

ĐỀ XUẤT KỸ THUẬT TƯ VẤN NHÓM CHO DỮ LIỆU PHIM DỰA VÀO PHÂN CỤM THÔ – MỜ

Phạm Thanh Huyền^{1*}

¹Khoa Công nghệ Thông tin, Trường Đại học Hạ Long

*Email: phamthanhuyen@daihochalong.edu.vn

Ngày nhận bài: 24/09/2021 Ngày nhận bài sửa sau phản biện: 13/12/2021 Ngày chấp nhận đăng: 27/12/2021

TÓM TẮT

Tư vấn nhóm là một kỹ thuật trong hệ tư vấn nhằm hướng tới sở thích của một nhóm người dùng. Từ nghiên cứu khái niệm xấp xỉ dưới và xấp xỉ của lý thuyết tập thô mờ, bài báo trình bày mô hình tư vấn nhóm có sử dụng thuật toán tìm kiếm láng giềng gần nhất dựa vào tập thô mờ để cải thiện kết quả tư vấn. Mục đích của nghiên cứu này nhằm thực hiện phân cụm các tập mục và lựa chọn các nhóm tập mục được nhiều người dùng quan tâm. Để thực hiện mục đích, tác giả tiến hành thực nghiệm trên tập dữ liệu phim cụ thể và so sánh kết quả với việc sử dụng tìm kiếm láng giềng truyền thống, từ đó, chỉ ra hiệu quả của mô hình đề xuất.

Từ khóa: hệ tư vấn, tập thô – mờ, tư vấn nhóm

PROPOSING A NEW TECHNIQUE OF GROUP RECOMMENDATION FOR MOVIE DATA BASED ON FUZZY ROUGH CLUSTERING

ABSTRACT

Group Recommendation is a technique in the Recommender Systems aiming at a group of users' preferences. Via the study results of the concept of lower approximation and approximation of the fuzzy rough sets theory, the paper presents a Group Recommender Model that uses the nearest neighbor search algorithm based on the fuzzy rough sets to improve the recommender results. Also, the study is trying to perform clustering of items and select groups of itemsets that are of interest to a wide range of users. We have experimented on a specific movie data set and compared the results with the method based on searching traditional k-neighbors, thereby showing the effectiveness of our proposed model.

Keywords: fuzzy rough sets, group recommendation, recommendation systems.

1. ĐẶT VẤN ĐỀ

1.1. Hệ tư vấn và tư vấn nhóm

Hệ tư vấn (RS – Recommendation Systems) là một thành phần quan trọng trong khai phá dữ liệu (Koren & Bell, 2015). Hệ

thống tư vấn có mặt khắp mọi nơi và được xác định như một chiến lược ra quyết định cho người dùng (user) trong các môi trường thông tin phức tạp. Từ góc nhìn của kinh tế, nó là một phương tiện tự động cung cấp các tư vấn về các “mục tư vấn” (hay “mục”, gọi

tất là items) mà người dùng quan tâm (hoặc thuộc sở thích riêng của người dùng) như sản phẩm (hàng hóa/ dịch vụ), bài viết, trang web, bản nhạc, bộ phim, video, v.v.. Tuy nhiên, hầu hết các hệ thống mạng xã hội chỉ tư vấn cho từng người cụ thể. Thực tế, nhiều hoạt động hàng ngày được thực hiện bởi nhiều người từ các địa chỉ khác nhau như xem phim, xem các chương trình tivi, chọn nhà hàng, đi du lịch, lựa chọn các dịch vụ công cộng. Do vậy, hệ thống cần xem xét các gợi ý cho một số lượng người dùng nhất định, đó chính là tư vấn nhóm (GR – Group Recommendation). Phương pháp này có ý nghĩa đặc biệt khi phạm vi số lượng mục tư vấn cần tìm kiếm lớn (có thể là hàng trăm nghìn, hàng triệu mục) mà việc mô tả các mục cần tìm kiếm là một việc làm khó khăn đối với người dùng.

