phương pháp thực hiện kiểm thử, sau đó so sánh các kết quả kiểm thử thực tế để có thể phân tích những ưu, nhược điểm của từng thuật toán. Bên cạnh đó, việc so sánh các phương pháp xử lý ngữ cảnh cũng sẽ được trình bày trong nội dung này.

Chương 5 – Ứng dụng minh họa

Trong chương này, nội dung sẽ trình bày một cách tổng quát kiến trúc của dịch vụ khuyến nghị dựa trên ngữ cảnh và các ứng dụng minh họa tương tác với dịch vụ này. Cụ thể là hệ thống nghe nhạc HMusic và ứng dụng khuyến nghị quà tặng HGift. Nội dung chương chỉ trình bày một cách tổng quát các ứng dụng minh họa mà không đi sâu vào quá trình xây dựng và triển khai ứng dụng.

Chương 6 – Kết luận và hướng phát triển

Nội dung trình bày tổng quan những nội dung được thực hiện trong khóa luận, những thành quả cũng như những hạn chế khóa luận gặp phải. Bên cạnh đó đưa ra những hướng phát triển của đề tài trong tương lai.

HỆ KHUYẾN NGHỊ DỰA TRÊN NGỮ CẢNH

Hệ khuyến nghị trước đây là những ứng dụng có khả năng tự nhận biết và đưa ra những nội dung phù hợp cho người dùng với những khái niệm cơ bản như người dùng, sản phẩm, đánh giá. Qua thời gian, người ta nhận thấy có nhiều thông tin khác tác động đến hệ khuyến nghị gọi là **ngữ cảnh**, điều này cũng mang đến nhiều vấn đề mới cho hệ khuyến nghị kết hợp với yếu tố ngữ cảnh.

Phần đầu tiên của chương sẽ trình bày tổng thể chung về hệ khuyến nghị, những vấn đề tồn tại trong hệ khuyến nghị và tổng quát hóa các kỹ thuật giải quyết trong khuyến nghị. Phần tiếp theo sẽ trình bày về hệ khuyến nghị dựa trên ngữ cảnh bao gồm: những định nghĩa, quan điểm khác nhau nhìn nhận về ngữ cảnh, phương pháp thu thập ngữ cảnh, đánh giá tác động của ngữ cảnh. Quan trọng hơn đó là những phương pháp tiếp cận giải quyết bài toán khuyến nghị kết hợp với yếu tố ngữ cảnh. Các hướng tiếp cận bao gồm: Tiền xử lý ngữ cảnh (Contextual Pre-Filtering), Hậu xử lý ngữ cảnh (Contextual Post-Filtering) và Mô hình hóa ngữ cảnh (Contextual Modeling) [3], [9].

2.1 Hệ khuyến nghị

Hệ khuyến nghị (Recommender Systems) là những công cụ phần mềm và những kỹ thuật đưa ra những lời gợi ý về sản phẩm cho người dùng [9]. Hệ khuyến nghị thu thập và khai thác thông tin về người dùng, sản phẩm, sự tương tác giữa người dùng và sản phẩm để đưa ra một danh sách sản phẩm được cá nhân hóa phù hợp nhất với nhu cầu của người dùng. Những thông tin này có thể được thu thập trong quá trình tương

tác của người dùng với hệ thống hoặc từ hệ thống khác người dùng đã tương tác. Thông tin tương tác của người dùng được sử dụng phổ biến nhất là những đánh giá của người dùng về những sản phẩm. Sau đó hệ thống sẽ sử dụng những đánh giá này để dự đoán những đánh giá trên những sản phẩm mà người dùng chưa đánh giá, những sản phẩm có những đánh giá được dự đoán cao nhất sẽ được **khuyến nghị** cho người dùng [1]. Đánh giá của người dùng được biểu diễn khác nhau tùy thuộc vào nhu cầu của ứng dụng. Giá trị đánh giá có thể là một dãy số 1-5 phản ánh mức độ yêu thích của người dùng đến sản phẩm, giá trị nhị phân 0-1 (thích-không thích) hoặc có thể là tần suất một sản phẩm được mua, giao dịch [5], [6].

2.1.1 Chức năng

Theo Ricci và cộng sự [9], hệ khuyến nghị có những chức năng sau:

Tăng số lượng sản phẩm được bán ra là chức năng hết sức quan trọng trong các hệ khuyến nghị thương mại bằng cách khuyến nghị những sản phẩm mà có thể đáp ứng nhu cầu người dùng nhằm tăng lượng sản phẩm tiêu thụ. Tương tự với các hệ thống khuyến nghị phi thương mại, khuyến nghị giúp người dùng tiếp cận nhanh hơn đến những đối tượng, nội dung phù hợp nhất.

Tăng số loại sản phẩm bán ra cũng là một chức năng quan trọng trong hệ khuyến nghị. Sự đa dạng trong chủng loại sản phẩm cho phép người dùng lựa chọn sản phẩm phù hợp với sở thích và nhu cầu. Tuy nhiên, việc tiếp cận nhiều chủng loại sản phẩm khác nhau gây ra khó khăn cho người dùng so với việc hệ thống đưa ra khuyến nghị vì các hệ thống thương mại thường có số lượng chủng loại mặt hàng tương đối lớn.

Tăng sự hài lòng của người dùng, người dùng sẽ thích thú trong việc sử dụng những hệ thống đưa ra những lời gợi ý, lời khuyên hữu ích và liên quan đến nhu cầu thực tế, điều này đồng nghĩa với việc hệ khuyến nghị làm tăng sự hài lòng của người dùng.

Tăng sự trung thành của người dùng, người dùng có xu hướng sử dụng hệ thống lâu dài nếu nó đáp ứng được nhu cầu người dùng, tức là nó đưa ra cho người dùng những lời khuyên lựa chọn sản phẩm thực sử hữu ích.

2.1.2 Vấn đề của hệ khuyến nghị

Theo Ricci và cộng sự [9], hệ khuyến nghị tồn tại những vấn đề sau:

Khả năng đáp ứng dữ liệu lớn, một số giải thuật vận hành tương đối tốt và hiệu quả với dữ liệu nhỏ nhưng lại trở thành một giải thuật kém hiệu quả khi thực hiện trên dữ liệu lớn và không ngừng tăng trưởng.

Khuyến nghị một cách chủ động nhằm đưa ra những gợi ý cho người dùng mà không cần bất kỳ yêu cầu tường minh nào từ người dùng, điều này đòi hỏi hệ khuyến nghị phải nhận định nhu cầu một cách chính xác, nếu nhận định và đưa ra khuyến nghị không hợp lý sẽ gây ra hiện tượng làm phiền đến người dùng.

Tính riêng tư trong khuyến nghị, hệ khuyến nghị tập trung khai thác thông tin người dùng để đưa ra khuyến nghị, tuy nhiên việc thu thập quá nhiều thông tin người dùng có thể gây ra ảnh hưởng tiêu cực đến tính riêng tư của người dùng. Vì vậy, các hệ khuyến nghị cũng cần quan tâm đến việc bảo vệ thông tin của người dùng.

Đa dạng khuyến nghị, sự hài lòng của người dùng sẽ tăng lên nếu họ tiếp cận được nhiều loại sản phẩm mới hơn là nhận được lời khuyên trùng lặp và thiếu tính đa dạng về mặt chủng loại sản phẩm.

Ngoài ra còn một số vấn đề trong hệ khuyến nghị như: triển khai khuyến nghị trên nền tảng di động với tài nguyên hạn chế, xây dựng trình chủ khuyến nghị (recommender server), khuyến nghị trong môi trường hợp tác của người dùng...

2.1.3 Kỹ thuật tiếp cận

Theo Ricci [9], các kỹ thuật giải quyết trong hệ khuyến nghị được chia thành sáu loại chính bao gồm: Lọc cộng tác (Collaborative Filtering), Lọc nội dung (Content-Based Filtering), Dựa trên tri thức (Knowledge Based), Dựa trên cộng đồng (Community-Based), Demographic và Hệ thống lai (Hybrid Systems).

Collaborative Filtering (CF): là kỹ thuật⁷ được sử dụng phổ biến và rộng rãi nhất trong các hệ thống về khuyến nghị hiện nay, ý tưởng chính của kỹ thuật này là sử dụng những đánh giá của người dùng trên một số sản phẩm để đưa ra dự đoán và khuyến

⁷ Thuật ngữ Kỹ thuật, Phương pháp, Hướng tiếp cận đều có ý nghĩa tương đương trong khóa luận này.

những sản phẩm mới. Kỹ thuật CF được chia thành hai loại: dựa trên bộ nhớ (memory-based) và dựa trên mô hình (model-based).

- **Memory-based CF** sử dụng một phần hoặc toàn bộ dữ liệu về đánh giá của người dùng trên các sản phẩm để đưa ra dự đoán. Mỗi người dùng sẽ được xem như là một thành phần của một nhóm người dùng có cùng chung sở thích (xu hướng), nhóm người dùng đó gọi là “láng giềng”. Khi đó hệ thống sẽ đưa ra khuyến nghị là những sản phẩm đã được quan tâm bởi “láng giềng” mà người dùng mới (hoặc người dùng hiện tại) chưa quan tâm trước đó [4].
- **Model-based CF** không sử dụng toàn bộ dữ liệu để đưa ra đánh giá, thay vào đó, kỹ thuật này sẽ xây dựng mô hình (mô hình máy học, mô hình khai phá dữ liệu) phản ánh mức độ quan tâm của người dùng đến sản phẩm bằng cách tìm ra quy tắc, quy luật dựa trên dữ liệu huấn luyện (training data) [4], sau đó sẽ dựa vào mô hình này đưa ra dự đoán. Kỹ thuật này giúp khắc phục một số hạn chế trong kỹ thuật memory-based.

Content-Based Filtering (CBF): Kỹ thuật này đưa ra khuyến nghị là những sản phẩm tương đồng với những sản phẩm mà người dùng đã quan tâm trước đó. Mức độ tương đồng của các sản phẩm được tính toán dựa trên những đặc tính của sản phẩm (mô tả về sản phẩm). Lấy ví dụ trong khuyến nghị âm nhạc, một người dùng thích nghe một bài hát của một ca sĩ nào đó thì hệ thống sẽ đưa ra các khuyến nghị là những bài hát khác cũng được trình bày bởi ca sĩ đó.

Knowledge-Based (KB): Kỹ thuật này đưa ra khuyến nghị dựa trên những suy luận tri thức về những đặc tính của sản phẩm đáp ứng nhu cầu, sở thích và hữu ích với người dùng. Hầu hết khuyến nghị dựa trên tri thức thuộc loại dựa trên trường hợp (case-based), trong những hệ thống này, một hàm đánh giá mức độ nhu cầu của người dùng tương đương với khuyến nghị, ở đây, giá trị của hàm đánh giá được xem như giá trị dùng để khuyến nghị cho người dùng. Constraint-based Systems (hệ thống dựa trên ràng buộc) là một loại khác của Knowledge-Based Systems. Cả hai loại này đều tương tự nhau, yêu cầu của người dùng được thu thập, phục vụ cho các yêu cầu được đề xuất trong các tình huống khác, sau đó kết quả đề nghị được giải thích. Điểm khác nhau của hai loại này dựa trên cách xử lý: khuyến nghị dựa trên trường hợp sử dụng giá trị

tương đồng trong khi khuyến nghị dựa trên ràng buộc khai phá những tri thức được định nghĩa thông qua các luật thể hiện sự liên quan của yêu cầu người dùng đến những tính năng, đặc tính của sản phẩm.

Khuyến nghị dựa trên tri thức có xu hướng làm việc tốt hơn những kỹ thuật trước đó nhưng nếu kỹ thuật này không được trang bị tốt những thành phần “học” của một hệ tri thức, kỹ thuật này dễ dàng kém hiệu quả hơn so với những kỹ thuật khai phá tương tác của người dùng (như CF).

Community-Based: Kỹ thuật này đưa ra khuyến nghị dựa trên những thông tin bạn bè của người dùng, ý tưởng chung là “hãy nói cho tôi biết bạn bè của bạn là ai và tôi sẽ nói cho bạn biết bạn là ai”. Thực tế cho thấy người dùng thường nghe theo những khuyến nghị từ bạn bè của họ hơn là những khuyến nghị tương tự từ những cá nhân khác. Kết hợp với sự gia tăng của các mạng xã hội hiện nay, kỹ thuật này đang được quan tâm nhiều hơn trong các hệ thống khuyến nghị xã hội (social recommender systems), nó lấy những thông tin về quan hệ xã hội của người dùng và bạn bè của người dùng, sau đó khuyến nghị cho người dùng được đưa ra dựa trên những đánh giá từ bạn bè của người dùng đó.

Demographic: Khuyến nghị những sản phẩm dựa trên thông tin người dùng (user’s profile). Một số website đã áp dụng một cách đơn giản và có hiệu quả kỹ thuật này, chẳng hạn một người dùng truy cập vào một trang web cụ thể dựa trên vùng miền hoặc ngôn ngữ hoặc những đề nghị được điều chỉnh theo độ tuổi, giới tính. Kỹ thuật này đang dần phổ biến trong lĩnh vực marketing, tuy nhiên, kỹ thuật này ít được quan tâm nghiên cứu trong hệ khuyến nghị hiện nay.

Hybird Systems: Kỹ thuật này dựa trên sự kết hợp của nhiều kỹ thuật, kỹ thuật này kết hợp nhiều phương pháp để tận dụng được các ưu điểm và hạn chế các nhược điểm của từng phương pháp riêng rẽ, nói cách khác kỹ thuật này dùng ưu điểm của phương pháp này để khắc phục nhược điểm của phương pháp kia. Lấy ví dụ, trong kỹ thuật CF, một sản phẩm mới sẽ không bao giờ được khuyến nghị nếu như sản phẩm đó chưa được bất kỳ người dùng nào đánh giá, trong khi kỹ thuật CBF có thể đưa ra

khuyến nghị cho sản phẩm đó vì nó dựa trên những đặc tính của sản phẩm đó mà không quan tâm đến đánh giá của người dùng.

2.2 Ngữ cảnh

2.2.1 Định nghĩa

Ngữ cảnh (hay còn gọi là bối cảnh) là một đa khái niệm đã được nghiên cứu trên các lĩnh vực khác nhau, bao gồm cả khoa học máy tính, ngôn ngữ học, triết học, tâm lý học, và tổ chức khoa học...[3]. Do tính chất ngữ cảnh được sử dụng trong nhiều lĩnh vực khác nhau, mỗi lĩnh vực lại có một quan điểm khác biệt nên ngữ cảnh cần được định nghĩa cụ thể hơn so với định nghĩa chung “*ngữ cảnh được xem là những điều kiện hoặc những hoàn cảnh tác động đến một yếu tố nào đó*” [10]. Trên thực tế, tồn tại rất nhiều định nghĩa về ngữ cảnh trong các lĩnh vực khác nhau thậm chí kể cả trong một lĩnh vực cụ thể. Bazire và Brézillon [11] đã trình bày và đánh giá 150 định nghĩa ngữ cảnh thuộc nhiều lĩnh vực khác nhau để cho thấy tính phức tạp và đa khái niệm của ngữ cảnh, Bazire và Brézillon đã nhận định “*thật khó để định nghĩa ngữ cảnh có thể thỏa mãn hết mọi lĩnh vực*”.

Lĩnh vực khoa học máy tính cũng tồn tại rất nhiều định nghĩa về ngữ cảnh, Schilit và Theimer [12] định nghĩa ngữ cảnh như “*vị trí và định danh của người và đối tượng gần địa điểm đó*”. Sau đó, định nghĩa đó được mở rộng bởi Schilit và cộng sự [13] “*ngữ cảnh bao gồm nhiều hơn là vị trí của người dùng, ngữ cảnh còn bao gồm ánh sáng, độ nhiễu, kết nối mạng, chi phí truyền dữ liệu, băng thông và kể cả tình huống xã hội như việc đang ở cùng quản lý hay đồng nghiệp*”. Dey [2] đã định nghĩa một cách tổng quát hơn “*ngữ cảnh là bất kỳ thông tin nào được sử dụng để mô tả những yếu tố hoàn cảnh của một thực thể. Một thực thể là một người, một địa điểm hoặc một đối tượng liên quan đến tương tác giữa người dùng và ứng dụng*”. Đây là định nghĩa đầu tiên về ngữ cảnh phù hợp được áp dụng rộng rãi trong các ngành khoa học tính toán. Trong hệ khuyến nghị, thực thể là người dùng, sản phẩm và đánh giá từ người dùng. Ví dụ như, một người dùng cần tìm một món quà tặng thì người dùng đó cần quan tâm đến các yếu tố như: tặng cho ai, độ tuổi, giới tính và tặng vào dịp gì để tìm ra món quà phù hợp. Khi đó ngữ cảnh sẽ là đối tượng tặng quà, giới tính, độ tuổi

của người được tặng và lý do tặng món quà đó vì nó tác động trực tiếp đến việc lựa chọn quà của người dùng cũng như đưa ra khuyến nghị của hệ thống. Hoặc một ví dụ khác, người dùng đang buồn thì sẽ nghe bài hát gì, khi đó tâm trạng đóng vai trò là ngữ cảnh vì nó sẽ là cơ sở để một ứng dụng nghe nhạc đưa ra lời khuyên là các bài hát có thể phù hợp trong tâm trạng buồn.

Dourish [14] đã đưa ra phân tích về ngữ cảnh trong môi trường tính toán và đưa ra hai cách nhìn về ngữ cảnh: cách nhìn đại diện (representational view) và cách nhìn tương tác (interactional view). Theo đó, cách nhìn đại diện tách ngữ cảnh độc lập với hành động và cung cấp thông tin về hành động đó, ngữ cảnh của một hành động sẽ được định nghĩa trước và nó chỉ có một ý nghĩa duy nhất trong tất cả các hành động. Ngược lại, cách nhìn tương tác định nghĩa ngữ cảnh như là một thuộc tính quan hệ (relational property), phạm vi giá trị của ngữ cảnh sẽ luôn luôn thay đổi, nghĩa là sẽ không tồn tại một tập giá trị cho ngữ cảnh trước đó. Dourish [14] giải thích rằng: ngữ cảnh (theo cách nhìn tương tác) phát sinh từ hành động, ngữ cảnh không chỉ ở hành động đó nhưng được tạo ra, duy trì và hoạt động một cách tích cực trong hành động đó. Trong khóa luận này, ngữ cảnh sẽ mang cách nhìn đại diện (representational view), tức là ngữ cảnh sẽ được định nghĩa trước và có những miền giá trị xác định được đưa ra bởi những chuyên gia phát triển hệ thống [1], cách nhìn này đơn giản hơn và khả thi hơn trong việc tiếp cận và tính toán so với cách nhìn tương tác (interactional view).

Những ứng dụng có sử dụng ngữ cảnh được gọi là những hệ thống dựa trên ngữ cảnh (context-aware systems). Vì vậy hệ khuyến nghị sử dụng yếu tố ngữ được gọi là hệ khuyến nghị dựa trên ngữ cảnh, cụ thể ngữ cảnh trong hệ khuyến nghị đóng vai trò như những thông tin mở rộng (bên cạnh người dùng và sản phẩm) liên quan và ảnh hưởng đến việc đưa ra khuyến nghị [1].

2.2.2 Thu thập ngữ cảnh

Thông tin ngữ cảnh được thu thập bằng nhiều cách, bao gồm các cách sau đây:

Tường minh bằng cách tiếp cận trực tiếp với những đối tượng liên quan hoặc nguồn ngữ cảnh, ngữ cảnh có thể được thu thập trực tiếp thông qua những câu hỏi hoặc gợi ý những thông tin này thông qua những công cụ khác. Ví dụ như, một website có thể lấy

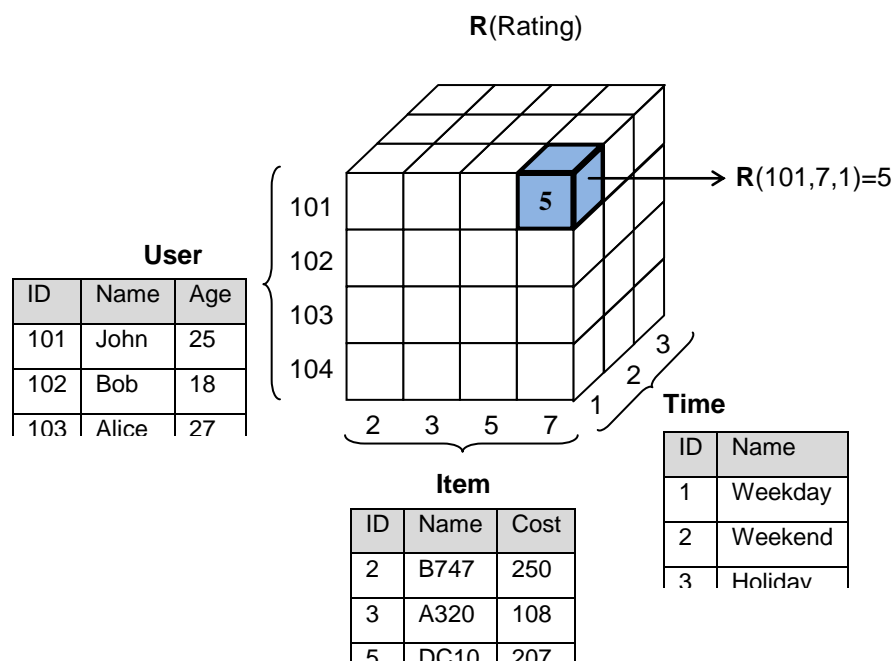
thông tin ngữ cảnh bằng cách yêu cầu người dùng điền thông tin vào một bảng nhập liệu hoặc trả lời một số câu hỏi cụ thể trước khi người dùng đó có thể truy cập vào website.

Tiềm ẩn từ dữ liệu hoặc từ môi trường phát triển ứng dụng, ví dụ như là địa điểm của người dùng được thu thập thông qua một nhà cung cấp dịch vụ di động hoặc thời gian có thể thu thập một cách tiềm ẩn thông qua nhãn thời gian thực hiện. Theo cách này, thông tin ngữ cảnh hoàn toàn được thu thập mà không cần phải thực hiện một hành động tương tác trực tiếp nào giữa người dùng hoặc nguồn dữ liệu ngữ cảnh.

Suy diễn ngữ cảnh sử dụng các phương pháp thống kê hoặc khai phá dữ liệu. Ví dụ như, nhà cung cấp dịch vụ truyền hình không thể thu thập một cách trực tiếp những loại kênh truyền hình được xem bởi một thành viên cụ thể trong gia đình (chồng, vợ, con...) nhưng có thể được thu thập bằng cách suy diễn thông qua việc khai thác những chương trình hoặc kênh truyền hình mà tất cả các thành viên đó đã lựa chọn. Để làm được điều đó, việc xây dựng một mô hình dự đoán (phân lớp, gom cụm...) là hết sức cần thiết, việc suy diễn hiệu quả còn phụ thuộc vào mô hình dự đoán cũng như đặc trưng dữ liệu.

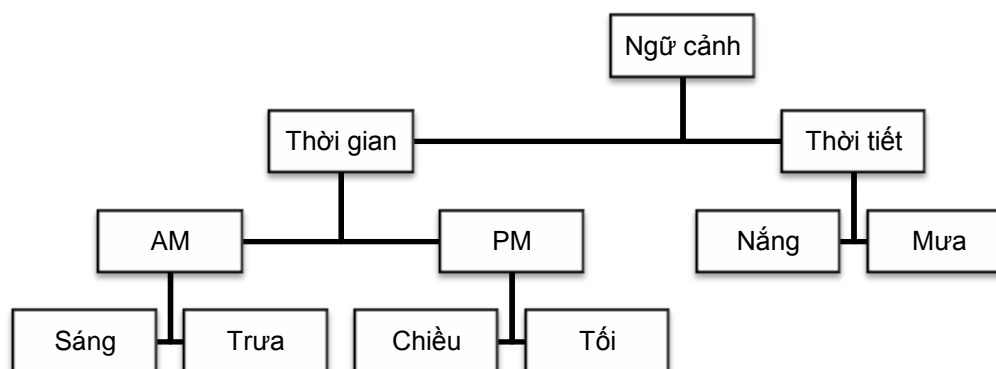
2.2.3 Biểu diễn ngữ cảnh trong hệ khuyến nghị

Ngữ cảnh được biểu diễn cùng với thông tin người dùng, sản phẩm dựa trên mô hình nhiều chiều tương tự OLAP [15], ngoài hai chiều người dùng (User), sản phẩm (Item), các thông tin ngữ cảnh được xem như các chiều mở rộng. Đặt D_1, D_2, \dots, D_n là các chiều, trong đó có hai chiều là người dùng và sản phẩm, các chiều còn lại là ngữ cảnh với mỗi chiều D_i là tập con của tích Descartes các thuộc tính A_{ij} , $j=(1, 2, \dots, k_i)$, $D_i \subseteq A_{i1} \times A_{i2} \times \dots \times A_{ik_i}$, trong đó thuộc tính định nghĩa bởi một tập các giá trị. Trong một số trường hợp, mỗi chiều được định nghĩa bằng một thuộc tính duy nhất và $k_i=1$ trong hầu hết các trường hợp. Lấy ví dụ thông tin ngữ cảnh là thời gian (Time). Khi đó đánh giá của người dùng dựa vào thời gian là giá trị của một ô trong khối (cube) lập phương được biểu diễn bởi **Hình 2.1**.



Hình 2.1: Không gian 3 chiều User×Item×Time

Thực chất, thông tin ngữ cảnh chính là một không gian nhiều chiều bao gồm nhiều ngữ cảnh, mỗi ngữ cảnh là một chiều và có nhiều giá trị tương ứng. Khi đó tập ngữ cảnh C thực chất là tập hợp ngữ cảnh C_i , trong đó mỗi ngữ cảnh được định nghĩa trước và có giá trị rời rạc theo cách nhìn đại diện (representational view), ngữ cảnh c được hiểu ngầm định là $c = (c_1, c_2, \dots, c_k), c_i \in C_i$. Giá trị $c_i \in C_i$ là giá trị đơn và thuộc về một cấp của ngữ cảnh C_i nếu ngữ cảnh đó có các giá trị phân cấp, giá trị phân cấp ở mức cao hơn còn gọi là tổng quát của mức thấp hơn. Lấy ví dụ, ta có ngữ cảnh là thời gian và thời tiết. Thời gian = {sáng, trưa, chiều, tối}, thời tiết = {nắng, mưa}. Các giá trị của thời gian có thể gom thành hai nhóm AM = {sáng, trưa}, PM = {chiều, tối}. Khi đó, giá trị của ngữ cảnh sẽ được biểu diễn thành một cấu trúc phân cấp như sau (Hình 2.2).



Hình 2.2: Biểu diễn thông tin ngữ cảnh tổng quát bằng cấu trúc phân cấp

2.2.4 Lựa chọn ngữ cảnh trong hệ khuyến nghị

Ngữ cảnh có tác động đến việc đánh giá, lựa chọn sản phẩm của người dùng, tuy nhiên không phải tất cả các ngữ cảnh đều liên quan hoặc hữu ích (relevant context) trong hệ khuyến nghị. Lấy ví dụ đối với một hệ khuyến nghị quà tặng, các thông tin ngữ cảnh có khả năng ảnh hưởng đến việc khuyến nghị có thể là: đối tượng, độ tuổi người nhận, giới tính người nhận, lý do tặng quà và thời gian mua quà. Rõ ràng, không phải các thông tin ngữ cảnh trên đều liên quan đến việc lựa chọn quà tặng. Thông tin về thời gian mua quà dường như chẳng tác động đến việc lựa chọn một món quà. Có hai cách tiếp cận để lựa chọn các thông tin ngữ cảnh liên quan là thủ công (manual) và tự động (automatic). Trong đó cách lựa chọn thủ công được thực hiện bởi các chuyên gia, nhà phân tích từ các lĩnh vực liên quan, cách lựa chọn tự động được thực hiện thông qua các kỹ thuật trong máy học, khai phá dữ liệu và xác suất thống kê ở giai đoạn tiền xử lý dữ liệu.

