ỨNG DỤNG SỰ THÍCH NGHI CỘNG ĐỒNG TRONG CÁC HỆ THỐNG ĐÀO TẠO THÔNG MINH

Nguyễn An Tế
Đặng Thế Khoa

ABSTRACT

Nowadays, personalized systems are increasingly applied in diverse fields including e-Learning. In general, Intelligent Tutoring Systems build learners' profiles in order to give them adaptive access to information and other learning resources. Recently, some researchers propose using Collaborative Filtering as an alternative approach for personalizing information access in Intelligent Tutoring Systems. With this technique, a user receives recommendations on the basis of the evaluation of his/her community.
In this paper, we present a new approach of multi-criteria communities for Collaborative Filtering in Intelligent Tutoring Systems in which criteria can be extracted from features/ characteristics in learner profiles.

KEYWORDS:

Intelligent Tutoring Systems, User Profile, Collaborative Filtering, and Rule-based Induction

I. MỞ ĐẦU

Hiện nay các hệ thống thích nghi (Personalized Systems) đang phát triển rất mạnh và được ứng dụng trong nhiều lãnh vực như: tìm kiếm thông tin (Information Retrieval), thương mại điện tử (e-Commerce) hay đào tạo từ xa (e-Learning). Trong những hệ thống này, mỗi người sử dụng sở hữu một thành phần mô tả đặc trưng (profile) mà tùy theo lãnh vực ứng dụng sẽ bao gồm những thông tin khác nhau mô tả về mình như: thông tin cá nhân (họ tên, tuổi, nghề nghiệp, …), sở thích, nhu cầu hay mục tiêu, trình độ hay nền tảng kiến thức, … và đặc biệt là quá trình giao tiếp với hệ thống (interaction history) (Amato & Straccia, 1999; Brusilovski, 2001). Trong lãnh vực e-Learning, các hệ thống đào tạo thông minh (Intelligent Tutoring Systems – ITS, hay Adaptive Hypermedia Systems – AHS) khai thác các profiles để xây dựng những bài học và hỗ trợ người học trong việc khai thác thông tin cùng với những tài nguyên học tập khác một cách phù hợp với từng cá nhân người học (Brooks và cộng sự, 2006; Brusilovsky, 2001; Madhour & Forte, 2006). Ví dụ, hệ thống sẽ xây dựng tập hợp các luật thích nghi (adaptation rule) về nội dung, cách trình bày và cách duyệt xem (navigation). Bên cạnh đó, những thông tin hay kiến thức về một khái niệm nào đó trong một môn học sẽ được phân chia thành nhiều phần và mỗi phần sẽ được liên kết với một điều kiện cụ thể về trình độ của người học. Sau đó, tùy theo trình độ được thể hiện trong profile của từng cá nhân mà hệ thống sẽ áp dụng những luật thích nghi tương ứng nhằm cung cấp những phần nội dung thông tin hay kiến thức về khái niệm cùng với cách trình bày một cách phù hợp nhất.

1.1. Thành phần mô tả đặc trưng và sự thích nghi cá nhân

Trong những hệ thống thích nghi nói chung và ITS nói riêng, sự thích nghi được thực hiện một cách trực tiếp bằng cách đối sánh giữa profile của một người sử dụng với các tài nguyên sẵn có của hệ thống, ví dụ như thông tin, nhằm đáp ứng nhu cầu riêng biệt của cá nhân này. Chúng tôi gọi đây là sự thích nghi cá nhân. Nhìn chung, chất lượng của sự thích nghi cá nhân có thể được đánh giá dựa trên ba yếu tố cơ bản là:

mô hình hóa những người sử dụng qua các profiles, mô hình tổ chức các tài nguyên và các phương pháp thích nghi giữa profiles và tài nguyên (thường được dựa trên các kỹ thuật trong lãnh vực trí tuệ nhân tạo hay thông minh nhân tạo – Artificial Intelligence). Trong phạm vi bài báo này, chúng tôi quan tâm đặc biệt đến yếu tố thứ nhất bao gồm những vấn đề liên quan đến ba tiến trình: khởi tạo, cập nhật và khai thác profiles (Montaner và cộng sự, 2003).