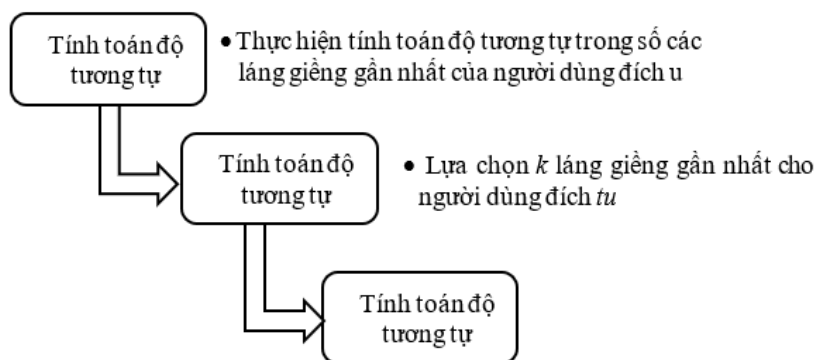
GR tìm sự liên quan về sở thích giữa những người dùng từ các đánh giá của họ về các mục (items). Nói cách khác, mục đích của nó để gợi ý các mục được yêu thích tới một nhóm người dùng có cùng sở thích. Thực chất, GR là một kỹ thuật trong RS nhằm hướng tới sở thích của một nhóm người dùng. Số lượng của GR nhằm xác định các phim theo sở thích của nhóm người dùng đã được phát triển nhiều trong những năm trước đây (Zhang và nnk., 2020). Một số thành viên trong nhóm được tư vấn có sở thích cùng nhau và ảnh hưởng với nhau trong việc ra quyết định trên các mục được tư vấn tới nhóm. Nhiều chuyên gia, nhà khoa học đã xem xét một số yếu tố nhóm từ sự tương tác xã hội giữa các người dùng để lựa chọn thành viên trong nhóm. Sự tương đồng về các ý kiến

và sở thích giữa các thành viên trong nhóm là căn cứ để tư vấn các phim hấp dẫn tới nhóm (Fang và nnk., 2018).

1.2. Lý thuyết tập thô – mờ và ứng dụng

Lý thuyết tập thô – mờ đã được công bố và ứng dụng nhiều trong phân lớp, khai phá dữ liệu và nhiều lĩnh vực khác. Lý thuyết tập thô – mờ, sự kết hợp giữa lý thuyết tập thô và lý thuyết tập mờ, là một trong những mở rộng ý nghĩa cho tập thô tổng quát. Thực tế, các nhà nghiên cứu tin rằng hai lý thuyết này có thể bổ sung cho nhau khi xử lý với các loại thông tin không chắc chắn (Sarkar, 2007; Kumar & Yadav, 2015). Hơn nữa, mô hình thô – mờ và công cụ học máy thực tế có kết nối mạnh với nhau (Vluymans và nnk., 2015). Nhiều mở rộng của phân hoạch hoặc quan hệ tương đương trên tập thô – mờ liên tục được đề xuất bởi các nhà nghiên cứu. Các toán tử xấp xỉ thô – mờ cơ bản được dựa vào một quan hệ tương tự và có liên quan với các phương pháp láng giềng gần nhất.

Sử dụng lý thuyết tập thô trong xử lý các bài hệ tư vấn còn nhiều tiềm năng (Zhang và nnk., 2020). Hơn nữa, việc ứng dụng các mở rộng của tập thô như tập thô – mờ, tập thô phủ, tập thô – mờ bức tranh .v.v. vào các bài toán khai phá dữ liệu nói chung, bài toán phân lớp và bài toán tư vấn nói riêng còn nhiều vùng trống (Sudha & Kumaravel, 2017). Với các kỹ thuật của hệ tư vấn, các nhà nghiên cứu có sử dụng thuật toán láng giềng gần nhất k-NN (k-Nearest Neighbour) trong thực hiện phân cụm mục (người dùng) (Park và nnk., 2014; Zhang và nnk., 2020).



