

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
KHOA CÔNG NGHỆ THÔNG TIN**

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC



**MÔ HÌNH KẾT HỢP HÀNH VI
ĐÁNH GIÁ VÀ BÌNH LUẬN CHO
TƯ VẤN KHÁCH SẠN**

VŨ QUANG SƠN

KỸ SƯ NGÀNH HỆ THỐNG THÔNG TIN

HÀ NỘI, NĂM 2021

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
KHOA CÔNG NGHỆ THÔNG TIN**

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC



**MÔ HÌNH KẾT HỢP HÀNH VI
ĐÁNH GIÁ VÀ BÌNH LUẬN CHO
TƯ VẤN KHÁCH SẠN**

VŨ QUANG SƠN - B17DCCN545

**GIẢNG VIÊN HƯỚNG DẪN
TRẦN ĐÌNH QUÊ**

HÀ NỘI, NĂM 2021

Lời cảm ơn

Lời đầu tiên, em xin gửi lời cảm ơn chân thành tới tất cả thầy cô đang giảng dạy trong mái trường Học viện Công nghệ Bưu chính Viễn thông đã tận tình truyền đạt những kinh nghiệm và kiến thức quý báu giúp em hoàn thành nhiệm vụ học tập trong suốt khoảng thời gian hơn 4 năm là sinh viên của học viện. Em xin gửi lời biết ơn sâu sắc đến thầy PGS.TS Trần Đình Quế, người đã tận tình hướng dẫn, chỉ bảo, định hướng và nhắc nhở em trong suốt quá trình học tập cũng như hoàn thành đồ án này.

Cho con gửi lời cảm ơn chân thành đến bố mẹ, ông bà, anh chị em đã luôn động viên, ủng hộ, cổ vũ và tạo điều kiện tốt nhất cho con trong suốt những năm tháng ngồi trên ghế nhà trường.

Cuối cùng, cho tôi gửi lời cảm ơn đến những người bạn, người anh, người chị của tôi, những người luôn chia sẻ, động viên, giúp đỡ và ở bên tôi mỗi khi tôi gặp khó khăn nhất!

Em xin chân thành cảm ơn!

Hà Nội, ngày 10 tháng 12 năm 2021

Sinh viên thực hiện

Vũ Quang Sơn

Mục lục

Lời cảm ơn	i
Mục lục	ii
Danh sách hình vẽ	iv
Danh sách bảng	v
Danh sách đoạn mã	vi
Danh mục từ viết tắt và tạm dịch	vii
Danh mục từ tạm dịch	ix
Tóm tắt khóa luận	1
1 Đặt vấn đề	5
2 Cơ sở lý thuyết	7
2.1 Phân loại mô hình khuyến nghị	8
2.2 Loại cộng tác	9
2.2.1 Loại cộng tác dựa trên bộ nhớ	10
3 Phương pháp giải quyết vấn đề	13
4 Thực nghiệm, kết quả, so sánh và đánh giá	15

5	Kết luận	17
	Tài liệu tham khảo	19

Danh sách hình vẽ

2.1	Tổng quan hệ thống khuyến nghị	7
2.2	Mô hình thuật toán lọc cộng tác	11

Danh sách bảng

Danh sách đoạn mã

Danh mục từ viết tắt và tạm dịch

Từ viết tắt	Tiếng Anh	Tạm dịch
CF	Collaborative Filtering	Lọc cộng tác
SVD	Singular Value Decomposition	Phân tích giá trị đơn vị đặc biệt
PCA	Principal Component Analysis	Phân tích thành phần chính
MBCF	Memory-based Collaborative Filtering	Lọc cộng tác dựa trên bộ nhớ
UBCF	User-based Collaborative Filtering	Lọc cộng tác dựa trên người dùng
IBCF	Item-based Collaborative Filtering	Lọc cộng tác dựa trên sản phẩm
KNN	K-Nearest Neighbors	K láng giềng gần nhất