Adomavicius và cộng sự [15] đã áp dụng phương pháp tự động sử dụng xác suất thống kê để nhận diện các chiều ngữ cảnh nào là liên quan và thực sự ảnh hưởng đến khuyến nghị bằng cách kiểm tra hai phân phối xác suất. Để xác định có lựa chọn ngữ cảnh đó hay không, Adomavicius và cộng sự [15] đã chia dữ liệu liên quan đến ngữ cảnh đó thành nhiều nhóm, mỗi nhóm sẽ là các dữ liệu tương ứng với từng giá trị của ngữ cảnh. Lấy ví dụ, ta có ngữ cảnh thời tiết có hai giá trị là nắng và mưa, ta chia dữ liệu đánh giá thành hai nhóm tương ứng. Ứng với mỗi nhóm, tính giá trị trung bình

đánh giá cho từng người dùng, sau đó áp dụng kiểm thử t -test trên hai mẫu theo từng cặp (pairwised t -test) để kiểm chứng ý nghĩa thống kê trên hai mẫu đó. Đối với các ngữ cảnh chỉ có hai giá trị, ta có thể nhận biết được ngữ cảnh liên quan với một lần kiểm định, tuy nhiên đối với ngữ cảnh có nhiều giá trị ta lần lượt phải thực hiện tất cả kiểm định theo từng cặp tương đương với các cặp giá trị của ngữ cảnh đó. Phương pháp này đã được sử dụng và mang lại hiệu quả trong một số thử nghiệm [5], [6], [15]. Ngoài ra còn có một số phương pháp khác để nhận biết tác động của ngữ cảnh như Liu và cộng sự [16] sử dụng kiểm định chi bình phương hay Odic và cộng sự [17] sử dụng kiểm định Freeman-Halton để nhận biết ngữ cảnh liên quan. Bên cạnh đó, Baltrunas và cộng sự [18] đánh giá các ngữ cảnh liên quan bằng cách yêu cầu người dùng đánh giá trực tiếp tác động của ngữ cảnh có ảnh hưởng đến hành vi cá nhân người dùng hay không, sau đó dữ liệu đánh giá được thu thập trong trường hợp không có ngữ cảnh và có ngữ cảnh, tiếp theo áp dụng kiểm định t -test dựa trên đánh giá trung bình của từng sản phẩm trong trường hợp có ngữ cảnh và không có ngữ cảnh để kết luận ngữ cảnh liên quan. Tuy nhiên, phương pháp của Baltrunas có thể không chính xác vì người dùng có thể có hành vi khác nhau khi đánh giá sản phẩm với trường hợp ngữ cảnh thực tế và trường hợp ngữ cảnh giả định [17], [19]. Do thời gian hạn chế, nội dung khóa luận chỉ trình bày kết quả kiểm định thực tế đối với phương pháp được đề xuất bởi Adomavicius và cộng sự [15], chi tiết kiểm thử thực tế được trình bày trong chương 4, mục 4.5.5.

2.3 Hệ khuyến nghị dựa trên ngữ cảnh

Hệ khuyến nghị được xem như một lĩnh vực nghiên cứu độc lập từ những năm 1990 khi các nhà nghiên cứu bắt đầu tập trung vào vấn đề khuyến nghị như một cách nắm bắt sở thích người dùng (user) trên các sản phẩm (item) khác nhau. Quá trình được bắt đầu với những giá trị đánh giá (rating) trực tiếp từ người dùng hoặc suy diễn từ hệ thống. Khi những giá trị đánh giá ban đầu được xác định, hệ khuyến nghị sẽ tính toán hàm r cho mỗi cặp người dùng u , sản phẩm i .

$$r(u, i) : U \times I \rightarrow R \quad (2.1)$$

Trong đó, U là tập các người dùng, I là tập sản phẩm và R là tập các giá trị đánh giá được sắp thứ tự. Hàm r sẽ được tính toán trên toàn bộ ma trận $U \times I$, sau đó hệ khuyến nghị sẽ đưa ra sản phẩm (chưa được đánh giá bởi người dùng) có giá trị đánh giá cao nhất được suy ra từ hàm r . Nói cách khác, bài toán khuyến nghị đã được mô hình hóa thành hàm r , $\forall u \in U$ ta cần tìm $i \in I$ sao cho $r(u, i)$ đạt giá trị lớn nhất, khi đó i chính là sản phẩm khuyến nghị cho người dùng. Hệ khuyến nghị sử dụng ma trận $(U \times I)$ hai chiều được gọi là hệ khuyến nghị truyền thống hay hệ khuyến nghị hai chiều (2D – two dimensions)⁸.

Việc đưa ra khuyến nghị trong hệ khuyến nghị 2D chỉ dựa trên người dùng, sản phẩm và đánh giá của người dùng lên sản phẩm và không quan tâm đến những thông tin ngữ cảnh. Hệ khuyến nghị dựa trên ngữ cảnh (CARS) mô hình hóa và đưa ra dự đoán của người dùng bằng cách kết hợp thông tin ngữ cảnh vào việc khuyến nghị. Khi đó hàm r được mô hình hóa thành hàm dự đoán cho mỗi cặp người dùng u , sản phẩm i , ngữ cảnh c .

$$r(u, i, c) : U \times I \times C \rightarrow R \quad (2.2)$$

Trong đó U là tập người dùng, I là tập sản phẩm, R là tập giá trị đánh giá và C là tập những thông tin ngữ cảnh. Như vậy, bài toán khuyến nghị dựa trên ngữ cảnh cũng được mô hình hóa thành hàm r , $\forall u \in U, \forall c \in C$ ta cần tìm $i \in I$ sao cho $r(u, i, c)$ đạt giá trị lớn nhất. Cần chú ý rằng, ký hiệu $r(u, i, c)$ và ký hiệu $r(u, i, c_1, c_2, \dots, c_k)$ hoàn toàn tương đương nhau, khi đó giá trị c chính là tổ hợp của các giá trị của từng ngữ cảnh $c = (c_1, c_2, \dots, c_k), c_i \in C_i, i = 1..k$.

2.4 Các phương pháp tiếp cận trong hệ khuyến nghị dựa trên ngữ cảnh

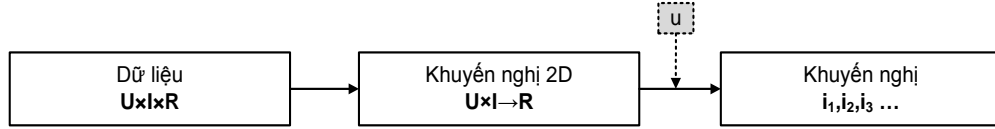
Kỹ thuật xử lý khuyến nghị dựa trên ngữ cảnh được chia thành hai nhóm chính: (1) Khuyến nghị dựa vào ngữ cảnh để truy vấn và tìm kiếm nội dung, (2) Khuyến nghị dựa vào ngữ cảnh để khai thác và đánh giá sở thích, nhu cầu người dùng.

⁸ Trong nội dung luận văn, các thuật ngữ khuyến nghị 2D, khuyến nghị truyền thống hoặc khuyến nghị 2D truyền thống đều có ý nghĩa tương đương.

Khuyến nghị dựa vào ngữ cảnh để truy vấn và tìm kiếm nội dung thường được dùng trong các hệ khuyến nghị trong lĩnh vực du lịch và trên nền tảng di động. Các hệ thống dùng kỹ thuật này thường lấy thông tin ngữ cảnh trực tiếp từ người dùng như tâm trạng, sở thích... hoặc thông qua môi trường ứng dụng lấy ngữ cảnh như định vị vị trí, địa điểm, thời gian, thời tiết... để truy vấn và tìm kiếm những đối tượng cần khuyến nghị và đưa ra gợi ý cho người dùng những đối tượng liên quan đến những ngữ cảnh đó. Hệ thống MyMap [20] đưa ra gợi ý thông tin về nhà hàng cho người dùng bằng cách truy vấn và tìm kiếm những nhà hàng thỏa mãn ngữ cảnh về thời tiết, mùa và khuôn viên của nhà hàng (trong nhà, ngoài trời).

Khuyến nghị dựa vào ngữ cảnh để khai thác và đánh giá sở thích, nhu cầu người dùng là xu hướng hiện nay trong hệ khuyến nghị dựa trên ngữ cảnh. Khác với cách hướng tiếp cận thứ nhất, hướng tiếp cận này đưa ra mô hình và nhận biết sở thích của người dùng thông qua những tương tác với hệ thống hoặc dựa trên những đánh giá của người dùng trên các sản phẩm trước đó. Để đưa ra dự đoán và khuyến nghị, hướng tiếp cận này áp dụng các kỹ thuật trong khuyến nghị (xem 2.1.3) như lọc cộng tác, lọc nội dung, dựa trên tri thức... hoặc có thể áp dụng những kỹ thuật phân tích dữ liệu trong khai phá dữ liệu (data mining) hoặc máy học (machine learning) như SVM, phân lớp Bayesian...

Ngoài ra, có thể kết hợp cả hai phương pháp trên tích hợp trong một hệ thống. Trong giới hạn của khóa luận, cách giải quyết bài toán hệ khuyến nghị kết hợp với yếu tố ngữ cảnh theo hướng giải quyết thứ hai được trình bày chi tiết. Như đã giới thiệu (xem 2.3), hệ khuyến nghị 2D xem ma trận $U \times I$ như là dữ liệu của hàm khuyến nghị $U \times I \rightarrow R$ và kết quả khuyến nghị cho người dùng là những sản phẩm i_1, i_2, i_3, \dots được xếp hạng dựa vào hàm khuyến nghị cho người dùng u . Quy trình khuyến nghị bao gồm 3 thành phần đó được biểu diễn một cách tổng quát như **Hình 2.3**.



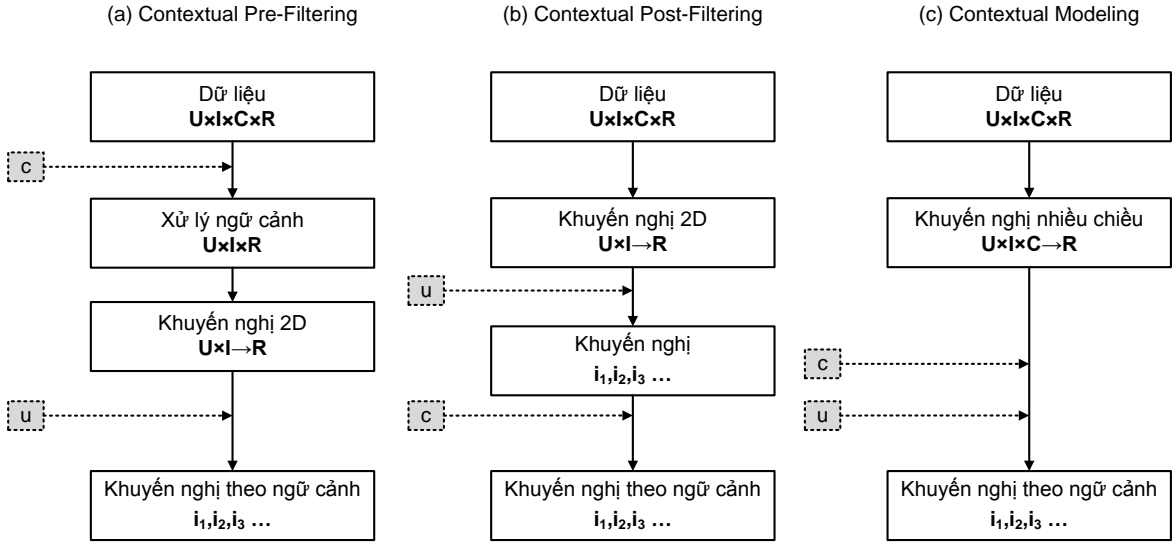
Hình 2.3: Quy trình khuyến nghị 2D

Hệ khuyến nghị dựa trên ngữ cảnh với ma trận $U \times I \times C$ là dữ liệu được mở rộng từ $U \times I$ kết hợp với ngữ cảnh, để đưa ra khuyến nghị cho người dùng u trong ngữ cảnh c , quy trình này được giải quyết với ba kỹ thuật chính như sau:

Contextual Pre-Filtering (Tiền xử lý ngữ cảnh): Thông tin ngữ cảnh được dùng để điều chỉnh dữ liệu phù hợp với ngữ cảnh đó sau đó áp dụng các kỹ thuật trong hệ khuyến nghị 2D truyền thống để đưa ra khuyến nghị (minh họa bởi **Hình 2.4a**).

Contextual Post-Filtering (Hậu xử lý ngữ cảnh): Thông tin ngữ cảnh được bỏ qua và khuyến nghị được đưa ra dựa vào các kỹ thuật trong hệ khuyến nghị 2D truyền thống. Sau đó, kết quả khuyến nghị sẽ được điều chỉnh cho phù hợp với từng người dùng dựa trên những thông tin ngữ cảnh (minh họa bởi **Hình 2.4b**).

Contextual Modeling (Mô hình hóa ngữ cảnh): Thông tin ngữ cảnh được sử dụng trực tiếp trong kỹ thuật khuyến nghị như là một phần của dự đoán đánh giá (minh họa bởi **Hình 2.4c**).

**Hình 2.4:** Các kỹ thuật khuyến nghị dựa trên ngữ cảnh**2.4.1 Tiền xử lý ngữ cảnh (Contextual Pre-Filtering)**

Contextual Pre-Filtering là kỹ thuật sử dụng yếu tố ngữ cảnh để chọn lọc và điều chỉnh dữ liệu cho hệ khuyến nghị 2D. Lợi ích của kỹ thuật này là có thể áp dụng được số lượng lớn những giải thuật khuyến nghị 2D truyền thống đã được phát triển trước đó, ngữ cảnh đóng vai trò như là một điều kiện để truy vấn (query) và chọn lọc (filter) các thông tin đánh giá.

Exact Pre-Filtering (EPF – lọc dữ liệu chính xác) là kỹ thuật truy vấn và lọc dữ liệu liên quan một ngữ cảnh cụ thể, lấy ví dụ như hệ khuyến nghị âm nhạc, một người dùng muốn nghe nhạc trong tâm trạng buồn thì hệ thống sẽ truy vấn và lọc tất cả các dữ liệu liên quan đến tâm trạng buồn. Adomavicius và cộng sự [3] đã đề xuất phương pháp giảm chiều dựa trên kỹ thuật EPF đưa ma trận $U \times I \times C$ thành ma trận $U \times I$ sau đó sử dụng khuyến nghị 2D để đưa ra khuyến nghị. Tuy nhiên, kỹ thuật này có một số hạn chế: thứ nhất, dữ liệu trong một ngữ cảnh cụ thể có thể chưa có tác động một cách tích cực đến đánh giá. Lấy ví dụ, người dùng sẽ có đánh giá tương tự khi xem phim vào thứ bảy và chủ nhật nhưng khác nhau vào các ngày trong tuần, khi đó lựa chọn ngữ cảnh thứ bảy chưa thực sự hiệu quả hơn là lựa chọn cả thứ bảy và chủ nhật. Thứ hai, việc lọc dữ liệu liên quan đến một ngữ cảnh cụ thể có thể gây ra hiện tượng dữ liệu thừa (xem 3.2.1) và gây khó khăn trong việc đưa ra dự đoán đánh giá một cách chính xác.

Context Generalization (tổng quát hóa ngữ cảnh) trình bày một kỹ thuật lọc ngữ cảnh tổng quát cho một ngữ cảnh cụ thể. Gọi $c' = \{c'_1, c'_2, \dots, c'_k\}$ là một tổng quát của ngữ cảnh $c = \{c_1, c_2, \dots, c_k\}$ nếu và chỉ nếu $c_i \rightarrow c'_i$ với $i=1..k$. Khi đó c' được dùng để truy vấn lọc dữ liệu thay cho c . Lấy ví dụ, cuối tuần là một tổng quát của thứ bảy và chủ nhật, $\{\text{thứ bảy, chủ nhật}\} \rightarrow \text{cuối tuần}$, Giờ AM là tổng quát của sáng và trưa, ký hiệu $\{\text{sáng, trưa}\} \rightarrow \text{AM}$. Cách biểu diễn thông tin ngữ cảnh theo cấu trúc phân cấp thể hiện trực quan về ngữ cảnh tổng quát (xem 2.2.3), khi đó giá trị mức cao hơn sẽ là tổng quát của mức thấp hơn. Hạn chế chính của kỹ thuật này là việc đưa ra ngữ cảnh tổng quát như thế nào là chính xác và hiệu quả. Cách thứ nhất để đưa ra ngữ cảnh tổng quát là xem xét ý nghĩa của ngữ cảnh. Cách thứ hai thực hiện một cách tự động dựa vào đánh giá hiệu quả của các ngữ cảnh tổng quát sau đó lựa chọn những ngữ cảnh tổng quát mang lại hiệu quả dự đoán cao để lọc dữ liệu, đây là vấn đề đã và đang được nghiên cứu hiện nay.

Như vậy, dữ liệu được sử dụng trong kỹ thuật này (Pre-Filtering) được chia thành nhiều phần ứng với từng thông tin ngữ cảnh, hay nói cách khác, dữ liệu đã được giảm chiều (reduction) theo từng ngữ cảnh cụ thể. Sau đó khuyến nghị 2D sẽ được áp dụng trên từng phần của dữ liệu (dữ liệu cục bộ) và đưa ra cho người dùng những sản phẩm khuyến nghị với ngữ cảnh tương ứng. Hiệu quả của việc sử dụng dữ liệu cục bộ và dữ liệu toàn cục bỏ qua thông tin ngữ cảnh cũng là một vấn đề của kỹ thuật này. Đây là lý do giải thích tại sao phương pháp giảm chiều hiệu quả ở một số trường hợp và ngược lại, Adomavicius và cộng sự đã đề xuất một kỹ thuật kết hợp giải quyết vấn đề này (xem 2.4.4).

2.4.2 Hậu xử lý ngữ cảnh (Contextual Post-Filtering)

Kỹ thuật hậu xử lý ngữ cảnh bỏ qua thông tin ngữ cảnh và áp dụng khuyến nghị 2D truyền thống để đưa ra danh sách các sản phẩm khuyến nghị. Sau đó, với từng người dùng cụ thể, danh sách này sẽ được “ngữ cảnh hóa”, tức điều chỉnh sao cho phù hợp với người dùng bằng hai cách: (i) **loại bỏ** (filtering) những sản phẩm liên quan trong ngữ cảnh cụ thể hoặc (ii) **điều chỉnh** (adjusting) danh sách sản phẩm bằng cách xếp hạng giảm dần danh sách dựa vào ngữ cảnh cụ thể.

Theo Adomavicius và cộng sự [3], kỹ thuật này được chia thành hai hướng tiếp cận là dựa trên kinh nghiệm (heuristic-based) và dựa trên mô hình (model-based). Hậu xử lý ngữ cảnh dựa trên kinh nghiệm (heuristic-based contextual post-filtering) tìm ra những đặc tính sản phẩm mà người dùng quan tâm trong một ngữ cảnh cụ thể, sau đó dùng những đặc tính này để điều chỉnh trên danh sách sản phẩm khuyến nghị bằng hai cách.

- **Loại bỏ** những sản phẩm không có những đặc tính liên quan đến ngữ cảnh.
- **Điều chỉnh** xếp hạng giảm dần sản phẩm dựa trên những đặc tính liên quan đến ngữ cảnh.

Ngược lại, hậu xử lý ngữ cảnh dựa trên mô hình (model-based contextual post-filtering) đưa ra mô hình dự đoán. Mô hình có thể đưa ra xác suất lựa chọn sản phẩm của người dùng trong ngữ cảnh cụ thể, sau đó xác suất này tiếp tục được dùng để điều chỉnh danh sách sản phẩm khuyến nghị bằng hai cách:

- **Loại bỏ** những sản phẩm có xác suất được người dùng lựa chọn trong ngữ cảnh cụ thể có giá trị nhỏ hơn một ngưỡng định sẵn.
- **Điều chỉnh** xếp hạng giảm dần các sản phẩm có xác suất được người dùng lựa chọn trong ngữ cảnh cụ thể có giá trị lớn hơn ngưỡng định sẵn.

Như vậy, việc “ngữ cảnh hóa” được áp dụng trên danh sách các sản phẩm khuyến nghị được đưa ra bởi khuyến nghị 2D vì thế kỹ thuật này thừa kế được những ưu điểm của phương pháp khuyến nghị này.

2.4.3 Mô hình hóa ngữ cảnh (Contextual Modeling)

Khác với hai kỹ thuật trước đó, tiền xử lý ngữ cảnh và hậu xử lý ngữ cảnh đều sử dụng khuyến nghị 2D truyền thống để giải quyết bài toán ngữ cảnh. Ở kỹ thuật này, thông tin ngữ cảnh được sử dụng trực tiếp như một phần của việc xử lý để đưa ra dự đoán đánh giá. Theo Adomavicius và cộng sự [3], một số giải thuật sử dụng kỹ thuật này được chia thành hai loại: Dựa trên kinh nghiệm và dựa vào mô hình.

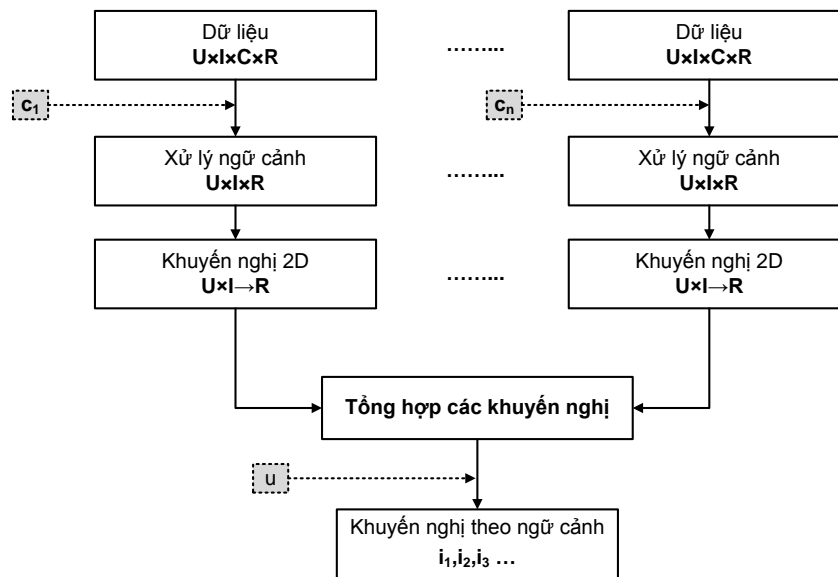
Giải thuật mô hình hóa ngữ cảnh **dựa trên kinh nghiệm** mở rộng các giải thuật sử dụng độ tương đồng để tính trọng số hoặc độ tương tự giữa những người dùng trong

một ngữ cảnh cụ thể [3] hoặc trong nhiều ngữ cảnh tương tự [6]. Trong khi giải thuật mô hình hóa ngữ cảnh **dựa trên mô hình** áp dụng các giải thuật trong khai phá dữ liệu, máy học hoặc xác suất thống kê để đưa ra mô hình dự đoán đánh giá. Một số mô hình áp dụng phương pháp phân tích ma trận và hồi quy [8], [21], [22] hoặc mô hình dựa trên phân phối xác suất kết hợp với mạng Bayesian [19].

2.4.4 Kết hợp các kỹ thuật

Một cách khác để xử lý bài toán khuyến nghị dựa trên ngữ cảnh là kết hợp các kỹ thuật để nâng cao hiệu quả khuyến nghị, mục đích của kỹ thuật này là kết hợp những ưu điểm của từng kỹ thuật khác nhau và dùng ưu điểm của kỹ thuật này để giải quyết nhược điểm của kỹ thuật kia.

Adomavicius và cộng sự [3] đưa ra một kỹ thuật kết hợp nhiều mô hình cùng loại để đưa ra khuyến nghị. Cụ thể, giải thuật thực hiện áp dụng khuyến nghị 2D trên từng phân khúc của ngữ cảnh – *segment* (mỗi phân khúc có thể là từng giá trị riêng rẽ, tổ hợp các ngữ cảnh hoặc một cấu trúc phân cấp) để tìm ra các phân khúc có dự đoán tốt hơn, sau đó để khuyến nghị sản phẩm trong ngữ cảnh cụ thể, ta chọn phân khúc tốt nhất có chứa ngữ cảnh cần khuyến nghị và áp dụng phương pháp khuyến nghị 2D để đưa ra kết quả cuối cùng (**Hình 2.5**).



Hình 2.5: Khuyến nghị dựa trên ngữ cảnh với kỹ thuật kết hợp

KỸ THUẬT LỌC CỘNG TÁC

Hệ khuyến nghị kết hợp với yếu tố ngữ cảnh được giải quyết thông qua ba kỹ thuật là tiền xử lý ngữ cảnh, hậu xử lý ngữ cảnh và mô hình hóa ngữ cảnh. Tuy nhiên, tiền xử lý ngữ cảnh và hậu xử lý ngữ cảnh đều sử dụng hệ khuyến nghị 2D như một giai đoạn chính trong quá trình giải quyết, nói cách khác khuyến nghị 2D là “nền” của hai kỹ thuật này, phần còn lại là xử lý ngữ cảnh phù hợp với khuyến nghị. Như đã trình bày trong chương 2, có rất nhiều kỹ thuật giải quyết vấn đề khuyến nghị 2D, trong đó giải quyết dựa trên lọc cộng tác (Collaborative Filtering – CF) được các nhà nghiên cứu đặc biệt quan tâm và đạt được những thành công nhất định, vì vậy nội dung chương này sẽ tập trung trình bày về lọc cộng tác được sử dụng trong khuyến nghị 2D. Bên cạnh đó, nội dung cũng sẽ trình bày kỹ thuật kết hợp ngữ cảnh trong lọc cộng tác (Context-Aware Collaborative Filtering – CACF) thường được sử dụng trong các hệ khuyến nghị dựa trên ngữ cảnh.

Phần đầu tiên sẽ trình bày về lọc cộng tác [4], kết hợp ngữ cảnh trong lọc cộng tác [1] và những vấn đề tồn tại trong lọc cộng tác. Tiếp theo sẽ trình bày phương pháp đánh giá hiệu quả của giải thuật lọc cộng tác thông qua một số chỉ số đánh giá độ chính xác dự đoán như MAE, RMSE và một số chỉ số đánh giá khuyến nghị như Precision, Recall. Phần còn lại trình bày chi tiết một số giải thuật tiêu biểu gồm khuyến nghị CF (User-Based, Baseline Predictor, Matrix Factorization) và CACF, trong đó giải thuật CACF gồm các giải thuật theo ba hướng tiếp cận với các giải thuật tương ứng:

- Tiền xử lý ngữ cảnh: Phân chia sản phẩm (Item Splitting) [7].
- Hậu xử lý ngữ cảnh: Weigh POF và Filter POF [5], [6].
- Mô hình hóa ngữ cảnh: Ngữ cảnh láng giềng (Contextual Neighbors) [6] và Context-Aware Matrix Factorization [8].

3.1 Kỹ thuật lọc cộng tác

3.1.1 Lọc cộng tác

Giả định cơ bản của lọc cộng tác (collaborative filtering) là nếu người dùng X và Y đánh giá N sản phẩm tương tự nhau hoặc có hành vi (xem, nghe, mua...) tương tự nhau thì sẽ có những đánh giá hoặc hành động lên những sản phẩm khác cũng tương tự nhau [4].

CF sử dụng những dữ liệu đánh giá của người dùng và sản phẩm ưa thích của người dùng đó để đưa ra dự đoán cho người dùng mới những mặt hàng mà họ có thể thích. Trong kỹ thuật CF , dữ liệu đánh giá là một ma trận gồm danh sách m người dùng $\{u_1, u_2, \dots, u_m\}$ và danh sách n sản phẩm $\{i_1, i_2, \dots, i_n\}$, giá trị đánh giá của người dùng thông qua thang giá trị 1-5 hoặc lượt mua, lượt bình chọn... phản ánh mức độ yêu thích của người dùng trên sản phẩm. Tuy nhiên, tồn tại một số giá trị đánh giá bị thiếu trong ma trận do người dùng không cung cấp sở thích của họ trên những mặt hàng đó.

