

**Đại học Dundee****Lọc công tác và hệ thống khuyến nghị dựa trên học tập sâu cho khởi đầu lạnh  
mặt hàng**

Nguy, Jian; Ông, Gianhua; Chen, Kai; Chu, Yi; Đuờng, Zuoyin

*Xuất bản năm:*

Hệ thống chuyên gia với các ứng dụng

*DOI:*[10.1016/j.eswa.2016.09.040](https://doi.org/10.1016/j.eswa.2016.09.040)*Ngày xuất bản:*

2017

*Phiên bản tài liệu*

Phiên bản đánh giá ngang hàng

[Liên kết đến xuất bản trên Discovery Research Portal](#)*Trích dẫn cho phiên bản đã xuất bản (APA):*Nguy, J., He, J., Chen, K., Zhou, Y., & Tang, Z. (2017). Lọc công tác và học tập sâu  
hệ thống khuyến nghị cho các mặt hàng bắt đầu lạnh. *Hệ thống chuyên gia với các ứng dụng*, 69, 29-39.  
<https://doi.org/10.1016/j.eswa.2016.09.040>**Quyền chung**

Bản quyền và quyền đạo đức cho các án phẩm có thể truy cập được trong Công thông tin nghiên cứu Discovery được các tác giả và / hoặc người khác giữ lại  
chủ sở hữu bản quyền và đó là điều kiện truy cập các án phẩm mà người dùng nhận ra và tuân thủ các yêu cầu pháp lý liên quan đến  
những quyền này.

- Người dùng có thể tải xuống và in một bản sao của bất kỳ án phẩm nào từ Discovery Research Portal cho mục đích nghiên cứu hoặc nghiên cứu riêng tư.
- Bạn không được phân phối thêm tài liệu hoặc sử dụng tài liệu đó cho bất kỳ hoạt động kiếm lợi nhuận hoặc lợi nhuận thương mại nào.
- Bạn có thể tự do phân phối URL xác định án phẩm trong công thông tin công cộng.

**Bố chính sách**

Nếu bạn tin rằng tài liệu này vi phạm bản quyền, vui lòng liên hệ với chúng tôi để cung cấp chi tiết và chúng tôi sẽ xóa quyền truy cập vào tác phẩm ngay lập tức  
và điều tra yêu cầu của bạn.

Ngày tải xuống: 09. Tháng 1 năm 2020

## Bản thảo được chấp nhận

Lọc cộng tác và khuyến nghị dựa trên học tập sâu  
Hệ thống cho các mặt hàng bắt đầu lạnh

Jian Wei, Jianhua He, Kai Chen, Yi Zhou, Zuoyin Tang

PII: S0957-4174 (16) 30530-9  
 DOI: [10.1016/j.eswa.2016.09.040](https://doi.org/10.1016/j.eswa.2016.09.040)  
 Tài liệu tham khảo: ESWA 10904

Đề xuất hiện trong: *Hệ thống chuyên gia với các ứng dụng*

Ngày nhận: 21 tháng 7 năm 2016  
 Ngày sửa đổi: 17 tháng 9 năm 2016  
 Ngày được chấp nhận: 28 tháng 9 năm 2016

Xin trích dẫn bài viết này như: Jian Wei, Jianhua He, Kai Chen, Yi Zhou, Zuoyin Tang, Collaborative Filter-Hệ thống khuyến nghị dựa trên học tập sâu cho các mặt hàng bắt đầu lạnh, *hệ thống chuyên gia với Ứng dụng* (2016), doi: [10.1016/j.eswa.2016.09.040](https://doi.org/10.1016/j.eswa.2016.09.040)

Đây là một tệp PDF của một bản thảo chưa được chỉnh sửa đã được chấp nhận để xuất bản. Như một dịch vụ cho khách hàng của chúng tôi, chúng tôi đang cung cấp phiên bản đầu tiên của bản thảo này. Bản thảo sẽ trải qua sao chép, sappId và xem xét bằng chứng kết quả trước khi nó được xuất bản ở dạng cuối cùng. Xin vui lòng lưu ý rằng trong quá trình sản xuất, các lỗi có thể được phát hiện có thể ảnh hưởng đến nội dung và tất cả các khuyến cáo pháp lý áp dụng cho các tạp chí liên quan.

© 2016. Phiên bản bản thảo này được cung cấp theo giấy phép CC-BY-NC-ND 4.0  
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

### Điểm nổi bật

- . Hai mô hình đề xuất đã được đề xuất cho các mặt hàng bắt đầu lạnh.

- Các mô hình kết hợp lọc cộng tác nhận thức thời gian và học tập sâu.
- Thủ nghiệm trên bộ dữ liệu Netflix cho thấy sự cải thiện lớn so với các phương pháp tiếp cận hiện có.

# BẢN THẢO ĐƯỢC CHẤP NHẬN

---

Trang 4

BẢN THẢO ĐƯỢC CHẤP NHẬN

2

## Lọc cộng tác và học sâu

## Hệ thống khuyến nghị dựa trên khởi đầu lạnh

### Mặt hàng

Jian Wei<sup>1</sup>, Jianhua He<sup>1</sup>, Kai Chen<sup>2</sup>, Yi Zhou<sup>2</sup>, Zuoyin Tang<sup>1</sup>

<sup>1</sup> Trường Kỹ thuật và Khoa học Úng dụng, Đại học Aston, Birmingham, B4 7ET, Vương quốc Anh.

<sup>2</sup> Khoa Kỹ thuật Điện tử, Đại học Giao thông Thượng Hải, Thượng Hải, Trung Quốc.

Email: {weij2, j.he7, z.tang1} Bolog@aston.ac.uk,

{kchen, zy\_21th Bolog@sjtu.edu.cn.

Tác giả tương ứng: Tiến sĩ Jianhua He.

**trùu tượng**

Hệ thống đề xuất là một loại hệ thống thông minh cụ thể, khai thác xếp hạng người dùng lịch sử trên các mục và / hoặc thông tin phụ trợ để đưa ra khuyến nghị về các mục cho người dùng. Nó đóng một vai trò quan trọng trong vai trò trong một loạt các mua sắm trực tuyến, dịch vụ thương mại điện tử và các ứng dụng mạng xã hội.

Lọc công tác (CF) là phương pháp phổ biến nhất được sử dụng cho các hệ thống đề xuất, nhưng nó bị ảnh hưởng từ vấn đề khởi động nguội hoàn toàn (CCS) trong đó không có hồ sơ xếp hạng có sẵn và khởi động lạnh không hoàn chỉnh (ICS) có vấn đề khi chỉ có một số lượng nhỏ hồ sơ xếp hạng có sẵn cho một số mặt hàng hoặc người dùng mới trong hệ thống. Trong bài viết này, chúng tôi đề xuất hai mô hình đề xuất để giải quyết các vấn đề CCS và ICS cho các mục mới, dựa trên khung phương pháp CF kết hợp chặt chẽ và học tập thần kinh sâu mang. Một mạng thần kinh sâu cụ thể SADE được sử dụng để trích xuất các tính năng nội dung của các mục. Các mô hình CF hiện đại, timeSVD ++, mô hình và sử dụng động lực học tạm thời theo sở thích của người dùng và các tính năng của mục, được sửa đổi để đưa các tính năng nội dung vào dự đoán xếp hạng cho các mục bắt đầu lạnh.

Các thử nghiệm mở rộng trên bộ dữ liệu xếp hạng phim Netflix lớn được thực hiện, cho thấy rằng chúng tôi các mô hình đề xuất được đề xuất phản hồi vượt trội so với các mô hình cơ sở dữ liệu xếp hạng cảm lạnh bắt đầu cao mục. Hai mô hình đề xuất được đề xuất cũng được đánh giá và so sánh trên các mục của ICS, và một kế hoạch linh hoạt về đào tạo lại và chuyển đổi mô hình được đề xuất để đối phó với việc chuyển đổi các hạng mục từ bắt đầu lạnh đến trạng thái bắt đầu không lạnh. Kết quả thử nghiệm trên chương trình đề xuất phim Netflix sự kết hợp chặt chẽ giữa phương pháp CF và mạng lưới thần kinh học tập sâu là khả thi và rất hiệu quả đối với đề nghị bắt đầu lạnh. Thiết kế nói chung và có thể được áp dụng cho nhiều người giới thiệu khác

**BẢN THẢO ĐƯỢC CHẤP NHẬN****Trang 5****BẢN THẢO ĐƯỢC CHẤP NHẬN**

3

hệ thống mua sắm trực tuyến và các ứng dụng mạng xã hội. Giải pháp cho vấn đề bắt đầu lạnh Phần lớn có thể cải thiện trải nghiệm người dùng và sự tin tưởng của các hệ thống đề xuất, và thúc đẩy lạnh hiệu quả bắt đầu các mục.

**Từ khóa:** Hệ thống khuyến nghị; Khai thác dữ liệu; Mạng lưới thần kinh học tập sâu; Collabo-  
Lọc hợp lý; Vấn đề bắt đầu lạnh

**I. TÔI XÁC NHẬN**

Hệ thống khuyến nghị đóng vai trò trung tâm cho nhiều ứng dụng trực tuyến và thương mại điện tử các dịch vụ, như mạng xã hội, giới thiệu các sản phẩm như phim, nhạc và bài viết (Shi, Larson, & Hanjalic, 2014; Campos, Dez, & Cantador, 2014; Linden, Smith, & York, 2003). Nhiều công ty lớn như Amazon, eBay và Netflix đã áp dụng khuyến nghị kỹ thuật cho hệ thống của họ để ước tính sở thích tiềm năng của khách hàng và đề xuất sản phẩm liên quan hoặc các mặt hàng cho người dùng. Biểu diễn khuyến nghị có tác động rất lớn đến thành công thương mại của các công ty này về doanh thu và người dùng thỏa đáng.

Theo loại dữ liệu được thu thập và cách sử dụng chúng trong khuyến nghị-hệ thống dation, cách tiếp cận cho khuyến nghị có thể được phân loại là dựa trên nội dung (CB),

lọc cộng tác (CF) và lai một (Koren, Bell, & Volinsky, 2009).

Lọc CB được sử dụng rộng rãi cho thiết kế hệ thống khuyến nghị, sử dụng nội dung của các mục để tạo các tính năng và thuộc tính để phù hợp với hồ sơ người dùng. Vật phẩm được so sánh với vật phẩm người dùng thích trước đó và các mục phù hợp nhất sau đó được đề xuất. Một vấn đề lớn Phương pháp lọc CB là RS cần tìm hiểu sở thích của người dùng đối với một số loại mặt hàng và áp dụng những điều này cho các loại mặt hàng khác.

Cách tiếp cận CF là cách tiếp cận phổ biến nhất cho thiết kế hệ thống khuyến nghị. Nó sử dụng một số lượng lớn dữ liệu được thu thập từ hành vi người dùng trong quá khứ và dự đoán những mục nào người dùng sẽ giống. Nó không cần phải phân tích nội dung của các mục. Thay vào đó, nó dựa vào mối quan hệ giữa người dùng và vật phẩm, thường được mã hóa trong ma trận phản hồi xếp hạng với từng loại yếu tố đại diện cho một đánh giá người dùng cụ thể trên một mặt hàng cụ thể. Một minh họa về CF dựa trên khuyến nghị được hiển thị trong Hình 1. Bên trái của Hình 1 hiển thị biểu đồ mối quan hệ của 3 người dùng và 4 phim, được kết nối bởi 5 cạnh. Mỗi cạnh được liên kết với xếp hạng từ 1 đến 5 sao, đại diện cho mức độ ưa thích của người dùng của bộ phim được kết nối. Ma trận bên phải

Hình 1 được tạo theo biểu đồ mối quan hệ. Nhiệm vụ đề xuất CF chung

# BẢN THẢO ĐƯỢC CHẤP NHẬN

## Trang 6

### BẢN THẢO ĐƯỢC CHẤP NHẬN

4

là để dự đoán các xếp hạng còn thiếu (chẳng hạn như các xếp hạng được biểu thị bằng biểu tượng người dùng đã cho hoặc cho các mục nhất định bằng cách khai thác dữ liệu và khám phá ma trận xếp hạng mục người dùng.