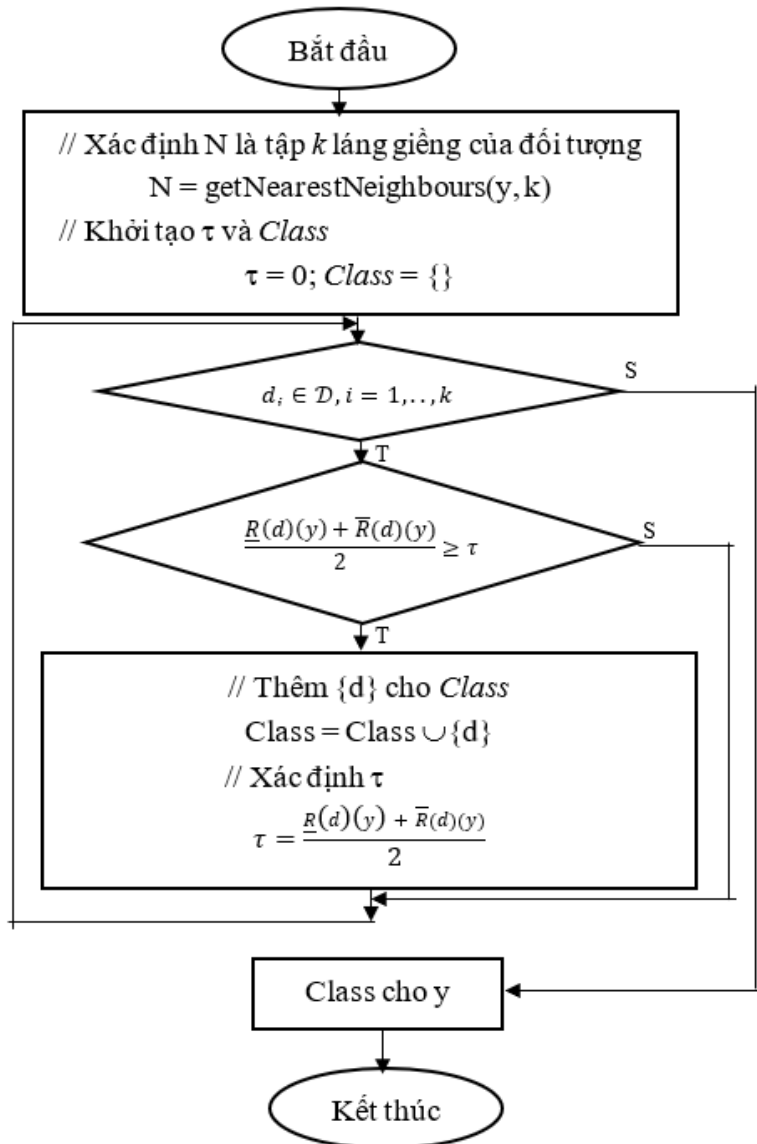
Hình 1. Mô hình tư vấn lọc cộng tác dựa vào láng giềng gần

Việc khai thác mở rộng lý thuyết tập thô như tập thô – mờ, tập thô phủ, v.v. nhằm cải thiện chất lượng phân cụm, nâng cao chất lượng kết quả tư vấn vẫn cần tiếp tục được nghiên cứu (Zhang và nnk., 2020). Tác giả đã mô hình hóa các bước tìm kiếm tư vấn lọc công tác và thuật toán phân cụm láng giềng gần nhất (Jensen & Cornelis, 2011) như Hình 1; từ đó tiến hành nghiên cứu và thực nghiệm láng giềng gần nhất dựa vào tập thô – mờ (FRNN – Fuzzy Rough Nearest Neighbour) trong hệ tư vấn, so sánh với kết quả thực

nghiệm k-NN trong hệ tư vấn.

1.3. Láng giềng gần nhất dựa vào tập thô – mờ

Trong không gian xấp xỉ (U, A, \mathbb{R}) , $U = \{u_1, u_2, \dots, u_n\}$ là một tập hữu hạn khác rỗng, các thuộc tính $A = C \cup D$ trên U , C là tập thuộc tính điều kiện, D là tập thuộc tính quyết định và $\mathbb{R} = \{R_1, R_2, \dots, R_m\}$ là một tập của các quan hệ tương đương. Giả sử rằng có một tập huấn luyện X ($X \subseteq U$) và một thể hiện mục tiêu un được dự đoán trong một lớp.



(Nguồn: Kumar & Yadav, 2015)