Bảng 3.1: Ma trận đánh giá

	i_1	i_2	i_3	...	i_n
u_1	1	4	5		1
u_2	2	3	4		3
u_3	4	3	2		?
...					
u_m	2	1	?	4	4

Các hệ thống sử dụng lọc cộng tác thế hệ đầu tiên như GroupLens⁹ sử dụng dữ liệu để tính toán độ tương tự giữa người dùng hoặc sản phẩm để đưa ra dự đoán và khuyến nghị, kỹ thuật này gọi là kỹ thuật *dựa trên bộ nhớ* hay còn gọi là dựa trên kinh

⁹ <http://www.grouplens.org/>, truy cập lần cuối 22/02/2013

nghiệm (memory-based) và được ứng dụng trong một số hệ thống thương mại điện tử như Amazon¹⁰, Barnes&Noble¹¹. Ưu điểm của kỹ thuật này là dễ triển khai và có hiệu quả cao nhưng lại mắc phải một số hạn chế như: độ tương tự giữa các mặt hàng không đáng tin cậy khi dữ liệu thưa và số lượng các sản phẩm tương tự nhau hạn chế.

Để khắc phục những hạn chế của kỹ thuật dựa trên bộ nhớ, kỹ thuật lọc cộng tác *dựa trên mô hình* (model-based) sử dụng những dữ liệu đánh giá để đưa ra mô hình dự đoán bằng những mô hình khai thác dữ liệu hoặc máy học như Bayesian Belief Nets [23], [24], mô hình gom cụm [25], mô hình ngữ nghĩa tiềm ẩn (latent semantic) [26] và kỹ thuật giảm chiều ma trận (SVD, PCA) [27].

Kỹ thuật lọc nội dung (Content-Based Filtering – CBF) là một hướng tiếp cận khác đưa ra khuyến nghị bằng các phân tích thông tin nội dung và các quy luật. Điểm khác biệt của kỹ thuật này so với CF là việc sử dụng những đặc trưng của người dùng và sản phẩm để đưa ra khuyến nghị. Cả CF và CBF đều có những hạn chế riêng, trong khi các hệ CF không khai thác thông tin đặc trưng của người dùng hay mặt hàng thì các hệ CBF không sử dụng những thông tin về độ tương tự giữa người dùng và sản phẩm. Kỹ thuật lai CF (Hybrid CF) như thuật toán Content-Boosted CF [28] và Personality Diagnosis [29] kết hợp kỹ thuật CF và CBF với mục đích cải thiện những hạn chế và hiệu suất khuyến nghị của các cách tiếp cận trên.

3.1.2 Kết hợp ngữ cảnh trong lọc cộng tác

Lọc cộng tác với ngữ cảnh (Context-Aware Collaborative Filtering – CACF) là kỹ thuật được mở rộng từ kỹ thuật lọc cộng tác CF kết hợp với thông tin ngữ cảnh. Kỹ thuật này mang những tính chất, đặc điểm cũng như hạn chế tương tự kỹ thuật CF. Dữ liệu đánh giá trong CACF là một ma trận gồm danh sách m người dùng $\{u_1, u_2, \dots, u_m\}$ và n sản phẩm $\{i_1, i_2, \dots, i_n\}$ trong đó giá trị đánh giá sẽ được lưu trữ cùng với k thông tin ngữ cảnh $c(u, i)$, $c(u, i) = (c_1, c_2, \dots, c_k)$, $c_j \in C_j$, trong đó c_j là giá trị của ngữ cảnh C_j trong danh sách k ngữ cảnh $\{C_1, C_2, \dots, C_k\}$ [1], [30]. Với cách biểu diễn này, ma trận

¹⁰ <http://www.amazon.com/>, truy cập lần cuối 22/02/2013

¹¹ <http://www.barnesandnoble.com/>, truy cập lần cuối 22/02/2013

đánh giá trở thành ma trận nhiều chiều trong đó có hai chiều là người dùng, sản phẩm và các chiều còn lại là ngữ cảnh (xem 2.2.3).

Bảng 3.2: Ma trận đánh giá mở rộng với ngữ cảnh

	i_1	i_2	i_3	...	i_n
u_1					
u_2					
u_3					
...					
u_m					

	i_3				
	c_1	c_2	...	c_k	rating
u_3	x	y	...	z	5

Các kỹ thuật trong lọc cộng tác ngữ cảnh cũng bao gồm các kỹ thuật dựa trên bộ nhớ (memory-based CACF), dựa trên mô hình (model-based CACF) và lai (hybrid CACF) như kỹ thuật CF truyền thống.

Kỹ thuật dựa trên bộ nhớ tính độ tương tự (hoặc trọng số) của người dùng (hoặc sản phẩm) theo nhiều cách khác nhau. Độ tương tự có thể được tính bằng cách điều chỉnh dữ liệu thành ma trận 2 chiều trong từng giá trị ngữ cảnh (phương pháp giảm chiều) sau đó áp dụng các chỉ số tương tự trong CF như Cosin hoặc Pearson Correlation [3]. Một số hướng tiếp cận mở rộng phương pháp “láng giềng” kết hợp ngữ cảnh bằng cách tính trọng số giữa những cặp (*người dùng, sản phẩm, ngữ cảnh*) thông qua khoảng cách Euclidian hoặc Manhattan [9] trong khi một số giải thuật tính toán độ tương tự giữa những cặp (*người dùng, ngữ cảnh*) như thuật toán Contextual Neighbors [6]. Kỹ thuật dựa trên mô hình cũng áp dụng những mô hình trong máy học và khai phá dữ liệu như mở rộng mô hình phân tích ma trận (Matrix Factorization) [1],[31],[22], mô hình dựa trên phân phối xác suất kết hợp với mạng Bayesian [19] hoặc sử dụng mô hình hồi quy như thuật toán *Tensor Factorization* [21].

Kỹ thuật lai kết hợp giữa kỹ thuật CACF và phân tích ngữ nghĩa (semantic) sử dụng ontology cũng được nghiên cứu và phát triển và mang lại một số hiệu quả nhất định trong việc đưa ra khuyến nghị nhằm khắc phục những hạn chế của những giải thuật trước đó cũng được triển khai thông qua một số ứng dụng [32],[33],[34].

3.2 Vấn đề trong lọc cộng tác

3.2.1 Dữ liệu thưa

Một số hệ khuyến nghị trong thương mại trong lọc cộng tác CF phải sử dụng một số lượng lớn người dùng và sản phẩm để đưa ra đánh giá, vì vậy ma trận đánh giá trở nên rất thưa (sparsity) do thiếu nhiều đánh giá của người dùng lên sản phẩm. Hiện tượng dữ liệu thưa có thể xuất hiện trong nhiều trường hợp khác nhau. Trường hợp “cold start” xuất hiện khi một người dùng mới (sản phẩm mới) được đưa vào hệ thống, khi đó rất khó để tìm những người dùng (sản phẩm) tương tự để đưa ra khuyến nghị vì thiếu thông tin đánh giá, hiện tượng này còn gọi là hiện tượng “người dùng mới” (“sản phẩm mới”). Một trường hợp khác của dữ liệu thưa là hiện tượng “reduced coverage” khi số lượng đánh giá của người dùng quá nhỏ so với tổng số lượng lớn sản phẩm vì vậy giải thuật không thể đưa ra khuyến nghị cho những người dùng đó. Bên cạnh đó, việc những người dùng không cùng đánh giá trên những sản phẩm giống nhau (sản phẩm đồng đánh giá) sẽ gây ra hiện tượng không tìm được những người dùng tương tự trong những giải thuật liên quan đến độ tương tự.

Đối với lọc cộng tác ngữ cảnh CACF, hiện tượng dữ liệu thưa cũng xảy ra khi số lượng người dùng, sản phẩm và ngữ cảnh lớn, ma trận đánh giá cũng trở nên thưa hay việc xuất hiện những sản phẩm mới, người dùng mới cũng gây ra hiện tượng “cold start”. Các trường hợp như số lượng đánh giá quá ít trên sản phẩm trong ngữ cảnh cụ thể so với một số lượng lớn sản phẩm và ngữ cảnh cũng gây ra hiện tượng “reduced coverage” hay những người dùng không có những đánh giá trên cùng sản phẩm trên ngữ cảnh cụ thể cũng gây khó khăn trong những giải thuật liên quan đến độ tương tự.

3.2.2 Khả năng đáp ứng

Khi một lượng lớn số lượng người dùng và sản phẩm ngày càng tăng trưởng trong các hệ khuyến nghị thương mại, kỹ thuật CF gặp phải vấn đề về khả năng đáp ứng (scalability) tài nguyên hay vấn đề thời gian xử lý để đưa ra khuyến nghị. Mặt khác, việc đáp ứng yêu cầu tức thời cho người dùng đặt ra vấn đề về khả năng phản hồi của hệ khuyến nghị CF. Vấn đề này cũng xảy ra hoàn toàn tương tự trong các hệ CACF khi ngoài sản phẩm và người dùng, số lượng ngữ cảnh cũng tăng trưởng nhanh.

3.2.3 Đồng nghĩa

Đồng nghĩa (synonymn) là hiện tượng các sản phẩm tương tự nhau nhưng khác tên hoặc định danh. Hầu hết hệ khuyến nghị không có khả năng giải quyết hiện tượng này, đồng nghĩa với việc giảm hiệu suất của các hệ khuyến nghị (bao gồm CF và CACF).

3.2.4 Vấn đề khác

“Gray Sheep” là hiện tượng những người dùng mà sở thích (hành vi) của họ không hoàn toàn thuộc hoặc không thuộc về một nhóm người dùng nào, vì thế CF (cũng như CACF) không thể áp dụng lên những người dùng này. “Black Sheep” là hiện tượng mà sở thích (hành vi) của những người dùng “cô lập” khác biệt hoàn toàn so với người dùng khác và rất khó để đưa ra dự đoán. Đây là vấn đề chung của hệ khuyến nghị không chỉ riêng cho các hệ khuyến nghị CF hay CACF.

“Shilling attacks” là hiện tượng mà những người dùng cung cấp những đánh giá tích cực cho những sản phẩm có lợi cho họ và đưa ra những đánh giá bất lợi cho những sản phẩm khác của đối thủ. Ngoài ra còn một số vấn đề khác như vấn đề dữ liệu nhiễu, các vấn đề bảo mật, bảo vệ riêng tư cũng như sự phù hợp của các mặt hàng khuyến nghị.

3.3 Đánh giá kỹ thuật lọc cộng tác

Trong hầu hết các hệ khuyến nghị sử dụng kỹ thuật lọc cộng tác, các giải thuật thường phải đưa ra các dự đoán đánh giá cho các sản phẩm chưa được đánh giá bởi người dùng, sau đó dựa trên những dự đoán đánh giá này đưa ra danh sách các sản phẩm khuyến nghị cho người dùng. Nói cách khác, kỹ thuật lọc cộng tác thường phải thực hiện hai nhiệm vụ chính đó là *dự đoán đánh giá* và *khuyến nghị sản phẩm*. Vì vậy, để đánh giá mức độ hiệu quả của giải thuật lọc cộng tác, việc *đánh giá độ chính xác của dự đoán* và *đánh giá mức độ hiệu quả của khuyến nghị* thường được các nhà nghiên cứu quan tâm nhiều nhất.

3.3.1 Đánh giá dự đoán

Hệ khuyến nghị thường phải đưa ra tập hợp các giá trị dự đoán cho các sản phẩm mà người dùng chưa đánh giá, khi đó độ chênh lệch (hay còn gọi là độ lỗi) giữa giá trị dự đoán và giá trị đánh giá thực phản ánh mức độ *chính xác dự đoán* của giải thuật trong hệ khuyến nghị. Chỉ số đánh giá độ lỗi bao gồm chỉ số sai số trung bình *MAE* (Mean Absolute Error) và chỉ số căn bậc hai sai số bình phương trung bình *RMSE* (Root Mean Square Error) là các chỉ số được dùng phổ biến nhất để đánh giá độ lỗi của các giải thuật *CF* [4]. Gọi N là kích thước của tập dữ liệu kiểm thử, p_i và r_i lần lượt là giá trị đánh giá dự đoán từ giải thuật và giá trị đánh giá thực tế từ người dùng, khi đó *MAE* và *RMSE* được tính theo công thức.

$$MAE = \frac{1}{N} \sum_{i=1}^N |p_i - r_i| \quad (3.1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - r_i)^2} \quad (3.2)$$

Giá trị *MAE* chỉ ra sai lệch trung bình giữa giá trị dự đoán trong khi giá trị *RMSE* xác định rõ độ lớn giữa các độ sai lệch đó bằng cách sử dụng bình phương. Khi đó, giá trị *RMSE* càng lớn càng cho thấy mức độ chênh lệch càng lớn giữa giá trị dự đoán và giá trị thực. Hai chỉ số này thường được dùng bổ trợ cho nhau trong việc đánh giá độ chính xác dự đoán của khuyến nghị.

3.3.2 Đánh giá khuyến nghị

Hệ khuyến nghị sau khi thực hiện việc đưa ra dự đoán đánh giá, danh sách sản phẩm phù hợp sẽ được khuyến nghị cho người dùng thông qua hai hướng tiếp cận chính như sau:

Top-K Recommendation: Đưa ra danh sách K sản phẩm có xếp hạng dự đoán cao nhất cho người dùng. Trong trường hợp này, chỉ một phần nhỏ sản phẩm được khuyến nghị vì các yếu tố khách quan như thời gian để xem tất cả sản phẩm hạn chế hay việc tránh tiếp cận với nguồn thông tin quá tải [35]. Giá trị của K phụ thuộc vào từng ứng dụng và số lượng sản phẩm trong từng ứng dụng riêng biệt.

Find All Good Items: Trong hướng tiếp cận này, tất cả các sản phẩm được dự đoán như “sản phẩm tốt” – tức là sản phẩm có giá trị dự đoán lớn hơn một **ngưỡng định sẵn** và ngưỡng này đặc trưng theo từng ứng dụng và thang đánh giá sẽ được khuyến nghị. Đối với các ứng dụng thang đánh giá 1-5, thông thường sản phẩm tốt sẽ có giá trị dự đoán lớn hơn 3... Hướng tiếp cận này đưa ra tất cả các sản phẩm tiềm năng (có thể phù hợp) cho người dùng để tránh bỏ sót các sản phẩm phù hợp [35].

Khi đó, để đánh giá mức độ hiệu quả của giải thuật khuyến nghị các sản phẩm có phù hợp với người dùng hay không, chỉ số *Precision* và *Recall* được dùng để phản ánh điều này. Các chỉ số này được dùng phổ biến trong lĩnh vực tìm kiếm thông tin (information retrieval) và áp dụng rộng rãi trong các hệ khuyến nghị. Trong hệ khuyến nghị, các sản phẩm sẽ thuộc một trong hai loại, phù hợp (*relevant*) và không phù hợp (*irrelevant*) với nhu cầu, sở thích người dùng, vì vậy thang đánh giá phải được chuyển về dạng nhị phân tương ứng với phù hợp và không phù hợp. Đối với dữ liệu có thang đánh giá 1-5, thông thường những sản phẩm có đánh giá 4-5 là sản phẩm phù hợp và 1-3 là sản phẩm không phù hợp [36]. Bên cạnh đó, tập sản phẩm được chọn (*selected*) khuyến nghị cho người dùng và tập không khuyến nghị cho người dùng (*not selected*). Các chỉ số *Precision* và *Recall* được tính dựa vào bảng sau:

Bảng 3.3: Các tham số liên quan đến chỉ số Precision và Recall

	Khuyến nghị (Selected - s)	Không Khuyến nghị (Not Selected - n)	Tổng
Phù hợp (Relevant - r)	N_{rs}	N_{rn}	N_r
Không phù hợp (Irrelevant - i)	N_{is}	N_{in}	N_i
Tổng	N_s	N_n	N

Trong đó

- N_s là tổng số sản phẩm được khuyến nghị trên toàn bộ dữ liệu.

- N_n là tổng số sản phẩm không được khuyến nghị trên toàn bộ dữ liệu.
- N_r là tổng số sản phẩm phù hợp trên toàn bộ dữ liệu.
- N_i là tổng số sản phẩm không phù hợp trên toàn bộ dữ liệu.
- N là tổng số sản phẩm, $N=N_r+N_i=N_n+N_s$.
- N_{rs} là tổng số sản phẩm khuyến nghị phù hợp được khuyến nghị.
- N_{is} là tổng số sản phẩm khuyến nghị không phù hợp được khuyến nghị.
- N_{rn} là tổng số sản phẩm khuyến nghị phù hợp không được khuyến nghị.
- N_{in} là tổng số sản phẩm khuyến nghị không phù hợp không được khuyến nghị.

Chỉ số *Precision* thể hiện tỉ số sản phẩm khuyến nghị phù hợp trên tổng số các sản phẩm được khuyến nghị. Chỉ số *Recall* thể hiện tỉ số sản phẩm khuyến nghị phù hợp trên tổng số sản phẩm phù hợp.

$$Precision = \frac{N_{rs}}{N_s} \quad (3.3)$$

$$Recall = \frac{N_{rs}}{N_r} \quad (3.4)$$

Khó khăn chính của việc tính chỉ số *Precision* và *Recall* là việc nhận biết những sản phẩm khuyến nghị đưa ra phù hợp với sở thích người dùng. Việc nhận biết này có thể thực hiện bằng cách thu thập phản hồi trực tiếp từ người dùng, tuy nhiên cách thức này không thực sự khả quan khi lượng người dùng quá lớn (ví dụ trong bộ dữ liệu Movielens¹² số lượng người dùng hơn 6000) và hầu hết việc đánh giá những chỉ số này không có sự tham gia của người dùng thực [9]. Bên cạnh đó, một số dữ liệu có số lượng sản phẩm lớn đến hàng ngàn thậm chí hàng trăm ngàn mặt hàng và người dùng chỉ thực hiện đánh giá trên một số lượng sản phẩm giới hạn. Khi đó, hệ khuyến nghị có thể đưa ra một số sản phẩm mà người dùng đó chưa thực hiện đánh giá hay đơn thuần là người dùng hoàn toàn không biết về sản phẩm đó do số lượng sản phẩm quá lớn, do đó việc tính chỉ số *Precision* và *Recall* cũng gặp trở ngại lớn vì hầu hết mọi dữ liệu thu thập đánh giá từ người dùng đều không được thực hiện bằng một hệ khuyến nghị thực thụ.

¹² <http://www.grouplens.org/node/73>, truy cập lần cuối 22/02/2013

Có rất nhiều phương pháp tính toán hai chỉ số trên được đưa ra. Dựa trên những tổng hợp từ Herlocker và cộng sự [36], Ricci và cộng sự [9], chỉ số *Precision* và *Recall* được tính như sau:

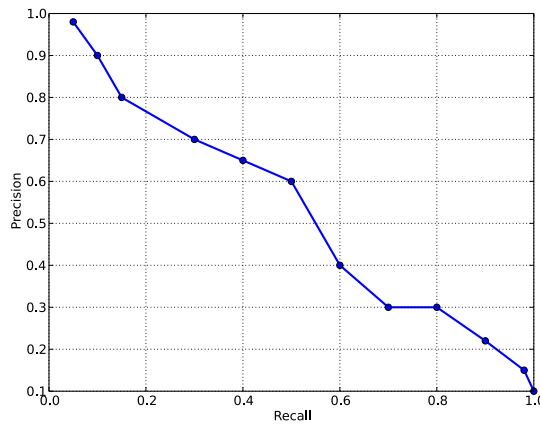
- Thực hiện phân chia tập dữ liệu thành hai tập riêng biệt: Huấn luyện (Train) và Kiểm thử (Test) với tỉ lệ xác định, thông thường là 80% – 20% và giao giữa hai tập $Train \cap Test = \emptyset$.
- Trong tập kiểm thử Test, đối với khuyến nghị 2D sẽ bao gồm các người dùng u và các sản phẩm theo từng người dùng đó. Khi đó gọi tập U sẽ là tập các người dùng trong Test, với mỗi người dùng $u \in U$, tập các sản phẩm tương ứng với người dùng đó trong tập kiểm thử là $T(u)$. Đối với khuyến nghị ngữ cảnh, tập U sẽ là tập chứa các cặp người dùng u với ngữ cảnh cụ thể và $T(u)$ là các sản phẩm tương ứng với người dùng và ngữ cảnh đó. Chú ý rằng $T(u)$ chỉ chứa những sản phẩm phù hợp, tức là sản phẩm có giá trị đánh giá đạt một ngưỡng giá trị định trước.
- Sử dụng tập huấn luyện để đưa ra dự đoán cho các sản phẩm, sau đó thực hiện khuyến nghị K sản phẩm (theo hướng tiếp cận *Top K Recommendation* hoặc *Find All Good Items*) cho từng người dùng u thuộc tập U . Trong đó K thường có giá trị từ 1-10, giá trị của K còn phụ thuộc vào số lượng sản phẩm của dữ liệu. Gọi $L(u)$ là tập các sản phẩm được khuyến nghị cho người dùng u .
- Thực hiện so sánh các sản phẩm có trong $L(u)$ và $T(u)$, gọi $H(u)$ là tập chứa những sản phẩm xuất hiện trong cả hai tập $L(u)$ và $T(u)$, $L(u) \cap T(u) = H(u)$.
- Khi đó giá trị *Precision* và *Recall* được tính bằng công thức:

$$Precision = \frac{1}{|U|} \sum_{u \in U} \frac{|L(u) \cap T(u)|}{|L(u)|} = \frac{1}{|U|} \sum_{u \in U} \frac{|H(u)|}{|L(u)|} \quad (3.5)$$

$$Recall = \frac{1}{|U|} \sum_{u \in U} \frac{|L(u) \cap T(u)|}{|T(u)|} = \frac{1}{|U|} \sum_{u \in U} \frac{|H(u)|}{|T(u)|} \quad (3.6)$$

Trong đó $|H(u)|, |T(u)|, |L(u)|, |U|$ lần lượt là kích thước của tập $H(u)$, $T(u)$, $L(u)$ và U . Các chỉ số này phụ thuộc vào giá trị của N , vì thế các chỉ số này còn được gọi là **Precision-at-N** [9] và **Recall-at-N** [37].

Các chỉ số Precision và Recall còn được biểu diễn bằng đồ thị PR (Precision-Recall curve). Để xây dựng đồ thị PR, thay đổi số lượng khuyến nghị cho người dùng bằng các tăng số lượng K sản phẩm với *Top-K Recommendation* [9], [35] hoặc thay đổi ngưỡng giá trị quy định sản phẩm phù hợp/không phù hợp với *Find All Good Items* [1], khi đó ta sẽ có giá trị Precision thay đổi tại những giá trị Recall khác nhau tương ứng. Khi tất cả các sản phẩm trong hệ thống được khuyến nghị cho người dùng, ta có giá trị Recall đạt giá trị cao nhất. **Hình 3.1** biểu diễn một trường hợp của đồ thị PR, đồ thị cho thấy một giải thuật có thể có các giá trị Precision tại các mức giá trị Recall khác nhau.



Hình 3.1: Đồ thị Precision – Recall (PR)

3.4 Giải thuật lọc cộng tác trong khuyến nghị truyền thống

3.4.1 User-Based

Giải thuật User-Based (UB) biểu diễn người dùng bằng vector m chiều (tương ứng với m sản phẩm) và giá trị mỗi chiều trong vector là đánh giá của người dùng lên sản phẩm. Sau đó, giải thuật tìm n láng giềng (người dùng) tương tự với người dùng hiện tại (active user) sử dụng độ đo Pearson Correlation hoặc Cosin Similitary trong không gian vector [4], giá trị của độ đo (còn gọi độ tương tự) có giá trị trong khoảng $[0,1]$ và giá trị này càng lớn phản ánh mức độ tương tự giữa các người dùng càng cao. Giá trị của n phụ thuộc vào số lượng người dùng trên từng ứng dụng cụ thể và việc sử dụng độ đo nào phù hợp tùy thuộc vào ứng dụng cũng như thang đánh giá 1-5 hay 0-1.

Độ đo Pearson Correlation giữa hai người dùng u, v được tính bằng công thức:

$$sim(u, v) = \frac{\sum_{i \in I} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{vi} - \bar{r}_v)^2}} \quad (3.7)$$

Độ đo Cosin giữa hai người dùng u, v tính bằng công thức:

$$sim(u, v) = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \bullet \vec{v}}{\|\vec{u}\| \times \|\vec{v}\|} = \frac{\sum_{i \in I} r_{ui} \times r_{vi}}{\sqrt{\sum_{i \in I} r_{ui}^2} \times \sqrt{\sum_{i \in I} r_{vi}^2}} \quad (3.8)$$

Trong đó $i \in I$ là tập những sản phẩm đã được đánh giá bởi hai người dùng u, v , \bar{r}_u và \bar{r}_v là đánh giá trung bình của người dùng u, v trong I . Giá trị $sim(u, v)$ càng tiến về 1 phản ánh mức độ tương tự nhiều nhất giữa hai người dùng u, v .

Sau khi tìm được n láng giềng, giải thuật sẽ tìm K sản phẩm được đánh giá nhiều nhất từ n láng giềng mà người dùng hiện tại chưa đánh giá và sẽ khuyến nghị cho người dùng những sản phẩm có giá trị dự đoán đánh giá cao. Giá trị dự đoán đánh giá cho người dùng u , sản phẩm i được tính bằng các công thức sau:

$$\begin{aligned} (a) \quad \hat{r}_{u,i} &= \frac{1}{N} \sum_{u' \in \hat{U}} r_{u'i} \\ (b) \quad \hat{r}_{u,i} &= k \sum_{u' \in \hat{U}} sim(u, u') \times r_{u'i} \\ (c) \quad \hat{r}_{u,i} &= \bar{r}_u + k \sum_{u' \in \hat{U}} sim(u, u') \times (r_{u'i} - \bar{r}_{u'}) \end{aligned} \quad (3.9)$$

Trong đó: \hat{U} là tập người dùng láng giềng, k là tham số chuẩn hóa $k = \frac{1}{\sum_{u \in \hat{U}} |sim(u, u')|}$, $sim(u, u')$ là độ tương tự *cosin* hoặc *Pearson*, \bar{r}_u và $\bar{r}_{u'}$ là đánh giá trung bình của người dùng u, u' .

Ưu điểm của giải thuật này là dễ cài đặt, dễ dàng mở rộng người dùng và sản phẩm mới và hoạt động tốt với các sản phẩm được đồng đánh giá bởi đa số người dùng. Tuy nhiên, nhược điểm của giải thuật là phụ thuộc vào đánh giá của người dùng tương tự, giải thuật không thể khuyến nghị cho người dùng mới cũng như không đưa

ra được dự đoán đánh giá cho các sản phẩm mới, khó khăn trong việc đáp ứng với số lượng dữ liệu lớn và hiệu suất giảm khi dữ liệu quá thưa [4].

3.4.2 Baseline Predictor

Kỹ thuật CF tập trung vào phân tích các tương tác (đánh giá) giữa người dùng và sản phẩm để đưa ra giá trị đánh giá dự đoán khác nhau, tuy nhiên trong một số trường hợp, đánh giá dự đoán đó còn chịu tác động từ đặc tính của người dùng lẫn sản phẩm. Một ví dụ điển hình của vấn đề này đó là tính thành kiến hay thiên vị (bias) của người dùng như người dùng khó tính có xu hướng đưa ra đánh giá thấp hơn những người dùng dễ tính, hay người dùng không thích nghe nhạc của một ca sĩ nào đó thì luôn cho những đánh giá thấp trong khi bài hát đó có thể rất hay từ đa số người dùng.