Danh mục từ tạm dịch

Machine Learning	Học máy
Deep Learning	Học sâu
Reinforcement learning	Học tăng cường
Federated Learning	Học liên kết

Mở đầu

Cuộc sống của con người ngày càng phát triển, các nhu cầu cá nhân như: giao lưu, kết bạn, tiêu dùng, du lịch, ... ngày tăng. Nhu cầu tiêu dùng ngày càng tăng, cùng với sự phát triển của công nghệ thông tin, các hệ thống thương mại điện tử ra đời và ngày càng lớn mạnh, tiêu biểu như: Facebook, Youtube, Tripadvisor, ... Những trang thương mại điện tử này hỗ trợ doanh nghiệp quảng bá sản phẩm tới tay người tiêu dùng nhanh hơn so với bán hàng truyền thống. Tuy nhiên, khi người dùng được tiếp cận sản phẩm, dịch vụ một cách nhanh chóng thì họ cũng phải đối mặt với vấn đề có quá nhiều sản phẩm và dịch vụ và đâu thực sự là sản phẩm họ cần. Đây là tình trạng quá tải thông tin, khi người dùng có quá nhiều lựa chọn. Tuy nhiên, đôi khi họ cũng phải đối mặt với tình huống nghịch lý rằng có rất nhiều thông tin, nhưng thường rất khó để có thông tin phù hợp [2]. Với hiện trạng nêu trên, nhu cầu cấp thiết đặt ra cần có các hệ thống tự động hóa, hỗ trợ người dùng lọc thông tin cũng như cá nhân hóa đối với từng người dùng.

Hệ tư vấn ra đời nhằm giải quyết vấn đề quá tải thông tin từ người dùng, giúp họ khám phá những sản phẩm khác nhau nằm trong sở thích của mình. Có rất nhiều trang thương mại điện tử lớn sử dụng hệ tư vấn nhằm cải thiện doanh thu và tăng sự thân thiện với người dùng, một trong số đó là Youtube. Youtube, ra đời vào tháng 2, 2005 với sự phát triển nhanh chóng đã trở thành nền tảng chia sẻ video trực tuyến lớn nhất hiện nay với hơn 1 tỷ lượt xem mỗi ngày từ hàng triệu người dùng và mỗi phút có hơn 24 giờ thời lượng video được tải lên nền tảng này. Hệ tư vấn là một phần trong sự thành công của Youtube khi đóng góp 60% lượt bấm xem

video từ trang chủ và các video được gợi ý từ hệ thống có tỷ lệ bấm xem gấp 2 lần những video được nhiều người xem nhất và được đánh giá cao nhất [1].

Một trong các thuật toán tư vấn điển hình và phổ biến là lọc cộng tác và hoạt động rất hiệu quả. Các hệ tư vấn truyền thống thường sử dụng dữ liệu điểm đánh giá để làm cơ sở tư vấn. Tuy nhiên, theo [4], thuật toán này vẫn còn những vấn đề còn tồn tại như:

- Vấn đề người dùng mới, sản phẩm mới (Cold Start)
- Vấn đề thừa thớt dữ liệu

Do thói quen lười đánh giá từ người dùng, gây ra những vấn đề trên ảnh hưởng tới độ chính xác của hệ tư vấn lọc cộng tác.

Với sự bùng nổ của các trang thương mại điện tử, các hành vi bày tỏ quan điểm ngày càng đa dạng và phong phú. Do đó, các phương pháp phân loại văn bản ngày càng được cải thiện và trở nên chính xác hơn. Những dữ liệu văn bản này cũng mang ý nghĩa bày tỏ quan điểm đối với sản phẩm.

Để hệ tư vấn có những đề xuất chính xác hơn cũng như tận dụng dữ liệu văn bản cùng các kỹ thuật phân loại được phát triển, đề án lựa chọn đề tài "**Mô hình kết hợp hành vi đánh giá và bình luận cho tư vấn khách sạn**" với mục tiêu nghiên cứu lý thuyết về hệ tư vấn, các kỹ thuật tư vấn, tiền xử lý văn bản và phân loại văn bản về lĩnh vực cụ thể là gợi ý các khách sạn trên các bộ dữ liệu thu thập được.