Hình 1. Một đại diện đơn giản hóa cho các hệ thống đề xuất CF phim.

Tuy nhiên, mọi người đều biết rằng phương pháp CF bị ảnh hưởng bởi sự thua thót và khởi đầu lạnh (CS) các vấn đề. Trong ma trận đánh giá chỉ có một tỷ lệ nhỏ các yếu tố nhận được giá trị. Ngay cả nhiều nhất các mặt hàng phổ biến có thể chỉ có một vài xếp hạng. Ví dụ: trong một tập dữ liệu xếp hạng Netflix lớn được cung cấp cho cuộc thi Giải thưởng Netflix (Bennett, & Lanning, 2007), có khoảng 100 triệu xếp hạng được cung cấp bởi hơn 480.000 người dùng cho khoảng 18.000 bộ phim. Chỉ có khoảng 1% ma trận đánh giá yếu tố nhận xếp hạng. Với một ma trận đánh giá thua thót, rất khó để ước tính mối quan hệ giữa các mục và người dùng và đưa ra khuyến nghị hiệu quả. Một người nổi tiếng khác vấn đề đối với phương pháp CF là vấn đề CS, có thể xảy ra với người dùng mới hoặc vật phẩm mới. CF

Cách tiếp cận đòi hỏi một số lượng lớn xếp hạng từ người dùng hoặc xếp hạng trên một mặt hàng để có hiệu quả đề xuất, sẽ không hoạt động cho người dùng mới, mặt hàng mới hoặc cả hai do một vài xếp hạng có sẵn trong hệ thống. Ngoài ra, vấn đề CS có thể được chia thành vấn đề CCS và ICS

vấn đề bởi số lượng hồ sơ xếp hạng bằng không hay không. Nói chung, sự thua thót của xếp hạng

đối với các mục CS cao hơn 85% (Zhang và cộng sự, 2014) và mức độ xếp hạng của các mục CCS rất thấp là 100%. Hình 2 trình bày một minh họa đơn giản về việc phân loại các mặt hàng CCS, ICS và không CS

trong các hệ thống khuyến nghị.

Phương pháp lai là phương pháp kết hợp phương pháp lọc CB và phương pháp CF cố gắng để khắc phục nhược điểm của họ và cung cấp một kết quả hiệu quả hơn (Agarwal, & Chen, 2009; Chen et al., 2012; Hu và cộng sự, 2013). Cần lưu ý rằng phần lớn các công trình về khuyến nghị CS vẫn đề đang cố gắng đưa ra khuyến nghị về các mặt hàng có thể thú vị cho người dùng cụ thể.

Mặc dù rất nhiều công việc đã được thực hiện với phương pháp lai để giải quyết sự thua thót và lạnh lẽo

# BẢN THẢO ĐƯỢC CHẤP NHẬN

Trang 7

## BẢN THẢO ĐƯỢC CHẤP NHẬN

5

Hình 2. Minh họa về mặt hàng không CS (a), mặt hàng ICS (b) và mặt hàng CCS (c), trong đó / chỉ ra xếp hạng đã biết.

bắt đầu các vấn đề, đề xuất của các mục CS vẫn là một vấn đề nghiên cứu mở.

Trong bài báo này, chúng tôi điều tra vấn đề khuyến nghị của CS về việc cung cấp dự đoán về mức độ phổ biến của các mục CS đã cung cấp cho người dùng phổ thông và đưa ra giải pháp để dự đoán mức độ phổ biến các mặt hàng CCS và các mặt hàng ICS. Có hai động lực chính cho công việc này:

- Các mục CS cần được đề xuất để có xếp hạng cho đề xuất được cải thiện và chúng nên được đề xuất chính xác để cung cấp cho người dùng trải nghiệm tốt hơn với đề xuất hệ thống. Nếu không, các mục CS có thể đi đến một chu kỳ không mong muốn là không nhận được xếp hạng.

- Xếp hạng ước tính cho các mục CS hoặc các mục vẫn đang được lên kế hoạch có thể đưa ra một biện pháp về mức độ phổ biến của các mặt hàng đó ngay cả trước khi chúng được đưa ra thị trường (như sách, phim, v.v.), do đó giúp đưa ra quyết định đúng đắn về kế hoạch bán hàng và chiến lược bán hàng. Sự chính xác ước tính như vậy là rất quan trọng cho loại mục đích này.

Chúng tôi thiết kế hai mô hình đề xuất tích hợp, trong đó các tính năng mục được học từ

một kiến trúc học tập sâu SDAE (Vincent, Larochelle, Lajoie, Bengio, & Manzagol, 2010).

sử dụng các mô tả của các mục được truy xuất trực tuyến, sau đó các tính năng này được khai thác và tích hợp vào mô hình CF thời gian SVD++ (Koren, 2010). timeSVD++ là một trong những CF hoạt động tốt nhất

các mô hình theo dõi hành vi thay đổi thời gian trong dữ liệu và đưa động lực học tạm thời vào

tài khoản.

# BẢN THẢO ĐƯỢC CHẤP NHẬN

**Trang 8****BẢN THẢO ĐƯỢC CHẤP NHẬN**

6

Đóng góp của chúng tôi được tóm tắt như sau.

- . Chúng tôi đã đề xuất một khung chung về tích hợp phương pháp CF và học máy các thuật toán để cải thiện hiệu suất đề xuất cho các mục CS. Trong các mô hình đề xuất của chúng tôi tính năng nội dung được trích xuất từ mô tả nội dung (như cốt truyện phim) bằng cách học sâu mạng nơ ron được sử dụng làm vectơ nhân tố mục chính trong mô hình đề xuất cho các mục CCS và được xấp xỉ bởi các vectơ hệ số mục trong mô hình cho các mục ICS. Các tính năng nội dung không chỉ được sử dụng một cách lỏng lẻo để xác định độ tương tự của mục như được thực hiện trong phương pháp lai hiện có cho các mặt hàng bắt đầu lạnh, nhưng cũng trở thành phần chính của mô hình đề xuất, ảnh hưởng đến cả việc đào tạo mô hình và dự đoán xếp hạng không xác định cho các mục CS.
- . Khung tích hợp phương pháp CF và thuật toán học máy cho mục CS khuyến nghị là chung chung. Các cách tiếp cận CF khác nhau và các thuật toán học máy có thể được sử dụng cho các hệ thống đề xuất chung. Điểm tích hợp quan trọng là khai thác về các tính năng vật phẩm bằng thuật toán học máy và nhúng các tính năng vật phẩm vào Mô hình đề xuất CF.
- . Dựa trên các mô hình và thiết kế hệ thống cụ thể của khung chung được trình bày, trong đó trạng thái của mô hình CF hiện đại, timeSVD ++ và mạng lưới thần kinh học sâu tiên tiến mô hình, SADE, được sử dụng cho đề xuất các mục CS. Áp dụng các mô hình cho Netflix đề nghị phim với gần 100 triệu xếp hạng đã được điều tra. Các thí nghiệm kết quả cho thấy rằng sự kết hợp chặt chẽ giữa phương pháp CF và cách tiếp cận dựa trên nội dung cho khuyến nghị là khả thi và rất hiệu quả. Ví dụ: dự đoán xếp hạng RMSE của mô hình đề xuất IRCD-CCS cho đề xuất mục CCS thấp hơn 0,045 so với phương pháp thực hiện tốt thứ hai, đại diện cho một cải tiến hiệu suất đáng kể trong lĩnh vực nghiên cứu thiết kế hệ thống khuyến nghị.
- . Ngoài việc thiết kế và đánh giá các mô hình đề xuất cho các hạng mục CCS và Các mục ICS riêng biệt, chúng tôi cũng so sánh hiệu suất của mô hình IRCD-CCS và IRCD-Mô hình ICS về dự đoán xếp hạng cho các mục ICS. Trong thực tế, các hệ thống khuyến nghị giữ giới thiệu các mục mới vào hệ thống theo thời gian. Nếu một mục mới được giới thiệu là CCS mục, mô hình CF không thể cung cấp dự đoán xếp hạng cho nó. Nếu mục đó là một mục ICS, các mô hình CF có thể không đưa ra khuyến nghị tốt. Có thể có lợi khi áp dụng CCS

# **BẢN THẢO ĐƯỢC CHẤP NHẬN**

**BẢN THẢO ĐƯỢC CHẤP NHẬN**

7

mô hình đề xuất cho dự đoán xếp hạng mục ICS. Chúng tôi đề xuất một kế hoạch chuyển đổi mô hình đề xuất cho các mục ICS và đào tạo lại các mô hình đề đối phó với thực tế vấn đề chuyển trạng thái mục từ CS sang không CS. Theo hiểu biết tốt nhất của chúng tôi, điều này vấn đề thực tế chưa được nghiên cứu trước đây trong tài liệu.

Phần còn lại của bài báo được tổ chức như sau. Phần II mô tả các công việc liên quan.

Phần III trình bày vấn đề nghiên cứu và các mô hình đề xuất được đề xuất. Mục IV

trình bày phương pháp đánh giá và kết quả thí nghiệm. Cuối cùng Phần V kết luận.

**II. RĂNG BẮT ĐẦU WORK**

Về mặt kỹ thuật, phương pháp nhân tố ma trận (MF) đã được áp dụng cho CF bởi nhiều công việc khác nhau.

MF tập trung vào việc nhân tố ma trận đánh giá thành các vectơ tiềm ẩn của người dùng có kích thước thấp và độ trễ của mục vectơ. Đào tạo một mô hình như vậy có thể được giải quyết một cách hiệu quả bằng cách sử dụng SGD (Koren, Bell, & Volinsky, 2009) hoặc xen kẽ các ô vuông nhỏ nhất (ALS) (Zhou, Wilkinson, Schreiber, & Pan, 2008) để giảm thiểu khoảng cách bình phương tổng. Các tác giả (Salakhutdinov, & Mnih, 2007) giới thiệu xác suất hệ số ma trận (PMF) chia tỷ lệ tuyến tính trên các tập dữ liệu lớn và vượt trội so với tiêu chuẩn mô hình phân rã giá trị số ít (SVD). Dựa trên PMF, một số biến thể và khái quát hóa được đề xuất như PMF Bayes (Salakhutdinov, & Mnih, 2008), PMF tổng quát (Shan, & Banerjee, 2010).

Vì thuật toán CF truyền thống chỉ dựa vào mối quan hệ giữa người dùng và vật phẩm, đó là thường được mã hóa trong ma trận UI, hiệu suất đề xuất về vấn đề thưa thớt và CS vấn đề phần lớn bị hạn chế. Một số lượng lớn các phương pháp kết hợp thông tin bổ sung các nguồn ngoài ma trận UI đã được phát triển để khắc phục các vấn đề. Đặc biệt phụ trợ thông tin của người dùng hoặc vật phẩm và thông tin liên quan đến tương tác được khai thác để cải thiện khuyến nghị độ chính xác (Shi, Larson, & Hanjalic, 2014).

