

GIẢI PHÁP TÍCH HỢP MẠNG XÃ HỘI TRONG XÂY DỰNG HỆ TRỢ GIẢNG THÔNG MINH

Huỳnh Lý Thanh Nhân^{1,3}, Lê Huy Thập¹, Nguyễn Thái Nghe²

¹Khoa Công nghệ Thông tin, Trường Đại học Lạc Hồng

²Khoa Công nghệ Thông tin & Truyền Thông, Trường Đại học Cần Thơ.

³Khoa Kỹ thuật Công nghệ Môi trường, Trường Đại học An Giang.

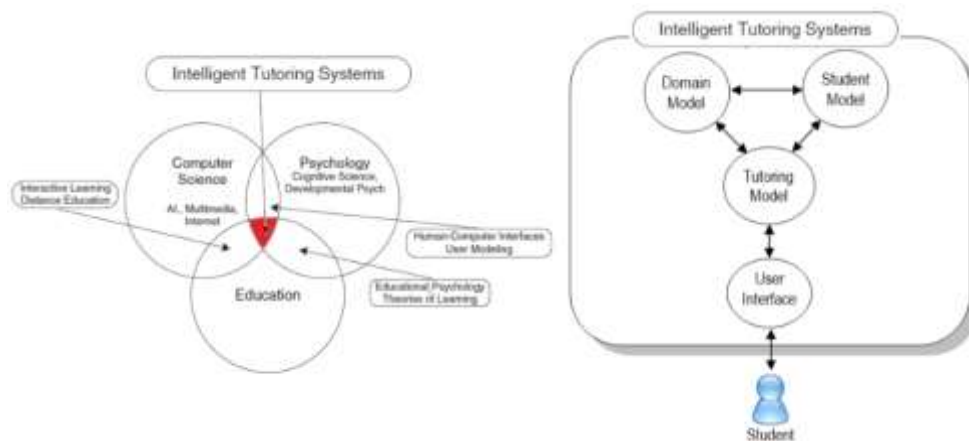
hltnhan@agu.edu.com, lththap1007@gmail.com, ntngh@cit.ctu.edu.vn

TÓM TẮT — Bài viết này đề xuất giải pháp dùng kỹ thuật phân rã ma trận tích hợp mạng xã hội (Social Matrix Factorization) trong xây dựng Hệ trợ giảng thông minh (Intelligent Tutoring System - ITS). Trước hết chúng tôi điểm lại các khái niệm chính trong ITS cũng như Kỹ thuật phân rã ma trận trong Hệ thống gợi ý, sau đó đề xuất mô hình dự đoán kết quả học tập của người học trong ITS. Thực nghiệm trên tập dữ liệu thu thập từ thực tế cho thấy việc tích hợp thêm mối quan hệ xã hội của người học có khả năng làm tăng độ chính xác của mô hình dự đoán trong ITS.

Từ khoá — Hệ thống gợi ý, hệ trợ giảng thông minh, khai phá dữ liệu giáo dục, mạng xã hội trong giáo dục.

I. GIỚI THIỆU

Hệ trợ giảng thông minh là một hệ thống có khả năng tùy chỉnh và phản hồi kết quả tức thì cho người học mà không cần sự can thiệp hay trợ giúp của giáo viên. Hệ trợ giảng thông minh (ITS)¹ được cấu thành từ ba lĩnh vực khác nhau: Khoa học máy tính, giáo dục, và tâm lý học. Theo từng hệ thống mà mỗi một ITS có thể có những thành phần khác nhau, nhưng về bản chất thì một ITS có bốn thành phần chính: (1) Mô hình liên quan đến việc học tập của sinh viên (Student Model), (2) Khối kiến thức liên quan trong lĩnh vực đang học (Domain Knowledge), (3) Mô hình liên quan đến việc dạy của giáo viên (Instructor/Tutoring Model) và (4) Giao diện người dùng (User Interface), như được minh họa trong Hình 1 [6].



Hình 1. Sự cấu thành và các thành phần chính của ITS

Để có những phản hồi thích hợp cho người học, một trong những công việc quan trọng nhất của “Student Model” là dự đoán kết quả học tập của sinh viên khi người học đang giải quyết vấn đề (bài tập) cụ thể nào đó. Tuy vậy, trong nghiên cứu này, chúng tôi chưa thảo luận về việc xây dựng một hệ trợ giảng thông minh cụ thể, mà chỉ tập trung vào công việc quan trọng nhất của “Student Model” là sử dụng mô hình nào để dự đoán được năng lực học tập của sinh viên với kết quả càng chính xác càng tốt. Dự đoán được năng lực học tập của sinh viên có vai trò rất quan trọng và lợi ích của nó cũng được thảo luận trong nhiều nghiên cứu trước đây [1][2][4].

Nhiều nghiên cứu đã xuất bản về dự đoán năng lực học tập sinh viên từ các phương pháp kỹ thuật truyền thống như cây quyết định, máy học vector hỗ trợ SVM, mô hình mạng Bayes, được trình bày trong tài liệu [9][10], đến những kỹ thuật gần đây như hệ thống gợi ý (Recommender Systems – RS) được tìm thấy trong tài liệu [3][7][11]. Tuy nhiên, đa số các giải thuật thuộc nhóm RS đều chỉ tập trung khai thác với những thuật toán cơ sở, do đó các giải thuật chưa tận dụng được hết các vấn đề mở rộng và nâng cao của giải thuật như mối quan hệ giữa các user và thông tin bổ sung cho item.