Về nguyên tắc, các hệ thống thích nghi cá nhân hoàn toàn "không có lợi về công", ai làm nấy hưởng. Điều đó có nghĩa là người sử dụng không chia sẻ công sức của mình với bất kỳ người nào khác trong hệ thống thích nghi cá nhân. Trong suốt quá trình khai thác, người sử dụng sẽ giao tiếp với hệ thống, cung cấp những thông tin cần thiết cho việc khởi tạo và cập nhật profile của chính mình, và sau đó họ sẽ khai thác và thụ hưởng những gì xuất phát từ công sức đã bỏ ra. Điều này dẫn đến một số vấn đề khó khăn trong ba tiến trình kể trên liên quan đến profile:

a) Chất lượng của profile ban đầu: Khi mới đăng ký (registration), người sử dụng phải cung cấp một số thông tin cần thiết giúp cho hệ thống có thể khởi tạo profile. Tuy nhiên, những thông tin mà người sử dụng cung cấp thường không chính xác và không đầy đủ, ngay cả khi được hỗ trợ bằng những giao diện và công cụ hiệu quả (ví dụ: người sử dụng gặp rất nhiều khó khăn khi mô tả nền tảng kiến thức hay mức độ quan tâm đến các chủ đề).

b) Cập nhật profile: Trong quá trình khai thác, người sử dụng sẽ nhận được những tài nguyên do hệ thống cung cấp dựa trên profile đang sở hữu, và đồng thời phải cung cấp ngược lại cho hệ thống những phản hồi (feedback), đánh giá những gì mà mình đã nhận được để hệ thống có thể cập nhật profile một cách phù hợp. Như vậy, những feedbacks của người sử dụng cũng được xem là một thành phần quan trọng tích hợp trong profile. Ở đây, chúng ta có thể nhận diện một số vấn đề chính sẽ nảy sinh trong tiến trình cập nhật profile dựa trên feedbacks.

Thứ nhất, tiến trình cập nhật diễn ra chậm (vòng quay: hệ thống tính toán thích nghi, người sử dụng nhận và đánh giá tài nguyên, hệ thống cập nhật profile). Hơn nữa, người sử dụng thường có thói quen chỉ chấm điểm những gì ưa thích và đôi khi là những điều chán ghét ("điểm số" nằm gần khu vực hai đầu mút). Do đó, hệ thống cần phải tự bổ sung thêm những đánh giá "gián tiếp" không tường minh (implicit feedback) bằng cách suy diễn dựa trên quá trình khai thác và giao tiếp của người sử dụng (browse, print, delete, send, …).

Kế đến, tiến trình cập nhật được thực hiện dựa trên giả thiết là chỉ với feedbacks (đánh giá trực tiếp hay suy diễn gián tiếp) của người sử dụng thì cũng hoàn toàn đủ cho việc cập nhật profile. Vì vậy, tiến trình cũng chỉ chủ yếu dựa trên nguồn dữ liệu này. Trên thực tế, điều này không phải lúc nào cũng đúng. Ví dụ, profile hiện thời của người sử dụng đang thể hiện mối quan tâm về Trí tuệ nhân tạo, và hệ thống sẽ thường xuyên cung cấp những tài liệu liên quan đến chủ đề này. Sau một thời gian, nhu cầu của người sử dụng có thể thay đổi, và hệ thống liên tục nhận được những feedbacks là những điểm số đánh giá thấp của người sử dụng về những tài liệu đã nhận thuộc chủ đề Trí tuệ nhân tạo. Nếu chỉ dựa trên feedbacks, hệ thống chỉ có thể đoán biết được rằng người sử dụng không còn quan tâm đến chủ đề cũ nhưng gần như bất lực trong việc xác định chủ đề mới đang được quan tâm đặc biệt.

c) "Lối mòn" trong khai thác: Qua ví dụ ở trên, chúng ta cũng có thể thấy rằng một khi profile đã "ổn định" sau một quá trình cập nhật thì người sử dụng chỉ nhận toàn những gì được mô tả trong profile của mình và không có cơ hội khám phá những lãnh vực hay chủ đề thú vị mới.