Thông tin phụ trợ để cập nhật các thuộc tính về người dùng và các mặt hàng. Đối với thuộc tính người dùng, tin tưởng mạng được kết hợp vào xếp hạng thô để dự đoán (Victor, Cornelis, Teredesai, & De Con gá trong 2008). Các tác giả (Má, King, & Lyu, 2011) đề xuất khung phân tích nhân tố xác suất trong đó có quan hệ tin cậy xã hội của người dùng vào tài khoản. Các tác giả (Zhang, Liu, Zhang, & Zhou, 2010) sử dụng các thẻ xã hội của người dùng và thiết kế thuật toán đề xuất dựa trên khuếch tán chỉ được sử dụng trong các hệ thống gắn thẻ xã hội. Các tác giả (Lika, Kolomvatsos, & Hadjiefthymiades, 2014) chấp nhận dữ liệu nhân khẩu học của người dùng và áp dụng quy tắc dự đoán đơn giản bằng cách tính tổng xếp hạng có trọng số

**BẢN THẢO ĐƯỢC CHẤP NHẬN**

được thực hiện bởi những người dùng tương tự để tạo ra xếp hạng cho người dùng mới. Tự động (Ocepka, Rugeljb, & Bosnica, 2015) kết hợp lựa chọn thuộc tính và học tập địa phương vào mô hình đề xuất cho CS

người dùng. Cả (Lika, Kolomvatsos, & Hadjiefthymiades, 2014) và (Ocepka, Rugeljb, & Bosnica, 2015) đã sử dụng một trong những cách tiếp cận cơ bản của chúng tôi (phương pháp ToU) trong các mô hình đề giới thiệu sản phẩm cho người dùng CS. Các tác giả (Zhou, Yang, & Zha, 2011) cố gắng tìm hiểu hồ sơ người dùng thông qua một bổ sung quá trình phỏng vấn. Đối với các thuộc tính vật phẩm, mô hình chủ đề hợp tác (TLB) (Wang, & Blei, 2011) áp dụng mô hình chủ đề và phân bô Dirichlet tiềm ẩn (LDA) để tìm hiểu tính năng nội dung mục. Tuy nhiên, mô hình này chỉ hoạt động trên vấn đề dự đoán xếp hạng ngầm định và đại diện tiềm ẩn thì không học hiệu quả với thông tin nội dung rất thưa thớt.

Với những thành công lớn trong lĩnh vực hình ảnh, video và trí tuệ nhân tạo, học sâu công nghệ thu hút lợi ích lớn trong lĩnh vực hệ thống khuyến nghị (Salakhutdinov, Mnih, & Hinton, 2007; Georgiev, & Nakov, 2013; Sainath, Kingsbury, Sindhwani, Arisoy, & Ramabhad-chay, 2013). Học sâu cộng tác (CDL) (Wang et al., 2015) là một ví dụ điển hình cho thấy áp dụng học tập sâu cho các hệ thống khuyến nghị bằng cách tích hợp bộ mã hóa tự động khử nhiễu xếp chồng (SDAE) thành một mô hình CF dựa trên yếu tố tiềm ẩn đơn giản cho đề xuất phim và bài viết. Tuy nhiên, CDL chỉ tập trung vào tình huống của những người dùng hiếm hoi và các tương tác ngầm giữa người dùng và vật phẩm, và mô hình CF rất đơn giản được xem xét. Mục tiêu chính của CDL là đề xuất các mục hàng đầu N, không cho dự đoán xếp hạng rõ ràng.

Mặt khác, thông tin liên quan đến tương tác đề cập đến thông tin liên quan với hành vi tương tác UI, như dấu thời gian và vị trí xếp hạng được thực hiện. Gần đây có mối quan tâm lớn đến việc sử dụng thông tin thời gian cho CF, điều này thể hiện sự vượt trội hiệu suất đề xuất (Koren, 2010; Xiong, Chen, Huang, Schneider, & Carbonell, 2010; Zhang, Wang, Yu, Sun, & Lim, 2014). TimeSVD ++ (Koren, 2010) là mô hình mô phỏng động lực tạm thời của lợi ích người dùng bằng cách thay đổi xu hướng tĩnh và các yếu tố tiềm ẩn thành thời gian-những người phụ thuộc. Các tác giả (Xiong, Chen, Huang, Schneider, & Carbonell, 2010) giới thiệu một bộ của vector tính năng thời gian bổ sung và sử dụng hệ số tenuo để tìm hiểu các tính năng. Một sự khác biệt sơ đồ mô hình hóa theo sở thích của người dùng được trình bày trong (Zhang, Wang, Yu, Sun, & Lim, 2014), trong đó một ma trận chuyển tiếp tiềm ẩn được sử dụng để tóm tắt các tùy chọn phát triển cho mỗi người dùng. Các tác giả (Xiao, Ai, Hsu, Wang, & Jiao, 2015) đề xuất một phương pháp phụ thuộc vào thời gian để tính toán sự tương đồng giữa những người dùng khác nhau. Nhưng chúng không thể áp dụng trực tiếp cho vấn đề CS.

Nói chung, vấn đề CS có thể được phân loại thành vấn đề người dùng CS và vấn đề mục CS theo

# BẢN THẢO ĐƯỢC CHẤP NHẬN

đề xếp hạng hoàn toàn thiếu cho người dùng hoặc các mặt hàng. Đối với vấn đề người dùng CS, như hệ thống thông tin như địa điểm và giới tính không mô tả sự quan tâm của người dùng một cách hiệu quả, một số gần đây nghiên cứu cố gắng làm phong phú hồ sơ người dùng với thông tin từ các kênh khác, chẳng hạn như niềm tin xã hội mạng (Victor, Cornelis, Teredesai, & De Cock, 2008; Ma, King, & Lyu, 2011), hệ thống gắn thẻ (Zhang, Liu, Zhang, & Zhou, 2010), quá trình phỏng vấn (Zhou, Yang, & Zha, 2011). Nhưng những các loại thông tin khó thu thập trong điều kiện bình thường. Ngoài ra, khó khăn hơn để có được thông tin cá nhân của người dùng mới vì vấn đề riêng tư. Ngược lại, trọng tâm của bài viết này là về vấn đề mục CS với các động lực được mô tả trước đó. Để mà làm giảm bớt sự khan hiếm thông tin của các mặt hàng CS, hầu hết các nỗ lực nghiên cứu cho đến nay đã được dành cho

dè định hình các mục mới với thông tin bổ sung (ví dụ: thu thập các thuộc tính của mục). Tuy nhiên, vẫn còn tồn tại một số hạn chế trong các công trình nghiên cứu hiện có. Thứ nhất, khó đào ra các tính năng cụ thể của các mặt hàng mới với các thuộc tính không hạn chế. Và thu thập hạt mìn các thuộc tính như thẻ, từ khóa và danh mục luôn tồn thời gian và chi phí. Thứ hai, hầu hết các nghiên cứu kết hợp thông tin nội dung mục với dữ liệu xếp hạng (Wang, & Blei, 2011; Wang et al., 2015; Schein, Popescul, Ungar, & Pennock, 2001) áp dụng các mô hình xác suất tổng quát và có xu hướng dễ dàng vượt qua các tình huống mục CS. Vấn đề cuối cùng với những tác phẩm này là chúng không mất thời gian thông tin vào tài khoản. Trong bài báo này, một giải pháp tích hợp học tập sâu và phương pháp lọc hợp tác được đề xuất để giải quyết những hạn chế này và phần lớn cải thiện hiệu suất đề xuất cho các mục CS.

### III. PROPOSED READING MODEL

Trong phần này, chúng tôi đề xuất hai mô hình đề xuất tích hợp với CF và học sâu, được gọi là IRCD-CCS và IRCD-ICS cho các mục CCS và các mục ICS tương ứng. Một lời dè nghị hệ thống được giả định với người dùng U và các mục V không CS. Ngoài ra, nó được giả định là có J CCS các mục không nhận được xếp hạng từ người dùng cho đến thời điểm điều tra và các mục I ICS, chỉ nhận được một vài đánh giá từ người dùng. Chúng tôi dè xếp hạng  $r_{ui}(t)$  biểu thị tỷ lệ theo người dùng u trên mục i tại thời điểm t. Nhiệm vụ đề xuất được xem xét trong bài viết này là ước tính những điều chưa biết xếp hạng cho cả các mục CCS và ICS dựa trên các mục đã biết. Chúng tôi dè  $r_{ui}(t)$  biểu thị dự đoán các giá trị của  $r_{ui}(t)$ .

# BẢN THẢO ĐƯỢC CHẤP NHẬN

---

Trang 12

## BẢN THẢO ĐƯỢC CHẤP NHẬN

10

### A. Học sâu về các tính năng nội dung

Vì các mô hình CF truyền thống không thể ước tính xếp hạng cho các mục CS, nội dung bổ sung mô tả cho các mục được lấy cho mô hình đề xuất. Các tính năng mục được trích xuất từ các mô tả nội dung và được sử dụng với mô hình CF để ước tính xếp hạng mục CS.

Đầu tiên, thông tin nội dung thô của tất cả các mục được xử lý để tạo vecto dựa trên các túi từ tiếp cận. Các vecto liên quan này sau đó được SDAE học để có được các tính năng nội dung mục, sau đó được sử dụng trong các mô hình CF. SDAE là một mạng lưới sâu xếp chồng lên nhau bằng nhiều bộ khử nhiễu tự động (DAEs). Mỗi lớp SDAE được đào tạo như một DAE bằng cách giảm thiểu lỗi trong việc xây dựng lại đầu vào của nó (là đầu ra của lớp trước).

Thông thường chúng tôi coi nửa lớp đầu của mạng là phần mã hóa và nửa cuối các lớp như là một phần giải mã. Phần mã hóa cố gắng tìm hiểu các biểu diễn tính năng của nhiều đầu vào bị hỏng và phần giải mã cố gắng tự xây dựng lại đầu vào sạch trong đầu ra. An cấu trúc ví dụ của SDAE được hiển thị trong Hình 3.

Hình 3. Một cấu trúc đồ họa của SDAE.

Chính thức, đưa ra một tập hợp các vectơ dưới dạng thông tin nội dung thô của tất cả các mục, SDAE lớp L giải quyết vấn đề tối ưu hóa sau:

$$\text{tối thiểu}_{W_1, b_1} - C_L^{-2} + \frac{1}{W_1^{-2}}, \quad (1)$$

Trong đó  $C_L$  biểu thị đầu ra của lớp L của mạng và  $W_1$  và  $b_1$  biểu thị trọng số ma trận và vectơ sai lệch của lớp L của mạng. Thêm chi tiết về cấu trúc và đào tạo SDAE

### Trang 13

## BẢN THẢO ĐƯỢC CHẤP NHẬN

11

dược đề cập đến (Wang et al., 2015). Khi mô hình được đào tạo, các tính năng nội dung mục có thể là thu được từ lớp ẩn  $C_{L/2}$  của mạng. Đối với một mục i đã cho, biểu diễn tính năng, ký hiệu là  $\theta_i$ , là một vectơ có kích thước thấp.

Cần lưu ý rằng ngoài mục tiêu tìm hiểu các tính năng từ hồ sơ xếp hạng, một mục tiêu khác Mục tiêu của việc sử dụng SDAE là giảm tính chiều của các vectơ dựa trên nội dung vật phẩm giống nhau với các vectơ nhân tố tiềm ẩn, sau đó có thể được hợp nhất vào quy trình CF.