Đề án được chia thành 4 chương với nội dung như sau:

Chương 1: Tổng quan về hệ tư vấn - Nội dung trong Chương 1 giới thiệu tổng quan về hệ tư vấn và các kỹ thuật lọc cộng tác. Ngoài ra, Chương 1 còn trình bày ngắn gọn các vấn đề còn tồn tại của hệ tư vấn lọc cộng tác.

Chương 2: Tư vấn dựa trên mô hình kết hợp - Trong chương này, đề án trình bày về mô hình kết hợp giữa hành vi đánh giá và hành vi bình luận và cách ứng dụng mô hình kết hợp vào hệ tư vấn lọc cộng tác.

Ngoài ra, nội dung Chương 2 còn trình bày về các kỹ thuật tiền xử lý dữ liệu văn bản cùng với 3 kỹ thuật phân loại văn bản: Naïve Bayes, Logistic Regression, SVM.

Chương 3: Thử nghiệm và đánh giá - Chương 3 tập trung trình bày về bộ dữ liệu được thử nghiệm, phương pháp thực nghiệm, bộ dữ liệu được sử dụng và kết quả thực nghiệm và đánh giá.

Chương 1

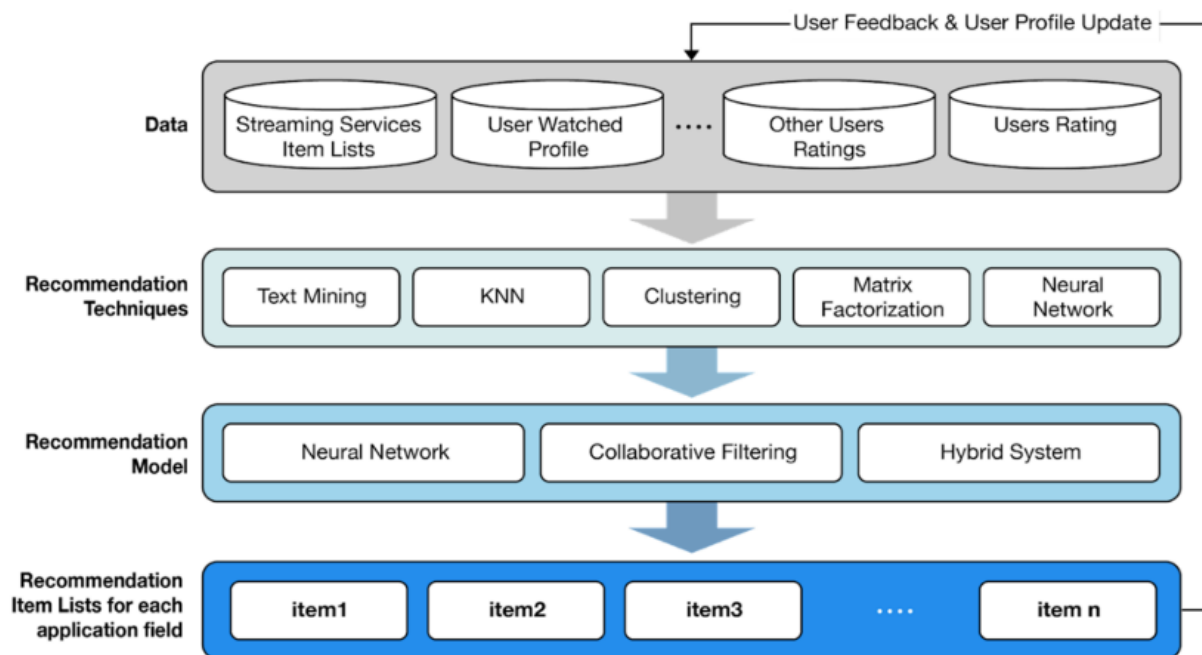
Đặt vấn đề

Trong Chương 1, đề án trình bày một cách tổng quan về hệ tư vấn. Ngoài ra, vai trò, lợi ích của hệ tư vấn đối với thương mại điện tử cũng được trình bày trong chương này. Nội dung của Chương 1 được phân chia như sau:

Chương 2

Cơ sở lý thuyết

Hệ thống khuyến nghị là 1 công nghệ hỗ trợ đắc lực cho con người, giúp phân tích lượng dữ liệu khổng lồ được cung cấp bởi người dùng. Hệ thống dự đoán điểm của các sản phẩm, tạo 1 danh sách sắp xếp thứ các sản phẩm này cho mỗi người dùng, và giới thiệu tới người dùng những sản phẩm mà họ có thể thích. Nội dung trong phần này trình bày tổng quan về hệ khuyến nghị, các mô hình thuật toán cùng với các kỹ thuật thuật khai phá dữ liệu trong hệ khuyến nghị hiện có. Hình 2.1 là tổng quan luồng



Hình 2.1: Tổng quan hệ thống khuyến nghị

2.1. Phân loại mô hình khuyến nghị

hoạt động của 1 hệ khuyến nghị, gồm có các bước xử lý: (1) thu thập dữ liệu, (2) khai phá dữ liệu, (3) mô hình hóa dữ liệu, (4) và đưa ra gợi ý. Dữ liệu sử dụng trong hệ khuyến nghị có thể là các đánh giá, bình luận về sản phẩm, danh sách sản phẩm mà người dùng theo dõi, v.v. Các kỹ thuật khai phá dữ liệu truyền thống, có thể kể đến như: phân cụm, khai phá dữ liệu văn bản, KNN hay là học sâu, sử dụng mạng nơ-ron. Tiếp đó, các mô hình khuyến nghị sử dụng các đặc trưng đã được trích chọn để có thể mô hình hóa dữ liệu, từ đó đưa ra các khuyến nghị phù hợp tới người dùng.

Nội dung trong chương này tập trung giới thiệu và phân loại một cách tổng quát về các mô hình khuyến nghị hiện nay, có thể áp dụng vào bất kỳ 1 hệ thống khuyến nghị nào.

2.1 Phân loại mô hình khuyến nghị

Các mô hình khuyến nghị có thể được chia thành 3 nhóm chính [3]:

- **Lọc dựa trên nội dung:** Trong cách tiếp cận này, hệ thống sẽ thu thập các dữ liệu rõ ràng (điểm đánh giá sản phẩm) hoặc dữ liệu ngầm (bấm vào một đường dẫn) và tạo ra hồ sơ người dùng. Hệ thống sẽ thực hiện tư vấn những sản phẩm dựa trên những sản phẩm và hành vi liên quan tới hồ sơ người dùng. Do sở thích của người dùng thường được chia thành vài nhóm cơ bản, việc chỉ sử dụng hồ sơ của 1 người dùng khiến hệ thống không tận dụng được thông tin từ những người dùng khác, từ đó hạn chế sự linh hoạt của hệ tư vấn.
- **Lọc cộng tác:** Không giống với lọc dựa trên nội dung, lọc cộng tác tìm kiếm những người dùng có sở thích tương tự nhau. Từ giả định những người dùng A có sở thích giống với người dùng B, hệ thống sẽ tiến hành tư vấn cho người dùng B những sản phẩm phù hợp người dùng A. Lọc cộng tác có 2 hướng tiếp cận: dựa trên bộ nhớ và dựa trên mô hình. Hướng tiếp cận dựa trên bộ nhớ tính toán độ tương tự giữa các người dùng từ đó thực hiện tư vấn. Nhược điểm của hướng

tiếp cận này là sự tốn kém tài nguyên khi số lượng người dùng và sản phẩm tăng lên. Hướng tiếp cận dựa trên mô hình sử dụng các mô hình đã được huấn luyện thông qua các thuật toán học máy hoặc khai phá dữ liệu để thực hiện tư vấn.