1.2. Sự thích nghi cộng đồng

Theo nhu cầu thực tế, từ hơn một thập kỷ nay cũng đã xuất hiện nhiều công trình nghiên cứu và ứng dụng công nghệ thông tin nhằm hỗ trợ và khai thác sự hợp tác giữa cộng đồng trong các hệ thống thích nghi. Trong lãnh vực e-Learning, đó chính là nghiên cứu ứng dụng những kỹ thuật Tin học hỗ trợ sự hợp tác giữa những người học trong học tập (Computer-Supported Collaborative Learning) (Kumar, 1996; Stahl và cộng sự, 2006), hay trong lãnh vực tìm kiếm thông tin, hướng nghiên cứu về kỹ thuật tinh lọc dựa trên sự hợp tác (Collaborative Filtering – CF) cũng đang phát triển rất mạnh (Herlocker, 2002; Resnick và cộng sự, 1994). Chúng tôi gọi chung đây là sự thích nghi cá nhân dựa trên cộng đồng, hay vắn tắt hơn là sự thích nghi cộng đồng. Từ đây về sau, cụm từ thu gọn "sự thích nghi" dùng để chỉ chung cả hai loại thích nghi "cá nhân" và "cộng đồng".

Gần đây, nhiều nhóm nghiên cứu đã công bố những công trình ứng dụng CF vào trong các ITS (Kiared và cộng sự, 2006; Pintwart và cộng sự, 2006). Mục tiêu chính của bài báo này cũng liên quan đến hướng nghiên cứu ứng dụng CF để giải quyết một số vấn đề của sự thích nghi cá nhân trong các ITS. Điều đó có nghĩa là quan điểm của chúng tôi về sự thích nghi cộng đồng chỉ liên quan chủ yếu đến nhu cầu chia sẻ tài nguyên của những người học trong các ITS hơn là những vấn đề về phương pháp sư phạm hay giao tiếp xã hội giữa các cộng đồng người dạy và người học.

Nguyên tắc cơ bản của CF là người sử dụng có thể nhận được những tài nguyên hay thông tin thú vị xuất phát từ cộng đồng những người "gần gũi" với mình (Resnick và cộng sự, 1994), không kể những tài nguyên đã nhận được từ sự thích nghi cá nhân hoàn toàn dựa trên profile của mình. Ví dụ, trong một khóa học, mỗi sinh viên sở hữu một profile và nhận được những tài liệu thích hợp dựa trên profile của mình. Trong số đó, có thể có một tài liệu mà đa số những sinh viên nhận được đều đánh giá rất cao thì tài liệu này cũng sẽ được hệ thống gởi luôn đến những sinh viên còn lại (trước đây họ chưa nhận được tài liệu này). Có thể nói rằng kỹ thuật CF đã mô phỏng một hiện tượng khá phổ biến trong thực tế khi chúng ta thường xuyên nhận được nhiều loại thông tin bổ ích từ những người thân trong gia đình, họ hàng, đồng nghiệp, bạn bè, …

Như vậy, ưu điểm đầu tiên của kỹ thuật CF là sự chia sẻ tài nguyên giữa những người sử dụng, và ở từng lúc, hệ thống có thể bỏ qua các công đoạn phân tích nội dung, so khớp giữa profiles và tài nguyên. Từ đó, CF đã phần nào giải quyết được vấn đề lối mòn trong sự thích nghi cá nhân vì bây giờ, thông qua cộng đồng, người sử dụng sẽ có khả năng khám phá thêm những lãnh vực hay chủ đề mới, chưa được thể hiện trong profile của mình, mà có thể chúng rất thú vị và đáng quan tâm.