Baseline Predictor [9] đưa ra cách giải quyết tính thành kiến (bias) và đưa ra dự đoán đánh giá \hat{r}_{ui} của người dùng u cho sản phẩm i theo công thức sau:

$$\hat{r}_{ui} = \mu + b_u + b_i \quad (3.10)$$

Trong đó μ là giá trị đánh giá trung bình của người dùng trên toàn bộ sản phẩm, b_u là phương sai của các đánh giá của người dùng u và b_i là phương sai của đánh giá trên sản phẩm i . Nhiệm vụ của *Baseline Predictor* là tìm ra các tham số μ, b_u, b_i thỏa mãn giá trị nhỏ nhất của hàm:

$$\min \sum_{(u,i) \in K} \left[(\hat{r}_{ui} - \mu - b_u - b_i)^2 + \lambda (b_u^2 + b_i^2) \right] \quad (3.11)$$

Trong đó λ là tham số chuẩn hóa nhằm hạn chế hiện tượng “quá vừa dữ liệu” – *Overfitting*, là hiện tượng mô hình cho kết quả tốt (chính xác) trên dữ liệu huấn luyện nhưng cho kết quả không tốt (không chính xác) trên dữ liệu kiểm thử. Giá trị của μ có thể được tính dựa vào thống kê, b_u, b_i được tính bằng cách áp dụng phương pháp giảm dần (Stochastic Gradient Descent) [38]. Ý tưởng chung của phương pháp giảm dần là xác định vị trí một điểm bất kỳ trên hàm mục tiêu, sau đó tìm các vị trí lân cận thông qua tham số bước nhảy đến khi hàm mục tiêu đạt giá trị thấp hơn. Đến một ngưỡng nào đó không thể tìm điểm lân cận mới, khi đó xem hàm mục tiêu đạt giá trị nhỏ nhất.

3.4.3 Matrix Factorization

Mô hình phân tích ma trận (Matrix factorization) ánh xạ người dùng và sản phẩm vào một không gian nhân tố tiềm ẩn f chiều, trong đó mỗi sản phẩm được biểu diễn bằng vector $q_i \in \mathbb{R}^f$ và mỗi người dùng được biểu diễn bằng vector $p_u \in \mathbb{R}^f$, mỗi giá trị trong vector q_i thể hiện đặc tính của sản phẩm so với f nhân tố và giá trị trong vector p_u thể hiện sự yêu thích của người dùng lên f nhân tố đó. Giá trị của tích $q_i^\top p_u$ thể hiện sự tương tác của người dùng lên sản phẩm và có giá trị tương đương với giá trị đánh giá r_{ui} [39]. Mô hình này liên quan đến kỹ thuật SVD (Singular Vector Decomposition) được dùng trong việc nhận biết những nhân tố ngữ nghĩa tìm ẩn trong lĩnh vực tìm kiếm thông tin (information retrieval) [40], vì thế mô hình này còn được gọi là mô hình SVD [9].

Ngoài ra, giá trị đánh giá còn chịu ảnh hưởng bởi tính thành kiến – thiên vị (bias), khi đó giá trị đánh giá dự đoán \hat{r}_{ui} mở rộng từ phương pháp Baseline Predictor (xem 3.4.2) được tính theo công thức sau:

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^\top p_u \quad (3.12)$$

Trong đó μ , b_u , b_i lần lượt là giá trị đánh giá trung bình của người dùng trên toàn bộ sản phẩm, giá trị thiên vị của người dùng u và giá trị thiên vị của sản phẩm i . Việc tìm tham số b_u, b_i thỏa mãn giá trị cực tiểu của hàm:

$$\min \sum_{(u,i) \in K} \left[\left(\hat{r}_{ui} - \mu - b_u - b_i - q_i^\top p_u \right)^2 + \lambda \left(b_u^2 + b_i^2 + \|p_u\|^2 + \|q_i\|^2 \right) \right] \quad (3.13)$$

Phương pháp giảm đồi (xem 3.4.2) được áp dụng để tìm các tham số thỏa mãn cực tiểu của hàm trên bằng cách thực hiện lặp nhiều lần để tính các tham số theo công thức:

$$\begin{aligned} b_u &\leftarrow b_u + \gamma(e_{ui} - \lambda b_u) \\ b_i &\leftarrow b_i + \gamma(e_{ui} - \lambda b_i) \\ q_i &\leftarrow q_i + \gamma(e_{ui} \cdot p_u - \lambda q_i) \\ p_u &\leftarrow p_u + \gamma(e_{ui} \cdot q_i - \lambda p_u) \end{aligned} \quad (3.14)$$

Trong đó λ là tham số chuẩn hóa nhằm hạn chế hiện tượng *overfitting* (xem 3.4.2) và γ là tham số bước nhảy, $e_{ui} = r_{ui} - \hat{r}_{ui}$, r_{ui} là giá trị đánh giá thực. Giá trị γ , λ được điều chỉnh thông qua phương pháp *m-Fold Cross Validation* (xem 4.2), thông thường giá trị của chúng rất nhỏ. $\gamma=0.005$, $\lambda=0.02$ được dùng khi thử nghiệm với dữ liệu Netflix¹³. Quá trình trên thực hiện lặp đi lặp lại đến khi tìm được giá trị cực tiểu hay thực chất là các tham số b_u, b_i, q_i, p_u không thay đổi sau nhiều lần lặp. Thông thường, thuật toán dừng lại đến khi không tìm được giá trị tối ưu của lần lặp $k+1$ so với k . Mặt khác, việc lựa chọn số lần lặp quá lớn có thể gây lãng phí tài nguyên của ứng dụng trong khi giá trị cực tiểu của hàm mục tiêu không thay đổi nhiều. Thực hiện số lần lặp tối thiểu là 10 và tối đa là 1000, kết quả dự đoán của giải thuật thay đổi không đáng kể với 2 lần lặp là 100 và 1000 với bộ dữ liệu Food, kết quả được minh họa bởi **Bảng 3.4**.

Bảng 3.4: Kết quả giải thuật MF với số lần lặp thay đổi

Số lần lặp	MAE	RMSE	Thời gian huấn luyện (s)
10	0.901	1.117	0.405
100	0.888	1.114	1
1000	0.881	1.116	10

Bên cạnh đó, việc lựa chọn số lượng nhân tố f cũng tác động đến việc huấn luyện mô hình, thực nghiệm số lượng nhân tố $f=10$ trên dữ liệu Food cho kết quả tốt nhất, tăng số lượng nhân tố f không làm thay đổi kết quả hoặc làm thay đổi không đáng kể, ngược lại tiêu tốn nhiều thời gian huấn luyện.

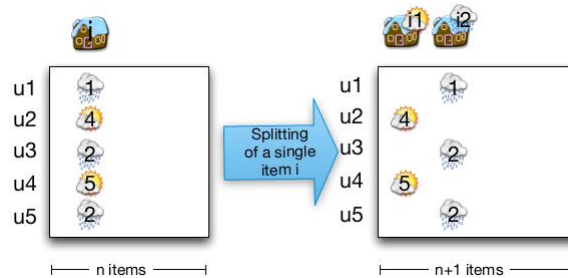
Ưu điểm của giải thuật là có thể giải quyết vấn đề dữ liệu thưa, làm việc hiệu quả với dữ liệu lớn và khả năng dự đoán đánh giá tương đối tốt. Tuy nhiên, chi phí cài đặt phức tạp, tiêu hao tài nguyên và không gian lưu trữ. Mặt khác do quá trình huấn luyện được thực hiện ngoại tuyến (off-line) định kỳ nên khi số lượng người dùng, sản phẩm thay đổi liên tục thì giải thuật phải đưa ra mô hình mới, vì vậy giải thuật chỉ có khả năng giải quyết một phần nhỏ vấn đề người dùng mới, sản phẩm mới, tức là giải thuật chưa thể đưa ra dự đoán một cách chính xác cho những đối tượng này dựa trên mô hình từ dữ liệu huấn luyện đó chưa được cập nhật.

¹³ <http://www.netflixprize.com/>, truy cập lần cuối 22/02/2013

3.5 Giải thuật lọc cộng tác trong hệ khuyến nghị dựa trên ngữ cảnh

3.5.1 Item Splitting

Baltrunas và Ricci [7] đã giới thiệu một giải thuật là Item Splitting (phân chia sản phẩm) dựa trên phương pháp giảm chiều [3], giải thuật này được mở rộng dựa trên CF và theo hướng tiếp cận tiền xử lý ngữ cảnh (Contextual Pre-Filtering). Ý tưởng của giải thuật này là tìm ra những sản phẩm có sự khác biệt trong việc đánh giá sản phẩm của người dùng trên ngữ cảnh thông qua giá trị $r(u, i, c)$, $c = \{c_1, c_2, \dots, c_k\}$. Với mỗi sản phẩm, dữ liệu đánh giá liên quan được chia thành hai tập ứng với hai sản phẩm mới, việc phân chia được thực hiện bằng cách tìm một giá trị cụ thể của một ngữ cảnh $c_j = v$, tập đánh giá thứ nhất sẽ là các giá trị đánh giá khi $c_j = v$ và tập thứ hai là tất cả các đánh giá còn lại với $c_j \neq v$. Nếu giá trị đánh giá giữa hai tập đó mang ý nghĩa thống kê, sản phẩm đó sẽ được chia thành hai sản phẩm mới tương ứng với $c_j = v$ và $c_j \neq v$. Khi đó, trong quá trình kiểm thử, giá trị của dự đoán đánh giá cho một sản phẩm sẽ được tính bằng giá trị dự đoán của một trong những sản phẩm mới. Lấy ví dụ, sản phẩm i được chia thành hai sản phẩm là i_1 và i_2 chứa những đánh giá tương ứng với $c_j = v$ và $c_j \neq v$, để đưa ra giá trị dự đoán cho sản phẩm i trong ngữ cảnh $c_j = x$, giá trị dự đoán sẽ được tính trên i_1 nếu $v = x$ và i_2 nếu $v \neq x$. **Hình 3.2** minh họa việc phân chia một sản phẩm i thành hai sản phẩm i_1 và i_2 tương ứng với ngữ cảnh thời tiết mang giá trị nắng và mưa, sau khi phân chia, số sản phẩm tăng lên tuy nhiên số lượng đánh giá vẫn không thay đổi trong trường hợp này.



Hình 3.2: Phân chia sản phẩm

Giá trị đánh giá giữa hai tập mang ý nghĩa thống kê có thể được tính bằng tiêu chuẩn t , gọi s là một phân chia của sản phẩm i bằng cách chọn một giá trị của một ngữ cảnh và chia dữ liệu đánh giá thành hai tập, $s \in S$, S là tập các phân chia có thể của

sản phẩm i . Ta chọn phân chia $s \in S$ cho sản phẩm i nếu $t(i,s)$ đạt giá trị lớn nhất so với các phân chia trong s . Giá trị $t(i,s)$ có thể là t_{mean} , t_{prop} , t_{IG} , t_{chi} , t_{random} .

- $t_{mean}(i,s)$ được định nghĩa bằng kiểm định t -test hai mẫu để kiểm định sự khác nhau giữa các giá trị trung bình trong hai mẫu khi phân chia s được sử dụng, giá trị t_{mean} càng lớn phản ánh sự khác nhau của các giá trị đánh giá trung bình sản phẩm trong hai mẫu càng cao.
- $t_{prop}(i,s)$ sử dụng kiểm định z -test để kiểm định sự khác nhau giữa hai tỉ lệ đánh giá cao và thấp trong hai sản phẩm mới i_1, i_2 được phân chia bởi s . Đánh giá cao có thể là (3-5) và thấp là (1-2) với thang đánh giá là 1-5.
- $t_{IG}(i,s)$ sử dụng độ đo IG (information gain) [41] phản ánh mức độ hiệu quả phân chia của s .
- $t_{chi}(i,s)$ sử dụng kiểm định *chi bình phương* để kiểm định sự khác nhau giữa hai tỉ lệ đánh giá cao và thấp trong hai sản phẩm mới i_1, i_2 phân chia bởi s . $t_{chi}(i,s)$ tương tự với $t_{prop}(i,s)$ nhưng khác nhau về phương pháp kiểm định.
- $t_{random}(i,s)$ trả về một giá trị ngẫu nhiên cho mỗi phân chia s .

Theo đánh giá thực tế từ Baltrunas và Ricci [7], $t_{mean}(i,s)$ cho kết quả tốt nhất so với các loại còn lại. Sau khi thực hiện phân chia sản phẩm, ngữ cảnh đã được biểu diễn một cách ngầm định thông qua các sản phẩm mới, khuyến nghị ngữ cảnh trở về khuyến nghị 2D truyền thống và áp dụng các giải thuật trong lọc cộng tác CF như User-Based (xem 3.4.1) hoặc Matrix Factorization (xem 3.4.3),... để đưa ra dự đoán và khuyến nghị. Theo Baltrunas và Ricci, việc phân chia sản phẩm thành hai nhóm có thể giải quyết được trường hợp dữ liệu thưa, tuy nhiên nếu ngữ cảnh chưa thực sự tác động một cách hiệu quả đến đánh giá, trong một số trường hợp, giải thuật có thể làm tăng độ lỗi khi dự đoán đánh giá [7]. Bên cạnh đó, giải thuật sử dụng khuyến nghị 2D như một giai đoạn trong quá trình đưa ra dự đoán đánh giá, vì thế giải thuật sẽ mang những ưu điểm và nhược điểm của chính giải thuật 2D này.

3.5.2 Weigh POF và Filter POF

Panniello và Gorgoglione [6] đã đề xuất hai kỹ thuật gọi là Weigh POF và Filter POF theo kỹ thuật hậu xử lý ngữ cảnh (Contextual Post-Filtering). Cả hai kỹ thuật đều phân

tích dữ liệu người dùng và ngữ cảnh cụ thể để tính xác suất lựa chọn sản phẩm của người dùng trong ngữ cảnh đó, sau đó xác suất này dùng để điều chỉnh các sản phẩm được đưa ra bằng khuyến nghị 2D.

Theo Panniello và Gorgoglione, giá trị $P_c(u,i)$ là xác suất lựa chọn sản phẩm i bởi người dùng u trong ngữ cảnh c được tính bằng số lượng những láng giềng (người dùng tương tự) của người dùng u lựa chọn sản phẩm i trong ngữ cảnh c chia cho tổng số lượng người dùng tương tự với u [6]. Trong đó, những người dùng tương tự với người dùng u được xác định thông qua những đánh giá trong ngữ cảnh c . Số lượng người dùng tương tự được lựa chọn thông qua thực nghiệm cụ thể trên từng ứng dụng. Sự khác nhau giữa Weigh POF và Filter POF là việc “ngữ cảnh hóa” các danh sách sản phẩm từ khuyến nghị 2D.

Weigh POF *xếp hạng* danh sách sản phẩm dựa trên giá trị dự đoán đánh giá được tính bằng công thức:

$$\hat{r}_{uic} = \hat{r}_{ui} \times P_c(u,i) \quad (3.15)$$

Trong khi đó Filter POF *loại bỏ* những sản phẩm trong danh sách khuyến nghị dựa trên một ngưỡng quy định trước bằng công thức:

$$\hat{r}_{uic} = \begin{cases} \hat{r}_{ui} & \text{khi } P_c(u,i) \geq P^* \\ 0 & \text{khi } P_c(u,i) < P^* \end{cases} \quad (3.16)$$

Trong đó \hat{r}_{uic} là giá trị dự đoán đánh giá, \hat{r}_{ui} là giá trị dự đoán đánh giá từ khuyến nghị 2D, P^* là một ngưỡng giá trị định trước, theo quá trình thực nghiệm bởi Panniello và Gorgoglione [5], [6], giá trị của P^* thay đổi từ 0.1 đến 0.9 và giá trị 0.1 cho kết quả tốt nhất. Việc loại bỏ các sản phẩm với Filter POF với mục đích “nếu chỉ ít người dùng chọn sản phẩm trong một ngữ cảnh thì tốt hơn là không khuyến nghị những sản phẩm đó cho người dùng mặc dù khuyến nghị 2D cho kết quả dự đoán cao và vẫn khuyến nghị”.

Việc tính $P_c(u,i)$ dựa trên tần số lựa chọn sản phẩm i trong ngữ cảnh c dựa trên số người dùng tương tự với u được Panniello và Gorgoglione tính toán với dữ liệu dựa trên giao dịch (*transaction-based*) và giá trị đánh giá chính là tần số mua sản phẩm,

khi đó nếu tần số lớn hơn 1 được xem như người dùng yêu thích sản phẩm đó [5], [6]. Tuy nhiên, đối với thang giá trị 1-5, người dùng lựa chọn (yêu thích) sản phẩm được quy định dựa trên một ngưỡng định sẵn của giá trị đánh giá.

Bảng 3.5: Kết quả so sánh giải thuật MF, Weigh MF và Filter MF trên dữ liệu Comoda

	MF	Weigh MF	Filter MF
Precision	0.00354	–	–
Recall	0.03337	–	–

Giải thuật Filter POF và Weigh POF chưa thực sự hiệu quả trong trường hợp dữ liệu thưa và việc chọn lọc hay xếp hạng các sản phẩm dựa vào xác suất $P_c(u,i)$ chưa hiệu quả khi số lượng sản phẩm đồng đánh giá trên ngữ cảnh hạn chế. Thực nghiệm trên dữ liệu Comoda với giải thuật MF, Weigh MF và Filter MF, kết quả **Bảng 3.5** cho thấy giải thuật Weigh MF và Filter MF không thể đưa ra các sản phẩm khuyến nghị so với MF vì xác suất $P_c(u,i)$ phụ thuộc vào sản phẩm đồng đánh giá trong ngữ cảnh c , khi giá trị này tiến về 0, điều đó đồng nghĩa với việc sản phẩm i không thể được khuyến nghị cho người dùng u . Mặt khác, giải thuật Post-Filtering điều chỉnh danh sách khuyến nghị sao cho phù hợp, vì thế khả năng dự đoán giá trị đánh giá xem như tương đương với giải thuật 2D. Điều đó đồng nghĩa với việc các chỉ số MAE và RMSE *hoàn toàn tương đương* với giải thuật 2D. Mặt khác, do giải thuật dựa trên khuyến nghị 2D nên Filter POF và Weigh POF sẽ có những ưu điểm lẫn nhược điểm của chính giải thuật khuyến nghị 2D đó.

3.5.3 Contextual Neighbors

Một giải thuật được gọi là Ngữ cảnh láng giềng (Contextual Neighbors – CN) được đề xuất bởi Panniello và Gorgoglione [6] theo hướng tiếp cận mô hình hóa ngữ cảnh (Contextual Modeling) như sau:

Đầu tiên, với mỗi người dùng u và ngữ cảnh c , một hồ sơ người dùng (user's profile) trong ngữ cảnh c được định nghĩa là một hồ sơ ngữ cảnh (contextual profile) $Prof(u,c)$, lưu ý rằng nếu c có nhiều giá trị thì ta phải định nghĩa $Prof(u,c)$ cho từng giá trị của c , $Prof(u,c) = (r_{u1c}, r_{u2c}, \dots, r_{uic}, \dots, r_{ukc})$, $i=1..k$ trong đó r_{uic} được biểu diễn nhiều

cách khác nhau như là giá trị đánh giá của người dùng hoặc là tần số lựa chọn sản phẩm i trong ngữ cảnh c [5], [6].

Sau đó những thông tin *profile* sẽ được sử dụng để tính độ tương tự giữa những người dùng để tìm ra “láng giềng” thông qua độ tương tự cosin d giữa các *profile*.

$$d[Prof(u,c), Prof(u',c')] = \frac{Prof(u,c) \bullet Prof(u',c')}{\|Prof(u,c)\| \times \|Prof(u',c')\|} = \frac{\sum_{i \in S} r_{uic} \times r_{u'ic'}}{\sqrt{\sum_{i \in S} r_{uic}^2} \times \sqrt{\sum_{i \in S} r_{u'ic'}^2}} \quad (3.17)$$

Trong đó r_{uic} và $r_{u'ic'}$ là giá trị đánh giá của người dùng u , u' trên sản phẩm i trong ngữ cảnh c, c' tương ứng. $S = \{i \in I \mid r_{uic} \neq \emptyset \wedge r_{u'ic'} \neq \emptyset\}$ là tập hợp những sản phẩm đã được đánh giá bởi người dùng u, u' trong ngữ cảnh c, c' .

Sau đó, lựa chọn N láng giềng cho cặp (u, c) thông qua các cặp (u', c') có giá trị $d[Prof(u, c), Prof(u', c')]$ lớn theo những ràng buộc sau:

- **Mdl₁**: Không có ràng buộc trên các cặp (u', c') , chọn N cặp (u', c') tương tự với (u, c) (có giá trị lớn giảm dần).
- **Mdl₂**: Chọn số cặp (u', c') theo từng giá trị cụ thể của c' đều nhau. Ví dụ c' có 2 giá trị và $N=10$, ta chọn 5 cặp ứng với giá trị thứ nhất và 5 cặp ứng với giá trị thứ hai.
- **Mdl₃**: Chọn N cặp (u', c') tương tự với cặp (u, c) trong cùng một mức phân cấp ngữ cảnh.
- **Mdl₄**: Chọn số cặp (u', c') theo từng giá trị cụ thể của c' đều nhau và cùng một mức phân cấp ngữ cảnh. Ví dụ c' cùng cấp ngữ cảnh với c và có 2 giá trị và $N=10$, ta chọn 5 cặp ứng với giá trị thứ nhất và 5 cặp ứng với giá trị thứ hai.

Giá trị dự đoán đánh giá cho người dùng u , sản phẩm i , ngữ cảnh c được tính theo công thức:

$$\hat{r}_{uic} = k \sum_{Prof(u', c') \in \hat{P}} d[Prof(u, c), Prof(u', c')] \times r_{u'ic'} \quad (3.18)$$

Trong đó \hat{P} là tập những “láng giềng” của cặp (u, c) và k là tham số chuẩn hóa.

$$k = \frac{1}{\sum_{Prof(u',c') \in \hat{P}} |d[Prof(u,c), Prof(u',c')]|} \quad (3.19)$$

Theo quá trình đánh giá được thực hiện bởi Panniello và Gorgoglione[5], [6], các ràng buộc trên làm thay đổi kết quả khuyến nghị không đáng kể.

Ưu điểm của giải thuật là dễ cài đặt, thêm người dùng mới và sản phẩm mới tương đối linh hoạt, khả năng dự đoán đánh giá và khuyến nghị sản phẩm đa dạng vì sử dụng những sản phẩm đã được đánh giá từ những ngữ cảnh tương đồng. Vì thế giải thuật giải quyết một phần vấn đề **dữ liệu thừa theo từng ngữ cảnh** (khi so sánh với EPF) vì sử dụng đánh giá trong ngữ cảnh tương đồng. Cụ thể, thực nghiệm với dữ liệu Comoda là dữ liệu có số tổ hợp ngữ cảnh lớn và độ thừa của dữ liệu lên đến 99.99%. Giải thuật EPF-UB không thể cho kết quả khuyến nghị vì dữ liệu được phân hoạch theo từng ngữ cảnh *quá thừa* trong khi giải thuật CN có khả năng đưa ra khuyến nghị nhiều hơn và tốt hơn giải thuật UB. Kết quả được trình bày ở **Bảng 3.6**.

Bảng 3.6: Kết quả so sánh giải thuật UB, EPF-UB, CN trên dữ liệu Comoda

	UB	EPF-UB	CN
Precision	0.0035	—	0.0922
Recall	0.0314	—	0.0494

Tuy nhiên, giải thuật chưa thực sự hiệu quả với trường hợp trường hợp người dùng mới và sản phẩm mới chưa có đánh giá, việc xây dựng *profile* hoàn toàn không thực hiện được với trường hợp này. Mặt khác, việc tính toán độ tương tự giữa các *profile* còn chịu tác động của những sản phẩm đồng đánh giá trong ngữ cảnh cụ thể nên vấn đề về sản phẩm đồng đánh giá trong các ngữ cảnh cũng là một thách thức của giải thuật này.

3.5.4 Context-Aware Matrix Factorization

Koren và cộng sự [39] đã đề xuất mô hình sử dụng kỹ thuật phân tích ma trận (matrix factorization) kết hợp với phương pháp Baseline Predictor để đưa ra dự đoán đánh giá trong khuyến nghị 2D dựa trên *CF* (xem 3.4.3). Trong mô hình trên, các tham số baseline (các tham số b_w , b_i thể hiện sự thành kiến – thiên vị) dùng cho người dùng và

sản phẩm thể hiện sự chênh lệch (deviation) đánh giá của người dùng và sản phẩm so với giá trị trung bình đánh giá toàn cục. Các tham số trên làm tăng tính phức tạp của mô hình nhưng cũng làm tăng tính chính xác của mô hình [1]. Bên cạnh đó, Koren và cộng sự [39] đã trình bày sự tác động của thời gian vào đánh giá thông qua tham số baseline. Từ cơ sở đó, Baltrunas [1] đã mở rộng phương pháp phân tích ma trận bằng cách kết hợp nhiều thông tin ngữ cảnh vào mô hình theo hướng tiếp cận mô hình hóa ngữ cảnh (contextual modeling) với kỹ thuật CACF. Với mỗi ngữ cảnh, một tham số sẽ được thêm vào mô hình biểu hiện sự chênh lệch (deviation) đánh giá bởi tác động của ngữ cảnh. Mô hình phân rã ma trận biểu diễn người dùng và sản phẩm là những vector p_u và q_i trong không gian f nhân tố tiềm ẩn (xem 3.4.3), gọi $r_{uic_1...c_k}$ là giá trị đánh giá của người dùng u lên sản phẩm i trong ngữ cảnh $c=(u,i,c_1,...,c_k)$, những bộ $(u,i,c_1,...,c_k)$ tồn tại giá trị đánh giá được lưu trong tập $R = \{(u,i,c_1,...,c_k) | r_{uic_1...c_k} \neq \emptyset\}$. Khi đó, giá trị đánh giá dự đoán được tính bằng công thức sau:

$$\hat{r}_{uic_1...c_k} = q_i^\top p_u + \bar{i} + b_u + \sum_{j=1}^k b_{ic_j} \quad (3.20)$$

Trong đó p_u , q_i là vector biểu diễn cho người dùng và sản phẩm, \bar{i} là giá trị trung bình của đánh giá của sản phẩm i trong R , b_u là tham số baseline cho người dùng u và b_{ic_j} là tham số baseline cho ngữ cảnh c_j và sản phẩm i . Các tham số p_u , q_i , b_u , b_{ic_j} có giá trị thỏa mãn giá trị cực tiểu của hàm:

$$\min \sum_{r \in R} \left[\left(r_{uic_1...c_k} - q_i^\top p_u - \bar{i} - b_u - \sum_{j=1}^k b_{ic_j} \right)^2 + \lambda \left(b_u^2 + \|p_u\|^2 + \|q_i\|^2 + \sum_{j=1}^k b_{ic_j}^2 \right) \right] \quad (3.21)$$

Áp dụng phương pháp giảm đồi (xem 3.4.2) để thực hiện tìm cực tiểu của hàm trên, trong đó, các tham số p_u , q_i , b_u , b_{ic_j} được tính như sau:

$$\begin{aligned} b_u &\leftarrow b_u + \gamma_{b_u} (err - \lambda b_u) \\ b_{ic_j} &\leftarrow b_{ic_j} + \gamma_{b_{ic_j}} (err - \lambda b_{ic_j}), j = 1..k \\ q_i &\leftarrow q_i + \gamma_{q_i} (err \cdot p_u - \lambda q_i) \\ p_u &\leftarrow p_u + \gamma_{p_u} (err \cdot q_i - \lambda p_u) \end{aligned} \quad (3.22)$$

Trong đó λ là tham số chuẩn hóa nhằm hạn chế hiện tượng *overfitting* (xem 3.4.2) và γ_* là tham số bước nhảy, $err = r_{uic_1...c_k} - \hat{r}_{uic_1...c_k}$, $r_{uic_1...c_k}$ là giá trị đánh giá thực. Trong mô hình trên, một số tham số sẽ thay đổi giá trị thường xuyên hơn bởi vì nhiều người dùng có số lượng đánh giá nhiều hoặc một số ngữ cảnh sẽ xuất hiện nhiều hơn, điều này có thể làm mất cân bằng của mô hình bởi vì mô hình chỉ tập trung huấn luyện cho những dữ liệu người dùng có đánh giá nhiều và những ngữ cảnh xuất hiện nhiều. Để giải quyết vấn đề đó, giá trị của tham số bước nhảy γ_* (ví dụ như γ_{b_u}) được tính bằng giá trị γ chia cho logarith của số lượng những đánh giá làm thay đổi tham số đó. Lấy ví dụ, một người dùng có 20 đánh giá, khi đó trong một lần huấn luyện, giá trị của b_u sẽ thay đổi 20 lần, khi đó $\gamma_{b_u} = \gamma / \log(20)$, giá trị γ rất nhỏ, thông thường là 0.001 và được điều chỉnh thông qua phương pháp m-Fold Cross Validation.