- **Hệ tư vấn lai** Lọc dựa trên nội dung và lọc cộng tác đều có ưu điểm và nhược điểm riêng. Để giải quyết vấn đề này, hệ tư vấn lai được sinh ra, là sự kết hợp của 2 kỹ thuật trên.

Trong phần tiếp theo, đề án sẽ tập trung vào việc trình bày mô hình 1 số lọc cộng tác tiêu biểu.

2.2 Lọc cộng tác

Lọc cộng tác là một mô hình lọc thông tin, xây dựng 1 cơ sở dữ liệu sở thích người dùng thông qua dữ liệu tương tác giữa họ với sản phẩm để dự đoán các sản phẩm phù hợp với sở thích của họ, từ đó đưa ra các khuyến nghị về sản phẩm. Ý tưởng của mô hình lọc cộng tác là từ dữ liệu hành vi tương tác giữa người dùng và sản phẩm, hệ thống sẽ tính toán mức độ tương đồng giữa các người dùng hoặc giữa các sản phẩm, tạo cơ sở thực hiện khuyến nghị. Những người dùng có mức độ tương đồng cao sẽ có xu hướng mua những sản phẩm giống nhau. Với mỗi cách tính độ tương đồng sẽ cho một mô hình lọc cộng tác khác nhau.

Các mô hình lọc cộng tác có thể được chia ra thông qua 2 hướng tiếp cận: lọc cộng tác dựa trên bộ nhớ và lọc cộng tác dựa trên mô hình. Hướng tiếp cận dựa trên bộ nhớ tính toán độ tương tự giữa các người dùng từ đó thực hiện tư vấn. Nhược điểm của hướng tiếp cận này là sự tốn kém tài nguyên khi số lượng người dùng và sản phẩm tăng lên. Ngoài ra, hệ thống cần tính toán tại thời điểm khuyến nghị, điều này sẽ ảnh hưởng tới thời gian đưa ra dự đoán. Hướng tiếp cận dựa trên mô hình sử dụng các mô hình đã được huấn luyện thông qua các thuật toán học máy hoặc khai phá dữ liệu để thực hiện tư vấn. Với hướng tiếp cận này, mô hình sẽ cần