Vấn đề then chốt của CF chính là việc tạo lập và quản lý các cộng đồng người sử dụng (xem phần 2). Dưới góc nhìn của người sử dụng về tính hiệu quả thì những phần việc liên quan đến vấn đề này phải được đẩy chủ yếu về phía hệ thống, và người sử dụng không phải tốn thêm nhiều công sức ngoài việc cung cấp những feedbacks luôn phải cần thiết cho sự thích nghi cá nhân.

1.3. Mục tiêu

Mục tiêu của chúng tôi trong bài báo này là đề xuất một mô hình quản lý các cộng đồng, dùng làm nền tảng cho kỹ thuật CF nhằm đa dạng hóa tài nguyên phục vụ người học trong các ITS. Mô hình đề xuất bao gồm ba phần chính: a) mô hình các không gian cộng đồng đa tiêu chuẩn, b) phương pháp thích nghi dựa trên cộng đồng đa tiêu chuẩn, và c) cách tiếp cận suy diễn cộng đồng nhằm tích hợp những người sử dụng mới vào trong các cộng đồng đang tồn tại.

Về bố cục của bài báo, trong phần 2 kế tiếp, chúng tôi sẽ giới thiệu những nét chính

về hiện trạng những nghiên cứu có liên quan đến sự thích nghi cộng đồng. Phần 3 là nội dung chính của bài báo, giới thiệu đề xuất của chúng tôi về một mô hình quản lý cộng đồng đa tiêu chuẩn cho các hệ thống thích nghi như ITS. Cuối cùng, những kết luận về mô hình đề xuất và những hướng nghiên cứu mở rộng sẽ được trình bày trong phần 4.

II. HIỆN TRẠNG CÁC NGHIÊN CỨU LIÊN QUAN

Dưới góc độ khai thác các cộng đồng trong một hệ thống thích nghi có sử dụng CF, khái niệm cộng đồng được xem như một tập hợp những cá nhân gần gũi với nhau theo một tiêu chuẩn nào đó như: tuổi tác, nghề nghiệp, trình độ, chủ đề quan tâm, … và chúng tôi tìm hiểu những nghiên cứu xoay quanh các khía cạnh chính về sự thích nghi cộng đồng như: tạo lập cộng đồng, định vị người sử dụng trong các cộng đồng đã có, và cung cấp tài nguyên thích nghi dựa trên cộng đồng.

2.1. Tạo lập cộng đồng

Tạo lập cộng đồng là một tác vụ cơ bản của CF nhằm cung cấp những tài nguyên hay thông tin thích hợp cho người sử dụng. Theo cách tiếp cận cổ điển và cũng rất phổ biến hiện nay, hệ thống sẽ khai thác một ma trận đánh giá (ratings matrix) R như trong Hình 1, trong đó các dòng và cột sẽ tương ứng với các tập hợp người sử dụng U và tài nguyên W (ví dụ: văn bản, đoạn phim video, …), và mỗi giá trị rij của ma trận chính là điểm số của người sử dụng ui đã đánh giá, nếu có, trên tài nguyên dj mà trước đây hệ thống đã cung cấp cho người này. Lưu ý, mỗi dòng R[i] cũng được xem như một thành phần nằm trong profile của ui. Hiện nay, trong cách tiếp cận cổ điển có hai phương pháp chính để tạo lập cộng đồng: phương pháp láng giềng gần nhất và phương pháp xác suất.

Phương pháp láng giềng gần nhất (Neighborhood-based Method). Đây là phương pháp được sử dụng nhiều nhất (Breese và cộng sự, 1998; Herlocker, 2002; Resnick và cộng sự, 1994). Trong phương pháp này, để tạo cộng đồng cho ui, trước tiên hệ thống sẽ tính toán độ tương đồng (hay khoảng cách) giữa dòng R[i] với tất cả các dòng còn lại, dựa trên hệ số tương quan Pearson (Pearson correlation) hoặc côsin. Sau đó, hệ thống sẽ chọn lọc và giới hạn lại những người sử dụng gần ui nhất, dựa trên một ngưỡng về khoảng cách (bán kính) hay dựa trên một kích thước cố định cho cộng đồng (N thành viên). Phương pháp này rất đơn giản và hiệu quả trong phần lớn trường hợp, ngoại trừ nhược điểm về thời gian và chi phí vì phải tính toán trên toàn bộ ma trận đánh giá R.