Một điểm hạn chế của giải thuật chính là việc giải quyết vấn đề sản phẩm mới. Công thức (3.21) sử dụng \bar{i} là giá trị trung bình của sản phẩm i để xây dựng mô hình. Khi đó, đối với sản phẩm mới i , mô hình huấn luyện chỉ dựa vào các vector người dùng và các tham số baseline, khi đó giá trị dự đoán cho sản phẩm mới rất nhỏ, điều này sẽ gây ra chênh lệch lớn so với giá trị đánh giá thực. Để hạn chế vấn đề này, ta có thể thay \bar{i} bằng \bar{r} là giá trị trung bình của toàn bộ sản phẩm trong R đối với trường hợp sản phẩm mới i . Mặt khác, giải thuật được mở rộng từ phương pháp Matrix Factorization nên giải thuật đều có những ưu điểm và nhược điểm hoàn toàn tương tự với phương pháp này (xem 3.4.3).

3.6 Kết luận chung

Các giải thuật được trình bày đều thuộc về hai nhóm giải thuật chính đó là nhóm giải thuật dựa trên bộ nhớ và nhóm giải thuật dựa trên mô hình. Dựa trên những tổng hợp bởi Su và cộng sự [4] kết hợp với những kiểm thử thực nghiệm ở mục 4.5, những ưu nhược điểm chung của hai nhóm giải thuật được tóm tắt dưới bảng sau:

Bảng 3.7: Phân tích ưu nhược điểm của kỹ thuật dựa trên bộ nhớ và dựa trên mô hình

Kỹ thuật	Giải thuật	Ưu điểm	Nhược điểm
Dựa trên bộ nhớ	User-Based (UB).	Dễ cài đặt.	Phụ thuộc vào sản phẩm đồng đánh giá.
	EPF-UB.	Dễ dàng thêm dữ liệu người dùng, sản phẩm mới.	Vấn đề dữ liệu thừa, sản phẩm mới, người dùng mới.
	ISUB.		
	Weigh UB, Filter UB.	Hiệu quả tăng khi số lượng sản phẩm đồng đánh giá lớn.	Không thể khuyến nghị sản phẩm mới chưa được đánh giá.
	Contextual Neighbors.		Khả năng đáp ứng giảm khi xử lý với dữ liệu lớn.
Dựa trên mô hình	Matrix Factorization.	Giải quyết được vấn đề dữ liệu thưa.	Chi phí xây dựng mô hình lớn.
	EPF-MF.	Cải thiện hiệu suất dự đoán tốt.	Có thể gây mất thông tin khi áp dụng các phương pháp giảm chiều ma trận.
	ISMF.		
	Weigh MF, Filter MF.	Có thể dự đoán và khuyến nghị những sản phẩm mới, người dùng mới.	Khó khăn khi thêm dữ liệu sản phẩm, người dùng vì mô hình phải được xây dựng lại.
	Context-Aware Matrix Factorization.	Khả năng đáp ứng nhanh.	

Bên cạnh đó, các giải thuật CACF trong lọc cộng tác đều tồn tại những ưu nhược điểm riêng được tổng hợp trình bày ở **Bảng 3.8**.

Bảng 3.8: Phân tích ưu nhược điểm của một số giải thuật CACF tiêu biểu

Giải thuật	Ưu điểm	Nhược điểm
EPF	Điều chỉnh ngữ cảnh để áp dụng, kế thừa các kỹ thuật khuyến nghị 2D.	Số lượng mô hình tăng tương ứng với số tổ hợp ngữ cảnh khi áp dụng giải thuật 2D (dựa trên mô hình). Điều chỉnh ngữ cảnh có thể gây ra hiện tượng dữ liệu thừa trên phạm vi trên từng ngữ cảnh. Lựa chọn ngữ cảnh không phù hợp để điều giảm chiều dữ liệu có thể làm giảm hiệu suất [3]. Nhược điểm của khuyến nghị 2D sử dụng.
Item Splitting	Thừa kế các giải thuật khuyến nghị 2D. Giải quyết được vấn đề dữ liệu thừa [7].	Giải thuật có thể làm tăng độ lỗi nếu ngữ cảnh chưa thực sự tác động đến đánh giá [1]. Nhược điểm của khuyến nghị 2D sử dụng.
Weigh PoF, Filter PoF	Ưu điểm của khuyến nghị 2D sử dụng. Dễ dàng cài đặt.	Nhược điểm của khuyến nghị 2D sử dụng. Vấn đề dữ liệu thừa, người dùng mới, sản phẩm mới. Phụ thuộc vào sản phẩm đồng đánh giá trong ngữ cảnh. Không thể khuyến nghị những sản phẩm mới hoặc những sản phẩm chưa được đánh giá trong ngữ cảnh cần khuyến nghị.
Contextual Neighbors	Dễ cài đặt, thêm mới dữ liệu. Khuyến nghị đa dạng vì sử dụng những sản phẩm đã được đánh giá từ những ngữ cảnh tương đồng. Giải quyết một phần vấn đề dữ liệu thừa trong từng ngữ cảnh so với phương pháp giảm chiều.	Không thể khuyến những sản phẩm mới, người dùng mới chưa có đánh giá.

Giải thuật	Ưu điểm	Nhược điểm
CAMF	<p>Giải quyết được vấn đề dữ liệu thưa trong từng ngữ cảnh so với phương pháp giảm chiều.</p> <p>Hiệu suất dự đoán tốt.</p>	<p>Có thể làm giảm khả năng dự đoán với những sản phẩm mới.</p> <p>Thời gian huấn luyện và chi phí xây dựng mô hình tốn kém.</p>

ĐÁNH GIÁ

Để kiểm chứng mức độ hiệu quả của giải thuật cũng như so sánh các giải thuật với nhau, các yêu cầu về thước đo đánh giá (metric evaluation), dữ liệu huấn luyện và dữ liệu kiểm thử phải tuân theo “chuẩn chung”. Trên thực tế, mức độ hiệu quả của các giải thuật phải được kiểm nghiệm trên các bộ dữ liệu (dataset) khác nhau và trên nhiều thước đo đánh giá (metric evaluation). Đặc biệt, đối với các giải thuật trong lọc cộng tác, các chỉ số đánh giá thường được dùng là MAE, RMSE, Precision và Recall để kiểm tra tính chính xác và mức độ hiệu quả của các giải thuật. Một vấn đề khác là việc phân chia dữ liệu kiểm thử cũng đòi hỏi yêu cầu cao không kém so với việc đưa ra giải thuật. Nội dung chương sẽ trình bày một cách chi tiết phương pháp phân chia dữ liệu đánh giá thành các dữ liệu huấn luyện và kiểm thử, kết quả kiểm thử thực nghiệm để đưa ra cái nhìn tổng quát mức độ hiệu quả của các giải thuật đã trình bày trong chương 3. Kết quả thực nghiệm sẽ được tiến hành trên nhiều bộ dữ liệu khác nhau là Food, HGift, Movielens Semi-Synthetic, Comoda và HMusic với số lượng người dùng, sản phẩm và ngữ cảnh khác nhau để kiểm chứng tính hiệu quả của giải thuật.

4.1 Những vấn đề trong kiểm thử đánh giá hệ khuyến nghị

Thứ nhất, các giải thuật khác nhau có thể hiệu quả cũng như kém hiệu quả khi thực hiện kiểm thử trên nhiều tập dữ liệu khác nhau. Hiện nay, hầu hết các dữ liệu kiểm thử có số lượng người dùng lớn hơn rất nhiều so với sản phẩm, khi đó giải thuật có thể kém hiệu quả khi thực hiện trên tập dữ liệu có số lượng người dùng nhỏ hơn số lượng sản phẩm. Tương tự, một số tính chất khác như mật độ phân bố đánh giá, thang giá trị

đánh giá hoặc các thuộc tính khác của dữ liệu cũng tác động không nhỏ đến việc đánh giá hiệu quả của một giải thuật [36].

Thứ hai, mục đích của việc đánh giá cũng ảnh hưởng đến việc xác định hiệu quả của giải thuật. Một số giải thuật tập trung vào việc đưa ra dự đoán đánh giá gần đúng nhất trong khi một số giải thuật tập trung vào việc nâng cao mức độ phù hợp khuyến nghị với sở thích của người dùng [36]. Hay nói cách khác là việc lựa chọn những tiêu chí đánh giá nào và kết hợp chúng ra sao cho phù hợp khi so sánh các thuật toán khác nhau cũng hết sức quan trọng.

Thứ ba, việc tìm kiếm, lựa chọn tập dữ liệu kiểm thử tin cậy còn gặp rất nhiều khó khăn. Các dữ liệu được phổ biến cho mục đích học tập, nghiên cứu có thể không đáp ứng đủ yêu cầu so với các dữ liệu nội bộ được thu thập từ các trung tâm, viện nghiên cứu. Đây cũng là vấn đề khó khăn khi thực hiện đánh giá nhiều giải thuật mới.

Thứ tư, việc phân chia dữ liệu kiểm thử cũng ảnh hưởng không nhỏ đến hiệu quả của giải thuật như tỉ lệ giữa tập huấn luyện và kiểm thử hay phương pháp phân chia dữ liệu. Trong hầu hết quá trình kiểm thử, dữ liệu được phân chia ngẫu nhiên và thường không theo quy tắc, khi đó những cách phân chia khác nhau cũng ảnh hưởng đến việc đánh giá mức độ hiệu quả của giải thuật một cách khách quan.

4.2 Các phương pháp phân chia dữ liệu

Để đánh giá chất lượng của một giải thuật khuyến nghị, dữ liệu cần được chia thành dữ liệu huấn luyện và dữ liệu kiểm thử. Dữ liệu thường được phân chia¹⁴ theo ba phương pháp sau đây: (1) holdout (2) leave-one-out và (3) m-fold-cross-validation [44].

Holdout là phương pháp phân chia dữ liệu thành hai tập huấn luyện và kiểm thử, các tập phải có tỉ lệ khác nhau. Thông thường, việc phân chia thực hiện bằng cách chọn ngẫu nhiên những đánh giá từ tất cả người dùng tạo thành tập huấn luyện, phần còn lại là tập kiểm thử. Phương pháp này còn gọi là leave-k-out. Trong một số đánh giá, dữ liệu thường được chia với tỉ lệ 80% – 20%. Trong một số thử nghiệm, một vài tỉ lệ giữa tập huấn luyện và kiểm thử được thay đổi và với mỗi tỉ lệ mới, thực nghiệm đánh

¹⁴ Thực chất phương pháp phân chia dữ liệu kiểm thử và phương pháp thực hiện kiểm thử hoàn toàn tương tự nhau.

giá được thực hiện 10 lần, sau đó kết quả được lấy trung bình kết quả của tất cả các lần kiểm thử. Một số đánh giá khác sử dụng 10% người dùng và 5 đánh giá liên quan để kiểm thử.

Leave-one-out là trường hợp đặc biệt của phương pháp leave-k-out với $k=1$. Với mỗi người dùng, một đánh giá được chọn để kiểm thử và phần còn lại dùng để huấn luyện. Phương pháp này phù hợp để đánh giá những giải thuật cho những người dùng mới và số lượng đánh giá hạn chế. Mặt hạn chế của phương pháp này là hiện tượng *overfitting* (xem 3.4.2) có thể xảy ra.

M-fold cross-validation là một biến thể của phương pháp Holdout. Phương pháp này chia dữ liệu thành m phần, mỗi phần chính là tập kiểm thử và những phần còn lại là dữ liệu huấn luyện, sau đó kết quả đánh giá của giải thuật được tính trung bình. Thông thường, $m=10$ được chọn để phân chia dữ liệu và thực hiện kiểm thử, tuy nhiên giá trị m còn phụ thuộc vào đặc trưng của từng bộ dữ liệu khác nhau. Phương pháp này phù hợp cho kiểm thử những dữ liệu có người dùng mới tham gia vào hệ thống.

4.3 Dữ liệu đánh giá

Để đánh giá mức độ hiệu quả của các giải thuật, khóa luận sử dụng nhiều bộ dữ liệu đánh giá khác nhau để đảm bảo tính khách quan của kết quả đánh giá. Các bộ dữ liệu đánh giá là Food, HGift và HMusic, Comoda, Movielens Semi-Synthetic.

Food là bộ dữ liệu được thu thập thực tế và được cung cấp bởi Hideki Asoh và cộng sự [19]. Bộ dữ liệu bao gồm những đánh giá thức ăn từ 212 người dùng trên 20 thực đơn thức ăn. Dữ liệu được thu thập thông qua khảo sát đánh giá thực đơn với thang đánh giá 1-5 dựa vào mức độ đói (đói, bình thường, no) thực tế và một số mức độ đói là trường hợp giả định. Với dữ liệu này, mức độ đói với miền giá trị là {no, đói, bình thường} và trường hợp {thực tế, giả định} là hai thông tin ngữ cảnh. Mặt khác, dữ liệu tồn tại một số đánh giá trùng lặp vì vậy giá trị của đánh giá sẽ được lấy ngẫu nhiên là giá trị cuối cùng, bên cạnh đó, chọn lọc các đánh giá của người dùng có số lượng đánh giá tối thiểu là 5 đánh giá. Sau khi được xử lý, số lượng đánh giá của bộ dữ liệu này là 5300 đánh giá. Khi áp dụng kiểm định t -test trên từng cặp giá trị trong từng ngữ cảnh đều thỏa mãn mức tin cậy (confidence level) tối thiểu 95%.

HGift là bộ dữ liệu thu thập thực tế thông qua ứng dụng Facebook. Bộ dữ liệu bao gồm những đánh giá về quà tặng từ người dùng trên 20 sản phẩm là các món quà với thang đánh giá 1-5. Với dữ liệu này, thông tin ngữ cảnh bao gồm {đối tượng, giới tính, độ tuổi, lý do} là thông tin của người được tặng quà. Bộ dữ liệu sau khi loại bỏ các người dùng đưa ra số lượng đánh giá nhỏ hơn 5, dữ liệu bao gồm 44 người dùng, 20 sản phẩm quà tặng và 1477 đánh giá. Khi áp dụng kiểm định t -test trong từng cặp giá trị trong từng ngữ cảnh, tất cả ngữ cảnh đều tác động đến đánh giá từ người dùng.

HMUSIC là bộ dữ liệu thu thập thực tế từ ứng dụng website nghe nhạc trực tuyến hmusic.us bao gồm những đánh giá của người dùng trên bài hát. Dữ liệu đánh giá được thu thập với yếu tố ngữ cảnh là tâm trạng với miền giá trị là {vui, buồn, đang yêu, thất tình, bình thường, thư giãn, tức giận} và thời gian với miền giá trị là {sáng, trưa, chiều, tối, khuya}. Bộ dữ liệu bao gồm 1480 đánh giá được chọn lọc những đánh giá từ 134 bài hát và 44 người dùng có lượng đánh giá tối thiểu là 5 đánh giá.

Comoda là bộ dữ liệu đánh giá phim được thu thập thực tế và được cung cấp bởi Kosir và cộng sự [45]. Bộ dữ liệu bao gồm 82 người dùng, 1225 bộ phim và 12 thông tin ngữ cảnh với những miền giá trị khác nhau. Số lượng đánh giá được cung cấp trong bộ dữ liệu là 2248. Giá trị ngữ cảnh của một số dòng dữ liệu bị khuyết và được thay thế bằng những giá trị ngữ cảnh xuất hiện nhiều nhất. Bộ dữ liệu được chọn lọc những người dùng có số lượng đánh giá tối thiểu là 3. Điểm đặc biệt của bộ dữ liệu này là mật độ đánh giá rất thưa 99.99% và số sản phẩm đồng đánh giá trên cùng tổ hợp ngữ cảnh rất hạn chế với số tổ hợp ngữ cảnh được sử dụng trong dữ liệu là 1849 bộ.

Movielens Semi-Synthetic là bộ dữ liệu được mở rộng từ bộ dữ liệu Movielens-100K trong đó thông tin giới tính (gender) và độ tuổi của người xem phim được xem như thông tin ngữ cảnh tác động đến đánh giá, độ tuổi được chia thành ba nhóm: nhỏ hơn 18, từ 18 đến 50, lớn hơn 50 ứng với giá trị của ngữ cảnh độ tuổi. Vì đặc thù của bộ dữ liệu không bao gồm yếu tố ngữ cảnh, nên dữ liệu được định nghĩa lại với một yếu tố ngữ cảnh c mới thay thế cho giới tính với $c=\{0,1\}$ và được phát sinh ngẫu nhiên, khi đó thông tin c sẽ tác động đến đánh giá người dùng [1], [7]. Dữ liệu được mở rộng bằng cách chọn α % sản phẩm ngẫu nhiên từ tập sản phẩm và chọn ngẫu nhiên β % các đánh giá từ sản phẩm vừa chọn. Trong tất cả các đánh giá được chọn,

phát sinh ngẫu nhiên c và tăng (giảm) giá trị đánh giá hiện tại một đơn vị nếu c bằng 1(0) nếu giá trị đánh giá khác 5(1). Ví dụ với $\alpha=0.9$, $\beta=0.9$, 90% đánh giá từ 90% sản phẩm sẽ được điều chỉnh, giá trị α, β tăng (giảm) càng cho thấy tác động của c đến đánh giá người dùng tăng (giảm). Trong thử nghiệm này, $\alpha=0.9$, $\beta=0.9$ được chọn để điều chỉnh bộ dữ liệu Movielens 100K¹⁵. Đối với bộ dữ liệu mới này, ta **giả định** rằng 2 thông tin ngữ cảnh có **tác động** đến đánh giá của người dùng.

Bảng 4.1: Bảng tổng hợp dữ liệu kiểm thử

Tên	Đánh giá	Người dùng	Sản phẩm	Ngữ cảnh	Độ thừa
Food	5300	212	20	2	79.20 %
HGift	1477	44	20	4	99.68 %
HMusic	1480	44	134	2	99.28 %
Comoda	2248	82	1225	12	99.99 %
Movielens	100.000	943	1682	2	98.94 %

4.4 Phương pháp phân chia dữ liệu đánh giá

Thông thường, phương pháp phân chia dữ liệu đánh giá và thực nghiệm phổ biến nhất là phương pháp m -fold cross-validation. Tuy nhiên, thông thường phương pháp này được thực hiện bằng cách chọn lựa những đánh giá ngẫu nhiên và không theo quy tắc. Việc lựa chọn ngẫu nhiên các đánh giá trong m phần có thể đưa ra nhiều kết quả phân chia khác nhau, điều đó đồng nghĩa với việc sẽ cho nhiều kết quả kiểm thử khác nhau.

Để giải quyết vấn đề đó, phần kiểm định thực tế này, các dữ liệu kiểm thử sẽ được phân chia theo quy tắc sau đây:

- **Bước 1:** Thực hiện gom nhóm các đánh giá theo từng tổ hợp ngữ cảnh có thể của dữ liệu.
- **Bước 2:** Sắp xếp tăng dần các giá trị đánh giá trong từng tổ hợp ngữ cảnh.
- **Bước 3:** Thực hiện duyệt danh sách giá trị đánh giá trên từng tổ hợp ngữ cảnh, ứng với giá trị duyệt hiện tại trong danh sách, phần tử hiện tại này sẽ được đưa vào phần có số lượng đánh giá nhỏ nhất trong m phần.

¹⁵ Dữ liệu Movielens Semi-Synthetic $\alpha=0.9$, $\beta=0.9$ sẽ được gọi ngắn gọn là Movielens, tuy nhiên cần chú ý rằng bộ dữ liệu này **hoàn toàn khác** với dữ liệu Movielens 100K (<http://www.grouplens.org/node/73>), truy cập lần cuối 22/02/2013.

Việc phân chia theo nguyên tắc này cho kết quả mật độ phân bố giá trị đánh giá trong các phần là tương đương nhau và số lượng đánh giá của các phần sẽ có chênh lệch là nhỏ nhất.

4.5 Kết quả đánh giá thực tế

Để mang lại tính khách quan cho các giải thuật, phương pháp m-Fold Cross Validation được thực hiện với $m=5$, khi đó dữ liệu huấn luyện và dữ liệu kiểm thử sẽ có tỷ lệ tương ứng là 80% – 20%. Kết quả kiểm thử sẽ là trung bình của các kết quả trong 5 lần kiểm thử.

Thử nghiệm đánh giá khả năng đưa ra dự đoán thông qua các chỉ số MAE, RMSE bằng công thức (3.1) và (3.2) trong đó giá trị dự đoán đánh giá được giữ nguyên là kết quả thập phân và **không làm tròn**, trong một số thử nghiệm cho thấy việc làm tròn có thể làm thấp giá trị MAE, tuy nhiên làm tăng giá trị RMSE. Bên cạnh đó, các chỉ số Precision (%), Recall (%) cũng được xem xét để đánh giá khả năng đưa ra khuyến nghị giữa các giải thuật với nhau, giá trị của hai chỉ số này được tính theo công thức (3.3) và (3.4) với hướng tiếp cận *Top-K Recommendation* (xem 3.3.2), giá trị k tùy thuộc vào số lượng sản phẩm trên từng bộ dữ liệu (**Bảng 4.2**).

Bảng 4.2: Số lượng sản phẩm khuyến nghị trên các dữ liệu kiểm thử khác nhau

	Food	MovieLens	Comoda	HMUSIC	HGift
k	5	11	10	10	5

Các giải thuật được sử dụng trong đánh giá thực tế bao gồm hai nhóm, nhóm thứ nhất là các giải thuật CF và CACF dựa trên bộ nhớ và nhóm thứ hai là các giải thuật CF và CACF dựa trên mô hình. Các giải thuật được tổng hợp theo **Bảng 4.3**.

Bảng 4.3: Bảng tổng hợp các giải thuật kiểm thử

Giải thuật	Giải thuật dựa trên bộ nhớ	Giải thuật dựa trên mô hình
2D	User-Based (UB)	Matrix Factorization (MF)
Pre-Filtering	Item Splitting với UB (ISUB) EPF với UB (EPF-UB)	Item Splitting với MF (ISMF) EPF với MF (EPF-MF)
Post-Filtering	Weigh POF với UB (Weigh UB) Filter POF với UB (Filter UB)	Weigh POF với MF (Weigh MF) Filter POF với MF (Filter MF)
Contextual Modeling	Contextual Neighbors Mdl ₁	Context-Aware Matrix Factorization (CAMF)

Kết quả kiểm thử được so sánh theo năm nhóm với các dữ liệu kiểm thử khác nhau được tổng hợp theo bảng sau:

Bảng 4.4: Bảng phân nhóm so sánh các giải thuật và dữ liệu liên quan

Nhóm	Nội dung so sánh	Dữ liệu
1	So sánh khuyến nghị 2D và Pre-Filtering	Food Movielens
2	So sánh khuyến nghị 2D và Post-Filtering	
3	So sánh khuyến nghị 2D và Contextual Modeling	
4	So sánh Pre-Filtering, Post-Filtering và Contextual Modeling	
5	So sánh giải thuật xử lý ngữ cảnh liên quan và không xử lý ngữ cảnh liên quan	Comoda HMusic HGift

Đối với khuyến nghị 2D, trong từng dữ liệu, thông tin ngữ cảnh được loại bỏ và giá trị đánh giá của người dùng trên sản phẩm được tính bằng giá trị **trung bình** của sản phẩm đó trên tất cả các đánh giá cho sản phẩm đó trong nhiều ngữ cảnh khác nhau. Giải thuật UB và MF được chọn làm giải thuật thực nghiệm, những kỹ thuật 2D này sẽ được kết hợp với kỹ thuật tiền xử lý ngữ cảnh với các giải thuật Item Splitting cũng như Exact Pre-Filtering và kỹ thuật hậu xử lý ngữ cảnh với giải thuật Weigh POF và Filter POF. Đối với kỹ thuật mô hình hóa ngữ cảnh, hai giải thuật được chọn là CN và CAMF. Để so sánh các giải thuật này, dữ liệu Food và Movielens sẽ được sử dụng, là các dữ liệu có số lượng đánh giá tương đối cao với giá trị đánh giá đáng tin cậy và số sản phẩm đồng đánh giá lớn. Mặt khác, để so sánh mức độ tác động của ngữ cảnh, dữ liệu Comoda, HMusic và HGift sẽ được áp dụng là các dữ liệu có số tổ hợp ngữ cảnh lớn và số lượng sản phẩm đồng đánh giá hạn chế, vì thế các dữ liệu này sẽ được áp dụng trên các giải thuật không bị ảnh hưởng bởi sản phẩm đồng đánh giá. Các tham số

sử dụng trong các thuật toán này trên các dữ liệu khác nhau được trình bày ở **Phụ lục A**.

4.5.1 So sánh khuyến nghị 2D và Pre-Filtering

Trong phần này, nội dung sẽ trình bày kết quả thực nghiệm giữa khuyến nghị 2D với giải thuật UB và MF so sánh với các giải thuật khuyến nghị theo hướng tiếp cận Pre-filtering (EPF và Item Splitting). Kiểm thử được thực hiện trên dữ liệu Food và Movielens với kết quả được trình bày¹⁶ ở **Bảng 4.5** sau đây:

Bảng 4.5: Kết quả kiểm thử giải thuật khuyến nghị 2D và Pre-Filtering

Dữ liệu	Chỉ số đánh giá	Dựa trên bộ nhớ			Dựa trên mô hình		
		UB	EPF-UB	ISUB	MF	EPF-MF	ISMF
Food	MAE	0.940	0.895	0.882	0.881	0.707	0.876
	RMSE	1.152	1.103	1.111	1.116	0.945	1.104
	Precision	0.0723	0.0701	0.0743	0.0656	0.0712	0.0680
	Recall	0.3403	0.3265	0.3483	0.3083	0.3362	0.3200
Movielens	MAE	0.903	0.867	0.817	0.905	0.778	0.776
	RMSE	1.115	1.118	1.025	1.122	0.982	0.973
	Precision	0.00046	0.00212	0.00131	0.00658	0.00786	0.00637
	Recall	0.00050	0.00333	0.00214	0.01019	0.01358	0.01140

Đối với nhóm giải thuật dựa trên bộ nhớ

- Giải thuật ISUB là giải thuật cho kết quả dự đoán tốt nhất: trên Food, ISUB cho $MAE=0.882$ so với EPF-UB ($MAE=0.895$) và UB ($MAE=0.940$) trong khi đó EPF-UB cho $RMSE=1.103$ thấp nhất so với ISUB ($RMSE=1.111$) và UB ($RMSE=1.152$). Tương tự, trên Movielens, ISUB là giải thuật cho độ lỗi ít nhất ($MAE=0.817$, $RMSE=1.025$), giải thuật cho độ lỗi tối ưu tiếp theo là giải thuật EPF-UB ($MAE=0.867$, $RMSE=1.118$) và giải thuật UB với ($MAE=0.903$, $RMSE=1.115$).
- Về hiệu suất khuyến nghị, ISUB với $Precision=0.0743$ và $Recall=0.3483$ là giải thuật tốt nhất so với UB với chỉ số ($Precision=0.0723$, $Recall=0.3403$) và EPF

¹⁶ Phần còn lại của báo cáo, các số liệu đánh giá được trình bày dưới dạng bảng và một số bảng được gom theo từng nhóm được tô khác màu. Trong từng nhóm, số liệu in đậm biểu thị cho giá trị tốt nhất trong khi số liệu in nghiêng biểu thị cho giá trị kém hiệu quả nhất cho từng chỉ số đánh giá.