¹ http://en.wikipedia.org/wiki/Intelligent_tutoring_system

Với sự phát triển của mạng xã hội trực tuyến ngày càng tăng làm cho việc tích hợp mạng xã hội vào RS ngày càng được nhiều nhà nghiên cứu quan tâm. Họ đã thực nghiệm và kết quả cho thấy việc gợi ý đánh giá của người dùng thông qua việc bổ sung thêm mối quan hệ bạn bè đã đạt được độ chính xác tốt hơn [13][14]. Giải pháp tích hợp này cũng là một hướng tiếp cận để giải quyết vấn đề người dùng mới (cold-start) trong RS. Sử dụng việc tích hợp mạng xã hội không giống việc tìm người dùng giống nhau (similar user) mà nó sử dụng thành phần quan hệ của hệ thống khác (facebook, twister, zalo...) bổ sung thông tin vào hệ thống RS làm cho việc dự đoán đạt kết quả tốt hơn.

Bài viết này đề xuất hướng tiếp cận tích hợp thông tin trong mạng xã hội cho hệ trợ giảng thông minh, đặc biệt là trong Student Model để dự đoán năng lực học tập của sinh viên nhằm tận dụng được thông tin từ mạng xã hội của họ (đơn giản nhất là mối quan hệ bạn bè trong cùng một lớp học, do ta có thể giả định rằng “Nếu A là bạn của B, và A học tốt cũng như chăm học thì có thể ảnh hưởng đến B” và ngược lại hoặc những mối quan hệ khác có thể có trong dữ liệu), từ đó xây dựng thuật toán cho hướng tiếp cận đã đề xuất này. Phương pháp đề xuất được thực nghiệm trên tập dữ liệu thực tế và được đánh giá thông qua độ đo lỗi RMSE (Root Mean Squared Error). Kết quả cho thấy tích hợp mối quan hệ xã hội của sinh viên đã giảm được độ lỗi RMSE, tức là mô hình có khả năng dự đoán hiệu quả hơn.

Sau khi giới thiệu ở phần một. Trong phần hai, bài báo trình bày các nghiên cứu liên quan mật thiết với giải pháp đề xuất. Phần ba là phần giới thiệu giải pháp tích hợp mạng xã hội của sinh viên, trong phần này có trình bày phương pháp cơ sở (Matrix Factorization) và chỉ ra cách thức tích hợp mạng xã hội vào nó. Giải pháp đề xuất được thử nghiệm qua tập dữ liệu thực tế và kết quả thực nghiệm cũng được trình bày trong phần bốn. Phần cuối cùng là kết luận.

II. NHỮNG NGHIÊN CỨU LIÊN QUAN

Trong nghiên cứu [3] và [11], các tác giả đã ứng dụng kỹ thuật phân rã ma trận (Matrix Factorization - MF) vào bài toán dự đoán năng lực học tập của sinh viên (Student-Model), kỹ thuật này đã mang lại kết quả khả quan cho việc ứng dụng RS vào khai phá dữ liệu giáo dục. Tuy nhiên, kết quả đo độ lỗi RMSE vẫn còn cao và chưa tận dụng được những vấn đề mở rộng có thể tích hợp vào giải thuật cơ sở để nâng cao hiệu quả dự đoán.

Nhiều nghiên cứu [13][14][15] đã nâng cao độ chính xác dự đoán của kỹ thuật MF nhờ tận dụng mối quan hệ xã hội của người dùng bằng nhiều cách. Trong tài liệu [12] tác giả đã làm phép so sánh các phương pháp tích hợp mạng xã hội vào MF và chọn ra phương pháp tốt nhất. Hạn chế của giải thuật này là phụ thuộc vào sự hỗ trợ mối quan hệ của tập dữ liệu. Kết quả thí nghiệm của những nghiên cứu trên cho thấy khi tích hợp mối quan hệ sẽ nâng cao hiệu quả của thuật toán. Tuy nhiên, những nghiên cứu này chỉ đưa ra giải quyết các ứng dụng cho các bài toán trong thương mại và giải trí.

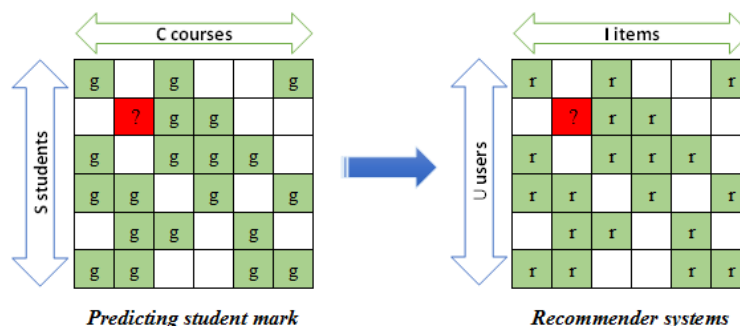
Vì vậy, bài báo này đã đưa ra hướng tiếp cận sử dụng mối quan hệ xã hội của người dùng, hay trong “Student-Model” thì nó là mối quan hệ bạn bè trong cùng một lớp học để tích hợp vào giải thuật Matrix Factorization để giải quyết bài toán dự đoán năng lực học tập của sinh viên trong Hệ trợ giảng thông minh.