### Mô hình B. timeSVD ++

Mô hình CF được sử dụng trong IRCD được đề xuất là timeSVD++. Có một số biến thể của Mô hình thời gianSVD++. Trong bài báo này, biến thể dựa trên yếu tố tiềm ẩn được xem xét. Đối với một yếu tố tiềm ẩn mô hình dựa trên, đánh giá của người dùng u trên một mặt hàng i được tính bằng sản phẩm bên trong của một vectơ  $q_i$  (hệ số mục cho mục i) và vectơ  $p_u$  (hệ số người dùng cho người dùng u) bằng cách sử dụng như sau công thức:

$$r_{ui}(t) = q_i^T p_u. \quad (2)$$

Để đưa các thành phần, thêm vào phản hồi ngầm và các hiệu ứng thời gian, timeSVD++ sử dụng quy tắc dự đoán đã sửa đổi bằng cách thêm một số dự đoán cơ sở vào (2) như sau:

$$r_{ui}(t) = \text{Gian} + b_i(t) + b_u(t) + q_i^T [p_u(t) + |N(u)| \sum_{j \in N(u)} y_j]. \quad (3)$$

Ở đây, Quảng cáo biểu thị xếp hạng trung bình tổng thể,  $b_i(t)$  và  $b_u(t)$  biểu thị các xu hướng nhận biết theo thời gian của mục i và người dùng u tương ứng. Các yếu tố vật phẩm không thay đổi theo thời gian vì chúng có tính chất tĩnh hơn con người. Tập  $N(u)$  chứa các mục được đánh giá bởi người dùng u. Yếu tố  $|N(u)|$  chỉ ra viễn cảnh của phản hồi ngầm, trong đó  $y_j$  là một vectơ cho mục j liên quan đến ẩn

thông tin phản hồi và được học từ quá trình đào tạo.

Các thiên vị  $b_i(t)$  và  $b_u(t)$  cho các mục và người dùng, tương ứng, được tính bằng cách sau

công thức:

# BẢN THẢO ĐƯỢC CHẤP NHẬN

Cần lưu ý rằng độ lệch mục nhận biết thời gian  $b_i(t)$  bao gồm một phần đứng yên  $b_i$  và thời gian

thay đổi phần  $b_{i, \text{Bin}(t)}$ , trong đó toàn bộ dòng thời gian được chia thành các thùng dựa trên thời gian Bin (t). Dành cho người dùng

## Trang 14

### BẢN THẢO ĐƯỢC CHẤP NHẬN

12

bias  $b_u(t)$ ,  $b_u$  đại diện cho phần đứng yên,  $\alpha_u \cdot \text{dev}_u(t)$  chụp một sự trôi dạt dần dần có thể, trong

mà độ lệch thời gian  $\text{dev}_u(t)$  được định nghĩa là:

$$\text{dev}_u(t) = \text{dấu}(t - t_u) \cdot |t - t_u|^\beta, \quad (6)$$

và  $b_{u,t}$  biểu thị sự trôi dạt đột ngột trong ngày. Tương tự như xu hướng người dùng, các yếu tố người dùng cũng trở thành nhận thức thời gian như  $p_u(t)$ .

$$p_u(t) = (p_{u1}(t), p_{u2}(t), \dots, p_{ud}(t)). \quad (7)$$

$$p_{uk}(t) = p_{uk} + \alpha_{uk} \cdot \text{dev}_u(t) + p_{uk,t} \quad k = 1, \dots, d. \quad (\text{số } 8)$$

Ở đây  $d$  là chiều của các yếu tố người dùng.

Để tìm hiểu các tham số mô hình, hệ thống giảm thiểu lỗi bình phương thường xuyên về xếp hạng đào tạo:

$$\begin{aligned} & \text{tối thiểu}_{q^*, p^*, b^*} \sum_{\text{điểm tối}, t} [ (r_{ui} - r_{ui})^2 + \lambda [q_i^2 + p_u^2 + p_{ut}^2 + p_{ut}^2 \\ & + \sum_{j \in N(u)} y_j^2 + b_{ut}^2 + b_{i, \text{Thùng}(t)}^2 + b_{u,t}^2 + a_{u,t}^2 + b_{u,t}^2 ] ]. \end{aligned} \quad (9)$$

Theo hàm lỗi bình phương được chuẩn hóa ở trên, phương pháp tối ưu hóa SGD được sử dụng

để lặp đi lặp lại tìm hiểu các tham số mô hình. Các vòng lặp thuật toán thông qua tất cả các xếp hạng trong đào tạo đặt lặp và cập nhật từng tham số theo độ dốc liên quan cho đến khi hệ thống hội tụ.

#### C. Mô hình dự đoán xếp hạng cho các mục CCS

Tiếp theo chúng tôi trình bày mô hình IRCD-CCS để dự đoán xếp hạng các mục CCS. Chúng tôi trình bày lần đầu tiên sự tính toán tương tự nội dung. Hai cách tiếp cận dự đoán xếp hạng cơ sở dựa trên nội dung sự tương tự sau đó được trình bày. Cuối cùng, mô hình IRCD-CCS cho các mục CCS được trình bày, trong đó tích hợp cách tiếp cận dựa trên nội dung tương tự và mô hình timeSVD++.

Để dự đoán xếp hạng cho các mục CCS, trước tiên chúng tôi sử dụng thước đo tương tự để liên kết các mục CCS với các mục không phải là CS và dự đoán xếp hạng cho các mục CCS từ các mục không liên quan đến CS nhất của chúng. Dựa trên các tính năng và phẩm thu được từ quá trình học sâu SDAE, chúng tôi sử dụng Pearson's công thức hộ số tương quan để tính toán độ tương tự giữa các mục CCS và các mục không CS.

Đối với bất kỳ hai vectơ đặc trưng  $i$  và  $\theta_j$  của các mục  $i$  và  $j$ , độ tương tự được tính như sau:

$$\Sigma_d$$

$$S_{ij} = \frac{\sqrt{\sum_d k=1 (\theta_{ik} - \bar{\theta}_i) \cdot (j_k - \bar{\theta}_j)}}{k=1 (\theta_{jk} - \bar{\theta}_j)^2} \quad (10)$$

Trang 15

## BẢN THẢO ĐƯỢC CHẤP NHẬN

13

trong đó  $\bar{\theta}_i$  và  $\bar{\theta}_j$  là giá trị trung bình của vectơ  $\theta_i$  và  $\theta_j$ .

Chúng tôi xem xét hai cách tiếp cận cơ bản để dự đoán xếp hạng cho các mục CCS. Cách tiếp cận đầu tiên dự đoán xếp hạng cho các mục CCS từ các mục M không giống CS nhất của chúng trong toàn bộ mục không CS được đặt sau khi xếp hạng bị thiếu cho các mục không CS được dự đoán, được gọi là mục Cách tiếp cận hàng đầu (ToA). Đặt  $S_M(j)$  biểu thị tập hợp các mục M không tương tự CS nhất với một mục CCS  $j$ ,  $j [1, J]$ . Đối với phương pháp ToA, công thức sau đây được sử dụng để dự đoán xếp hạng bởi người dùng u trên CCS mục j:

$$r_{uj} = \frac{\sum_{i \in S_M(j)} r_{ui} S_{ij}}{\sum_{i \in S_M(j)} S_{ij}} \quad (11)$$

Cần lưu ý rằng xếp hạng  $r_{ui}$  có thể là cả thực tế và dự đoán.

Ngoài ra, chúng tôi có thể sử dụng cách tiếp cận thứ hai để dự đoán xếp hạng của người dùng cho các mục CCS từ các mục không phải CS tương tự M của chúng trong tập hợp các mục không CS được người dùng xếp hạng, được gọi là phương pháp tiếp cận người dùng hàng đầu (ToU). Đặt  $S_M(u, j)$  biểu thị tập hợp M tương tự nhất các mục không CS trong số các mục không CS được u xếp hạng cho mục CCS  $j$ ,  $j [1, J]$ . Dành cho ToU

Cách tiếp cận, công thức sau đây được sử dụng để dự đoán xếp hạng của người dùng u trên mục CCS j:

$$r_{uj} = \frac{\sum_{i \in S_M(u, j)} r_{ui} S_{ij}}{\sum_{i \in S_M(u, j)} S_{ij}} \quad (12)$$

Lưu ý rằng trong trường hợp này, xếp hạng  $r_{ui}$  là xếp hạng thực trong tập huấn luyện.

Hai cách tiếp cận đơn giản và dễ hiểu ở trên giúp dự đoán xếp hạng hoàn toàn dựa trên trên các mục không phải CS tương tự và bỏ qua các thông tin khác trong ma trận xếp hạng. Theo chúng tôi các thí nghiệm, phương pháp ToU có độ chính xác cao hơn phương pháp ToA. Tiếp theo chúng tôi đề xuất một mô hình tích hợp bằng cách kết hợp cách tiếp cận ToU và timeSVD ++ với nhau.

Vì các phương thức CF thông thường không thể áp dụng trực tiếp cho các vấn đề CCS, chúng tôi đặt mục

yếu tố  $q_i$  trong mô hình timeSVD ++ thành tính năng nội dung mục  $\theta_i$  và thay thế xếp hạng trung bình chung

Xu hướng và vật phẩm  $b_u(t)$  với xếp hạng dự đoán, được tạo bởi phương pháp ToU. Chúng tôi

mở rộng cách tiếp cận ToU để tạo xếp hạng dự đoán cho tất cả các xếp hạng thực trong tập huấn luyện.

## BẢN THẢO ĐƯỢC CHẤP NHẬN

Quy tắc dự đoán sau đây được sử dụng trong mô hình IRCD-CCS được đề xuất:

$$r_{ui}(t) = b_u(t) + \theta_i^T [p_u(t) + |N(u)| \sum_{j \in R(u)} y_j] + \sum_{j \in S_M(u, i)} \frac{r_{uj} S_{ij}}{\sum_{j \in S_M(u, i)} S_{ij}} \quad (13)$$

và hàm lỗi bình phương chuẩn hóa tương ứng là:

$$\text{tối thiểu}_{p^*, b^*} \sum_{\text{tùn, tối, t}} [(r_{ui} - r_{ui})^2 + \lambda [p_u^2 + p_{uu}^2 + p_{ut}^2 + \sum_{j \in N(u)} |y_j|^2 + b_u^2 + a_{u-2}^2 + b_{u,t-2}^2]]. \quad (14)$$

Vì tất cả các tham số đều liên quan đến người dùng, nên nó có thể được sử dụng để dự đoán xếp hạng cho các mục CCS sau khi đào tạo.

Cần lưu ý rằng trong mô hình IRCD-CCS cho các mục CCS, các mục CCS không nhận được bất kỳ đánh giá từ người dùng. Do đó, các mục CCS không tham gia vào đào tạo mô hình timeSVD ++.

Tuy nhiên, để cho phép dự đoán theo quy tắc (13), chúng tôi huấn luyện mô hình IRCD-CCS với xếp hạng ma trận bằng cách đặt các yếu tố mục  $q_i$  của các mục không CCS thành tính năng nội dung của chúng theta  $i$ . Nó có nghĩa là rằng các tính năng nội dung học được từ SDAE được sử dụng thay vì các tính năng mục bị ẩn trong ma trận đánh giá trong mô hình IRCD-CCS cho các mục CCS.