($Precision=0.0701$, $Recall=0.3265$) trong khi đó EPF-UB kém hiệu quả nhất trên dữ liệu Food. EPF-UB và IS UB đều là những giải thuật tốt nhất về khuyến nghị trên Movielens với các chỉ số $Precision$ và $Recall$ tương ứng theo thứ tự EPF-MF (0.00212 và 0.00333), ISMF (0.00131 và 0.00214), MF (0.00046 và 0.00050).

Đối với nhóm giải thuật dựa trên mô hình

- Đối với dữ liệu Food, Giải thuật MF đều cho kết quả độ lỗi $MAE=0.881$, $RMSE=1.116$ cao hơn so với EPF-MF ($MAE=0.707$, $RMSE=0.945$) và ISMF ($MAE=0.876$, $RMSE=1.104$), EPF-MF cũng cho kết quả tốt hơn so với ISMF trên cả hai chỉ số này. Tương tự, trên dữ liệu Movielens, ISMF cho kết quả $MAE=0.776$, $RMSE=0.973$ tốt hơn EPF-MF ($MAE=0.778$, $RMSE=0.982$) và giải thuật MF cho độ lỗi nhiều nhất so với hai giải thuật còn lại.
- EPF-MF cho các chỉ số $Precision$ và $Recall$ tốt hơn hai giải thuật còn lại trên dữ liệu Food và Movielens. Tuy nhiên ISMF cho kết quả khuyến nghị tốt hơn MF trên Food nhưng kém hiệu quả hơn MF trên Movielens. Cụ thể, EPF-MF cho chỉ số $Precision=0.0712$ tốt hơn so với ISMF ($Precision=0.0680$) và kém nhất là MF với $Precision=0.0656$, kết quả với chỉ số $Recall$ với thứ tự giảm dần là EPF-MF, ISMF và MF với các giá trị tương ứng là 0.3362, 0.32 và 0.3083 trên dữ liệu Food. Tương tự trên Movielens, thứ tự giải thuật cho khả năng khuyến nghị giảm dần với các chỉ số $Precision$ và $Recall$ là EPF-MF với ($Precision=0.00768$, $Recall=0.01358$), tiếp theo MF với ($Precision=0.00658$, $Recall=0.01019$), cuối cùng là ISMF với ($Precision=0.00637$, $Recall=0.0114$).

Nhìn chung, các giải thuật đưa ra mô hình 2D và Pre-Filtering đều cho các kết quả về chỉ số MAE và $RMSE$ thấp hơn so với các giải thuật dựa trên bộ nhớ, bên cạnh đó phần lớn giải thuật Pre-Filtering cho khả năng khuyến nghị tốt hơn khuyến nghị 2D thông qua chỉ số $Precision$ và $Recall$.

Tóm lại, giải thuật Pre-Filtering thường cho kết quả tối ưu và hoạt động hiệu quả hơn so với các giải thuật 2D trong những hệ khuyến nghị dựa trên ngữ cảnh trong thử nghiệm này.

4.5.2 So sánh khuyến nghị 2D và Post-Filtering

Nội dung phần này sẽ so sánh mức độ hiệu quả của giải thuật khuyến nghị 2D truyền thống (UB và MF) so với các giải thuật Post-Filtering (Weigh POF và Filter POF).

Bảng 4.6 thể hiện kết quả kiểm thử giữa hai kỹ thuật thông qua các chỉ số đánh giá *MAE*, *RMSE*, *Precision*, *Recall* trên dữ liệu Food và Movielens.

Bảng 4.6: Kết quả kiểm thử giải thuật khuyến nghị 2D và Post-Filtering

Dữ liệu	Chỉ số đánh giá	Dựa trên bộ nhớ			Dựa trên mô hình		
		UB	Weigh UB	Filter UB	MF	Weigh MF	Filter MF
Food	MAE	0.940	0.940	0.940	0.881	0.881	0.881
	RMSE	1.152	1.152	1.152	1.116	1.116	1.116
	Precision	0.0723	0.0927	0.2021	0.0656	0.0959	0.1903
	Recall	0.3403	0.4364	0.1903	0.3083	0.4502	0.2021
Movielens	MAE	0.903	0.903	0.903	0.905	0.905	0.905
	RMSE	1.115	1.115	1.115	1.122	1.122	1.122
	Precision	0.00046	0.03799	0.02623	0.00658	0.05659	0.04436
	Recall	0.0005	0.12422	0.06453	0.01019	0.1185	0.083

Nhóm giải thuật Post-Filtering điều chỉnh danh sách khuyến nghị từ khuyến nghị 2D, do đó các chỉ số độ lỗi *MAE*, *RMSE* hoàn toàn tương đương với khuyến nghị 2D. Do đó, nội dung phần này sẽ không trình bày so sánh trên các chỉ số này.

Đối với nhóm giải thuật dựa trên bộ nhớ: giải thuật Filter UB và giải thuật Weigh UB đều cho khả năng khuyến nghị hoàn toàn tốt hơn khuyến nghị 2D. Cụ thể, trên Food, giải thuật Filter UB cho *Precision*=0.2021 tốt hơn so với Weigh UB (*Precision*=0.0927) và UB (*Precision*=0.0723). Trên Movielens, thứ tự giải thuật cho giá trị *Precision* tốt nhất theo thứ tự giảm dần là Weigh UB (0.03799), Filter UB (0.02623) và UB (0.00046). Đối với giá trị *Recall*, Filter UB cho kết quả thấp hơn so với UB với giá trị tương ứng là 0.1903 và 0.3403 trên dữ liệu Food, Weigh UB cho *Recall*=0.4364 tốt nhất (trên Food). Mặt khác, giải thuật Weigh UB đều cho kết quả *Precision* và *Recall* tốt hơn hai giải thuật còn lại trong khi đó UB là giải thuật cho các chỉ số này kém hiệu quả nhất trên Movielens.

Kết quả hoàn toàn tương đương với nhóm giải thuật dựa trên mô hình: giải thuật Weigh MF cho *Recall* tốt nhất trong khi đó Filter MF cho *Precision* tốt nhất (trên Food). Weigh MF với *Precision*=0.05659 và *Recall*=0.1185 hoàn toàn cho khả năng khuyến nghị tối ưu hơn Filter MF (*Precision*=0.04436, *Recall*=0.083) trên Movielens và giải thuật 2D MF với *Precision* và *Recall* tương ứng là 0.00658 và 0.01019 là giải thuật kém hiệu quả nhất trên cả hai dữ liệu thực nghiệm.

Nhìn chung, giải thuật Post-Filtering cho kết quả độ lỗi *MAE* và *RMSE* tương đương với giải thuật 2D. Điểm vượt trội của nhóm giải thuật này là khả năng đưa ra khuyến nghị cao (*Precision* và *Recall* tốt hơn 2D).

4.5.3 So sánh khuyến nghị 2D và Contextual Modeling

Trong phần so sánh này, giải thuật UB sẽ được so sánh với giải thuật CN với ràng buộc Mdl_1 và giải thuật MF sẽ được so sánh với giải thuật CAMF trên các bộ dữ liệu Food, Movielens, kết quả được minh họa bởi **Bảng 4.7**.

Bảng 4.7: Kết quả kiểm thử giải thuật khuyến nghị 2D và Contextual Modeling

Dữ liệu	Chỉ số đánh giá	Dựa trên bộ nhớ		Dựa trên mô hình	
		UB	CN	MF	CAMF
Food	MAE	0.940	0.959	0.881	0.843
	RMSE	1.152	1.165	1.116	1.065
	Precision	0.0723	0.0618	0.0656	0.0620
	Recall	0.3403	0.2910	0.3083	0.2928
Movielens	MAE	0.903	0.908	0.905	0.749
	RMSE	1.115	1.145	1.122	0.933
	Precision	0.00046	0.00057	0.00658	0.00167
	Recall	0.00050	0.00104	0.01019	0.00253

Đối với nhóm giải thuật dựa trên bộ nhớ, giải thuật CN hoàn toàn cho chỉ số *MAE*, *RMSE* đều kém hơn giải thuật UB trên cả hai dữ liệu Food và Movielens. Cụ thể CN cho kết quả (*MAE*=0.959, *RMSE*=1.165) kém hiệu quả hơn với UB (*MAE*=0.94, *RMSE*=1.152) trên Food và kết quả tương tự trên Movielens là CN (*MAE*=0.908, *RMSE*=1.145) và UB (*MAE*=0.903 và *RMSE*=1.115). Mặt khác giải thuật UB cho kết quả khuyến nghị với *Precision*=0.0723, *Recall*=0.3403 tốt hơn so với giải thuật CN

cho kết quả $Precision=0.0618$, $Recall=0.2910$ trên Food. Tuy nhiên CN với giá trị ($Precision=0.00057$, $Recall=0.00104$) cho kết quả khuyến nghị tốt hơn giải thuật UB ($Precision=0.00046$, $Recall=0.00050$) trên dữ liệu Movielens.

Đối với nhóm giải thuật dựa trên mô hình, giải thuật CAMF luôn cho kết quả độ lỗi MAE và $RMSE$ tốt hơn so với MF. Cụ thể, trên dữ liệu Food, CAMF cho độ lỗi $MAE=0.843$, $RMSE=1.065$ tốt hơn MF cho kết quả $MAE=0.882$, $RMSE=1.116$, trên dữ liệu Movielens kết quả cũng hoàn toàn tương tự CAMF ($MAE=0.749$, $RMSE=0.993$) so với giải thuật MF ($MAE=0.863$, $RMSE=1.059$). Tuy nhiên khả năng khuyến nghị của CAMF chưa hiệu quả so với MF khi CAMF cho giá trị $Precision$ và $Recall$ thấp hơn MF trên cả hai dữ liệu. Cụ thể với dữ liệu Food, CAMF cho kết quả $Precision=0.062$, $Recall=0.3038$ so sánh với MF ($Precision=0.656$, $Recall=0.3038$) cho chênh lệch không đáng kể, tương tự trên Movielens khi CAMF và MF lần lượt cho kết quả $Precision$ tương ứng theo thứ tự là 0.00167 và 0.00658, giá trị $Recall$ tương ứng theo thứ tự là 0.00253 và 0.01019.

Nhìn chung, giải thuật CN chưa cho kết quả dự đoán (MAE , $RMSE$) tốt hơn UB trong khi CAMF luôn cho kết quả dự đoán tốt hơn MF. Khả năng khuyến nghị ($Precision$, $Recall$) của giải thuật CAMF chưa tốt bằng MF trong khi đó CN trong một số trường hợp cho kết quả tốt hơn UB.

4.5.4 So sánh Pre-Filtering, Post-Filtering và Contextual Modeling

Pre-Filtering, Post-Filtering và Contextual Modeling là ba kỹ thuật chính trong hệ khuyến nghị dựa trên ngữ cảnh. Để kiểm chứng mức độ hiệu quả của các giải thuật thuộc các kỹ thuật này, nội dung phần này sẽ thực hiện so sánh các giải thuật thuộc hai nhóm: dựa trên bộ nhớ (EPF-UB, ISUB, Weigh UB, Filter UB, CN) và dựa trên mô hình (EPF-MF, ISMF, Weigh MF, Filter MF, CAMF) trên dữ liệu Food và Movielens thông qua các chỉ số đánh giá dự đoán (MAE , $RMSE$), đánh giá khả năng khuyến nghị ($Precision$, $Recall$).

Bảng 4.8: Kết quả kiểm thử giải thuật khuyến nghị Pre, Post-Filtering và Contextual Modeling dựa trên bộ nhớ

Dữ liệu	Chỉ số đánh giá	Dựa trên bộ nhớ				
		EPF-UB	ISUB	Weigh UB	Filter UB	CN
Food	MAE	0.895	0.882	0.940	0.940	0.959
	RMSE	1.103	1.111	1.152	1.152	1.165
	Precision	0.0701	0.0743	0.0927	0.2021	0.0618
	Recall	0.3265	0.3483	0.4364	0.1903	0.2910
Movielens	MAE	0.867	0.817	0.903	0.903	0.908
	RMSE	1.118	1.025	1.115	1.115	1.145
	Precision	0.00212	0.00131	0.03799	0.02623	0.00057
	Recall	0.00333	0.00214	0.12422	0.06453	0.00104

Kết quả thực nghiệm so sánh các giải thuật dựa trên bộ nhớ được minh họa bởi **Bảng 4.8**, cụ thể như sau:

- Giải thuật ISUB cho kết quả $MAE=0.882$ tốt nhất so với các giải thuật khác trong khi EMF UB cho kết quả tốt nhất với chỉ số $RMSE=1.103$ trên dữ liệu Food. Trên dữ liệu Movielens, ISUB cho kết quả tốt nhất với $MAE=0.817$ và $RMSE=1.025$ và là giải thuật cho kết quả dự đoán đánh giá tốt nhất. Hai giải thuật Weigh POF UB và Filter POF UB cho cùng kết quả độ lỗi và thấp hơn CN trên cả hai bộ dữ liệu Food và Movielens. Contextual Neighbors là giải thuật cho độ lỗi nhiều nhất so trên cả hai dữ liệu và là giải thuật kém hiệu quả nhất.
- Về khả năng khuyến nghị, nhóm giải thuật Post-Filtering (Weigh UB, Filter UB) vượt trội hơn các nhóm còn lại. Filter UB cho giá trị $Precision=0.2021$ tốt nhất và CN với $Precision=0.0618$ kém nhất, Weigh UB cho giá trị $Recall=0.4364$ tốt nhất so sánh với CN kém hiệu quả nhất với $Recall=0.291$ trên Food trong khi Weigh UB là giải thuật có khả năng khuyến nghị tốt nhất với các chỉ số $Precision$ và $Recall$ tối ưu trên Movielens với $Precision=0.03799$ và $Recall=0.12422$, tương tự CN cho giá trị $Precision=0.00057$ và giá trị $Recall=0.00104$. Các giải thuật EPF-UB và ISUB luôn cho kết quả $Precision$ và $Recall$ tốt hơn CN nhưng kém hiệu quả hơn Weigh UB và Filter UB.

Bảng 4.9: Kết quả kiểm thử giải thuật khuyến nghị Pre, Post-Filtering và Contextual Modeling dựa trên mô hình

Dữ liệu	Chỉ số đánh giá	Dựa trên mô hình				
		EPF-MF	ISMF	Weigh MF	Filter MF	CAMF
Food	MAE	0.707	0.876	0.881	0.881	0.843
	RMSE	0.945	1.104	1.116	1.116	1.065
	Precision	0.0712	0.0680	0.0959	0.1903	0.0620
	Recall	0.3362	0.3200	0.4502	0.2021	0.2928
Movielens	MAE	0.778	0.776	0.905	0.905	0.749
	RMSE	0.982	0.973	1.122	1.122	0.933
	Precision	0.00786	0.00637	0.05659	0.04436	0.00167
	Recall	0.01358	0.01140	0.11850	0.08300	0.00253

Kết quả so sánh nhóm giải thuật dựa trên mô hình được thể hiện ở **Bảng 4.9**, cụ thể như sau:

- Trên Food, EPF-MF cho $MAE=0.707$, $RMSE=0.945$ là giải thuật cho khả năng đánh giá tốt nhất, tiếp theo là giải thuật CAMF ($MAE=0.843$, $RMSE=1.065$) và giải thuật ISMF với ($MAE=0.876$, $RMSE=1.104$). Giải thuật Weigh POF MF và Filter POF MF cho kết quả độ lỗi tương đương và là những giải thuật có độ lỗi nhiều nhất. Trên Movielens, CAMF cho kết quả $MAE=0.749$, $RMSE=0.933$ là giải thuật cho khả năng dự đoán tốt nhất, tiếp theo là ISMF, EPF-MF, Weigh MF và Filter MF.
- Nhóm giải thuật Post-Filtering (Weigh MF, Filter MF) có khả năng khuyến nghị vượt trội hơn các nhóm còn lại. Filter MF và Weigh MF lần lượt đưa ra các chỉ số *Precision* và *Recall* tối ưu so với các nhóm còn lại (trên Food) trong khi Weigh MF là giải thuật có khả năng khuyến nghị tốt nhất (trên Movielens) với các chỉ số *Precision* và *Recall* đạt giá trị cao nhất. Cụ thể, trên dữ liệu Food, Filter MF cho $Precision=0.1903$ tốt nhất khi so sánh với Weigh MF ($Precision=0.0959$) và EPF-MF ($Precision=0.0712$), ba giải thuật cho giá trị Recall tối ưu theo thứ tự Weigh MF, EPF-MF và ISMF với giá trị lần lượt là 0.4503, 0.3362 và 0.32. Kết quả trên dữ liệu Movielens cho Weigh MF ($Precision=0.05659$, $Recall=0.1185$) tốt nhất khi so sánh với các giải thuật còn

lại. Giải thuật cho khả năng khuyến nghị hiệu quả tiếp theo là Filter MF với giá trị ($Precision=0.04436$, $Recall=0.08300$) và giải thuật EPF-MF với giá trị ($Precision=0.00786$, $Recall=0.01358$) trong khi đó CAMF ($Precision=0.00167$, $Recall=0.00253$) cho khả năng khuyến nghị chưa tối ưu so với các giải thuật còn lại.

Khi so sánh trên phạm vi cả hai nhóm giải thuật dựa trên bộ nhớ và mô hình. Nhìn chung nhóm giải thuật dựa trên mô hình cho kết quả độ lỗi MAE và RMSE thấp hơn so nhóm giải thuật dựa trên bộ nhớ. Nguyên nhân có thể giải thích cho kết quả này do độ tương tự của những người dùng (EPF-UB, ISUB, UB) hoặc các cặp (người dùng, ngữ cảnh) (giải thuật CN) chưa đáng tin cậy khi số lượng sản phẩm đồng đánh giá hạn chế.

Trong thử nghiệm này, phần lớn các giải thuật Pre-Filtering cho kết quả độ lỗi (MAE , $RMSE$) thấp nhất (trên cả hai nhóm giải thuật dựa trên mô hình và bộ nhớ), trong khi đó giải thuật Post-Filtering luôn cho kết quả khuyến nghị cao nhất ($Precision$ và $Recall$) và độ lỗi nhiều nhất (một số ít trường hợp thấp hơn Context Modeling). Giải thuật Contextual Modeling cho kết quả độ lỗi cao hơn (giải thuật CN) với Pre-Filtering hoặc thấp hơn (giải thuật CAMF) trong một số trường hợp và khả năng khuyến nghị chưa thực sự hiệu quả so với các nhóm khác.

Tóm lại, các giải thuật Pre-Filtering, Post-Filtering và Contextual Modeling đều có những ưu điểm và nhược điểm riêng khi so sánh trên những chỉ số đánh giá khác nhau. Việc kết luận kỹ thuật tiếp cận khuyến nghị dựa trên ngữ cảnh nào tốt hơn còn phụ thuộc vào rất nhiều yếu tố như nhu cầu của ứng dụng, trường hợp sử dụng của ứng dụng cũng như đặc trưng dữ liệu của từng ứng dụng.

4.5.5 So sánh giải thuật xử lý và không xử lý ngữ cảnh liên quan

Việc tác động của ngữ cảnh là yếu tố quyết định đến thành công của một thuật toán khuyến nghị kết hợp với yếu tố ngữ cảnh. Adomavicius và cộng sự [15] đã từng nhận xét rằng *không phải tất cả ngữ cảnh đều tác động đến đánh giá*. Để kiểm chứng cho nhận xét trên, phần nội dung này sẽ trình bày việc áp dụng giải thuật khuyến nghị với những dữ liệu có những yếu tố ngữ cảnh không liên quan dựa trên phương pháp của

Adomavicius và cộng sự [15] (xem 2.2.4). Dựa trên những phần kiểm chứng trước đó, giải thuật ISMF và CAMF là những giải thuật cho khả năng dự đoán tương đối tốt (cho giá trị MAE và RSME nhỏ) sẽ được áp dụng trên ba bộ dữ liệu là Comoda, HGift và HMusic. Thực nghiệm so sánh mức độ tác động của ngữ cảnh theo ba trường hợp sau:

- **TH₁**: Không loại bỏ bất kỳ một ngữ cảnh nào.
- **TH₂**: Loại bỏ ngữ cảnh khi **tồn tại ít nhất một** kiểm định t-test giữa các cặp giá trị của ngữ cảnh đó **không thỏa mãn** mức tin cậy 95%.
- **TH₃**: Loại bỏ ngữ cảnh khi **không tồn tại ít nhất một** kiểm định t-test giữa các cặp giá trị của ngữ cảnh đó **thỏa mãn** mức tin cậy 95%.

Dữ liệu Comoda, HMusic, HGift sẽ được so sánh trong 3 trường hợp. Sau khi những ngữ cảnh không liên quan được loại bỏ, giá trị đánh giá sẽ được lấy trung bình trong trường hợp dữ liệu đánh giá trùng lặp. Để đánh giá mức độ hiệu quả của giải thuật, các chỉ số *MAE*, *RMSE*, *Precision*, *Recall* sẽ được so sánh. Kết quả được tổng hợp dưới bảng sau đây:

Bảng 4.10: Kết quả so sánh giải thuật ISMF, CAMF khi xử lý ngữ cảnh liên quan trên dữ liệu Comoda, HMusic và HGift

Dữ liệu	Chỉ số đánh giá	ISMF			CAMF		
		TH ₁	TH ₂	TH ₃	TH ₁	TH ₂	TH ₃
Comoda	Ngữ cảnh	12	2	7	12	2	7
	MAE	0.841	0.841	0.835	0.884	0.871	0.874
	RMSE	1.067	1.078	1.063	1.172	1.144	1.160
	Precision	0.00348	0.00454	0.00401	0.00328	0.00320	0.00334
	Recall	0.03327	0.04391	0.03964	0.03141	0.03056	0.03200
HMusic	Ngữ cảnh	2	0	1	2	0	1
	MAE	0.806	–	0.809	0.844	–	0.888
	RMSE	1.055	–	1.059	1.121	–	1.142
	Precision	0.02305	–	0.02543	0.02202	–	0.01973
	Recall	0.11489	–	0.12891	0.09398	–	0.06520
HGift	Ngữ cảnh	4	1	4	4	1	4
	MAE	–	1.228	1.188	–	1.204	1.057
	RMSE	–	1.461	1.445	–	1.411	1.322
	Precision	–	0.1335	0.1347	–	0.1388	0.1599
	Recall	–	0.5623	0.5607	–	0.5838	0.6672

Đối với dữ liệu Comoda, sau khi xử lý ngưỡng cảnh với TH_2 và TH_3 , CAMF và ISMF cho kết quả tốt hơn so với TH_1 . Cụ thể, ISMF cho *MAE* giảm từ 0.841 xuống 0.835, *RMSE* giảm từ 1.067 xuống 1.063, giá trị *Precision*=0.00348, *Recall*=0.03327 tăng lên 0.00401 và 0.03964 với TH_3 so với TH_1 . Tương tự, CAMF phần lớn đều cho kết quả các chỉ số tốt hơn khi xử lý ngưỡng cảnh TH_2 , TH_3 so với TH_1 .

Đối với dữ liệu HMusic, khi thực hiện xử lý ngưỡng cảnh với TH_2 , tất cả ngưỡng cảnh đều bị loại bỏ vì thế giải thuật ISMF và CAMF không thể áp dụng trên dữ liệu thiếu yếu tố ngưỡng cảnh. Vì vậy, giải thuật ISMF và CAMF sẽ chỉ được so sánh khi xử lý ngưỡng cảnh với TH_1 và TH_3 . Đối với ISMF, mặc dù ISMF với TH_3 cho độ lỗi lớn hơn khi xử lý với TH_1 (*MAE* tăng 0.003 và *RMSE* tăng 0.004), tuy nhiên giá trị *Precision* và *Recall* được cải thiện (*Precision* tăng ~0.0024 và *Recall* tăng ~0.014). Ngược lại, CAMF với TH_3 cho hiệu quả kém hơn với TH_1 khi các chỉ số *MAE*, *RMSE* tăng trong khi *Precision* và *Recall* giảm..

Đối với dữ liệu HGift, ISMF và CAMF kém hiệu quả hơn khi xử lý dữ liệu với TH_2 so với TH_3 . Trong dữ liệu này, khi xử lý dữ liệu với TH_1 và TH_3 cho kết quả tương đương vì không loại bỏ bất kỳ ngưỡng cảnh nào. Cụ thể, *MAE* tăng từ 1.057 lên 1.204 và *Precision*=0.1599, *Recall*=0.6672 giảm xuống 0.1388 và 0.5838 với CAMF khi xử lý ngưỡng cảnh với TH_3 và TH_2 tương ứng. Tuy nhiên, với ISMF, *MAE* tăng từ 1.188 lên 1.228, *Precision* giảm (từ 0.1347 xuống 0.1335) trong khi *Recall* tăng (0.5607 lên 0.5623) với chênh lệch không đáng kể khi so sánh TH_3 và TH_2 .

Nhìn chung, phần lớn giải thuật lọc cộng tác CACF thực hiện với dữ liệu sau khi được xử lý ngưỡng cảnh đều hiệu quả hơn so với khi chưa được xử lý ngưỡng cảnh liên quan (Comoda, HGift) và một số ít kém hiệu quả hơn (HMusic). Cụ thể, trong thực nghiệm này, phần lớn giải thuật ISMF và CAMF cho khả năng dự đoán (chỉ số *MAE*, *RMSE*) và khả năng khuyến nghị (*Precision*, *Recall*) tốt hơn khi được áp dụng trên dữ liệu được xử lý ngưỡng cảnh liên quan. Qua đó, việc tác động của ngưỡng cảnh có vai trò hết sức quan trọng tác động trực tiếp đến hành vi của người dùng, cụ thể là việc đưa ra đánh giá cũng như quyết định đến hiệu quả của giải thuật trong hệ khuyến nghị dựa trên ngưỡng cảnh.

4.6 Kết luận chung

Thông qua những đánh giá trên, kết quả cho thấy giải thuật không chỉ phụ thuộc vào bản chất thuật toán mà còn phụ thuộc rất nhiều vào dữ liệu và không phải giải thuật nào tốt cũng hiệu quả trên mọi dữ liệu. Trong giới hạn của thử nghiệm này, nhìn chung, giải thuật kết hợp ngữ cảnh cho kết quả tốt hơn giải thuật khuyến nghị truyền thống và một số ít trường hợp chưa hiệu quả bằng khuyến nghị truyền thống. Cụ thể trong nhóm giải thuật kết hợp với ngữ cảnh, giải thuật Pre-Filtering và Post-Filtering sử dụng trong khóa luận đều cho thấy phần lớn tính ổn định và tốt hơn so với giải thuật 2D truyền thống, đặc biệt giải thuật Pre-Filtering cho kết quả tốt về dự đoán (phần lớn cho độ lỗi *MAE*, *RMSE* thấp nhất, đặc biệt là ISMF cho khả năng dự đoán ổn định và có khả năng giải quyết vấn đề dữ liệu thưa so với EPF) trong khi Post-Filtering (Weigh POF và Filter POF) cho kết quả tốt về mặt khuyến nghị (phần lớn cho chỉ số *Precision* và *Recall* cao nhất). Mặt khác, giải thuật Contextual Modeling dựa trên bộ nhớ (Contextual Neighbors) chưa thực sự tốt hơn so với Pre-Filtering lẫn 2D trong khi Contextual Modeling dựa trên mô hình (CAMF) cho kết quả dự đoán tốt hơn 2D và một số trường hợp tốt hơn các kỹ thuật còn lại. Mặc dù CAMF cho khả năng dự đoán tốt và ổn định, tuy nhiên CAMF cho khả năng khuyến nghị chưa hiệu quả so với các giải thuật còn lại.