W / U	d_1	...	d_j	...	d_n
u_1	r_{11}	...	r_{1j}	...	r_{12}
...					
u_i	r_{i1}	...	r_{ij}	...	r_{in}
...					
u_m	r_{m1}	...	r_{mj}	...	r_{mn}

Hình 4a: Bảng mã La Tinh G2 cho chữ cái và số đếm

T Phương pháp xác suất. Phương pháp này ít khi được sử dụng so với phương pháp trên. Theo qui trình tìm kiếm tài nguyên thích nghi, đầu tiên hệ thống sẽ áp dụng các phương pháp máy học trên một tập hợp ví dụ mẫu được rút trích từ ma trận R để xây dựng một mô hình xác suất, dùng trong việc dự đoán sự đánh giá của những người sử dụng trên các tài nguyên: prediction(u, d) = ? Nếu giá trị dự đoán vượt quá một ngưỡng nào đó thì hệ thống sẽ gởi d đến cho u. Về nguyên tắc, việc tạo lập các cộng đồng đã được thực hiện một cách gián tiếp thông qua quá trình xây dựng mô hình dự đoán bằng xác suất nói trên. Ví dụ, (Miyahara & Pazzani, 2000) đã áp dụng phương pháp phân lớp nhị phân Bayes: prediction(u, d) $\in$ {like, dislike}, trong đó những người sử dụng khác sẽ được xem như các thuộc tính dùng trong quá trình phân lớp. Như vậy, việc xác định tập hợp những thuộc tính

{ui} có ý nghĩa trong quá trình phân lớp (Feature Selection) sẽ tương đương với việc xác định cộng đồng của u. Phương pháp này khá phức tạp và hơn nữa, cộng đồng của một người sử dụng u sẽ phải phụ thuộc vào một tài nguyên d cụ thể, trong khi khái niệm cộng đồng nói chung phải tương đối độc lập với tài nguyên.

Tóm lại, chúng ta thấy các phương pháp tạo lập cộng đồng trong các hệ thống thích nghi đều dựa trên một tiêu chuẩn duy nhất là những feedbacks của người sử dụng (tiêu chuẩn Ratings). Tuy nhiên, trong thực tế thì mọi người thường hay nhận được những tài nguyên hay thông tin bổ ích từ nhiều cộng đồng khác nhau: người thân trong gia đình, họ hàng, bạn bè, đồng nghiệp, …

2.2. Định vị người sử dụng mới trong các cộng đồng

Trong kỹ thuật thích nghi CF cổ điển, người sử dụng sẽ nhận được tài nguyên hay thông tin thú vị từ cộng đồng theo tiêu chuẩn Ratings. Khi có một người mới đăng ký, hệ thống phải đối mặt với vấn đề khó khăn về khởi động (cold-start problem) vì cộng đồng theo tiêu chuẩn Ratings của người mới này vẫn còn chưa biết, và do đó, hệ thống không thể cung cấp những tài nguyên thích hợp (Rashid và cộng sự, 2002). Hiện nay có hai cách tiếp cận chính để giải quyết vấn đề này.

Khai thác thông tin thăm dò. Các hệ thống hiện nay thường vận dụng phương pháp phổ biến này để yêu cầu người mới đăng ký phải đánh giá, chấm điểm một số lượng tối thiểu tài nguyên. Hệ thống có thể đưa ra một danh sách tài nguyên đã được chọn một cách ngẫu nhiên tại mọi thời điểm hoặc chọn theo một tiêu chí thống kê nào đó để người sử dụng đánh giá. Tuy nhiên, phương pháp này có thể gây khó khăn cho người sử dụng nếu có những tài nguyên mà người sử dụng không thể đưa ra những đánh giá cụ thể vì nhiều lý do khác nhau. Vì vậy, người sử dụng mới cũng có thể tự mình chọn lấy những tài nguyên để chấm điểm.