Hình 2. Thuật toán FRNN

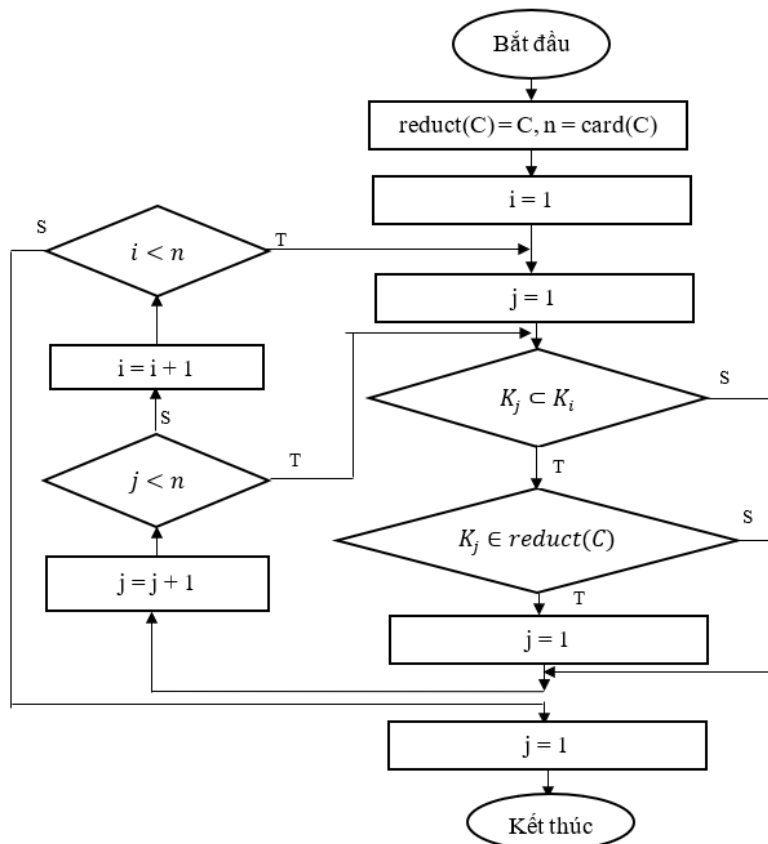
Phương pháp k-NN là một trong những kỹ thuật cụ thể để xác định k thể hiện trong một tập X gần với un và gần un với lớp đại diện trong số k láng giềng. Thực tế, k-NN được sử dụng trong nhiều ứng dụng để phát triển các phương pháp hiện đại bởi nó không cần tham số và đặc biệt nó không đưa ra giả định nào trên dữ liệu. Với việc lựa chọn một quan hệ dung sai mờ, Jensen và cộng sự đề xuất FRNN bằng sử dụng các xấp xỉ trên – xấp xỉ dưới của các lớp quyết định nhằm tăng hiệu quả phân lớp (Jensen & Cornelis, 2011). Thuật toán FRNN được biểu diễn bởi lưu đồ Hình 2. Tuy nhiên, Jensen và cộng sự chưa đề xuất cụ thể trong bài toán tư vấn.

Rõ ràng, FRNN có khả năng hơn để ứng dụng trong một số dữ liệu hiện đại. Trong nghiên cứu này, tác giả thực hiện kết hợp

FRNN để thực hiện phân cụm các tập mục và lựa chọn các nhóm tập mục được nhiều người dùng quan tâm bằng việc sử dụng GR. Việc nghiên cứu này là một đề xuất mới, mong muốn thu được hiệu quả tích cực cho khai phá dữ liệu phim MovieLens¹ trong tìm kiếm danh sách phim tư vấn phù hợp.

1.4. Rút gọn phủ

Với bài toán hệ tư vấn bằng sử dụng kỹ thuật lọc cộng tác và dựa vào kết quả thực nghiệm, Zhang và cộng sự đã cho thấy thuật toán CRA (CRA – Covering Reduction Algorithm) là một thuật toán rút gọn phủ tốt (Zhang và nnk., 2020). Nó được xây dựng dựa vào việc xây dựng thủ tục tìm tập quyết định phủ, từ đó, thu nhận kết quả tập phủ điều kiện tốt nhất. Các bước chi tiết của thuật toán CRA mô tả như trong Hình 3.



(Nguồn: Zhang và nnk., 2020)

Hình 3. Thuật toán CRA

¹ <https://grouplens.org/datasets/movielens/10m/>

Đề xuất tìm tập quyết định phủ trong lý thuyết tập thô là một căn cứ quan trọng để xác định các phần tử xấp xỉ dưới được rõ ràng. Với việc sử dụng thuật toán CRA, Zhang và cộng sự đã cố gắng loại bỏ càng nhiều láng giềng dư thừa càng tốt cho người dùng mới, đồng thời vẫn đảm bảo các tư vấn phim cho người dùng mới phải chính xác và đa dạng.