Nhìn chung, đa số các giải thuật dựa trên mô hình xây dựng mô hình ngoại tuyến (off-line) nên khả năng linh hoạt xử lý dữ liệu người dùng mới và sản phẩm mới của các giải thuật dựa trên mô hình chưa thực sự hiệu quả, vì khi những đối tượng này biến động một cách nhanh chóng và liên tục về số lượng, giải thuật phải xây dựng nhiều mô hình nhiều lần. Điều này đồng nghĩa với việc tiêu hao nhiều chi phí và tài nguyên, đặc biệt với các hệ thống có dữ liệu lớn và không ngừng tăng trưởng nhanh chóng. Bên cạnh đó, các giải thuật dựa trên bộ nhớ thường phải sử dụng một số lượng lớn đánh giá để tính toán các giá trị đánh giá, vì thế những giải thuật này có thể gặp có khăn trong việc phân bổ tài nguyên hệ thống xử lý.

Mặt khác, việc tác động “thực sự” của ngữ cảnh cũng ảnh hưởng rất nhiều đến hiệu quả của giải thuật. Một số thử nghiệm trong việc lựa chọn ngữ cảnh liên quan cũng cho thấy tầm quan trọng của việc lựa chọn ngữ cảnh trong hệ khuyến nghị. Việc

lựa chọn những ngữ cảnh không phù hợp hoàn toàn có thể làm giảm hiệu quả khuyến nghị. Đây cũng là một vấn đề hết sức quan trọng cần được chú trọng khi xây dựng các hệ khuyến nghị dựa trên ngữ cảnh.

Thông qua những thực nghiệm trên, những hệ khuyến nghị có sử dụng yếu tố ngữ cảnh tác động đến người dùng có thể sử dụng các thuật toán khác nhau tùy thuộc vào mức độ phức tạp của ứng dụng như số lượng người dùng, số lượng sản phẩm, số lượng ngữ cảnh. Bên cạnh đó, yêu cầu của ứng dụng như cần những dữ liệu dự đoán đánh giá (những ứng dụng tích hợp giải thích khuyến nghị) hay chỉ quan tâm đến sản phẩm khuyến nghị (danh sách khuyến nghị) cũng tác động đến việc lựa chọn thuật toán. Ngoài ra, môi trường phát triển, triển khai ứng dụng cũng đóng vai trò không nhỏ trong việc lựa chọn giải thuật khuyến nghị, đặc biệt là việc sử dụng không gian lưu trữ và việc đáp ứng tài nguyên. Việc kết luận một giải thuật nào đó là hoàn toàn tốt vẫn đang là một câu hỏi khó đặt ra nhiều thách thức cho các nhà nghiên cứu hiện nay.

Kết quả thực nghiệm này chưa thực sự giống với kết quả thực nghiệm bởi Panniello và cộng sự [5], [6] sử dụng dữ liệu giao dịch (*transaction-based dataset*) với tần số mua hàng thay cho thang đánh giá và thực hiện kiểm thử trên các giải thuật khuyến nghị dựa trên bộ nhớ (UB, EPF-UB, Weigh UB, Filter UB và CN Mdl₁). Kết quả thực nghiệm bởi Panniello và cộng sự [5], [6], như sau: Filter PoF tốt hơn giải thuật UB 2D truyền thống trong khi Weigh UB cho kết quả không tốt bằng giải thuật UB 2D. Mặt khác, giải thuật EPF-UB cho kết quả luôn tốt hơn giải thuật UB. Bên cạnh đó, CN Mdl₁ cho kết quả tốt hơn EPF-UB không đáng kể, chưa hiệu quả bằng Filter UB và tốt hơn Weigh UB. Nguyên nhân sử dụng nguồn dữ liệu khác nhau và bản chất dữ liệu khác nhau, đặc biệt là thang đánh giá 1-5 (có miền giá trị xác định) so với tần số mua hàng (không có miền giá trị xác định) nên kết quả kiểm thử so sánh khác nhau hoàn toàn hợp lý so với thực nghiệm kiểm thử trong khóa luận này. Mặt khác, giải thuật Item Splitting [7] và CAMF [8] do Baltrunas và cộng sự đề xuất không thực hiện bất kỳ so sánh nào trên cùng dữ liệu với khóa luận. Do đó, khóa luận không thể so sánh được với các thử nghiệm khác trên giải thuật Item Splitting và CAMF.

ỨNG DỤNG MINH HỌA

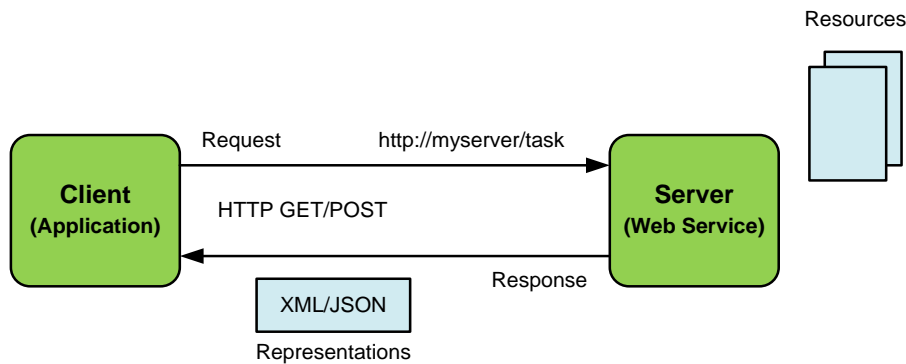
Ricci và cộng sự [9] đã tổng hợp và đưa ra rất nhiều vấn đề của khuyến nghị, trong đó bao gồm việc xây dựng trình chủ khuyến nghị (Recommender Server) dùng chung cho các hệ khuyến nghị trong môi trường Internet. Vì vậy, khóa luận sẽ thực hiện xây dựng một dịch vụ khuyến nghị có thể dùng chung cho nhiều ứng dụng. Trong chương này sẽ trình bày một cách tổng quát nhất về trình chủ khuyến nghị được xây dựng dưới dạng web service dựa trên kiến trúc RESTful thông qua các giao thức HTTP và hỗ trợ một số thuật toán khuyến nghị ngữ cảnh trong lọc cộng tác tiêu biểu. Mặt khác, để minh họa việc sử dụng dịch vụ khuyến nghị này, hai ứng dụng với quy mô khác nhau, nội dung khác nhau được triển khai để cho thấy khả năng tích hợp linh hoạt của dịch vụ này bất chấp các rào cản về ngôn ngữ lập trình và cấu trúc dữ liệu khác nhau.

Phần đầu tiên sẽ trình bày về dịch vụ khuyến nghị ngữ cảnh với những chức năng, thành phần và nguyên tắc hoạt động. Tiếp theo sẽ trình bày hai ứng dụng minh họa là hệ thống nghe nhạc HMusic và ứng dụng khuyến nghị quà tặng HGift. Nội dung chương chỉ tập trung vào các thành phần chính của các ứng dụng mà không đi sâu vào quá trình xây dựng ứng dụng cũng như việc xây dựng mã nguồn.

5.1 Dịch vụ khuyến nghị dựa trên ngữ cảnh

5.1.1 Kiến trúc của dịch vụ

Dịch vụ khuyến nghị dựa trên ngữ cảnh (context – aware recommender service) là một dịch vụ web¹⁷ (web service) xây dựng dựa trên kiến trúc RESTful [46] bao gồm bốn thành phần chính (resource, representation, uri và phương thức trong http) được minh họa bởi **Hình 5.1** như sau:



Hình 5.1: Kiến trúc tổng quát của RESTful Web Service

Resource (tài nguyên) là các thông tin được trao đổi thông qua môi trường Internet, được truy cập và vận chuyển giữa các ứng dụng client và server. Cụ thể trong dịch vụ khuyến nghị này, tài nguyên chính là dữ liệu đánh giá được gửi từ các ứng dụng client đến dịch vụ (server), thông tin yêu cầu khuyến nghị và danh sách khuyến nghị được server gửi cho ứng dụng client.

Representation (thể hiện) là thông tin trao đổi giữa ứng dụng client và dịch vụ, nó được xem như là dữ liệu hiện thời của tài nguyên (hay thể hiện của tài nguyên) tại thời điểm được yêu cầu. Thể hiện của tài nguyên thường được định nghĩa nhiều dạng khác nhau như: ký tự, hình ảnh, XML hoặc JSON. Trong dịch vụ khuyến nghị, thể hiện của tài nguyên được định nghĩa dưới định dạng JSON. Có ba định dạng thể hiện chính được định nghĩa trong dịch vụ khuyến nghị là dữ liệu đánh giá, thông tin yêu cầu khuyến nghị và danh sách khuyến nghị được minh họa bởi **Hình 5.2**.

¹⁷ Bản chất của dịch vụ web service là một ứng dụng chạy ở máy chủ (server). Các thuật ngữ dịch vụ khuyến nghị và server khuyến nghị mang ý nghĩa tương đương trong luận văn này.

```
{
  "appid": "i2s3j4j5z2",
  "data": [
    { "userid": "8", "context": "3-187", "itemid": "460", "rating": "4" },
    { "userid": "8", "context": "2-183", "itemid": "461", "rating": "4" },
    { "userid": "8", "context": "2-184", "itemid": "462", "rating": "4" },
    { "userid": "8", "context": "3-189", "itemid": "463", "rating": "4" }
  ]
}
```

(a) Dữ liệu đánh giá

```
{
  "userid": "8",
  "context": "3-183",
  "appid": "i2s3j4j5z2"
  "recommend": [
    { "itemid": "1548", "score": 4.94 },
    { "itemid": "501", "score": 4.85 }
  ]
}
```

(b) Thông tin yêu cầu khuyến nghị

(c) Danh sách khuyến nghị

Hình 5.2: Định dạng JSON các thể hiện trong dịch vụ khuyến nghị

URI trong RESTful được xem như là các liên kết đến tài nguyên được dùng để trao đổi các thể hiện. URI phải được định nghĩa trước, cố định và không thay đổi khi dịch vụ hoạt động (online) mặc dù có thể tồn tại nhiều thay đổi tại server như thay đổi thiết bị lưu trữ...

Phương thức trong HTTP được dùng để trao đổi dữ liệu giữa client và server, khi đó dữ liệu được xem như là các tham số truyền hay dữ liệu đính kèm. RESTful cho phép thao tác, trao đổi dữ liệu dưới 4 phương thức được định nghĩa trong HTTP đó là *get*, *post*, *put*, *delete* và quy định những chức năng cụ thể đối với từng phương thức. Trong dịch vụ khuyến nghị dựa trên ngữ cảnh phương thức *post* được dùng để trao đổi dữ liệu giữa ứng dụng client và server. Dữ liệu đính kèm trong phương thức *post* chính là các thể hiện (dữ liệu đánh giá, thông tin yêu cầu khuyến nghị hoặc danh sách sản phẩm khuyến nghị) được định nghĩa thông qua định dạng JSON.

5.1.2 Thành phần của dịch vụ

Dịch vụ khuyến nghị bao gồm ba thành phần chính

Thành phần tương tác client-server: Web Service được xây dựng với ngôn ngữ lập trình Java dựa trên Jersey Framework (JAX-RS) nên việc xử lý giao tiếp giữa client và server đã được đơn giản hóa thông qua các interface được định nghĩa trước đó.

Thành phần xử lý: Dịch vụ hỗ trợ các thuật toán khuyến nghị dựa trên ngữ cảnh với kỹ thuật lọc cộng tác. Trong dịch vụ này, giải thuật khuyến nghị bao gồm các giải thuật lọc cộng tác dựa trên mô hình kết hợp với yếu tố ngữ cảnh vì những giải thuật này giải quyết được các vấn đề như dữ liệu thưa, người dùng mới, sản phẩm mới và cho khả năng đáp ứng tốt nhờ sử dụng các tham số từ mô hình mà không phải sử dụng trên một số lượng lớn dữ liệu như lọc cộng tác dựa trên bộ nhớ. Cụ thể là giải thuật CAMF và ISMF vì những giải thuật này cho khả năng dự đoán tương đối tốt và ổn định. Các giải thuật có thể được xây dựng mở rộng dựa trên một cấu trúc trừu tượng được định nghĩa trước và dễ dàng tích hợp vào dịch vụ mà không ảnh hưởng đến các giải thuật khác.

Thành phần lưu trữ: Dịch vụ sử dụng hệ quản trị phi quan hệ (NoSQL) MongoDB để lưu trữ các thông tin về dữ liệu đánh giá cũng như mô hình (tham số) cho các ứng dụng. Việc sử dụng MongoDB mang lại tính linh hoạt cao hơn so với các hệ quản trị dữ liệu quan hệ, với mỗi ứng dụng sử dụng các thuật toán khác nhau, việc lưu trữ các mô hình (tham số) khác nhau, khi đó mỗi mô hình sẽ được lưu trữ độc lập với những *collection*¹⁸ có những cấu trúc khác nhau. Khi dịch vụ cung cấp thêm các giải thuật khác, việc lưu trữ sẽ trở nên rất dễ dàng với MongoDB thông qua các collection (cấu trúc của collection không phải định nghĩa trước) so với việc sử dụng các hệ quản trị cơ sở dữ liệu quan hệ. Mặt khác, tốc độ nhanh khi xử lý dữ liệu lớn cũng là một ưu thế của hệ quản trị này.

5.1.3 Hoạt động của dịch vụ

Dịch vụ bao gồm bốn chức năng chính:

- (i) Cung cấp mã sử dụng dịch vụ.
- (ii) Lưu trữ thông tin dữ liệu đánh giá.
- (iii) Xây dựng mô hình.
- (iv) Xử lý đưa ra danh sách khuyến nghị.

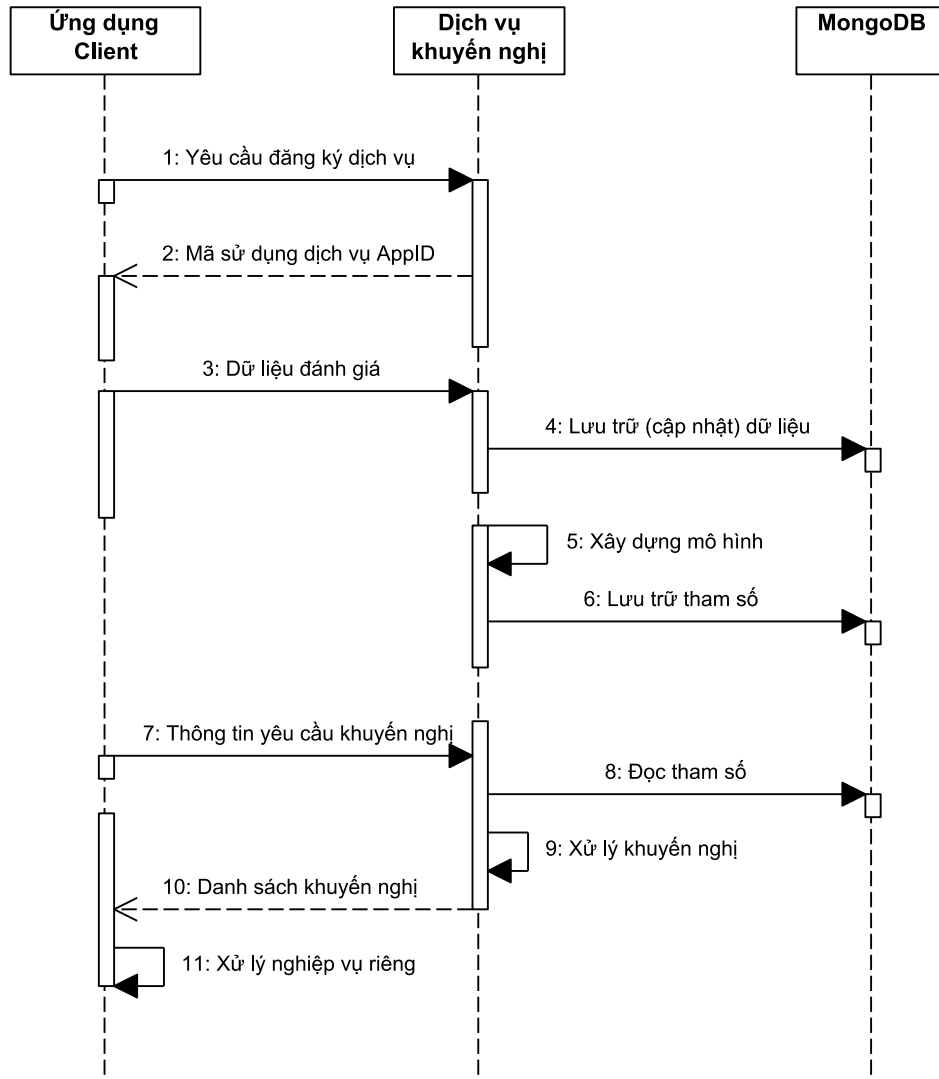
¹⁸ Collection là cấu trúc lưu trữ dữ liệu trong MongoDB, tương tự như một bảng trong các hệ quản trị cơ sở dữ liệu quan hệ. Chi tiết <http://docs.mongodb.org/manual/reference/glossary/#term-collection>, truy cập lần cuối 22/02/2013.

Trình tự hoạt động giữa dịch vụ khuyến nghị và các ứng dụng client được gom nhóm theo từng chức năng như sau:

- **Cung cấp mã sử dụng dịch vụ**
 - (1) Ứng dụng Client gửi yêu cầu đăng ký đến dịch vụ.
 - (2) Dịch vụ cung cấp cho client appid (mã ứng dụng) để sử dụng dịch vụ.
- **Lưu trữ thông tin dữ liệu đánh giá**
 - (3) Client gửi yêu cầu (request) dữ liệu đánh giá (định kỳ) dưới dạng JSON thông qua phương thức POST giao thức HTTP.
 - (4) Dịch vụ lưu trữ (cập nhật) dữ liệu mới của ứng dụng và gửi thông báo xác nhận đến ứng dụng client khi đã cập nhật dữ liệu thành công.
- **Xây dựng mô hình**
 - (5) Dịch vụ khuyến nghị xây dựng mô hình khuyến nghị cho từng ứng dụng client định kỳ.
 - (6) Lưu trữ lại các tham số.
- **Xử lý và đưa ra danh sách khuyến nghị**
 - (7) Client gửi yêu cầu khuyến nghị với các tham số là định danh người dùng và các định danh ngữ cảnh thông qua phương thức POST trong giao thức HTTP dưới định dạng JSON.
 - (8) Dịch vụ khuyến nghị đọc và xử lý giá trị các tham số đã được xây dựng trước đó.
 - (9) Dịch vụ khuyến nghị xử lý yêu cầu của ứng dụng xây dựng danh sách các sản phẩm khuyến nghị cùng với giá trị dự đoán đánh giá.
 - (10) Trả về danh sách khuyến nghị cho ứng dụng Client định dạng JSON được định nghĩa trước.
 - (11) Ứng dụng Client xử lý các nghiệp vụ phù hợp với nhu cầu của ứng dụng dựa trên kết quả danh sách sản phẩm từ dịch vụ.

Chú ý rằng, trong lần đầu tiên sử dụng dịch vụ, các chức năng (i) – (iv) phải được thực hiện tuần tự đối với từng ứng dụng. Trong những lần sử dụng tiếp theo, chức năng (ii), (iii), (iv) sẽ không bắt buộc thực hiện tuần tự. **Hình 5.3** minh họa sơ đồ

tuần tự các tương tác giữa dịch vụ khuyến nghị (Recommender Service) và ứng dụng client (Client Application).



Hình 5.3: Sơ đồ trình tự tương tác giữa ứng dụng client và dịch vụ khuyến nghị

5.2 Các ứng dụng tương tác với dịch vụ khuyến nghị dựa trên ngữ cảnh

5.2.1 Hệ thống nghe nhạc HMusic

Website nghe nhạc HMusic đưa ra các khuyến nghị cho người dùng dựa trên tâm trạng và thời gian, website bao gồm ba thành phần là người dùng, sản phẩm và ngữ cảnh.

Người dùng là tác nhân đưa ra các đánh giá cho sản phẩm cũng như là đối tượng chính mà hệ thống khuyến nghị hướng đến. Người dùng bao gồm hai loại là người

dùng đã đăng ký thành viên và người dùng không đăng ký thành viên. Tuy nhiên, để đánh giá cho sản phẩm và để nhận được những khuyến nghị phù hợp, người dùng phải đăng nhập vào hệ thống. Trong hệ thống này, người dùng đăng nhập sẽ được khuyến nghị thông qua giải thuật khuyến nghị dựa trên lọc cộng tác trong khi người dùng không đăng nhập sẽ được khuyến nghị dựa trên những thống kê cơ bản như giá trị trung bình các đánh giá hay số lượt nghe.

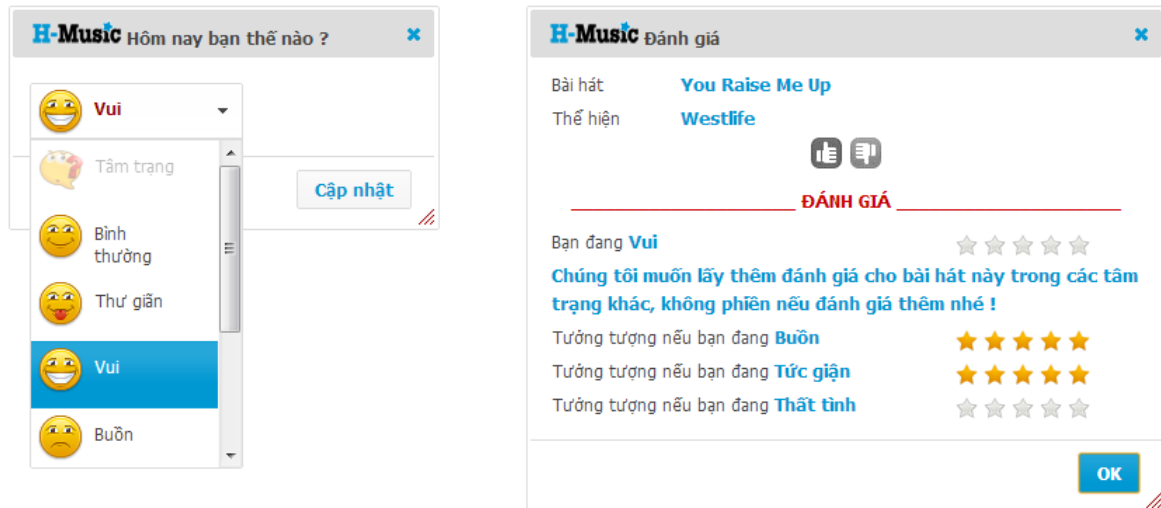
Sản phẩm là đối tượng được dùng để khuyến nghị cho người dùng. Trong hệ thống nghe nhạc HMusic, đối tượng cụ thể là các bài hát, album. Mặc dù mỗi loại sản phẩm có những thuộc tính khác nhau (thể loại, ca sĩ, ngôn ngữ), tuy nhiên trong hệ thống này, những thuộc tính đó không tham gia vào quá trình khuyến nghị.

Ngữ cảnh trong website nghe nhạc bao gồm hai yếu tố là thời gian nghe nhạc và tâm trạng người nghe và được thể hiện theo khía cạnh đại diện (representational view) (xem 2.2.1). Ngữ cảnh thời gian với miền giá trị là {sáng, trưa, chiều, tối, khuya}. Trong khi đó, ngữ cảnh tâm trạng với miền giá trị là {buồn, vui, thất tình, đang yêu, tức giận, thư giãn, bình thường}.

Bảng 5.1: Ngữ cảnh trong hệ thống nghe nhạc HMusic

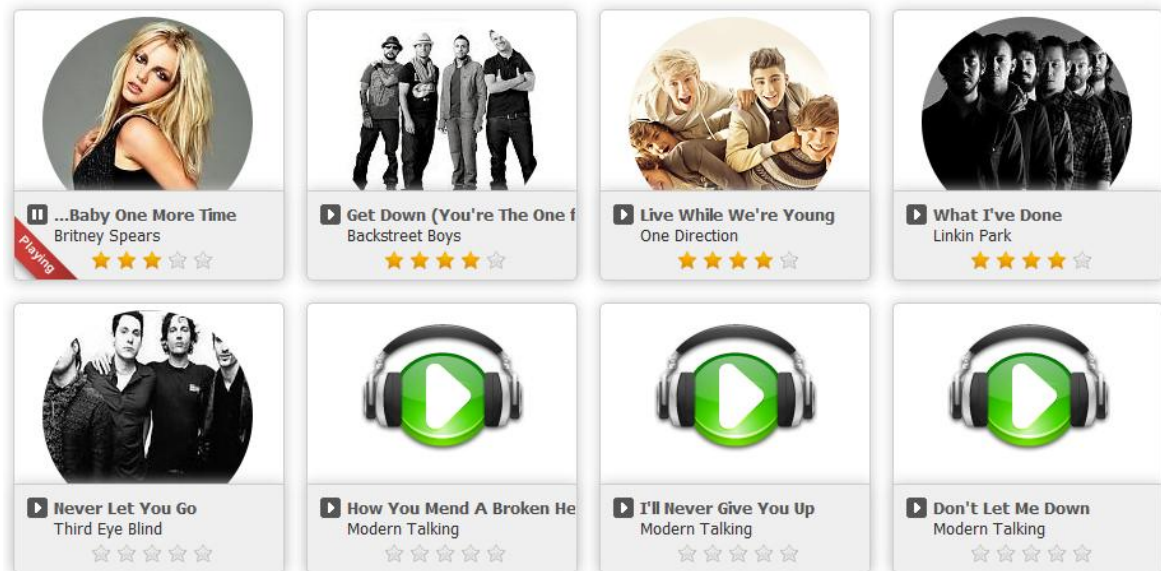
STT	Ngữ cảnh	Giá trị
1	Thời gian	Sáng, Trưa, Chiều, Tối, Khuya.
2	Tâm trạng	Bình thường, thư giãn, vui, đang yêu, buồn, thất tình, tức giận.

Thông tin đánh giá được thu thập trực tiếp từ người dùng với thang đánh giá 1-5 thể hiện sự yêu thích của người dùng trên một bài hát trong ngữ cảnh cụ thể. Ngữ cảnh tâm trạng được thu thập một cách tường minh bằng cách yêu cầu người dùng lựa chọn những tâm trạng phù hợp nhất với người dùng đó trong khi đó ngữ cảnh thời gian được thu thập tiềm ẩn thông qua nhãn thời gian mà người dùng nghe nhạc và đánh giá.