Cung cấp profile mẫu. Trước tiên, hệ thống sẽ xây dựng sẵn những profiles mẫu, và sau đó dựa trên những giá trị nhân khẩu học (tuổi, nghề nghiệp, …) hay trả lời một số câu hỏi mà người sử dụng sẽ được tự động gán cho một trong những profiles mẫu. Xuất phát từ profile này, hệ thống sẽ áp dụng sự thích nghi cá nhân nhằm cung cấp những tài nguyên phù hợp cho người sử dụng, và quá trình phản hồi feedbacks đã có thể được bắt đầu để sau đó được dùng trong việc định vị.

Nhìn chung, những phương pháp nêu trên đều yêu cầu người sử dụng phải tiêu tốn khá nhiều công sức để định vị vào một cộng đồng ban đầu mà cũng chưa chắc đã thật sự phù hợp.

2.3. Cung cấp tài nguyên thích nghi dựa trên cộng đồng

Về nguyên tắc, việc cung cấp những tài nguyên thích nghi cho một người sử dụng a chính là dự đoán sự đánh giá hay điểm số của người này đối với một tài nguyên d nào đó (Breese và cộng sự, 1998). Nếu điểm số dự đoán prediction(a, d) do hệ thống tính toán vượt qua một ngưỡng nào đó thì tài nguyên d sẽ được gởi đến cho người sử dụng a.

$$prediction(a,d) = \overline{r}_a + k\sum_{u \in C_a} w(a,u)(r_{u,d} - \overline{r}_u) \quad [1]$$

với $\overline{r}_a, \overline{r}_u$: điểm trung bình của a và u (giá trị trung bình của các dòng R[a] và R[u])

k : hệ số chuẩn hóa thang điểm

C_a : cộng đồng của người sử dụng a

$r_{u,d}$: điểm số mà u đánh giá trên d

w(a, u): hệ số tương quan Pearson giữa hai người sử dụng a và u:

$$Pearson_correlation(a,u) = \frac{\sum_{j \in D_{a,u}} (r_{a,j} - \overline{r}_a)(r_{u,j} - \overline{r}_u)}{\sqrt{\sum_{j \in D_{a,u}} (r_{a,j} - \overline{r}_a)^2 \sum_{j \in D_{a,u}} (r_{u,j} - \overline{r}_u)^2}}$$

Da, u: tập hợp tài nguyên mà a và u đều đã cho điểm

Dựa trên sự phân tích hiện trạng các nghiên cứu về thích nghi cộng đồng, chúng tôi nhận ra nhu cầu thật sự về một mô hình quản lý các cộng đồng trong các hệ thống thích nghi. Do đó, mục tiêu của chúng tôi trong bài báo này là đề xuất một mô hình quản lý cộng đồng thể hiện các đặc tính sau trong một hệ thống thích nghi: khai thác cộng đồng đa tiêu chuẩn nhằm nâng cao tính đa dạng trong sự thích nghi và cho phép định vị hiệu quả người sử dụng trong các cộng đồng đa tiêu chuẩn.

III. MÔ HÌNH QUẢN LÝ CỘNG ĐỒNG ĐA TIÊU CHUẨN

Mô hình đề xuất bao gồm ba phần: mô hình các không gian cộng đồng đa tiêu chuẩn, phương pháp thích nghi, và phương pháp suy diễn cộng đồng dựa trên mối quan hệ giữa các không gian.

3.1. Mô hình các không gian cộng đồng đa tiêu chuẩn (α-Community Spaces Model)

Dưới góc nhìn mang tính chức năng của việc cung cấp tài nguyên thích nghi, một cộng đồng trong mô hình của chúng tôi được định nghĩa như một tập hợp những người sử dụng gần gũi với nhau dựa trên một tiêu chuẩn so sánh nào đó. Từ đó, mỗi thuộc tính trong các profiles (tuổi tác, nghề nghiệp, nơi cư trú, trình độ, chủ đề quan tâm, feedbacks, …) đều có thể được sử dụng như một tiêu chuẩn để thành lập một không gian hay một tập hợp các cộng đồng . Như vậy, một người sử dụng có thể đồng thời thuộc nhiều cộng đồng khác nhau theo từng tiêu chuẩn riêng biệt.