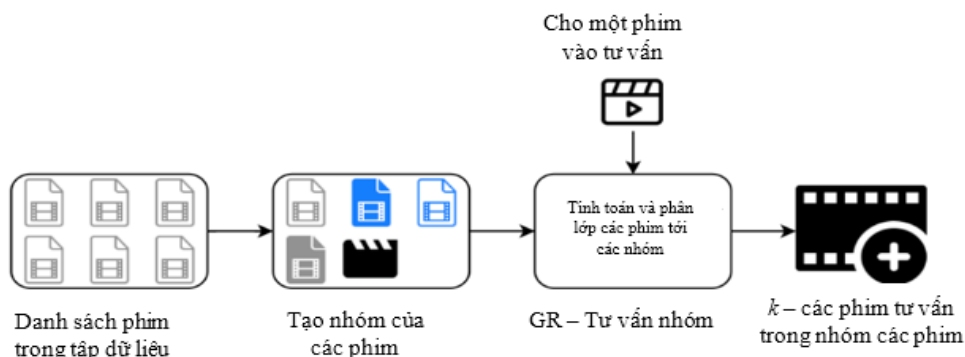
2. PHƯƠNG PHÁP NGHIÊN CỨU

Mô hình đề xuất đưa ra gợi ý các phim hay tới các thành viên trong một nhóm. Ý tưởng đề xuất sử dụng tư vấn nhóm dựa vào phân cụm thô – mờ nhằm dự đoán một tập các phim có người dùng quan tâm. Tư vấn nhóm xem xét riêng mỗi phim trong m phim từ một danh sách phim M cho trước dựa vào đánh giá đa dạng của người dùng. Tư vấn nhóm mục đích là dự đoán một nhóm các phim có người dùng quan tâm.

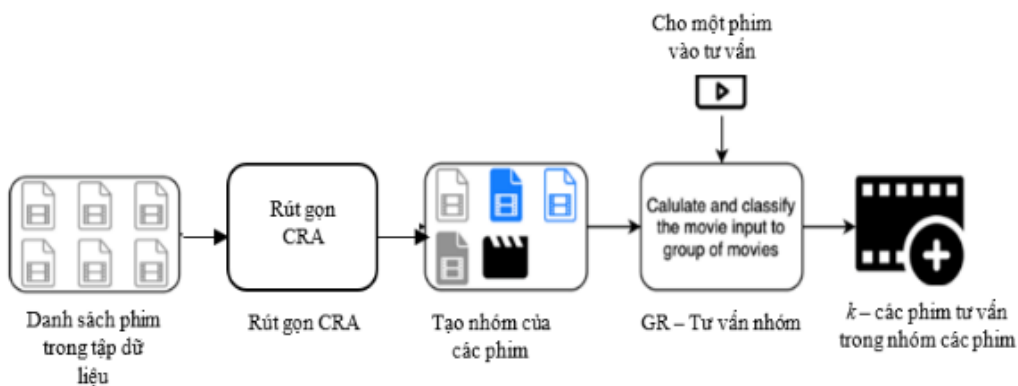
Để minh họa sức mạnh của phương pháp láng giềng mờ thô, hai thực nghiệm phân lớp được xây dựng. Đầu tiên, tác giả xây dựng mô

hình tư vấn nhóm bởi k -láng giềng theo trình tự cho hệ thống tư vấn. Sau đó, thực nghiệm thuật toán tìm láng giềng gần nhất (FRNN) trong lựa chọn, tìm kiếm nhóm phim để phân lớp cho m phim cho trước. Các đề xuất xử lý trên của tư vấn nhóm được minh họa.

Mô hình trong Hình 4 thể hiện 3 bước gồm tạo nhóm các phim, tính toán và phân lớp các phim tới các nhóm phim, lựa chọn k phim tư vấn từ nhóm các phim. Phương pháp sử dụng GR này vẫn tuân thủ quy trình thực hiện của kỹ thuật lọc cộng tác dựa vào láng giềng gần. Do vậy, để lựa chọn các phim tư vấn được tốt nhất, ở bước thứ hai, thực hiện tính toán và phân lớp các phim dựa vào các thuật toán tìm láng giềng gần nhất. Trong cài đặt thuật toán theo mô hình này, tác giả lựa chọn kiểm thử lần lượt các thuật toán phân cụm k -NN và FRNN để kiểm tra kết quả thu nhận bộ tư vấn tốt hơn. Với việc sử dụng lợi thế của FRNN, nghiên cứu mong muốn thu nhận kết quả phân cụm phù hợp và tốt hơn, đem lại kết quả tư vấn phim cao.



Hình 4. Đề xuất tư vấn nhóm cho dữ liệu phim MovieLens



Hình 5. Đề xuất tích hợp CRA cho tư vấn nhóm dựa FRNN

Thêm nữa, để nâng cao hiệu quả của nhóm phim từ vấn từ danh sách phim cho trước, tác giả đề xuất tích hợp thuật toán rút gọn phủ CRA của Zhang và cộng sự vào mô hình ở Hình 4. Tích hợp CRA vào mô hình đã được đề xuất như Hình 5 là một ý tưởng cải tiến mới. CRA hiệu quả với việc loại bỏ các người dùng lảng giềng dư thừa cho người dùng mới, trước khi thực hiện tạo nhóm phim để đưa vào tính toán và phân lớp tốt hơn. Phương pháp CRA đã được minh chứng hiệu quả đối với dữ liệu thưa thể hiện rõ trong thuật toán lọc cộng tác dựa trên phủ (Zhang và nnk., 2020).