III. GIẢI PHÁP ĐỀ XUẤT

Để tích hợp mạng xã hội vào kỹ thuật phân rã ma trận, trước tiên bài viết trình bày sự tương đồng và chuyển đổi dữ liệu sẵn có về định dạng cần thiết của thuật toán. Sau đó trình bày ngắn gọn kỹ thuật nền (baseline), kỹ thuật phân rã ma trận (xem thêm trong bài viết [11]) và giới thiệu chi tiết về hướng tiếp cận tích hợp mạng xã hội, mối quan hệ bạn bè của sinh viên trong một lớp.

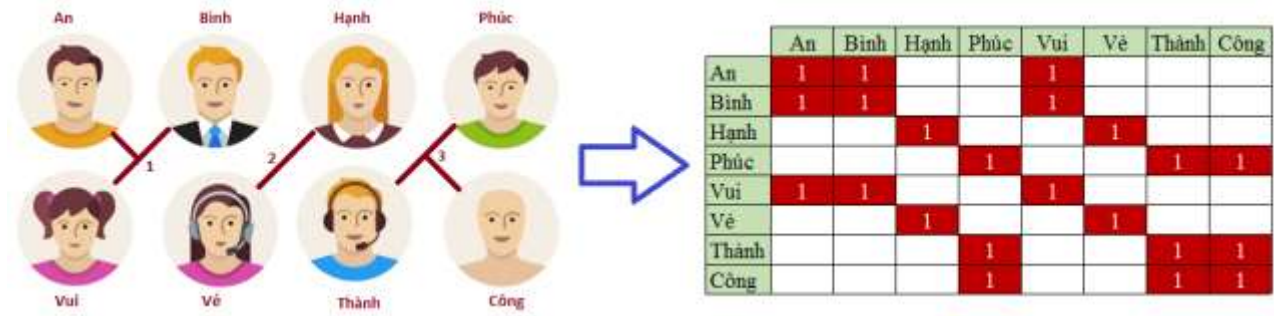
A. Chuyển đổi dữ liệu về định dạng của thuật toán

Trong hệ thống gợi ý (Recommender systems – RS) được cấu thành danh sách người dùng (user - u), danh sách các đối tượng như bài hát, bộ phim, sản phẩm (item - i) và các đánh giá (ratings - r) là chỉ số đánh giá của người dùng u trên đối tượng i . Tương tự, trong bài toán dự đoán điểm học tập của sinh viên trong mô hình “Student-Model” thì có danh sách các sinh viên s , danh sách các môn học c và điểm g . Như vậy, việc dự đoán đánh giá của người dùng trong bài toán xếp hạng (rating prediction) của RS sẽ tương đương với bài toán dự đoán điểm sinh viên [3][11] (xem hình 2). Sự ánh xạ này được biểu diễn như sau: {User \rightarrow Student} {Item \rightarrow Course} {Ratings \rightarrow Gradings}.



Hình 2. Sự tương đồng của bài toán dự đoán năng lực học tập sinh viên với bài toán xếp hạng

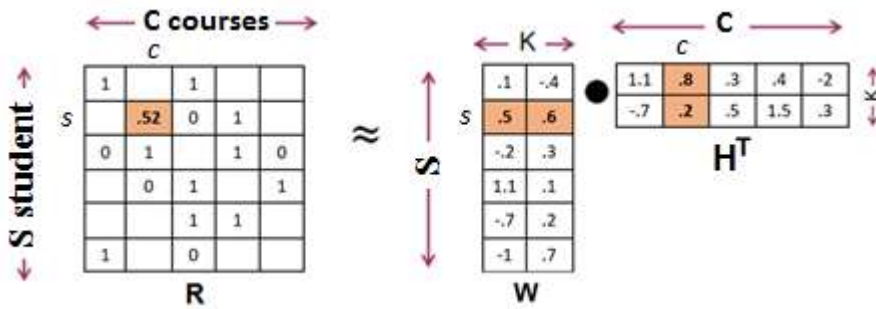
Ngoài ra, để tích hợp mạng xã hội của người học vào trong kỹ thuật phân rã ma trận, mối quan hệ bạn bè của các sinh viên trong cùng một lớp sẽ được chuyển thành ma trận quan hệ dạng nhị phân (như hình 3) để tích hợp vào giải thuật dự đoán. Nếu hai người bạn trong lớp thì sẽ thể hiện giá trị 1, nếu không cùng lớp sẽ có giá trị 0 trong ma trận.



Hình 3. Sự chuyển đổi mỗi quan hệ bạn bè thành ma trận mối quan hệ

B. Kỹ thuật phân rã ma trận (Matrix Factorization - MF)

Kỹ thuật phân rã ma trận là việc chia một ma trận lớn **R** thành hai ma trận **W** và **H** có kích thước nhỏ hơn rất nhiều so với ma trận **R**, sao cho **R** có thể được xây dựng lại từ hai ma trận nhỏ hơn này càng chính xác càng tốt [5], nghĩa là $R \approx WH^T$.