3.1.1. Không gian cộng đồng

Cho U là tập hợp những người sử dụng (|U| = m) và A là tập hợp các thuộc tính hay tiêu chuẩn (|A| = n). Với mọi $\alpha \in A$, ta định nghĩa một quan hệ tương đương R_α như sau:

$$\forall u, u' \in U, u\ R_\alpha\ u' \Leftrightarrow \alpha(u) = \alpha(u')$$

Ví dụ: $u\ R_{Location}\ u' \Leftrightarrow$ Location(u) = Location(u') (u và u' cư ngụ cùng thành phố).

Một cộng đồng thành lập dựa trên α (α-community), ký hiệu $G\alpha_k$, là một lớp tương đương của R_α. Không gian cộng đồng α (α-community space), ký hiệu Ω_α, chính là tập hợp thương của R_a:

$$\Omega_\alpha = U/R_\alpha = \{G\alpha_1, \ldots, G\alpha_r\}$$

Một tiêu chuẩn phức hợp là một tập hợp $P \subseteq A$ và có ít nhất hai tiêu chuẩn. Không gian cộng đồng phức hợp Ω_p cũng được định nghĩa dựa trên một quan hệ tương đương R_P tương tự như ở trên:

$$\forall u, u' \in U, u\ R_P\ u' \Leftrightarrow \forall (\alpha \in P, \alpha(u) = \alpha(u'))$$ [2]

Trong phạm vi bài báo này, chúng tôi khảo sát chủ yếu các không gian cộng đồng "đơn" Ω_α.

3.1.2. Vectơ định vị

Với mỗi tiêu chuẩn $\alpha_j \in A$, một người sử dụng u sẽ thuộc một cộng đồng nào đó trong không gian $\Omega\alpha_j$. Vectơ định vị (position vector) của u, ký hiệu P_u, là danh sách n cộng đồng của u trong các không gian tương ứng:

$$P : U \rightarrow V\alpha_1 \times \ldots \times V\alpha_n$$

$$P(u) = P_u = (pu_1, \ldots, pu_n)$$

với $V\alpha_j$ là miền giá trị của α_j.

3.1.3. Bảng cộng đồng

Bảng cộng đồng (α-community table) Tmxn (xem Hình 2) biểu diễn toàn bộ các không gian cộng đồng của hệ thống, và được đặc trưng bởi hai tập U và A: $T[u_i, \alpha_j] = \alpha_j(u_i)$

Về nguyên tắc, bảng T sẽ được hệ thống lấp đầy theo từng cột, qua việc phân lớp hay phân hoạch những người sử dụng theo từng tiêu chuẩn tương ứng. Tùy theo cấu trúc nội dung của các profiles trong các lãnh vực ứng dụng mà có thể có những tiêu chuẩn đơn giản cho sự phân hoạch: hệ thống sẽ phân nhóm những người sử dụng bằng cách so sánh trực tiếp các giá trị trong các

profiles của họ (cùng tuổi hay nhóm tuổi, cùng nghề nghiệp hay nhóm nghề nghiệp, cùng địa bàn cư trú, …). Ngược lại, cũng có những tiêu chuẩn phức tạp cho sự phân hoạch (ví dụ nền tảng kiến thức, các chủ đề quan tâm hay tiêu chuẩn Ratings), và hệ thống có thể phải vận dụng đến những phương pháp đã trình bày trong phần 2.1 hay những phương pháp khác mà chúng tôi sẽ không đi sâu vào chi tiết trong phạm vi bài báo này (Jain và cộng sự, 1999; Nguyen và cộng sự, 2005).