Mô hình đề xuất mới này đảm bảo tính chính xác và đa dạng của các mục cần tư vấn.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Dữ liệu

Để đánh giá tư vấn nhóm trong tập phim cần được tư vấn, trong bài báo này, tác giả xem xét tập dữ liệu MovieLens 10M² công bố trong hội thảo ACM HetRec năm 2011. Đây là tập dữ liệu xếp hạng các phim được thu thập từ một website³ gợi ý các bộ phim hay và nhiều người ưa thích. Các thông tin của dữ liệu được cập nhật trong nhiều khoảng thời gian khác nhau. Một số thông tin về tập dữ liệu này được mô tả trong Bảng 1.

Bảng 1. Thông tin tập dữ liệu MovieLens

Ratings	~10,000,000
Tags	~100,000
Movies	~10,000
Users	~72,000

Tập dữ liệu này là một ma trận với cột (các thuộc tính) là độ đánh giá (rating) về mỗi bộ phim của của người dùng (users) và dòng là mỗi bộ phim. Một bộ phim chưa được đánh giá bởi người dùng sẽ được đặt một giá trị Null.

Trên thực tế, bộ dữ liệu MovieLens không

được phát triển để đánh giá hiệu quả phương pháp tư vấn nhóm vì các nhóm người dùng trước không được cung cấp trong bộ dữ liệu này. Do vậy, tác giả thực hiện tạo các nhóm người dùng của mình để hỗ trợ đánh giá. Tỷ lệ dữ liệu huấn luyện và dữ liệu thử nghiệm là 7/3.

3.2. Nhóm phim

Đầu vào của mô hình nhóm phim là tên của phim và đầu ra là một tập bộ phim gần với phim ban đầu nhất. Việc xác định này, dựa cơ bản vào sự đánh giá của người dùng trên mỗi phim.

Để thu được các nhóm người dùng là tốt nhất, thực hiện chiến lược sử dụng phương pháp phân cụm k-means⁴. Đây là thuật toán phổ biến trong học máy không giám sát để tạo các nhóm người dùng. Từ đó, xác định số lượng nhóm tối thiểu để đánh giá chất lượng cụm. Dựa vào việc xác định số lượng nhóm khác nhau, tác giả có thể đánh giá mức độ khó trong tìm kiếm sự đồng quan điểm giữa các thành viên của một nhóm nhỏ trong các nhóm. Tác giả đã thực hiện phân chia các phim thành 10 hay 20 nhóm. Trình tự tiến hành phương pháp đề xuất được thực hiện như trong mô tả Hình 4 và Hình 5. Có ba kịch bản được tiến hành, cụ thể:

Kịch bản 1 là kịch bản có sử dụng k-NN thực hiện phân cụm các phim cần tư vấn.

- Bước 1. Thực hiện nhận danh sách phim đã được đánh giá để tạo nhóm các phim.

- Bước 2. Tạo GR, trong đó thực chất là tính toán và phân lớp các phim về các nhóm phim.

- Bước 3. Sử dụng phương pháp k-NN để lựa chọn k phim tư vấn từ nhóm các phim.

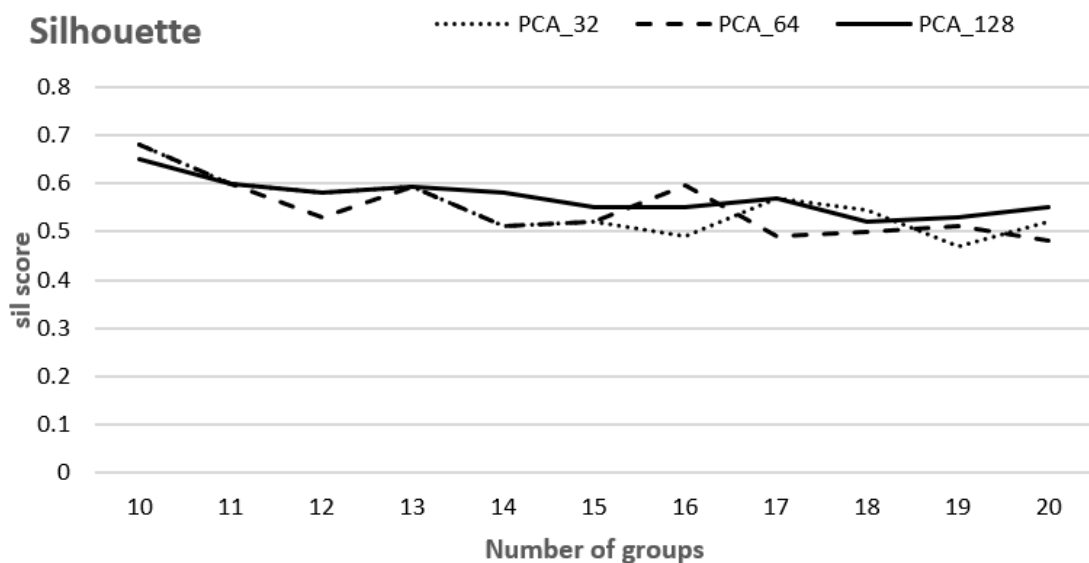
Kịch bản 2 là kịch bản có sử dụng FRNN thay cho k-NN như ở kịch bản 1. Theo kịch bản này, tác giả thực hiện tương tự bước 1 và bước 2 của kịch bản 1. Ở bước 3, sử dụng phương pháp FRNN thay cho k-NN nhằm lựa