Hình 4. Minh họa kỹ thuật phân rã ma trận

$W \in \mathbb{R}^{|S| \times |K|}$ là một ma trận mà ở đó mỗi dòng *s* là một véc-tơ bao gồm *K* nhân tố tiềm ẩn (latent factors) mô tả cho sinh viên *s*, và $H \in \mathbb{R}^{|C| \times |K|}$ là một ma trận mà ở đó mỗi dòng *c* là một véc-tơ bao gồm *K* nhân tố tiềm ẩn mô tả cho môn học *c*.

Gọi *w* và *h* là các phần tử tương ứng của hai ma trận **W** và **H**, hay *w* và *h* là các véc-tơ bao gồm *K* nhân tố tiềm ẩn mô tả cho sinh viên *s* và môn học *c*, khi đó điểm số *g* của sinh viên *s* trên môn học *c* được dự đoán bởi công thức:

$$\hat{g}_{sc} = \sum_{k=1}^K w_{sk} h_{ck} = w_s h_c^T \tag{1}$$

W và **H** là các tham số mô hình (còn gọi là các ma trận nhân tố tiềm ẩn) mà chúng ta cần phải xác định bằng cách tối ưu hóa tối thiểu (min) hàm mục tiêu (3) theo điều kiện nào đó, chẳng hạn RMSE (Root Mean Squared Error).

$$RMSE = \sqrt{\frac{1}{|D^{test}|} \sum_{s,c,g \in D^{test}} (g_{si} - \hat{g}_{si})^2} \tag{2}$$

Tối ưu hóa hàm mục tiêu theo phương pháp giảm dốc ngẫu nhiên (Stochastic Gradient Descent - SGD) [11]

$$O^{MF} = \sum_{(s,c,g) \in D^{train}} (g_{sc} - \sum_{k=1}^K w_{sk} h_{ck})^2 + \lambda (\|W\|_F^2 + \|H\|_F^2) \tag{3}$$

Với λ là hệ số chính tắc hóa ($0 \leq \lambda < 1$) và $\|\cdot\|_F^2$ là chuẩn Frobenius². Đại lượng $\lambda \cdot (\|W\|_F^2 + \|H\|_F^2)$ được dùng để ngăn ngừa sự quá khớp (over-fitting).

² https://en.wikipedia.org/wiki/Matrix_norm#Frobenius_norm

Với hàm đo độ lỗi mới này thì các giá trị w và h sẽ được cập nhật lại theo 2 công thức sau:

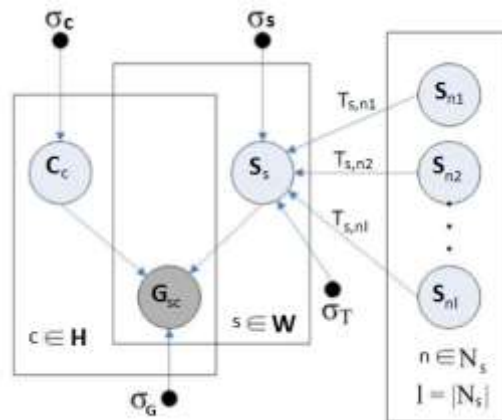
$$w'_{sk} = w_{sk} + \beta(2(g_{si} - \hat{g}_{si})h_{ck} - \lambda w_{sk}) \quad (4)$$

$$h'_{ck} = h_{ck} + \beta(2(g_{si} - \hat{g}_{si})w_{sk} - \lambda h_{ck}) \quad (5)$$

Trong đó: $e_{sc} = g_{sc} - \hat{g}_{sc}$ là độ lỗi của dự đoán, β là tốc độ học.

C. Kỹ thuật phân rã ma trận mạng xã hội (Social Network Matrix Factorization - SNMF)

Trong kỹ thuật phân rã ma trận cơ sở, chúng ta có ma trận S, C và ma trận đánh giá R (các giá trị của đánh giá có thể là [0..4] theo thang điểm 4 hoặc tùy hệ thống). Trong hướng tiếp cận tích hợp mạng xã hội, mỗi sinh viên s có một tập hợp N_s thể hiện các mối quan hệ bạn bè với s . Giá trị $T_{s,n}$, tỷ lệ xác suất thể hiện mối quan hệ của sinh viên s với sinh viên n , của tập hợp này có giá trị là một số thực 0 hoặc 1. Nếu $T=0$ thì hai người s và n khác lớp, nếu $T=1$ thì hai người này học chung lớp. Hình 5 [12] thể hiện mô hình tích hợp yếu tố mạng xã hội vào kỹ thuật MF tại vị trí người dùng.



Hình 5. Mô hình tích hợp yếu tố mạng xã hội vào kỹ thuật phân rã ma trận

Mối quan hệ của một sinh viên được biểu diễn bằng một vectơ chứa các giá trị quan hệ của n sinh viên chung lớp với sinh viên s , như vậy tập mối quan hệ của tất cả các sinh viên được biểu diễn bằng ma trận $T^{|S||S|}$. Ma trận này thường là ma trận không đối xứng nhưng với bài toán dự đoán năng lực học tập sinh viên thì ma trận này là ma trận đối xứng và nó là ma trận nhị phân.