(a) Lựa chọn ngữ cảnh

(b) Thu thập đánh giá khi nghe một bài hát



(c) Thu thập đánh giá khi nghe danh sách bài hát

Hình 5.4: Thu thập ngữ cảnh và đánh giá trong hệ thống HMusic

Khuyến nghị âm nhạc cho người dùng bao gồm khuyến nghị bài hát, khuyến nghị album. Trong hệ thống này những bài hát và album sẽ được khuyến nghị theo nhóm ngôn ngữ, tức là người dùng sẽ nghe một loại ngôn ngữ nhạc và các bài hát, album thuộc cùng ngôn ngữ đó sẽ được khuyến nghị. Cụ thể như sau:

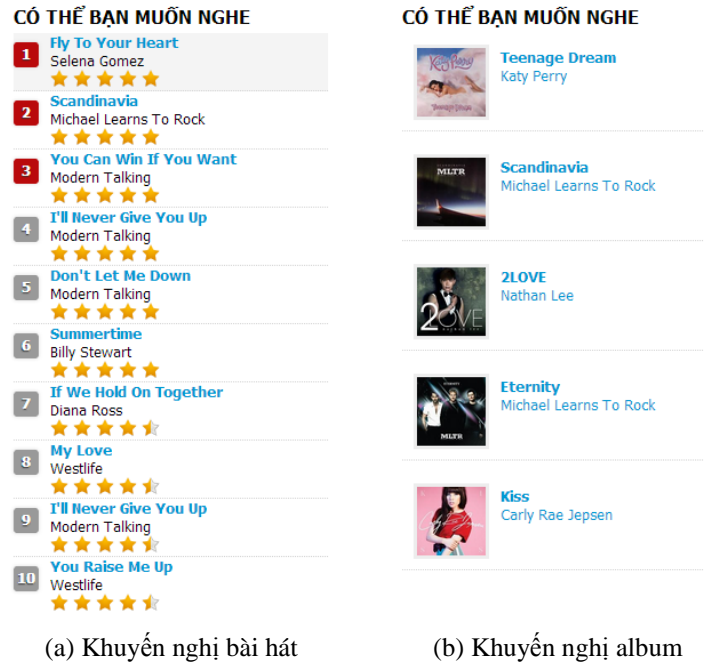
Đối với nhóm người dùng đăng nhập vào hệ thống:

- *Khuyến nghị 1 – Khuyến nghị bài hát:* Với ngữ cảnh cụ thể, hệ thống sẽ áp dụng thuật toán khuyến nghị *CF* để đưa ra 10-20 bài hát có giá trị dự đoán đánh giá cao nhất và xếp hạng giảm dần và khuyến nghị cho người dùng.

- *Khuyến nghị 2 – Khuyến nghị album*: Hệ thống thực hiện dự đoán những bài hát trong album sau đó giá trị đánh giá của album được tính trung bình trong ngữ cảnh cụ thể và khuyến nghị 5 album có giá trị trung bình dự đoán cao nhất.

Đối với nhóm người dùng không đăng nhập:

- *Khuyến nghị 1 – Khuyến nghị bài hát*: Hệ thống sẽ áp dụng thống kê tính giá trị trung bình của các bài hát theo tâm trạng, thời gian và đưa ra số lượng khuyến nghị 10-20 bài hát xếp hạng giảm dần cho người dùng.
- *Khuyến nghị 2 – Khuyến nghị album*: Hệ thống tính giá trị trung bình của từng bài hát trong album theo từng ngữ cảnh, sau đó giá trị dự đoán cho album sẽ tính trung bình, xếp hạng và đưa ra 5 khuyến nghị cho người dùng.



Hình 5.5: Khuyến nghị cho người dùng trên website hmusic.us

Giải thuật khuyến nghị *CACF* được cung cấp bởi dịch vụ khuyến nghị khi so sánh với các giải thuật MF, CAMF và ISMF¹⁹. Giải thuật ISMF cho kết quả tốt nhất với các chỉ số *MAE*, *Precision* và *Recall* (*Top-K Recommendation*, $k=10$), trong khi đó cho kết quả *RMSE* chênh lệch không đáng kể 0.04. Kết quả minh họa bởi **Bảng 5.2**.

¹⁹ ISMF và CAMF được áp dụng với trường hợp xử lý ngữ cảnh với TH₁ vì cho kết quả độ lỗi MAE, RMSE hơn so với trường hợp TH₃ (**Bảng 4.10**)

Bảng 5.2: Bảng so sánh giải thuật MF, ISMF và CAMF trong ứng dụng HMusic

Chỉ số đánh giá	MF	ISMF	CAMF
MAE	0.807	0.806	0.844
RMSE	1.051	1.055	1.121
Precision	0.02249	0.02305	0.02202
Recall	0.10841	0.11489	0.09398

5.2.2 Ứng dụng khuyến nghị quà tặng HGift

Khác với hệ thống website nghe nhạc HMusic, quy mô của ứng dụng HGift nhỏ hơn và là một ứng dụng được tích hợp trong mạng xã hội Facebook²⁰. Ứng dụng cũng bao gồm ba thành phần chính là người dùng, sản phẩm (quà tặng) và ngữ cảnh.

Người dùng trong ứng dụng này là các tài khoản Facebook, vì thế để sử dụng ứng dụng, người dùng phải đăng nhập vào tài khoản Facebook. Điều đó đồng nghĩa với việc người dùng luôn trong trạng thái đăng nhập và giải thuật khuyến nghị dựa trên lọc cộng tác *CACF* sẽ được áp dụng.

Sản phẩm là các mặt hàng phổ biến thường được sử dụng để làm quà tặng. Trong ứng dụng này, số lượng quà tặng giới hạn ở 20 quà tặng.

Ngữ cảnh bao gồm bốn thông tin là đối tượng tặng quà, độ tuổi, giới tính người nhận và lý do tặng quà được trình bày trong **Bảng 5.3**.

Bảng 5.3: Ngữ cảnh trong ứng dụng HGift

STT	Ngữ cảnh	Giá trị
1	Đối tượng	Bạn bè, người yêu, đồng nghiệp, cấp trên, người thân, trẻ em.
2	Giới tính	Nam, Nữ
3	Độ tuổi	Nhỏ hơn 15, từ 15 – nhỏ hơn 18, 18 – 25, lớn hơn 25.
4	Lý do	Sinh nhật, tốt nghiệp, tết, giáng sinh, cưới hỏi, valentine, ngày phụ nữ, đi làm, thăng chức, sức khỏe, lý do khác.

Thu thập đánh giá: Ngữ cảnh sau khi được thu thập trực tiếp từ bảng lựa chọn, các sản phẩm sẽ được đưa ra dựa trên giá trị trung bình đánh giá của tất cả người dùng xếp hạng giảm dần. Sau đó, người dùng đánh giá sản phẩm với giá trị từ 1-5 tương ứng với mức độ phù hợp của quà tặng.

²⁰ <http://facebook.com>, truy cập lần cuối 25/02/2013

Thông tin tặng quà

1. Tặng cho ai?

- ☒ Bạn bè
- ☐ Người yêu
- ☐ Đồng nghiệp
- ☐ Cấp trên
- ☐ Người thân
- ☐ Trẻ em
- ☐ Khác

2. Giới tính

- ☒ Nữ
- ☐ Nam

3. Độ tuổi

- ☐ Nhỏ hơn 15 tuổi
- ☐ Từ 15 đến nhỏ hơn 18 tuổi
- ☒ Từ 18 đến 25 tuổi
- ☐ Lớn hơn 25

4. Lý do

- ☒ Sinh nhật
- ☐ Tốt nghiệp
- ☐ Tết
- ☐ Cưới hỏi
- ☐ Valentine
- ☐ Ngày quốc tế Phụ Nữ
- ☐ Thăng chức
- ☐ Giáng sinh
- ☐ Khác

(a) Thu thập thông tin ngữ cảnh



(b) Thu thập đánh giá

Hình 5.6: Thu thập ngữ cảnh và đánh giá trong ứng dụng HGift

Khuyến nghị cho người dùng 5-10 sản phẩm quà tặng có giá trị dự đoán đánh giá cao từ giải thuật khuyến nghị. Việc lựa chọn giải thuật khuyến nghị dựa trên dữ liệu đánh giá được thu thập trước đó, giải thuật CAMF được áp dụng thông qua dịch vụ khuyến nghị khi so sánh với giải thuật ISMF và khuyến nghị 2D MF. Đây là những giải thuật cho giải quyết tốt các vấn đề về dữ liệu thưa, người dùng mới, sản phẩm mới và khả

năng đáp ứng dữ liệu lớn. Kết quả được minh họa bởi **Bảng 5.4** cho thấy giải thuật CAMF cho kết quả chỉ số MAE, RMSE, Precision, Recall (Top-K Recommendation với $k=5$) tốt hơn so với ISMF và MF.

Bảng 5.4: Bảng so sánh giải thuật MF, ISMF và CAMF trong ứng dụng HGift

Chỉ số đánh giá	MF	ISMF	CAMF
MAE	1.242	1.188	1.057
RMSE	1.482	1.445	1.322
Precision	0.1332	0.1347	0.1599
Recall	0.5584	0.5607	0.6672

5.3 Kết luận chung

Ứng dụng HMusic và HGift tương tác với dịch vụ khuyến nghị sử dụng các thuật toán ISMF và CAMF nhìn chung cho kết quả tốt hơn so với sử dụng khuyến nghị truyền thống (giải thuật MF). Điều này cho thấy yếu tố ngữ cảnh trong các ứng dụng thực sự ảnh hưởng đến hành vi người dùng.

Bên cạnh đó, dữ liệu thu thập từ hai ứng dụng tương đối hạn chế, cụ thể chỉ 1477 đánh giá từ ứng dụng HGift và 1480 đánh giá từ ứng dụng HMusic và thu thập trong một phạm vi người dùng hạn chế. Điều này cũng ảnh hưởng rất nhiều đến mức độ hiệu quả của giải thuật. Cụ thể, đối với ứng dụng HMusic, giải thuật *ISMF* mặc dù cho kết quả *MAE* tốt hơn MF **0.12%**, kết quả *RMSE* kém hiệu quả hơn MF **0.38%** tuy nhiên cho kết quả khuyến nghị với *Precision* cải thiện **2.6%** và *Recall* cải thiện **5.7%** so với giải thuật MF (**Bảng 5.3**). Điều này phản ánh ngữ cảnh trong ứng dụng HMusic chưa thực sự tác động mạnh mẽ đến hành vi của người dùng. Tương tự, ứng dụng tư vấn quà tặng HGift, giải thuật *CAMF* hiệu quả hơn giải thuật MF với **14.89% MAE**, **10.79% RMSE**, **16.69% Precision** và **16.3% Recall** (**Bảng 5.4**). Điều này phản ánh mức độ tác động thực sự của ngữ cảnh trong ứng dụng này.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Để kết thúc khóa luận tốt nghiệp đại học “Xây dựng hệ khuyến nghị dựa trên ngữ cảnh”, phần cuối cùng của báo cáo này sẽ trình bày một cách tóm tắt những kết quả, nội dung đã đạt được trong quá trình thực hiện. Mặt khác cũng đưa ra những khó khăn, hạn chế bên cạnh những vấn đề định hướng và phát triển cho đề tài trong tương lai.

6.1 Kết quả

Khóa luận đã đạt được những kết quả sau:

- *Nội dung lý thuyết*
 - Khái quát những vấn đề chính của hệ khuyến nghị bao gồm khái niệm, chức năng và các vấn đề trong khuyến nghị.
 - Tìm hiểu về ngữ cảnh: định nghĩa, phân loại, các phương pháp thu thập ngữ cảnh, biểu diễn ngữ cảnh trong hệ khuyến nghị và các kỹ thuật giải quyết.
 - Lọc cộng tác, các vấn đề về lọc cộng tác cùng với các tiêu chí đánh giá trong lọc cộng tác và phần mở rộng của lọc cộng tác kết hợp với yếu tố ngữ cảnh.
 - Các phương pháp phân chia dữ liệu, đánh giá trong hệ khuyến nghị và cách tính các chỉ số đánh giá.
- *Ứng dụng minh họa*
 - Xây dựng được hai ứng dụng thu thập dữ liệu đánh giá kết hợp với yếu tố ngữ cảnh là ứng dụng tặng quà trên facebook HGift và website nghe nhạc HMusic với số lượng ngữ cảnh khác nhau.

- Hiện thực chương trình cài đặt các thuật toán trong lọc cộng tác, cài đặt các kiểm thử thuật toán thông qua các chỉ số đánh giá MAE, RMSE, Precision, Recall với phương pháp kiểm thử *m-fold cross validation* thực hiện trên nhiều bộ dữ liệu khác nhau. Các thuật toán được xây dựng trong quá trình kiểm thử bao gồm:
 - Giải thuật User-Based, Matrix Factorization (SVD) 2D.
 - Giải thuật EPF, Item splitting trong kỹ thuật Pre-Filtering.
 - Giải thuật Weigh POF, Filter POF trong kỹ thuật Post-Filtering.
 - Giải thuật Contextual Neighbors, CAMF trong kỹ thuật Contextual Modeling.
- Số lượng dữ liệu được sử dụng trong khóa luận là 5 bộ: Food, Movielens Semi-Synthetic, Comoda, HGift và HMusic.
- Xây dựng thành công trình chủ khuyến nghị (Recommender Server) dưới dạng Web Service dựa trên kiến trúc RESTful và được áp dụng trong hai ứng dụng:
 - Website nghe nhạc dựa trên tâm trạng và thời gian HMusic.
 - Ứng dụng facebook tư vấn quà tặng HGift.
- Về mặt công nghệ: tìm hiểu và xây dựng ứng dụng bằng ngôn ngữ PHP với framework Codeigniter²¹. Thực hiện triển khai RESTful Web services trên nền tảng Java và hệ quản trị cơ sở dữ liệu NoSQL MongoDB²².

6.2 Khó khăn và hạn chế

Trong quá trình thực hiện đề tài, bên cạnh những mặt đạt được, những khó khăn và hạn chế trong quá trình thực hiện gặp phải như sau:

- **Khó khăn**
 - Khó khăn lớn nhất trong quá trình thực hiện khóa luận là việc tìm kiếm bộ dữ liệu kiểm thử có đầy đủ thông tin ngữ cảnh, đa số những bộ dữ liệu có chứa dữ

²¹ <http://www.ellislab.com/codeigniter>, truy cập lần cuối 22/02/2013

²² <http://www.mongodb.org/>, truy cập lần cuối 22/02/2013

liệu ngữ cảnh là dữ liệu được sử dụng nội bộ trong các phòng ban, viện nghiên cứu ở nước ngoài.

- Việc thu thập dữ liệu thực tế chỉ thực hiện trên một số lượng ít người dùng và việc đưa ra giá trị đánh giá còn mang tính chất ngẫu nhiên trên một số người dùng không thực sự quan tâm đến tính chính xác của dữ liệu thu thập.
- Quá trình tìm hiểu các phương pháp khuyến nghị dựa trên ngữ cảnh cùng với số lượng giải thuật lớn và việc tìm hiểu những phương pháp tính toán kiểm thử, tiêu chí đánh giá giải thuật tốn rất nhiều thời gian, đặc biệt là chỉ số *Precision* và *Recall* vì chỉ số này tồn tại rất nhiều cách tính và phương pháp đánh giá khác nhau.
- Bên cạnh đó, khó khăn trong nguồn tài liệu tham khảo cũng như thời gian hạn chế cũng ảnh hưởng đến quá trình thực hiện khóa luận tốt nghiệp.

- **Hạn chế**

- Dịch vụ khuyến nghị dựa trên ngữ cảnh chỉ dừng lại ở mức minh họa và số lượng ứng dụng sử dụng dịch vụ còn hạn chế. Một số vấn đề về việc bảo mật cho dịch vụ chưa được chú trọng triển khai.
- Số lượng ngữ cảnh trong các tập dữ liệu kiểm thử cũng như trong ứng dụng minh họa còn hạn chế ở một số lượng nhỏ, cụ thể đối với ứng dụng minh họa website nghe nhạc, số lượng ngữ cảnh bao gồm {thời gian, tâm trạng} của người nghe, ứng dụng HGift với bốn thông tin ngữ cảnh là {đối tượng, giới tính, độ tuổi, lý do} tặng quà.
- Khóa luận chưa đưa ra được nhiều sáng tạo, cải tiến hơn việc tổng hợp, nghiên cứu từ nhiều nguồn tài liệu khác nhau.

6.3 Hướng phát triển

- Nghiên cứu, phát triển một số thuật toán về xử lý ngữ cảnh liên quan cũng như các thuật toán kết hợp ngữ cảnh trong lọc cộng tác để nâng cao hiệu suất dự đoán và khuyến nghị.
- Thực hiện thu thập một số lượng lớn thông tin đánh giá từ người dùng cũng như mở rộng số lượng thông tin ngữ cảnh trong các ứng dụng.

- Mở rộng việc phát triển ứng dụng có thể tích hợp trong môi trường di động, tối ưu thời gian xử lý khuyến nghị.
- Mở rộng dịch vụ khuyến nghị với nhiều thuật toán và phát triển khả năng bảo mật cho dịch vụ.
- Xây dựng một bộ dữ liệu ngữ cảnh hoàn chỉnh phục vụ cho công tác nghiên cứu với số lượng người dùng, sản phẩm và ngữ cảnh lớn. Đặc biệt, dữ liệu cần có tính chính xác trong việc đưa ra giá trị dự đoán cũng như tăng số lượng sản phẩm đồng đánh giá.

6.4 Tổng kết

Khóa luận tốt nghiệp đại học với đề tài “Xây dựng hệ khuyến nghị dựa trên ngữ cảnh” đã thể hiện được một bức tranh toàn cảnh về hệ khuyến nghị nói chung cũng như hệ khuyến nghị dựa trên ngữ cảnh nói riêng. Khóa luận cũng đã thể hiện được ngữ cảnh là gì với những phân tích về nhiều định nghĩa khác nhau, những phương pháp xử lý ngữ cảnh trong hệ khuyến nghị cũng như các giải thuật liên quan trong lọc cộng tác. Mặt khác, việc thực hiện so sánh các giải thuật trong các dữ liệu ngữ cảnh khác nhau cũng cho thấy được những ưu nhược điểm chung của từng giải thuật, đó chính là cơ sở để giúp cho các nhà thiết kế và xây dựng hệ thống có thể lựa chọn cho những ứng dụng của mình những giải thuật phù hợp nhất và đạt hiệu quả cao nhất. Qua đó cũng nhận thấy các giải thuật không hoàn toàn là “tốt” trong mọi trường hợp, việc nhìn nhận ưu nhược điểm chỉ có thể đưa ra sự lựa chọn giải thuật “phù hợp” cho từng ứng dụng.

Do thời gian thực hiện khóa luận hạn chế nên khóa luận chỉ nêu ra được một số vấn đề chính trong hệ khuyến nghị dựa trên ngữ cảnh. Các vấn đề còn lại chưa được trình bày trong khóa luận sẽ là những đề tài định hướng và tiếp tục được nghiên cứu trong tương lai.

Tài liệu tham khảo

- [1] L. Baltrunas, “Context-Aware Collaborative Filtering Recommender Systems,” University of Bozen-Bolzano, Italy, 2011.
- [2] A. K. Dey, “Understanding and Using Context,” *Personal Ubiquitous Comput.*, vol. 5, no. 1, pp. 4–7, Jan. 2001.
- [3] G. Adomavicius and A. Tuzhilin, “Context-aware recommender systems,” in *Proceedings of the 2008 ACM conference on Recommender systems*, 2008, pp. 335–336.
- [4] X. Su and T. M. Khoshgoftaar, “A survey of collaborative filtering techniques,” *Adv. in Artif. Intell.*, vol. 2009, pp. 4:2–4:2, Jan. 2009.
- [5] U. Panniello, A. Tuzhilin, M. Gorgoglione, C. Palmisano, and A. Pedone, “Experimental comparison of pre- vs. post-filtering approaches in context-aware recommender systems,” in *Proceedings of the third ACM conference on Recommender systems RecSys 09*, 2009, p. 265.
- [6] U. Panniello and M. Gorgoglione, “Incorporating context into recommender systems: an empirical comparison of context-based approaches,” *Electronic Commerce Research*, vol. 12, no. 1, pp. 1–30, 2012.
- [7] L. Baltrunas and F. Ricci, “Context-based splitting of item ratings in collaborative filtering,” in *Proceedings of the third ACM conference on Recommender systems*, 2009, pp. 245–248.
- [8] L. Baltrunas, B. Ludwig, and F. Ricci, “Matrix factorization techniques for context aware recommendation,” in *Proceedings of the fifth ACM conference on Recommender systems*, 2011, pp. 301–304.
- [9] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, *Recommender Systems Handbook*, 1st ed. New York, NY, USA: Springer-Verlag New York, Inc., 2010.
- [10] N. Webster and J. L. McKechnie, *Webster’s New Twentieth Century Dictionary of the English Language, Unabridged: Based Upon the Broad Foundations Laid Down by Noah Webster*. Simon and Schuster, 1983.
- [11] M. Bazire and P. Brézillon, “Understanding context before using it,” in *Proceedings of the 5th international conference on Modeling and Using Context*, 2005, pp. 29–40.
- [12] B. Schilit and M. Theimer, “Disseminating Active Map Information to Mobile Hosts,” *IEEE Network*, vol. 8, pp. 22–32, 1994.

- [13] B. Schilit, N. Adams, and R. Want, "Context-Aware Computing Applications," in *Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications*, 1994, pp. 85–90.
- [14] P. Dourish, "What we talk about when we talk about context," *Personal Ubiquitous Comput.*, vol. 8, no. 1, pp. 19–30, Feb. 2004.
- [15] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin, "Incorporating contextual information in recommender systems using a multidimensional approach," *ACM Transactions on Information Systems*, vol. 23, no. 1, pp. 103–145, 2005.
- [16] L. Liu, F. Lecue, N. Mehandjiev, and L. Xu, "Using Context Similarity for Service Recommendation," in *Proceedings of the 2010 IEEE Fourth International Conference on Semantic Computing*, 2010, pp. 277–284.
- [17] A. Odić, M. Tkalčič, J. F. Tasič, and A. Košir, "Relevant Context in a Movie Recommender System: Users' Opinion vs. Statistical Detection," in *Proceedings of the 4th Workshop on Context-Aware Recommender Systems in conjunction with the 6th ACM Conference on Recommender Systems (RecSys 2012)*, 2012, vol. 889.
- [18] L. Baltrunas, B. Ludwig, S. Peer, and F. Ricci, "Context relevance assessment and exploitation in mobile recommender systems," *Personal Ubiquitous Comput.*, vol. 16, no. 5, pp. 507–526, Jun. 2012.
- [19] C. Ono, Y. Takishima, Y. Motomura, and H. Asoh, "Context-Aware Preference Model Based on a Study of Difference between Real and Supposed Situation Data," in *Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization: formerly UM and AH*, 2009, pp. 102–113.
- [20] B. De Carolis, I. Mazzotta, N. Novielli, and V. Silvestri, "Using common sense in providing personalized recommendations in the tourism domain," *Workshop on Context-Aware Recommender Systems (CARS 2009)*, New York, 2009.
- [21] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, "Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering," in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 79–86.
- [22] Y. Shi, M. Larson, and A. Hanjalic, "Mining mood-specific movie similarity with matrix factorization for context-aware recommendation," in *Proceedings of the Workshop on Context-Aware Movie Recommendation*, 2010, pp. 34–40.
- [23] X. Su and T. M. Khoshgoftaar, "Collaborative Filtering for Multi-class Data Using Belief Nets Algorithms," in *Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence*, 2006, pp. 497–504.
- [24] K. Miyahara and M. J. Pazzani, "Collaborative filtering with the simple Bayesian classifier," in *Proceedings of the 6th Pacific Rim international conference on Artificial intelligence*, 2000, pp. 679–689.

- [25] L. Ungar, D. Foster, E. Andre, S. Wars, F. S. Wars, D. S. Wars, and J. H. Whispers, “Clustering Methods for Collaborative Filtering,” 1998.
- [26] T. Hofmann, “Latent semantic models for collaborative filtering,” *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 89–115, Jan. 2004.
- [27] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, “Eigentaste: A Constant Time Collaborative Filtering Algorithm,” *Inf. Retr.*, vol. 4, no. 2, pp. 133–151, Jul. 2001.
- [28] P. Melville, R. J. Mooney, and R. Nagarajan, “Content-boosted collaborative filtering for improved recommendations,” in *Eighteenth national conference on Artificial intelligence*, 2002, pp. 187–192.
- [29] D. Pavlov, E. Manavoglu, D. M. Pennock, and C. L. Giles, “Collaborative Filtering with Maximum Entropy,” *IEEE Intelligent Systems*, vol. 19, no. 6, pp. 40–48, Nov. 2004.
- [30] H. Ahn, K. Kim, and I. Han, “Mobile advertisement recommender system using collaborative filtering: MAR-CF,” in *The Korea Society of Management Information Systems*, 2006, vol. 0, pp. 709–715.
- [31] L. Baltrunas, M. Kaminskas, B. Ludwig, O. Moling, F. Ricci, A. Aydin, K.-H. Lüke, and R. Schwaiger, “InCarMusic: Context-Aware Music Recommendations in a Car E-Commerce and Web Technologies,” vol. 85, C. Huemer, T. Setzer, W. Aalst, J. Mylopoulos, N. M. Sadeh, M. J. Shaw, and C. Szyperski, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 89–100.
- [32] A. Victor CODIN and I. Luigi CECCARON, “Extending Recommendation Systems with Semantics and Context-Awareness: Pre-Filtering Algorithms,” in *Proceedings of the 13th International Conference of the Catalan Association for Artificial Intelligence*, 2011.
- [33] S. S. Anand and B. Mobasher, “From Web to Social Web: Discovering and Deploying User and Content Profiles,” B. Berendt, A. Hotho, D. Mladenic, and G. Semeraro, Eds. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 142–160.
- [34] V. C. Ostuni, T. Di Noia, R. Mirizzi, D. Romito, and E. Di Sciascio, “Cinemappy: a Context-aware Mobile App for Movie Recommendations boosted by DBpedia,” in *1st International Workshop on Semantic Technologies meet Recommender Systems & Big Data (SeRSy 2012)*, 2012.
- [35] A. Gunawardana and G. Shani, “A Survey of Accuracy Evaluation Metrics of Recommendation Tasks,” *J. Mach. Learn. Res.*, vol. 10, pp. 2935–2962, Dec. 2009.
- [36] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, “Evaluating collaborative filtering recommender systems,” *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 5–53, Jan. 2004.

- [37] P. Cremonesi, Y. Koren, and R. Turrin, “Performance of recommender algorithms on top-n recommendation tasks,” in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 39–46.
- [38] T. Zhang, “Solving large scale linear prediction problems using stochastic gradient descent algorithms,” in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 116–.
- [39] Y. Koren, R. Bell, and C. Volinsky, “Matrix Factorization Techniques for Recommender Systems,” *Computer*, vol. 42, no. 8, pp. 30–37, Aug. 2009.
- [40] G. W. Furnas, S. Deerwester, S. T. Dumais, T. K. Landauer, R. A. Harshman, L. A. Streeter, and K. E. Lochbaum, “Information retrieval using a singular value decomposition model of latent semantic structure,” in *Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval*, 1988, pp. 465–480.
- [41] J. R. Quinlan, *C4.5: programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.
- [42] G. Hamerly and C. Elkan, “Learning the K in K-Means,” in *In Neural Information Processing Systems*, 2003, p. 2003.
- [43] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, “Top 10 algorithms in data mining,” *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, Dec. 2007.
- [44] P. Cremonesi, R. Turrin, E. Lentini, and M. Matteucci, “An Evaluation Methodology for Collaborative Recommender Systems,” in *Proceedings of the 2008 International Conference on Automated solutions for Cross Media Content and Multi-channel Distribution*, 2008, pp. 224–231.
- [45] A. Kosir, A. Odic, M. Kunaver, M. Tkalcic, and J. F. Tasic, “Database for contextual personalization,” *Elektrotehnicki Vestnik* 78, vol. 5, pp. 270–274, 2011.
- [46] R. T. Fielding, “Architectural styles and the design of network-based software architectures,” University of California, Irvine, 2000.

Phụ lục A. Giá trị tham số trong các giải thuật

Bảng A.1: Bảng giá trị tham số sử dụng trong các thuật toán trên các dữ liệu

Giải thuật	Tham số	Food	Movielens	Comoda	HGift	HMusic
UB ISUB EPF-UB	Số người dùng tương tự	120	650	80	–	–
CN	Số cặp <i>profile</i> tương tự	400	900	200	–	–
Weigh UB Filter UB	Số người dùng tương tự trong ngữ cảnh	160	700	–	–	–
	P^*	0.1	0.1	–	–	–
MF ISMF EPF-MF CAMF	Nhân tố f	10	10	10	10	10
	γ	0.001	0.001	0.001	0.001	0.001
	λ	0.05	0.05	0.05	0.05	0.05
	Số lần lặp	1000	1000	1000	1000	1000
Weigh MF Filter MF	Số người dùng tương tự trong ngữ cảnh	160	700	–	–	–
	P^*	0.1	0.1	–	–	–