² <https://grouplens.org/datasets/movielens/10m/>

³ <https://movielens.org>

⁴ [https://scikit-](https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html)

[learn.org/stable/modules/generated/sklearn.cluster.KMeans.html](https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html)



Hình 6. Chất lượng phân cụm

chọn có cân nhắc tới các dữ liệu mờ, tức quan tâm xét đến các phim có đánh giá của người dùng là giá trị mờ.

Kịch bản 3 là kịch bản kết hợp phương pháp rút gọn phủ CRA + FRNN. Mục đích của tác giả là thực hiện cải thiện thêm phương pháp ở kịch bản 2 bằng việc bổ sung một bước rút gọn phủ bằng phương pháp CRA tại bước 1. Tức là, để đảm bảo việc lựa chọn các phim được đa dạng, không chỉ chọn các phim có tính phổ biến, rút gọn phủ CRA sẽ cho một tập các phim có thể có ít người dùng đánh giá nhưng có mức độ liên quan gần với nhóm phim tư vấn. Do vậy, sau khi nhận danh sách phim đầu vào, phương pháp CRA thực hiện rút gọn phủ và cho ra một phủ tối ưu phù hợp, hiệu quả cho quá trình tạo nhóm phim không bị bỏ sót thuộc tính. Điều này giúp cho việc phân cụm, lựa chọn các phim cần tư vấn được đa dạng.

Ngoài số lượng nhóm, sự liên kết nhóm cũng là một tiêu chí quan trọng từ hầu hết các phim được đánh giá bởi nhiều người dùng khác nhau. Trong quá trình tiến hành, tác giả nhận thấy, bộ dữ liệu thực nghiệm có những vấn đề sau: “thiên lệch về những phim có tính phổ biến”, phát sinh “vấn đề khởi đầu nguội mục phim (Item cold-start problem)”. Đây là

thách thức điển hình của tư vấn lọc cộng tác. Phần lớn các phim nhận được ít hoặc không có đánh giá nào từ những người dùng. Do vậy, bộ dữ liệu MovieLens là một ma trận “thưa” với hơn 99% là giá trị chưa đánh giá. Để xử lý vấn đề này, có thể sử dụng phương pháp PCA (Principal Component Analysis) thể hiện dữ liệu trong không gian chiều lớn với một không gian chiều nhỏ hơn. Số lượng các nhân tố tiềm ẩn xác định được từ số lượng các thông tin ngắn, đây là điều mong muốn chọn được không không gian chiều nhỏ. Tác giả chọn thành phần trong PCA là 32, 64 và 128 vì sự tăng lên của số lượng các nhân tố tiềm ẩn sẽ cải thiện cá nhân hóa, nhưng nếu số lượng nhân tố quá cao thì mô hình bắt đầu bị overfitting.

Hình 6 cho chúng ta thấy một số thực nghiệm trong nhóm. Số lượng tối thiểu của các nhóm là 10 dựa vào phương pháp bóng trung bình và các thành phần trong PCA là 32. Tiếp tục thực hiện phương pháp đề xuất, tác giả có thể đánh giá k -NN và FRNN trong phân lớp nhóm phim tư vấn.