Để tích hợp được mối quan hệ trong lớp học của sinh viên vào MF, dựa vào bài báo [12] ta thay thế ma trận W bằng \hat{W} bằng công thức (6) và sau đó thực hiện tương tự kỹ thuật MF cơ sở.

$$\hat{w}_s = \frac{\sum_{n \in N_s} T_{s,n} w_n}{\sum_{n \in N_s} T_{s,n}} = \frac{1}{|N_s|} \sum_{n \in N_s} w_n \quad (6)$$

Do ma trận T là ma trận nhị phân nên mỗi phần tử $T_{s,n}$ có giá trị là 1 và $\sum_{n \in N_s} T_{s,n} = |N_s|$ nên có thể hiểu đơn giản

là \hat{W} bằng với giá trị trung bình của các người dùng có quan hệ. Việc dự đoán vẫn áp dụng theo công thức của giải thuật MF nhưng thay thế W bằng \hat{W} . Như vậy, công thức (1) trở thành:

$$\hat{g}_{sc} = \sum_{k=1}^K \hat{w}_{sk} h_{ck} = \hat{w}_s \bullet h_c^T \quad (7)$$

Hàm mục tiêu của kỹ thuật SocialMF vẫn được tính như cách làm của kỹ thuật MF nhưng có cộng thêm một hệ số của mối quan hệ (công thức 3 và 8). Tuy nhiên trong kỹ thuật chỉ cải tiến đơn giản từ kỹ thuật MF cơ sở mà chưa tích hợp phần độ lệch (biases) như tài liệu [3]. Hàm mục tiêu của kỹ thuật SocialMF được viết như sau:

$$O^{SocialMF} = \sum_{(s,c,g) \in D^{train}} (g_{sc} - \sum_{k=1}^K \hat{w}_{sk} h_{ck})^2 + \lambda (\|W\|_F^2 + \|H\|_F^2) + \lambda_T \sum_{s=1}^S \left(w_s - \frac{1}{|N_s|} \sum_{n \in N_s} w_n \right)^2 \quad (8)$$

Như vậy, các giá trị w và h cũng được cập nhật mới sau mỗi bước lặp được thực hiện như sau, trong đó $e_{sc} = g_{sc} - \hat{g}_{sc}$

$$w'_{sk} = w_{sk} + \beta(2e_{sc} h_{ck} - \lambda w_{sk}) + \lambda_T (w_{sk} - \frac{1}{|N_s|} \sum_{n \in N_s} w_{nk}) - \lambda_T \left(\sum_{t \in S_s} \frac{1}{|N_s|} (w_{tk} - \frac{1}{|N_s|} \sum_{w \in N_s} w_{wk}) \right) \quad (9)$$

$$h'_{ck} = h_{ck} + \beta(2e_{sc} w_{sk} - \lambda h_{ck}) \quad (10)$$

Với β là tốc độ học, λ là hệ số chính tắc cho ma trận \mathbf{W} và \mathbf{H} , λ_T là hệ số chính tắc cho ma trận quan hệ \mathbf{T} , Tập S_s là tập tất cả sinh viên, N_s là tập các sinh viên cùng lớp, $|N_s|$ là tổng số sinh viên cùng lớp với sinh viên s .

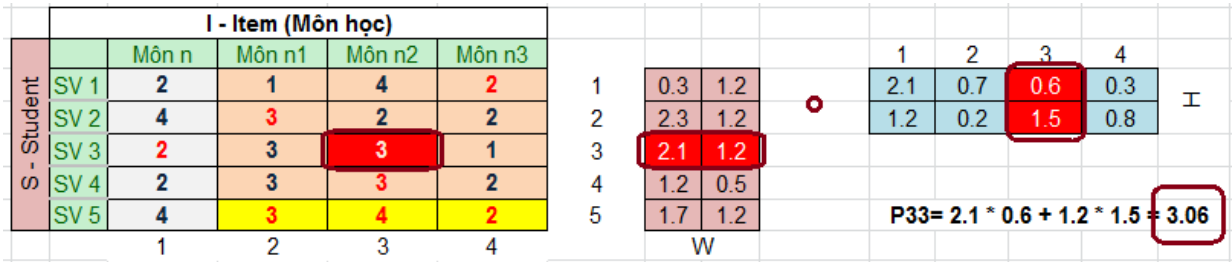
Chi tiết cho kỹ thuật này được tóm tắt trong thuật toán **PredictingStudentPerformance-SocialMF**. Trước tiên, chúng ta cần khởi tạo giá trị của các tham số một cách ngẫu nhiên theo chuẩn phân phối $\mathcal{N}(\mu, \sigma^2)$ với giá trị trung bình $\mu = 0$, độ lệch chuẩn $\sigma^2 = 0.01$. Khi điều kiện dừng chưa thỏa mãn, chẳng hạn như đạt đến số lần lặp tối đa hoặc tới điểm hội tụ thì các tham số vẫn còn cập nhật (converging: $O_{Iter(n-1)}^{socialMF} - O_{Itern}^{socialMF} < \epsilon$).