Hình .III-C cho thấy một khung đồ họa của mô hình MF truyền thống (phản bên trái) và IRCD-CCS mô hình (phản bên phải). Đối với mỗi người dùng  $u$  và mục  $i$ , mô hình MF truyền thống sẽ tính toán dự đoán đánh giá  $r_{ui}$  bằng cách thêm các yếu tố tiềm ẩn với yếu tố dự đoán sai lệch. Yếu tố tiềm ẩn là sản phẩm bên trong của yếu tố mục  $q_i$  và yếu tố người dùng  $p_u$ . Công cụ dự đoán thiên vị bao gồm xếp hạng trung bình tổng thể, độ lệch của vật phẩm  $b_i$  và thiên vị người dùng  $b_u$ . So với MF truyền thống, mô hình IRCD-CCS lần đầu tiên áp dụng SDAE để tìm hiểu tính năng nội dung mục  $\theta_i$  từ thông tin nội dung mục  $\theta_i$  C. Sau đó, ToU Cách tiếp cận được sử dụng để có được một đánh giá dự đoán sơ bộ  $r_{ui}$  dựa trên thước đo tương tự của tính năng nội dung mục. Trong hệ số mục đào tạo mô hình  $q_i$  được đặt thành tính năng nội dung mục  $\theta_i$ . Cuối cùng xếp hạng có thể được dự đoán bằng cách thêm hai phần cùng với quy tắc được hiển thị trong (13).

Cần lưu ý rằng mô hình IRCD-CCS cần xử lý trước để tạo dự đoán xếp hạng cho mọi xếp hạng thực trong tập huấn luyện ngoại tuyến, nhưng dự đoán trực tuyến có thể được tính ngay lập tức với quy tắc dự đoán.

#### D. Mô hình dự đoán xếp hạng cho các mục ICS

Mô hình IRCD-ICS sửa đổi timeSVD ++ bằng cách áp dụng các tính năng nội dung được học từ SDAE vào mục quá trình đào tạo nhân tố tiềm ẩn. Trong mô hình đề xuất IRCD-ICS cho các mục của chúng tôi, quy tắc dự đoán (3) được sử dụng lại, nhưng các tham số mô hình được học bằng cách giảm thiểu khác nhau chúc năng lỗi bình phương bình thường. Hàm lỗi bình phương đã sửa đổi được hiển thị

## BẢN THẢO ĐƯỢC CHẤP NHẬN

Hình 4. Sửa đổi đồ họa của khung IRCD-CCS.

phía dưới:

$$\begin{aligned} \text{tối thiểu } & (r_{ui} - r_{ui})^2 + \lambda [q_i - i] + p_u^2 + p_{uu}^2 + p_{ut}^2 \\ q^*, p^*, b^* & \sum_{\text{bạn, tôi, t}} \\ & + \sum_{j \in N(u)} y_j^2 + b_{i, \text{Thùng}(t)}^2 + b_{u, t}^2 + a_{u, t}^2 + b_{u, t}^2. \end{aligned} \quad (15)$$

Lưu ý rằng chiều của cả  $q_i$  và  $\theta_i$  là d.

Cần lưu ý rằng mô hình IRCD-ICS được đào tạo được sử dụng cho cả các mục ICS và phi CS nói chung (NCS) các mục (với số lượng xếp hạng tương đối lớn hơn). Chúng tôi không tạo đề xuất riêng các mô hình cho các mục ICS và các mục NCS chung. Đối với các mục ICS, bạn nên tìm hiểu yếu tố mục từ mô tả nội dung càng nhiều càng tốt, vì có rất ít hữu ích thông tin cần học cho các mục ICS từ ma trận xếp hạng. Do đó, các tính năng nội dung đóng một vai trò quan trọng trong khuyến nghị cho các mặt hàng ICS. Đối với các mục NCS, cả hai ma trận đánh giá và mô tả nội dung (tính năng nội dung) có thể được sử dụng và nên được sử dụng để cải thiện đánh giá hiệu suất dự đoán. Giới thiệu nhân tố  $q_i - i_2$  có thể giúp đạt được mục tiêu về các yếu tố mục học tập cho cả hai mục ICS và NCS.

So với timeSVD++, các thay đổi chính sau đây được thực hiện theo quy tắc cập nhật của vectơ yếu tố vật phẩm. Với mỗi xếp hạng nhất định  $r_{ui}(t)$ , lỗi dự đoán được tính bởi  $e_{ui}(t) = r_{ui}(t) - r_{ui}(t)$ . Phương trình cập nhật ban đầu của yếu tố vật phẩm  $q_i$  đã sử dụng trong timeSVD++ là:

$$q_i \leftarrow q_i + \gamma \{e_{ui}(t)[p_u(t) + |N(u)|^{-1} \sum_{j \in N(u)} y_j] - \lambda \cdot q_i\}, \quad (16)$$

Trong mô hình đề xuất, quy tắc cập nhật sau cho yếu tố vật phẩm  $q_i$  được sử dụng:

$$q_i \leftarrow q_i + \gamma \{e_{ui}(t)[p_u(t) + |N(u)|^{-1} \sum_{j \in N(u)} y_j] - \lambda \cdot (q_i - \theta_i)\}, \quad (17)$$

trong đó  $\gamma$  biểu thị tỷ lệ học tập. Các độc giả quan tâm được tham khảo (Koren, 2010) để biết thêm chi tiết về quy trình học mô hình timeSVD++. Hình III-D cho thấy một khung đồ họa cho

## Trang 18

### BẢN THẢO ĐƯỢC CHẤP NHẬN

16

mô hình MF truyền thống (phản bên trái) và mô hình IRCD-ICS (phản bên phải). Đối với mô hình IRCD-ICS, tính năng nội dung mục  $\theta_i$  được sử dụng để tìm hiểu yếu tố mục  $q_i$  trong quy trình đào tạo theo quy tắc (17).

Hình 5. Sửa đổi đồ họa của khung IRCD-ICS.

#### IV. P GIÁ TRI E GIÁ TRI

Trong phần này, hiệu suất đề xuất của các mô hình được đề xuất IRCD-CCS cho CCS các mục và IRCD-ICS cho các mục ICS được đánh giá. Việc chuẩn bị dữ liệu thí nghiệm và kết quả được trình bày và thảo luận.

##### A. Bộ dữ liệu và cài đặt thử nghiệm

Một bộ dữ liệu lớn trong thế giới thực được tạo bởi Giải thưởng Netflix được sử dụng để đánh giá các mô hình được đề xuất. Bộ dữ liệu Netflix chứa hơn 100 triệu xếp hạng rõ ràng trên thang điểm từ 1 đến 5 sao cho 17770 phim được xác định bởi 480189 người dùng ẩn danh. Khoảng thời gian tương ứng từ Ngày 12 tháng 12 năm 1999 đến ngày 12 tháng 12 năm 2005.

Để dự đoán xếp hạng cho các mục CS, chúng tôi cũng thu thập các lô phim từ IMDB để trích xuất thông tin nội dung mục. Đầu tiên chúng tôi đã thu thập các lô phim tương ứng của OMDb API<sup>1</sup>, một dịch vụ web miễn phí để lấy thông tin phim. Một chương trình dựa trên Python là được viết, đi qua các bộ phim trong bộ dữ liệu Netflix và tự động gửi yêu cầu tìm kiếm cốt truyện theo tiêu đề phim cho cơ sở dữ liệu OMDb. Sau đó, các lô phim thu thập được được lọc bằng cách loại bỏ các từ dừng, trong đó đề cập đến các từ phổ biến nhất trong dịch vụ ngôn ngữ ít thông tin hữu ích. Trong các thí nghiệm, danh sách các từ dừng được thu được với tích hợp

<sup>1</sup><http://www.omdbapi.com>

# BẢN THẢO ĐƯỢC CHẤP NHẬN

---

Trang 19

BẢN THẢO ĐƯỢC CHẤP NHẬN

17

Gói Python. Sau đó, chúng tôi tính toán tần số tài liệu nghịch đảo tần số (tf-idf) của từng chữ trong kho văn bản. Tf-idf là sản phẩm của tần số thuật ngữ và tần số tài liệu nghịch đảo. Tần suất thuật ngữ đề cập đến số lần xuất hiện của các từ trong tài liệu và tài liệu nghịch đảo tần suất được tính bằng cách chia tổng số tài liệu cho số lượng tài liệu chứa từ. Nói chung tf-idf được sử dụng như một yếu tố trọng số để phản ánh tầm quan trọng của một từ là một tài liệu. Dựa trên tf-idf, các từ S quan trọng nhất có giá trị tf-idf cao nhất được chọn để tạo thành một từ điển. S được đặt thành 20000 trong các thí nghiệm của chúng tôi. Mỗi cốt truyện phim là được biểu thi bằng một vectơ túi chữ S chiều, trong đó mỗi mục của vectơ biểu thi số lượng từ tương ứng xảy ra trong cốt truyện. Bước cuối cùng là chuẩn hóa cho tất cả vectơ. Sau khi các phim bị mất cốt truyện bị xóa, bộ dữ liệu cuối cùng có 476691 người dùng, 14657 phim và xếp hạng 95975845.

Một quy trình lấy và xử lý các sơ đồ phim được trình bày trong Hình 6.

Hình 6. Quy trình xử lý dữ liệu tiền xử lý trên các bộ phim.

**BẢN THẢO ĐƯỢC CHẤP NHẬN**

Vì chúng tôi quan tâm đến việc thực hiện dự đoán xếp hạng cho cả các mục CCS và ICS các mục, tập dữ liệu gốc được phân vùng thành một tập huấn luyện và một tập kiểm tra cho các mục CCS và Các mặt hàng ICS, tương ứng. Các bộ phim được sắp xếp theo dấu thời gian của xếp hạng nhận được đầu tiên của họ. Biểu đồ của phim vào ngày xếp hạng đầu tiên của chúng được hiển thị trong Hình 7. Chúng tôi chia toàn bộ

---

Trang 20**BẢN THẢO ĐƯỢC CHẤP NHẬN**

18

thời gian của bộ dữ liệu (2240 ngày) thành 100 khoảng thời gian và tính xem có bao nhiêu phim rơi vào mỗi khoảng. Nói chung, các bộ phim mới xuất hiện muộn, và do đó nằm ở phía bên phải trong hình. Để chuẩn bị tập huấn và kiểm tra cho các thí nghiệm CCS, chúng tôi chọn cách gần đây nhất Phim L cho tập thử nghiệm và các phim khác cho tập huấn luyện. Xếp hạng của phim L CCS được sử dụng để thử nghiệm và được đảm bảo rằng không ai có thể xem bất kỳ phim CCS nào trong số này người dùng trong tập huấn luyện. Để chuẩn bị các bộ kiểm tra và huấn luyện ICS, chúng tôi chọn bộ mới nhất Phim K cho tập thử nghiệm và các phim khác cho tập huấn luyện. Trong khi đó xếp hạng N sớm nhất của mỗi bộ phim được thử nghiệm được thêm vào tập huấn luyện và tất cả các xếp hạng còn lại của các K ICS này phim được sử dụng cho mục đích thử nghiệm.

Hình 7. Biểu đồ phim vào ngày xếp hạng đầu tiên của họ.

Đối với cài đặt thử nghiệm của một số tham số, chúng tôi đã chọn một số giá trị tiêu biểu cho hiệu suất đánh giá. Các mô hình đề xuất được đề xuất có thể hoạt động với các cấu hình chung của thông số. Vì có một khung giá trị lớn cho các tham số, nên rất tốn thời gian để kiểm tra tất cả các kết hợp cài đặt cho các tham số. Sau khi kiểm tra nhanh về recom-

hiệu suất sửa đổi với một số cài đặt tham số, chúng tôi đã xác định cài đặt cho  
tham số mô hình (ví dụ: cài đặt kích thước d của vectơ đặc trưng thành 50) với các tham số tốt  
hiệu suất.