3.3. Đánh giá thuật toán

Để đánh giá hiệu năng của GR, tác giả thực hiện tính độ chính xác ở độ đo k ($\text{acc}@k$) để đánh giá phân lớp dựa vào k -NN và độ

chính xác (Accuracy) của phân lớp dựa vào FRNN. Trong phân lớp k -NN, luật quyết định được dựa vào k -NN, lớp của mẫu thử nghiệm là một tập có số lượng bằng với tất cả các lớp trong k mẫu huấn luyện. Đối với FRNN, k là độ dài hiệu quả của các vector trọng số trên và dưới (số lượng các làng giếng gần nhất được xem xét). Công thức tính độ chính xác được xác định như sau:

$$\text{Độ chính xác} = \frac{\text{Số lượng của các dự đoán đúng}}{\text{Tổng số lượng của các dự đoán}}$$

Kết quả thực hiện thực nghiệm được thể hiện ở Bảng 2, trong đó, độ chính xác phân lớp ở mỗi phương pháp được chi tiết như sau:

Bảng 2. Độ chính xác phân lớp dựa vào k -NN và FRNN

	KNN	FRNN	CRA+FRNN
Acc@10	0.748	0.787	0.857
Acc@20	0.716	0.775	0.810

Theo kết quả này, tiếp cận dựa vào FRNN cho kết quả tốt hơn tiếp cận k -NN trong cả trường hợp 10 và 20 láng giếng gần nhất. Với FRNN, tác giả phân tích được vị trí dự đoán của đối tượng có mức độ thành viên mờ và k láng giếng của đối tượng, do vậy, nó có độ chính xác cao hơn phân lớp dựa vào k -NN. Ngoài ra, rõ ràng việc kết hợp CRA trong lựa chọn thuộc tính (bằng rút gọn phủ) và FRNN trong phân cụm phim về các nhóm phim tư vấn là hoàn toàn phù hợp với dữ liệu này. Cùng với việc nhúng PCA đã kiểm soát không gian dữ liệu khi phân cụm trong trường hợp dữ liệu thừa chiếm số lượng lớn (tức là số lượng đánh giá của người dùng về các phim là rất ít). Do vậy, đề xuất kết hợp CRA + FRNN có sử dụng kỹ thuật nhúng PCA vào phân cụm FRNN cho kết quả cao hơn hai phương pháp trước.

4. KẾT LUẬN

Phương pháp GR dựa vào phân lớp FRNN đảm bảo thu nhận kết quả có độ chính xác cao hơn phân lớp k -NN. Như vậy, có thể mở rộng lý thuyết tập mờ để xử lý tốt bài toán hệ tư

vấn. Trong thời gian tiếp theo, tác giả sẽ tiếp tục nghiên cứu các vấn đề mở rộng của lý thuyết tập mờ để cải thiện tính chính xác và đa dạng của dữ liệu tư vấn.

TÀI LIỆU THAM KHẢO

- Fang, G., Su, L., Jiang, D., & Wu, L. (2018). Group Recommendation Systems Based on External Social-Trust Networks. *Wireless Communications and Mobile Computing*, 2018.
- Jensen, R., & Cornelis, C. (2011). Fuzzy rough nearest neighbour classification and prediction. *Theoretical Computer Science*, 412, 5871 – 5884.
- Koren, Y., & Bell, R. (2015). Advances in Collaborative Filtering. In Ricci, F. L. Rokach, & Shapira, B (Eds), *Recommender Systems Handbook* (S. 77-118). Boston, MA: Springer.
- Kumar, M., & Yadav, N. (2015). Fuzzy Rough Sets and Its Application in Data Mining Field. *ACSIT*, 2(2), 237-240.
- Park, Y., Park, S., Lee, S., & Jung, W. (2014). Fast Collaborative Filtering with a k -nearest neighbor graph. *2014 International Conference on Big Data and Smart Computing (BIGCOMP)*.
- Sarkar, M. (2007). Fuzzy rough nearest neighbors algorithm. *Fuzzy Sets and Systems*, 158, 2123 – 2152.
- Sudha, M., & Kumaravel, A. (2017). Comparative Analysis between Rough Set Theory and Data Mining Algorithms on their Prediction. *Global Journal of Pure and Applied Mathematics*, 13(7), 3249-3260.
- Vluymans, S., D’eer, L., Saeys, Y., & Cornelis, C. (2015). Applications of Fuzzy Rough Set Theory in Machine Learning a Survey. *Fundamenta Informaticae*, 20, 1 -34.
- Zhang, Z., Kudo, Y., Murai, T., & Ren, Y. (2020). Improved covering-based collaborative filtering for new users’ personalized recommendations. *Knowledge and Information Systems*, 62, 3133–315.