Thuật toán: PredictingStudentPerformance-SocialMF (D^{train} , T-UserRelation, λ_T , \mathbf{K} , β , λ , điều kiện dừng)

1. **Đặt** $s \in S$ là một sinh viên, $c \in C$ là một môn học, $g \in G$ là điểm số
 2. **Đặt** $w[s][k]$ và $H[c][k]$ là các nhân tố tiềm ẩn của sinh viên và môn học
 3. $W \leftarrow \mathcal{N}(0, \sigma^2)$
 4. $H \leftarrow \mathcal{N}(0, \sigma^2)$
 5. **while** (Điều kiện dừng chưa thỏa) **do**
 6. Lấy ngẫu nhiên (s, c, g_{sc}) trong D^{train}
 7. $N_s =$ số sinh viên có quan hệ bạn bè với sinh viên s trong T-UserRelation;
 8. $\hat{g}_{sc} \leftarrow \sum_k^K (W[s][k] * H[c][k])$
 9. $e_{sc} = g_{sc} - \hat{g}_{sc}$
 10. **for** $k=1..K$ **do**
 11. **for** $n=1..N_s$ **do**
 12. $VT = VT - W[n][k] / N_s$
 13. **end for**
 14. $VT = W[s][k] - VT$
 15. **for** $t=1..S_s$ **do**
 16. **for** $w=1..N_s$ **do**
 17. $W[t][k] = W[t][k] - W[w][k] / N_s$
 18. **end for**
 19. $VP = VP + W[t][k] / N_s$
 20. **end for**
 21. $W[s][k] \leftarrow W[s][k] + \beta * (2e_{sc} * H[c][k] - \lambda * W[s][k]) + \lambda_T (VT - VP)$
 22. $H[c][k] \leftarrow H[c][k] + \beta * (2e_{sc} * W[s][k] - \lambda * H[c][k])$
 23. **end for**
 24. **end while**
 25. **return** $\{W, H\}$
 26. **end procedure.**
-

Sau quá trình tối ưu, ta nhận được các giá trị của ma trận **W** và **H** đã tối ưu. Khi đó, chúng ta có thể dễ dàng dự đoán kết quả thông qua công thức (11) và được biểu diễn minh họa ví dụ (như hình 6):

$$\hat{g}_{sc} = \sum_{k=1}^K w_{sk} h_{ck} = w_s \bullet h_c^T \tag{11}$$



Hình 6. Dự đoán cho sinh viên 3 học môn học n2

IV. KẾT QUẢ THỰC NGHIỆM

A. Dữ liệu

Trong nghiên cứu này chúng tôi sử dụng tập dữ liệu điểm của khoa CNTT&TT, trường Đại học Cần Thơ, được thu thập 10 năm từ năm 1994 đến năm 2004. Tập dữ liệu bao gồm 4017 sinh viên (4017 users) và 353 môn học (353 items) của 3 ngành học và gồm 279536 điểm chi tiết (279536 ratings). Để đánh giá hiệu quả của giải thuật, chúng tôi sử dụng nghi thức kiểm tra hold-out: lấy ngẫu nhiên 2/3 tập dữ liệu để học và 1/3 còn lại để kiểm tra.

B. Tiền xử lý dữ liệu

Chúng ta cần bốn thành phần dữ liệu để ứng dụng vào giải thuật SocialMF: dữ liệu sinh viên (User), dữ liệu môn học (Item), dữ liệu điểm số (Ratings) và cuối cùng là dữ liệu mối quan hệ bạn bè (UserRelation). Dữ liệu được xử lý từ hệ thống quản lý kết quả học tập của trường đại học sẵn có và chuyển chúng sang định dạng phù hợp với các giải thuật trong RS. Quá trình xử lý gồm bốn công đoạn sau:

Xử lý User_mapping, thông tin sinh viên bao gồm nhiều trường: mã số sinh viên, họ tên, phái, ngày tháng năm sinh, lớp, trình độ ngoại ngữ,... Bảng SINHVIEN được chuyển về dạng danh sách các {ID_SV, LOP} và không quan tâm đến các trường dữ liệu khác.

Xử lý Item_mapping, tương tự như user_mapping, thông tin môn học bao gồm: mã môn, tên môn học, số tín chỉ, số tiết lý thuyết, số tiết thực hành, thông tin đề cương, số tiết thời khóa biểu,... Khi ứng dụng vào giải thuật SocialMF, những môn học có sinh viên học và được chuyển thành danh sách các ID môn học.

Xử lý ma trận Ratings, sau khi chuyển được 2 bảng dữ liệu sinh viên và môn học. Bảng điểm được chọn lọc lại với các thông tin như: mã sinh viên, mã môn học, điểm, điểm thi lại. Điểm được chuyển đổi sang thang điểm 4. Tuy nhiên, trong tập dữ liệu này, khi có sinh viên thi lại thì điểm lần cuối cùng sẽ được sử dụng.

Ngoài ra, bảng SINHVIEN cũng được sử dụng để chuyển thành ma trận quan hệ giữa các sinh viên. Mỗi giá trị trong ma trận này sẽ có giá trị là 0 hoặc 1. Nếu sinh viên A và B học chung lớp sẽ có giá trị là 1 tại vị trí cột và dòng tương ứng của 2 sinh viên này, nếu họ không học chung lớp thì nó sẽ có giá trị 0 (như đã đề cập trong phần III.A).