Đối với các bộ dữ liệu thử nghiệm, số liệu thống kê xếp hạng cho các bộ huấn luyện và kiểm tra được trình bày trong  
Bảng I và Bảng II cho thí nghiệm CCS và thí nghiệm ICS, tương ứng.

Trước tiên, chúng tôi tiến hành thử nghiệm cho phim CCS và so sánh mô hình IRCD-CCS của chúng tôi với  
ba mô hình đề xuất: mô hình ToA, mô hình ToU và mô hình trung bình đơn giản (SA). ToA  
mô hình và mô hình ToU được trình bày trong Sec. III-C. Là các mô hình hiện có cho dự đoán mục CCS  
sử dụng các thông tin và tính năng nội dung khác nhau, chúng không thể so sánh trực tiếp với đề xuất

# BẢN THẢO ĐƯỢC CHẤP NHẬN

Trang 21

## BẢN THẢO ĐƯỢC CHẤP NHẬN

19

### BÀNG I

S TATISTICS CỦA CÁC ĐÀO TẠO VÀ KIỂM TRA KIỂM TRA CHO TRẠI NGHIỆM CCS MOVIE .

	L = 100	L = 300		
	Tập huấn luyện	Tập kiểm tra	Tập huấn luyện	Tập kiểm tra
Số lượng người dùng	476691	11764	476691	51171
Số lượng phim 14557		100	14357	300
Số lượng xếp hạng	95959733	16112	95874146	101699
Tỷ lệ trung bình	3,6042	3,4975	3,6043	3,5652

### BÀNG II

S TATISTICS CỦA ĐÀO TẠO VÀ THI Datasets CHO ICS MOVIE THÍ NGHIỆM (N = 5).

	K = 100	K = 200			K = 300
	Tập huấn luyện	Tập kiểm tra	Tập huấn luyện	Tập kiểm tra	Tập huấn luyện
Số lượng người dùng	476691	11603	476691	31725	476691
Số lượng phim 14657		100	14657	200	14657
Số lượng xếp hạng	95960233	15612	95924771	51074	95875646
Tỷ lệ trung bình	3,6042	3,5135	3,6043	3,4288	3,6043
					3,5718

mô hình. Do đó, mô hình dự đoán đơn giản SA cho các mục CCS được sử dụng: mọi xếp hạng mục CCS  
bởi người dùng được đặt ở mức trung bình xếp hạng của người dùng.

Mặt khác, các mô hình cơ sở cho thử nghiệm phim ICS bao gồm ALS, SGD,  
thời gianSVD ++ và CDL. ALS và SGD là hai thuật toán chính cho các tham số học tập. Trong  
thử nghiệm của chúng tôi, chúng tôi sử dụng chúng để giảm thiểu chức năng lỗi của (2). Như được mô tả trong phần trước  
phản, timeSVD ++ là một mô hình CF nhận biết thời gian, xem xét các hiệu ứng thời gian.

CDL là mô hình cùng thực hiện học tập đại diện sâu và CF, áp dụng ALS-  
thuật toán phong cách cho học tập. ALS và SGD cung cấp khuyến nghị dựa trên xếp hạng đơn giản,  
trong khi timeSVD ++ và CDL sử dụng thông tin thời gian bổ sung và thông tin nội dung mục  
tương ứng.

# BẢN THẢO ĐƯỢC CHẤP NHẬN

Quy trình thử nghiệm tổng thể có bốn bước chính, được mô tả dưới đây:

.Cấu hình các tham số hệ thống: các tham số sẽ được cấu hình bao gồm tốc độ học tập,

chính quy hóa và kích thước nhân tố cho từng mô hình như được mô tả trong Sec. IV-A. Những cài đặt này vẫn như vậy trong suốt các thí nghiệm.

**Trang 22****BẢN THẢO ĐƯỢC CHẤP NHẬN**

20

- . Chuẩn bị các bộ huấn luyện và kiểm tra: như được mô tả trong Sec. IV-A, tập huấn luyện và kiểm tra được tạo ra tương ứng theo các cấu hình với các kích thước thử nghiệm khác nhau (L cho thí nghiệm CCS và K cho thí nghiệm ICS). Hơn nữa, số lượng mẫu đào tạo cho các hạng mục ICS được thử nghiệm N cũng được cấu hình cho các giá trị khác nhau trong các thí nghiệm ICS.
- . Huấn luyện mô hình bằng cách khớp dữ liệu trong tập huấn luyện: cụ thể, ALS và CDL áp dụng ALS thuật toán để cập nhật các tham số mô hình, trong khi các mô hình IRCD, SGD và timeSVD ++ tìm hiểu tham số mô hình bằng thuật toán SGD. Phương pháp tiếp cận ToA và ToU tính toán dự đoán kết quả trực tiếp sử dụng phương trình. (11), (12).
- . Dự đoán xếp hạng trong bộ kiểm tra: điều này được thực hiện bằng cách sử dụng các mô hình được đào tạo. Dự đoán hiệu suất sau đó được tính toán và ghi lại.

***B. Kết quả và phân tích***

Trong tiêu mục này, các mô hình đề xuất được đánh giá theo RMSE và được so sánh với các mô hình đề xuất cơ bản khác. RMSE là một số liệu khách quan được sử dụng rộng rãi cho đánh giá hiệu suất của các mô hình hệ thống khuyến nghị, được định nghĩa là:

$$\text{RMSE} = \sqrt{\frac{1}{N_p} \sum_{\text{bạn, tôi}} (r_{ui} - \hat{r}_{ui})^2}, \quad (18)$$

Trong đó  $N_p$  biểu thị tổng số dự đoán.

Sau khi điều chỉnh các tham số trên bộ xác thực, chúng tôi so sánh các mô hình IRCD được đề xuất của chúng tôi với mô hình cơ sở khác. Kích thước d của các vectơ đặc trưng được đặt thành 50 cho tất cả các mô hình.

*I) Đánh giá hiệu suất của Mô hình IRCD-CCS trên Phim CCS:* Trước tiên, chúng tôi đánh giá hiệu suất dự đoán của mô hình IRCD-CCS cho các mục CCS. Bảng III trình bày dự đoán kết quả RMSE, so với số lượng các mục liên quan nhất M, kích thước của tập dữ liệu thử nghiệm cho CCS phim, L cho CCS phim mới. Mô hình SA, mô hình ToA, mô hình ToU và mô hình IRCD-CCS được so sánh. Số M của hầu hết các mục liên quan được định cấu hình là 20 và 100 và kích thước

Trong bộ dữ liệu thử nghiệm cho phim CCS được định cấu hình lần lượt là 100 và 300.

Cần lưu ý rằng mô hình IRCD-CCS thực hiện tốt nhất cho tất cả các kích thước được điều tra. Nó là hiệu suất tốt hơn đáng kể so với các mô hình cơ bản. Kết quả cho thấy một sự cải thiện khoảng 0,05 trên RMSE so với ToU mô hình tốt thứ hai. Hiệu suất của cả hai

Mô hình IRCD-CCS và mô hình ToU cải thiện phần lớn khi M tăng. Ngược lại, ToA mô hình hoạt động tốt với M nhỏ (ví dụ: chỉ có 20 mục ICS liên quan nhất để dự đoán xếp hạng),

**BẢN THẢO ĐƯỢC CHẤP NHẬN**

**Trang 23****BẢN THẢO ĐƯỢC CHẤP NHẬN**

21

nhưng với M lớn như 100 thì nó có hiệu suất kém, thậm chí còn tệ hơn SA  
mô hình. Điều này có thể được giải thích là ảnh hưởng của lỗi dự đoán được tích lũy là M  
tăng.

BÀNG III

PHƯƠNG PHÁP GIẢI QUYẾT CÁC MÔ HÌNH PREDICTION CHO PHIM CCS VỚI N ETFLIX DATASET .

Phương pháp tiếp cận	ToA	ToU		IRCD-CCS		SA
	M = 20	M = 100	M = 20	M = 100	M = 20	
RMSE (L = 100)	1.155	1.224	1.133	1.113	1.075	<b>1.053</b>
RMSE (L = 300)	1.134	1.218	1.140	1.127	1.096	<b>1.082</b>

2) *Dánh giá hiệu suất của Mô hình IRCD-ICS trên Phim ICS:* Tiếp theo, chúng tôi so sánh trước hiệu suất từ điển RMSE của mô hình IRCD-ICS với các mô hình hiện có cho phim ICS.  
Bảng IV trình bày kết quả thí nghiệm. Trong các thí nghiệm số lượng đào tạo liên quan xếp hạng N được đặt thành 5. Kích thước K của các phim ICS được thử nghiệm được đặt thành 100, 200 và 300. Đó là quan sát thấy rằng phương pháp được đề xuất IRCD-ICS đạt được độ chính xác tốt nhất trong mọi trường hợp. CDL thực hiện điều tồi tệ nhất mặc dù nó áp dụng SDAE cho việc học thông tin nội dung. Chính Lý do là CDL được đề xuất về việc sử dụng dữ liệu xếp hạng ngầm thay vì dữ liệu rõ ràng. Các khoảng cách lớn (khoảng 0,05) giữa SGD và ALS chứng tỏ rằng SGD có hiệu quả trong việc tạo ra dự đoán tốt hơn cho các mục ICS hơn ALS. Bằng cách mô hình hóa động lực học thời gian, timeSVD ++ vượt trội so với SGD với mức chênh lệch đáng kể từ 0,002 đến 0,017. So với timeSVD ++, IRCD-  
Mô hình ICS cho thấy sự cải thiện nhất quán hơn 0,004 với nội dung bao gồm thông tin và quá trình học tập sâu.

Để so sánh sâu về timeSVD ++ và mô hình IRCD-ICS được đề xuất, chúng tôi điều tra RMSE hiệu suất dự đoán thay đổi như thế nào với các lần lặp đào tạo. Tiêu biểu đường cong đào tạo được trình bày trong hình IV-B2. Đối với cả hai mô hình, RMSE giảm đơn điệu mà không có vấn đề quá mức. Mô hình IRCD-ICS hội tụ nhanh hơn thời gian SVD ++ trong quá trình đào tạo. Đối với trường hợp K là 100, mô hình IRCD-ICS chỉ mất 25 lần lặp để kết thúc trong khi timeSVD ++ cần nhiều hơn 35 lần lặp, có nghĩa là thời gian tính toán nhiều hơn.

3) *Dánh giá hiệu suất của cả hai mô hình trên phim ICS:* Nói chung nếu một mặt hàng có đủ số lượng xếp hạng mô hình IRCD-ICS cho mặt hàng này chắc chắn sẽ vượt trội so với

**BẢN THẢO ĐƯỢC CHẤP NHẬN****Trang 24****BẢN THẢO ĐƯỢC CHẤP NHẬN**

22

BÀNG IV

P HƯỚNG DẪN GIẢI QUYẾT CÁC MÔ HÌNH PREDICTION CHO PHIM ICS VỚI N Dataset ETFLIX .