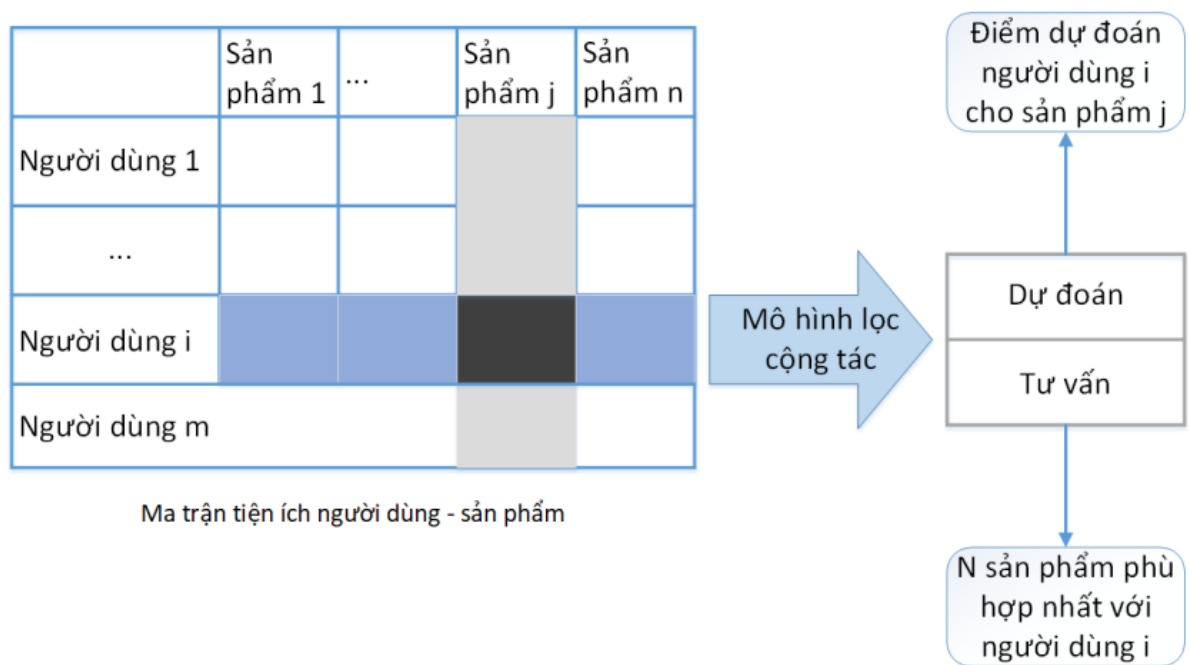
2.2. Lọc cộng tác

phải thực hiện huấn luyện trước, nhưng khi thực hiện khuyến nghị sẽ rất nhanh. Trong lọc cộng tác dựa trên bộ nhớ, ta có thể phân loại thành: lọc cộng tác dựa trên người dùng và lọc cộng tác dựa trên sản phẩm. Lọc cộng tác dựa trên người dùng là 1 mô hình so sánh sự tương đồng giữa các người dùng thông qua dữ liệu tương tác của họ lên các sản phẩm, từ đó khuyến nghị các sản phẩm phù hợp. Lọc cộng tác dựa trên sản phẩm dự đoán bằng cách sử dụng độ tương đồng giữa sản phẩm và sản phẩm được chọn bởi người dùng thông qua 1 ma trận tương tác của người dùng và sản phẩm. Nói cách khác, lọc cộng tác dựa trên bộ nhớ sử dụng các kỹ thuật như: độ tương quan Pearson, độ tương quan cô-sin, KNN để tạo các nhóm có đặc tính giống nhau, từ đó khuyến nghị các sản phẩm tới người dùng trong nhóm. Do cách hoạt động dựa trên dữ liệu đánh giá, nên mô hình khó có thể hoạt động tốt khi không có đủ dữ liệu cần thiết. Để khắc phục vấn đề này, lọc cộng tác dựa trên mô hình đưa ra khuyến nghị nhờ sử dụng các thuật toán như: phân cụm, SVD hay PCA.

2.2.1 Lọc cộng tác dựa trên bộ nhớ

Lọc cộng tác dựa trên người dùng

2.2. Lọc cộng tác



Hình 2.2: Mô hình thuật toán lọc cộng tác

Chương 3

Phương pháp giải quyết vấn đề

Phương pháp giải quyết vấn đề

Chương 4

Thực nghiệm, kết quả, so sánh và đánh giá

Thực nghiệm, kết quả, so sánh và đánh giá

Chương 5

Kết luận

Kết luận

Tài liệu tham khảo

- [1] James Davidson et al. “The YouTube video recommendation system”. In: *Proceedings of the fourth ACM conference on Recommender systems*. 2010, pp. 293–296.
- [2] Angela Edmunds and Anne Morris. “The problem of information overload in business organisations: a review of the literature”. In: *International journal of information management* 20.1 (2000), pp. 17–28.
- [3] Mahesh Goyani and Neha Chaurasiya. “A review of movie recommendation system: Limitations, Survey and Challenges”. In: *ELCVIA: electronic letters on computer vision and image analysis* 19.3 (2020), pp. 0018–37.
- [4] Folasade Olubusola Isinkaye, Yetunde O Folajimi, and Bolande Adefowoke Ojokoh. “Recommendation systems: Principles, methods and evaluation”. In: *Egyptian informatics journal* 16.3 (2015), pp. 261–273.