C. Xác định siêu tham số (hyper-parameters)

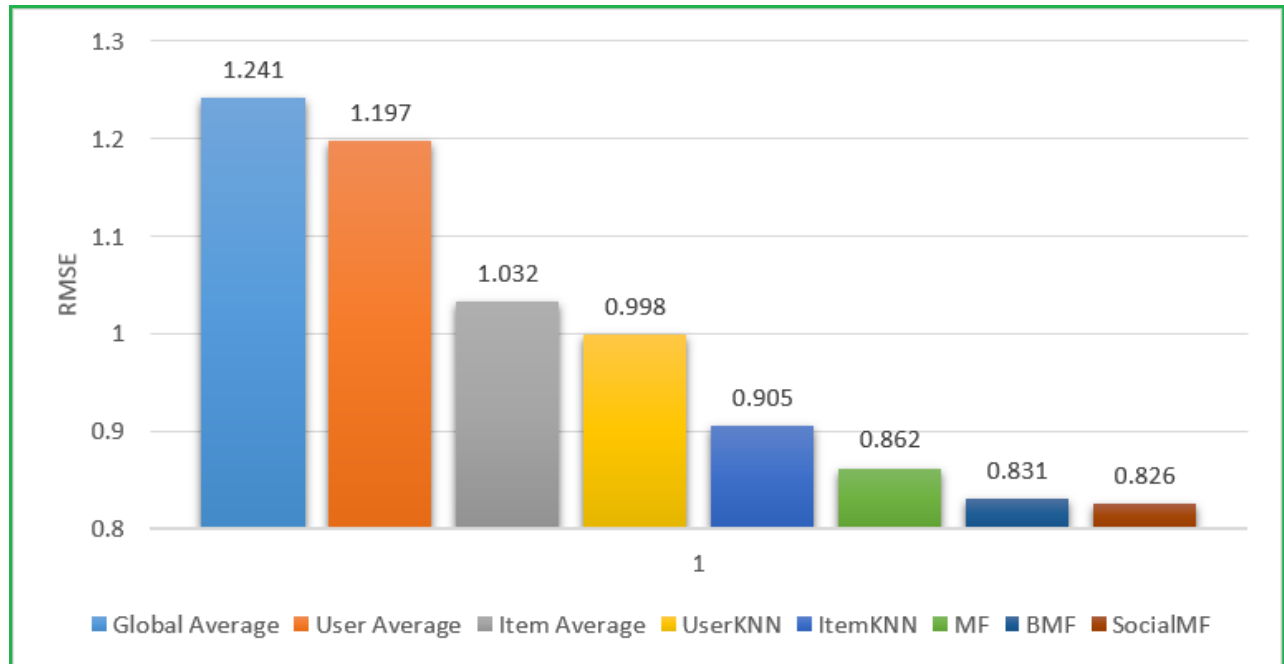
Các siêu tham số (hyper-parameters) của MF, BMF, SocialMF như số lần lặp (Iter), số nhân tố tiềm ẩn K, tốc độ học β , và hệ số chính tắc hóa λ được xác định bằng phương pháp tìm kiếm siêu tham số (hyper-parameter search) [10]. Tuy nhiên, do các tập dữ liệu khá lớn nên việc tìm kiếm bằng vét cạn sẽ mất nhiều thời gian, nên chúng tôi chỉ thực hiện việc tìm kiếm thô cho các phương pháp này: ví dụ: Iter \in (100, 200 . . . , 1000), K \in (5, 10, . . . ,30), $\beta \in$ (10^{-4} , 10^{-3} , 10^{-2} , $5 \cdot 10^{-5}$, $5 \cdot 10^{-4}$, $5 \cdot 10^{-3}$), $\lambda \in$ ($15 \cdot 10^{-4}$, $15 \cdot 10^{-3}$, $55 \cdot 10^{-5}$, $55 \cdot 10^{-4}$, $55 \cdot 10^{-3}$). Các tham số sử dụng trong thực nghiệm được mô tả chi tiết trong bảng 1.

Bảng 1. Các siêu tham số sử dụng trong thực nghiệm

Giải thuật	Tham số sử dụng	Thời gian training (s)
MF	K=10, $\lambda=0.045$, $\beta=0.1$, Iter=30	10
BMF	K=10, $\lambda=0.045$, $\beta=0.1$, Iter=30	12
SNMF	K=10, $\lambda=3.5$, $\beta=0.1$, SocialReg=7, Iter=232	4040

D. Kết quả thực nghiệm

Hình 5 trình bày kết quả thực nghiệm đánh giá bằng RMSE. Kết quả cho thấy rằng nếu tận dụng được thông tin từ mối quan hệ bạn bè của người học (thông tin từ mạng xã hội), cho kết quả cải thiện hơn so với những phương pháp đã có.