Thuật toán	RMSE (K = 100)	RMSE (K = 200)	RMSE (K = 300)
ALS	1.124	1.112	1.097
SGD	1.070	1.076	1.058
thời gianSVD ++	1.053	1.074	1.053
CDL	1.179	1.151	1.148
IRCD-ICS	<b>1,049</b>	<b>1.070</b>	<b>1.048</b>

Mô hình IRCD-CCS. Nhưng khi số lượng xếp hạng cho một mục ICS gần bằng 0, thì đó là không rõ liệu mô hình IRCD-ICS hay mô hình IRCD-CCS hoạt động tốt hơn. Do đó trong này tiêu mục chúng tôi so sánh hai mô hình IRCD-CCS và IRCD-ICS về đề xuất các mặt hàng ICS. Để đánh giá và so sánh các mô hình dưới các mức độ khác nhau của ma trận đánh giá, số lượng xếp hạng đào tạo liên quan N được đặt thành 1, 3, 5 và 7 cho phim ICS. Hình 9 cho thấy RMSE kết quả của timeSVD ++ và mô hình IRCD-ICS với N khác nhau cho các phim ICS. Mau xanh đường nét đứt cho biết RMSE của mô hình IRCD-CCS, không liên quan đến N. Như chúng ta có thể thấy từ Hình 9, giá trị RMSE của cả timeSVD ++ và IRCD-ICS giảm khi N tăng. Tuy nhiên, khi N nhỏ hơn 5, các mô hình dựa trên mục ICS bao gồm timeSVD ++ và IRCD-ICS thậm chí hoạt động kém hơn so với mô hình IRCD-CCS. Nó được chỉ ra rằng các mô hình dựa trên vật phẩm ICS không dự đoán tốt về xếp hạng cho các mục chỉ với một số lượng nhỏ xếp hạng, trong trường hợp đó mô hình dựa trên mục CCS được ưa thích.

### C. Thảo luận

Trong các thử nghiệm trên, các mô hình đề xuất khác nhau được sử dụng cho các mục CCS và các mặt hàng không CCS (ICS và NCS). Tuy nhiên, trong hoạt động thực tế của các hệ thống khuyến nghị, nếu xếp hạng cho một mục CCS được nhận từ người dùng, mục đó trở thành mục ICS. Trong trường hợp này, đổi với dữ liệu xếp hạng mục ICS mới này với mô hình được đào tạo và sử dụng cho các mục CCS có thể là tồi tệ hơn với một mô hình được sử dụng cho các mục ICS. Tuy nhiên, mô hình đề xuất hiện có cho các mục không phải CCS không sử dụng bất kỳ thông tin nào (xếp hạng và mô tả nội dung) liên quan đến điều này. Mục ICS, chỉ thay đổi trạng thái của nó từ mục CCS. Do đó, cần phải đào tạo lại

## BẢN THẢO ĐƯỢC CHẤP NHẬN

Trang 25

BẢN THẢO ĐƯỢC CHẤP NHẬN

23

mô hình đề xuất cho các mục không phải CCS có thêm thông tin mô tả nội dung và xếp hạng từ các mục không CCS mới.

Có hai vấn đề chính cần xem xét liên quan đến việc đào tạo lại mô hình cho người không CCS mục: 1) tần suất mô hình nên được đào tạo lại? 2) phần nào của mô hình nên đào tạo lại?

Đối với vấn đề tần suất đào tạo lại, chúng tôi không cần phải ngay lập tức đào tạo lại mô hình ICS. Bất cứ khi nào một số mặt hàng thay đổi trạng thái của họ từ CCS sang ICS. Theo khuyến nghị của các mặt hàng

không được thực hiện liên tục, không cần phải đào tạo lại mô hình ICS quá thường xuyên, điều này sẽ phát sinh chi phí tính toán rất cao. Ngay cả khi mô hình IRCD-ICS không được đào tạo lại khi CCS

thay đổi mục thành mục ICS, mô hình CCS có thể được sử dụng tạm thời để đề xuất điều đó

mục. Theo kết quả đánh giá, không có sự suy giảm hiệu suất lớn khi

Mô hình CCS được áp dụng cho các mục ICS. Do đó, chúng tôi có thể thiết kế một hệ thống khuyến nghị, trong đó

thu thập và lưu trữ các xếp hạng mới do người dùng thực hiện và thường xuyên đào tạo lại ICS và CCS

mô hình đề xuất, ví dụ, trong thang ngày hoặc tuần theo xếp hạng

các hoạt động.

Đối với các vấn đề mà các bộ phận của các mô hình sẽ được đào tạo, có rất ít tác động trong mô hình được đào tạo với việc đưa các xếp hạng mới vào tập huấn luyện, vì chúng chiếm tỷ lệ rất nhỏ trong tổng số xếp hạng trong tập huấn luyện. Để giảm tài tính toán, trong đào tạo lại mô hình, chúng tôi có thể giữ các giá trị cho các tham số mô hình được đào tạo (như  $q_i$ ,  $p_u$  và  $b_i$ ) được đào tạo tốt trong vòng trước của quá trình đào tạo và nên ổn định trong thời gian ngắn, và chỉ học giá trị tham số cho các mục mới.

#### V. CONCLUSION

Khuyến nghị về các mặt hàng bắt đầu lạnh là một thách thức và vẫn là một vấn đề nghiên cứu mở cho hệ thống khuyến nghị. Các mục bắt đầu lạnh có thể được phân loại để hoàn thành các mục bắt đầu lạnh (CCS)

không nhận được xếp hạng và các mục bắt đầu lạnh không hoàn chỉnh (ICS) nhận được nhiều hơn 0

nhưng rất ít xếp hạng. Trong bài báo này, chúng tôi đã đề xuất hai mô hình đề xuất để giải quyết vấn đề đề xuất cho các mục CCS và ICS, tương ứng. Các mô hình kết hợp một thời gian

nhận biết mô hình lọc cộng tác (CF) timeSVD ++ với kiến trúc SDAE học tập sâu.

Mạng lưới thần kinh học tập sâu SDAE chịu trách nhiệm khai thác các tính năng nội dung mục,

trong khi mô hình timeSVD ++ chịu trách nhiệm dự đoán các xếp hạng không xác định. Nó đã xem xét

# BẢN THẢO ĐƯỢC CHẤP NHẬN

Trang 26

## BẢN THẢO ĐƯỢC CHẤP NHẬN

24

động lực thời gian của sở thích người dùng và các tính năng mục. Một số lượng lớn các thí nghiệm là chạy để đánh giá các mô hình được đề xuất về mặt dự đoán lỗi RMSE trên

Tập dữ liệu Netflix. Kết quả cho thấy các mô hình của chúng tôi vượt trội so với các phương pháp cơ bản hiện có cho đề nghị bắt đầu lạnh. Từ phân tích và thí nghiệm của chúng tôi, tác động của bao gồm

thời gian và thông tin nội dung mục là rất lớn. Đặc biệt đối với vấn đề CCS, mô hình của chúng tôi

có thể tận dụng thành công lợi thế của các mô hình nhân tố tiềm ẩn CF để đạt được hiệu suất đáng kể

cải thiện. Ngoài ra, chúng tôi cũng so sánh các mô hình được đề xuất của chúng tôi về các mặt hàng mới của ICS

ommendation với mức độ khác nhau của ma trận đánh giá độ thưa thớt. Nó đã được tìm ra rằng ICS

mô hình dựa trên vật phẩm không đưa ra khuyến nghị tốt cho các mặt hàng nhận được rất ít xếp hạng

(ví dụ 3 xếp hạng). Trong trường hợp đó, nên sử dụng mô hình dựa trên mục CCS thay vì mục ICS

mô hình dựa trên.

Trong tương lai, chúng tôi có kế hoạch mở rộng các mô hình đề xuất cho các mặt hàng và công việc bắt đầu lạnh về các hướng nghiên cứu sau đây. Đầu tiên, chúng tôi quan tâm đến việc điều tra recom-  
hiệu suất sửa đổi cho các mục CCS và ICS với nhiều thông số và cấu hình hệ thống hơn

thiết lập, để tiết lộ thêm thông tin chi tiết về tác động của chúng đối với hiệu suất đề xuất và  
Tối ưu hóa hệ thống. Thứ hai, chúng tôi tạo và duy trì hai mô hình đề xuất riêng cho

Các mặt hàng CCS và các mặt hàng ICS, tương ứng. Cách tiếp cận này đòi hỏi lưu trữ và tính toán thêm

tài nguyên. Chúng tôi dự định thiết kế một mô hình đề xuất, áp dụng cho khuyến nghị

của cả hai mặt hàng CCS và ICS. Thứ ba, các mô hình đề xuất được đánh giá bởi RSME đánh giá

dự đoán, có thể không phản ánh hiệu quả hiệu suất của các hệ thống khuyến nghị thực sự.

Chúng tôi quan tâm đến việc thiết kế một phương pháp đánh giá hiệu suất bổ sung, có thể

xem xét các quyết định đề xuất mục và định lượng tác động của các quyết định

người dùng chấp nhận các mặt hàng được đề nghị. Cuối cùng, trong bài báo này, chúng tôi chạy thử nghiệm khởi đầu lạnh

đề xuất mục trên phim Netflix. Chúng tôi quan tâm đến việc áp dụng các mô hình đề

khuyến nghị của các sản phẩm khác như âm nhạc trực tuyến.

# BẢN THẢO ĐƯỢC CHẤP NHẬN

MỘT NỀN TẢNG

Công trình được hỗ trợ một phần bởi Quỹ khoa học tự nhiên quốc gia Trung Quốc (Cấp số 61221001) và Chương trình 111 (B07022) và Phòng thí nghiệm chính của Thượng Hải về Xử lý phương tiện kỹ thuật số và Truyền STCSM (14XD1402100).

Trang 27

## BẢN THẢO ĐƯỢC CHẤP NHẬN

25

### R NĂNG LỰC

Agarwal, D., & Chen, BC (2009). "Mô hình nhân tố tiềm ẩn Regression-based." *Trong Kỷ yếu*

*hội nghị quốc tế ACM SIGKDD lần thứ 15 về khám phá tri thức và khai thác dữ liệu*,

19-28.

Bennett, J., & Lanning, S. (2007). Giải thưởng Netflix. *Tiếp tục trong quá trình thực hiện cúp và hội thảo của KDD*,  
2007, 35.

Chen, T., Zhang, W., Lu, Q., Chen, K., Zheng, Z., & Yu, Y. (2012). Phần mềm SVDFeature: một bộ công cụ  
để lọc cộng tác dựa trên tính năng. *Tạp chí Nghiên cứu Máy học*, 13 (1),  
3619-3622.

Campos, PG, Dez, F., & Cantador, I. (2014). Các hệ thống giới thiệu nhận thức về thời gian:  
khảo sát và phân tích hệ thống các giao thức đánh giá hiện có. *Mô hình người dùng và người dùng thích nghi*  
*Tương tác*, 24 (1-2), 67-119.

Georgiev, K., & Nakov, P. (2013). Một khung không phải là iid để lọc cộng tác bị hạn chế  
máy Boltzmann. *Tiến hành hội thảo quốc tế lần thứ 30 về máy*  
*Học tập*, 1148-1156.

Hu, L., Cao, J., Xu, G., Cao, L., Gu, Z., & Zhu, C. (2013). Khuyến nghị cá nhân hóa qua  
nhân tố bộ ba tên miền chéo. *Trong các thủ tục của hội nghị quốc tế lần thứ 22 về*  
*World Wide Web*, 595-606.

Koren, Y., Bell, R., & Volinsky, C. (2009). Kỹ thuật nhân tố Matrix cho người giới thiệu  
các hệ thống. *Máy tính*, (8), 30-37.

Koren, Y. (2010). Bộ lọc cộng tác với các động lực thời gian. *Truyền thông của*

ACM , 53 (4), 89-97.

Linden, G., Smith, B., & York, J. (2003). Amazon Amazon. com đề xuất: Mục này đến mục khác

lọc cộng tác. *Điện toán Internet IEEE* , 7 (1), 76-80.