Hình 7. Kết quả RMSE cho thấy tích hợp thông tin từ mạng xã hội có khả năng cải thiện mô hình dự đoán

V. KẾT LUẬN

Trong bài viết này chúng tôi đã giới thiệu một hướng tiếp cận tích hợp mối quan hệ xã hội của người học vào Hệ trợ giảng thông minh, được minh họa thông qua kỹ thuật phân rã ma trận mạng xã hội (Social Matrix Factorization – SocialMF). Với tiếp cận này, ngoài những dữ liệu có sẵn, chúng ta có thể tận dụng thêm mối quan hệ giữa các user ở hệ thống khác để tích hợp vào xây dựng mô hình dự đoán, do đó cải thiện được kết quả dự đoán. Kết quả thực nghiệm trên tập dữ liệu thực tế cho thấy tiếp cận được đề xuất hoạt động tốt hơn so với các tiếp cận trước đây (chỉ sử dụng thông tin xếp hạng như trong Hệ thống gợi ý). Tuy vậy thời gian huấn luyện của giải thuật lâu hơn (khoảng 1h).

Mạng xã hội đang ngày càng phổ biến, đặc biệt là trong giới trẻ, sinh viên và học sinh. Vì vậy giải pháp tích hợp thông tin từ mạng xã hội vào Hệ trợ giảng thông minh là hoàn toàn khả thi trong thực tế. Trong tương lai, chúng tôi tiếp tục thử nghiệm mô hình trên những tập dữ liệu khác.

TÀI LIỆU THAM KHẢO

- [1] Nguyễn Duy Mộng Hà 2012, Đẩy mạnh công tác cố vấn học tập và tư vấn sinh viên góp phần nâng cao chất lượng dạy và học trong trường đại học theo học chế tín chỉ, *Tạp chí Giáo dục*, số 291, Kỳ 1, 2012: 32-35
- [2] Kỹ yếu hội thảo “Vai trò của cố vấn học tập trong đào tạo theo học chế tín chỉ tại các trường đại học, cao đẳng Việt Nam”, 24/12/2014. *Viện Nghiên cứu Giáo dục - Trường ĐHSP TP. Hồ Chí Minh*.
- [3] Huỳnh Lý Thanh Nhân và Nguyễn Thái Nghe. 2013. Hệ thống dự đoán kết quả học tập và gợi ý lựa chọn môn học. Kỹ yếu hội thảo quốc gia lần thứ XVI: *Một số vấn đề chọn lọc của CNTT&TT (@2013)*, trang 110-118. Nhà xuất bản Khoa học và Kỹ thuật. ISBN: 987-604-67-0251-1.
- [4] C. Romero, S. Ventura, M. Pechenizkiy, and R. S. Baker. (2010). Handbook of Educational Data Mining. *Chapman and Hall/CRC*, October 2010.
- [5] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Computer Society Press*, 42(8), 30-37.
- [6] Thai-NGhe, N., & Schmidt-Thieme, L. (2015, October). Multi-relational Factorization Models for Student Modeling in Intelligent Tutoring Systems. In *Knowledge and Systems Engineering (KSE), 2015 Seventh International Conference on* (pp. 61-66). IEEE.
- [7] Feng, M., Heffernan, N., & Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, 19(3), 243-266.
- [8] Singh, A. P., & Gordon, G. J. (2008). Relational learning via collective matrix factorization. In *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, ser. KDD '08. New York, NY, USA: ACM, pp. 650–658.

- [9] Nguyen Thai-Nghe, Paul Janecek, and Peter Haddawy. 2007. A comparative analysis of techniques for predicting academic performance. *Proceedings of the 37th ASEE/IEEE Frontiers in Education (FIE 2007)*, pp. T2G-7-T2G-12. ISSN: 0190-5848. E-ISBN: 978-1-4244-1084-2. Print ISBN: 978-1-4244-1083-5. *IEEE Xplore*
- [10] Mohammed Al-Sarem, 2015. Building a Decision Tree Model for Academic Advising Affairs Based on the Algorithm C4. 5. *IJCSI International Journal of Computer Science Issues*, Volume 12, Issue 5, September 2015 ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
- [11] Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., and Schmidt-Thieme, L. 2010. Recommender system for predicting student performance. In *Proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010)*. Vol. 1. Elsevier's Procedia CS, 2811-2819.
- [12] Jamali M., Ester M., 2010. A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks. *RecSys2010*, September 26–30, 2010.
- [13] H. Ma, I. King, and M. R. Lyu. 2009. Learning to recommend with social trust ensemble. In *SIGIR 2009*, pages 203–210.
- [14] H. Ma, H. Yang, M. R. Lyu, and I. King. 2008. Sorec: social recommendation using probabilistic matrix factorization. In *CIKM 2008*, pages 931–940.
- [15] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. In *KDD 2008*.

AN APPROACH FOR INTEGRATING SOCIAL NETWORKS INTO INTELLIGENT TUTORING SYSTEMS

Huynh Ly Thanh Nhan, Le Huy Thap, Nguyen Thai Nghe

ABSTRACT—This paper proposes an approach for integrating social networks into intelligent tutoring systems (ITS), especially for predicting student performance. Firstly, we review the main concepts of ITS as well as the matrix factorization technique and social matrix factorization in Recommender Systems (RS), then the predicting student performance model, which is integrated social networks, is proposed for building the ITS. Experimental results on a real dataset show that the integrated social relationships for students is able to increase the accuracy of predicting models in ITS.

Keywords —Recommender systems, educational mining, social matrix factorization, predicting student performance.