Lika, B., Kolomvatos, K., & Hadjiefthymiades, S. (2014). Đối mặt với vấn đề bắt đầu lạnh trong

hệ thống đề xuất. *Hệ thống chuyên gia với các ứng dụng* , 41 (4), 2065-2073.

Ma, H., King, I., & Lyu, MR. (2011). Học hỏi đề xuất với xã hội rõ ràng và tiềm ẩn

quan hệ. *Giao dịch ACM trên hệ thống và công nghệ thông minh (TIST)* , 2 (3), 29.

Salakhutdinov, R., Mnih, A., & Hinton, G. (2007). Các máy Boltzmann bị hạn chế cho

lọc cộng tác. *TIẾNG VIỆT trong Kỳ yếu hội thảo quốc tế lần thứ 24 về Máy*

# BẢN THẢO ĐƯỢC CHẤP NHẬN

**Trang 28**

## BẢN THẢO ĐƯỢC CHẤP NHẬN

26

học tập , 791-798.

Salakhutdinov, R., & Mnih, A. (2007). Yếu tố ma trận xác suất của cung. *Trong In NIPS* , 20, 1-8.

Salakhutdinov, R., & Mnih, A. (2008). Yếu tố ma trận xác suất Bayes sử dụng Markov

chuỗi Monte Carlo. *xông vào Kỳ yếu hội thảo quốc tế lần thứ 25 về Học máy* ,

880-887.

Shan, H., & Banerjee, A. (2010). Các yếu tố ma trận xác suất tổng quát của cộng đồng  
lọc .*trong khai thác dữ liệu (ICDM)*, Hội nghị quốc tế lần thứ 10 năm 2010 , 1025-1030.

Sainath, TN, Kingsbury, B., Sindhwani, V., Arisoy, E., & Ramabhadran, B. (2013). "Cấp bậc thấp  
nhân tố ma trận để đào tạo mạng lưới thần kinh sâu với các mục tiêu đầu ra chiều cao.

*Trong Âm học, Xử lý lời nói và tin hiệu (ICASSP)* , 6655-6659.

Shi, Y., Larson, M., & Hanjalic, A. (2014). Bộ lọc cộng tác của bộ lọc ngoài phạm vi ma trận mục người dùng:

Một cuộc khảo sát về tình trạng của nghệ thuật và những thách thức trong tương lai. *Khảo sát tính toán điện toán ACM (CSUR)* ,  
47 (1), 3.

Zhang, D., Hsu, CH, Chen, M., Chen, Q., Xiong, N., & Lloret, J. (2014). Khởi đầu lạnh  
khuyến nghị sử dụng phân cụm và hợp nhất cho các hệ thống đề xuất xã hội quy mô lớn.

*Giao dịch của IEEE về các chủ đề mới nổi trong máy tính* , 2 (2), 239-250.

Schein, AI, Popescul, A., Ungar, LH, & Pennock, DM (2001). Các mô hình tạo ra cho  
khuyến nghị bắt đầu lạnh. *Trong các thủ tục của Hội thảo SIGIR năm 2001 về Người giới thiệu  
Hệ thống* , 6.

Ocepeka, U., Rugeljb, J., & Bosnica, Z .. (2015). Khuyến khích cải thiện hệ số ma trận  
tions cho các ví dụ khi bắt đầu lạnh. *Hệ thống chuyên gia với các ứng dụng* , 42 (19), 6784-6794.

Victor, P., Cornelis, C., Teredesai, AM, & Gà trống, M. (2008). Tôi nên tin vào ai?

tác động của các số liệu quan trọng đến các khuyến nghị bắt đầu lạnh. *Trong các thủ tục của ACM 2008  
hội thảo chuyên đề về điện toán ứng dụng* , 2014-2018.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, PA (2010). Cấm xếp chồng lên nhau  
autoencoders: Học các biểu diễn hữu ích trong một mạng lưới sâu với việc khử nhiễu cục bộ  
tiêu chí. *Tạp chí Nghiên cứu máy học* , 11, 3371-3408.

Wang, C., & Blei, DM (2011). Mô hình chủ đề hợp tác của đề xuất cho đề xuất khoa học

# BẢN THẢO ĐƯỢC CHẤP NHẬN

bài viết. Sự kiện trong tiến trình của hội nghị quốc tế về kiến thức ACM SIGKDD lần thứ 17

*Khám phá và khai thác dữ liệu*, 448-456.

Wang, H., Wang, N., & Yeung, DY (2015). Học tập hợp tác sâu cho người giới thiệu

---

Trang 29

## BẢN THẢO ĐƯỢC CHẤP NHẬN

27

các hệ thống. Trong các thủ tục của Hội nghị quốc tế về kiến thức ACM SIGKDD lần thứ 21

*Khám phá và khai thác dữ liệu*, 1235-1244.

Xiong, L., Chen, X., Huang, TK, Schneider, JG, & Carbonell, JG (2010). Tạm thời

lọc cộng tác với nhân tố xác suất kéo căng Bayesian. Trong *In SDM*, 10, 211-222.

Xiao, Y., Ai, P., Hsu, CH, Wang, H., & Jiao, X. (2015). Bộ lọc cộng tác theo thứ tự thời gian

cho khuyến nghị tin tức. *Truyền thông Trung Quốc*, 12 (12), 53-62.

Chu, Y., Wilkinson, D., Schreiber, R., & Pan, R. (2008). Hợp tác song song quy mô lớn

lọc giải thưởng Netflix. Trong các khía cạnh thuật toán trong thông tin và quản lý, 337-

348.

Zhang, ZK, Liu, C., Zhang, YC, & Zhou, T. (2010). Giải quyết vấn đề khởi động lạnh trong  
hệ thống đề xuất với các thê xã hội. *EPL (Europhysics Letters)*, 92 (2), 28002.

Chu, K., Yang, SH, & Zha, H. (2011). Các yếu tố ma trận chức năng để khởi động nguội

khuyến nghị. Một số thủ tục của hội nghị ACM SIGIR quốc tế lần thứ 34 về

*Nghiên cứu và phát triển trong Truyền thông tin*, 315-324.

Zhang, C., Wang, K., Yu, H., Sun, J., & Lim, EP (2014). Chuyển tiếp yếu tố tiềm ẩn cho năng động

lọc cộng tác. Trong *In SDM*, 452-460.

# BẢN THẢO ĐƯỢC CHẤP NHẬN

Jian Wei Jian Wei nhận bằng cử nhân Kỹ thuật điện tử của Đại học Giao thông Thượng Hải

Trung Quốc, vào năm 2013. Ông hiện đang là nghiên cứu sinh tại Trường Kỹ thuật và Khoa học Ứng dụng ở Aston

Đại học, Vương quốc Anh. Quan tâm nghiên cứu của ông bao gồm học máy và khai thác dữ liệu, mô hình hóa và lập kế hoạch

trong các hệ thống máy tính phân tán và điện toán sương mù cho các phân tích dữ liệu lớn của IoT.

**Trang 30****BẢN THẢO ĐƯỢC CHẤP NHẬN**

28

**Tiến sĩ Jianhua Ông** là Giảng viên tại Đại học Aston, Vương quốc Anh. Ông đã nhận bằng Cử nhân và Thạc sĩ từ Đại học Khoa học và Công nghệ Huazhong (HUST), Trung Quốc và bằng tiến sĩ từ Nanyang Tech-Học viện khoa học và Công nghệ Singapore, vào năm 1995, 1998 và 2002, tương ứng. Tiến sĩ Ông là Phó giáo sư của Khoa Điện tử và Kỹ thuật Thông tin từ năm 2001. Ông đã làm việc với Đại học Bristol từ 2004 đến 2006 và với Đại học Essex năm 2007. Ông là Thành viên Ban Chỉ đạo và Phó biên tập viên của Giao dịch KSII trên Internet và Hệ thống thông tin, là Biên tập viên của Wireless Truyền thông và Điện toán di động, Tạp chí quốc tế về Hệ thống truyền thông, Biên tập viên khách hàng đầu cho bốn người đặc biệt Các vấn đề cho Tạp chí quốc tế về mạng cảm biến phân tán. Ông phục vụ như là vị trí chủ tịch cho một số hội nghị và Thành viên TPC cho nhiều hội nghị quốc tế bao gồm cả IEEE GLOBECOM và ICC. Lợi ích nghiên cứu chính của ông bao gồm lớn phân tích dữ liệu, điện toán sương mù / điện toán cạnh di động, phân tích cạnh, liên lạc từ máy đến máy, Internet của vạn vật hệ thống và công nghệ (ví dụ: đối với các thành phố nhỏ và giao thông thông minh) và công nghệ 4G / 5G. Ông đã là tác giả hoặc đồng tác giả hơn 150 bài báo kỹ thuật trong các tạp chí và hội nghị quốc tế lớn. Ông là thành viên cao cấp của IEEE ..

**Tiến sĩ Kai Chen** nhận bằng tiến sĩ, bằng cấp từ Đại học Giao thông Thượng Hải năm 2003 tại Trung Quốc. Anh ấy là một nhân viên học tập tại Viện Truyền thông hình ảnh và Kỹ thuật mạng, Thượng Hải Đại học Giao thông, Trung Quốc. Nghiên cứu chính của ông bao gồm lấy thông tin, nhận dạng đối tượng và lớn khai thác dữ liệu. Ông là thành viên chủ chốt của viện nghiên cứu kỹ thuật mạng. Ông là hiệu trưởng điều tra viên của một số dự án quốc gia quan trọng và nhiều dự án IAR (Công nghiệp-Học thuật-Nghiên cứu).

**Bác sĩ Yi Zhou** Yi Zhou nhận bằng tiến sĩ, bằng cấp từ Đại học Giao thông Thượng Hải năm 2010 tại Trung Quốc.

Cô là nhân viên tại Khoa Khoa học Máy tính của Đại học Giao thông Thượng Hải, Trung Quốc. Nghiên cứu chính của cô bao gồm nhận dạng đối tượng và khai thác dữ liệu lớn. Cô ấy đang làm việc trong một dự án của Trung Quốc Tự định nghĩa nhân vật, được hỗ trợ bởi Quỹ Khoa học quốc gia.

**BẢN THẢO ĐƯỢC CHẤP NHẬN****Trang 31****BẢN THẢO ĐƯỢC CHẤP NHẬN**

29

Tiến sĩ Zuoyin Tang hiện đang là Giảng viên của Trường Kỹ thuật và Khoa học Ứng dụng, Aston Đại học, Vương quốc Anh. Cô có bằng tiến sĩ tại Đại học Bath, Vương quốc Anh, vào năm 2008. Cô là tác giả và đồng tác giả của hơn 40 bài báo kỹ thuật trong các tạp chí và hội nghị quốc tế lớn. Tiến sĩ chính lợi ích nghiên cứu bao gồm quản lý tài nguyên cho các phân tích dữ liệu lớn, mạng di động, Internet của vạn vật và mạng cảm biến không dây.

# BẢN THẢO ĐƯỢC CHẤP NHẬN

---

Trang 32

BẢN THẢO ĐƯỢC CHẤP NHẬN

(a)  $K = 100$ .

(b)  $K = 200$ .

(c)  $K = 300$ .

Hình 8. Đường cong đào tạo của mô hình timeSVD ++ và IRCD-ICS.

---

Trang 33

BẢN THẢO ĐƯỢC CHẤP NHẬN

Hình 9. So sánh hiệu suất của các mô hình IRCD-CCS và IRCD-ICS để dự đoán xếp hạng của các mô hình ICS, K = 100.

# BẢN THẢO ĐƯỢC CHẤP NHẬN