	A / U	α_1	...	α_j	...	α_n
	u_1					
	...					
$T[u_i]$	u_i			$T[u_i, \alpha_j]$		
	...					
	u_m					
				$T\alpha_j$		

Hình 2. Bảng cộng đồng T_{mxn}.

3.2. Phương pháp thích nghi dựa trên cộng đồng đa tiêu chuẩn

Trong phần 2.3, chúng ta đã tìm hiểu phương pháp cung cấp tài nguyên thích nghi dựa trên những cộng đồng được tạo lập theo tiêu chuẩn Ratings. Phương pháp này sẽ không áp dụng được trong những trường hợp không tồn tại khoảng cách giữa những thành viên trong cộng đồng (w(a, u) trong [1]) hoặc tất cả các khoảng cách, nếu có, đều như nhau (ví dụ cộng đồng những sinh viên trung niên). Do đó, chúng tôi đề xuất phương pháp thích nghi mới dựa trên sự đồng thuận của cộng đồng (level of agreement) như sau.

Cho trước hai ngưỡng giá trị t_{Score}, $t_{Agreement}$. Giả sử $u_0 \in U$, $d \in W$, và G_α là cộng đồng của u_0 theo tiêu chuẩn $\alpha \in A$. Người sử dụng u_0 sẽ nhận được d nếu hai điều kiện sau đồng thời được thỏa:

$$\left(\frac{1}{|G_a|}\sum_{u \in G_a} r_{u,d}\right) \geq t_{Score}$$

$$|\{r_{u,d} \mid (u \in G_\alpha) \text{ và } (r_{u,d} \neq \text{null})\}| \geq t_{Agreement}$$

với $r_{u,d}$: đánh giá (điểm số) của u đối với d

Phương pháp này cũng có thể được cải biên bằng cách thu hẹp G_α, chỉ xét những "láng giềng gần nhất" của u_0 hoặc chỉ xét tập hợp những thành viên tiêu biểu/đặc trưng (trọng tâm) của G_α.

Phương pháp của chúng tôi sẽ góp phần làm gia tăng tính đa dạng của tài nguyên trong sự thích nghi cộng đồng. Ban đầu, quá trình thích nghi riêng lẻ theo từng cộng đồng $G\alpha_j$ sẽ tạo ra n tập hợp tài nguyên, và các tài nguyên d nằm trong tập kết quả cuối cùng sẽ được tính toán như sau:

$$prediction(u,d) = \sum_{j=1}^{n} w_j . prediction_j(u,d)$$

với w_j: trọng số của α_j, ban đầu sẽ được gán bằng nhau cho tất cả và sẽ được tinh chỉnh theo một phương pháp máy học nào đó (Claypool và cộng sự, 1999)

$prediction_j(u, d)$: đánh giá của cộng đồng của u theo α_j đối với d

3.3. Phương pháp suy diễn cộng đồng

Trong suốt quá trình khai thác, vectơ định vị của một người sử dụng thường không hoàn hảo, đặc biệt là vào khoảng thời gian đầu mới vừa đăng ký hoặc khi có sự tiến hóa đột biến trong nhu cầu. Khi đó, vectơ vẫn còn có thể thiếu một số cộng đồng chưa xác định được hoặc chứa những cộng đồng không chính xác. Nguyên nhân chủ yếu của sự không hoàn hảo này là do sự thiếu sót hay do chất lượng kém của dữ liệu trong profile. Như vậy, hệ thống cần phải có một phương pháp hiệu quả để hiệu chỉnh vectơ định vị nhằm nâng cao chất lượng của sự thích nghi cộng đồng.

Chúng tôi đề xuất cách tiếp cận suy diễn

một cộng đồng không hoàn chỉnh (thiếu hoặc sai) trong một vectơ P_u dựa trên các mối quan hệ cấu trúc (các cột) trong bảng cộng đồng T (Nguyen và cộng sự, 2007). Nói một cách cụ thể hơn, chúng tôi muốn giải đáp loại câu hỏi: "Có thể định vị hay không một người sử dụng trong $\Omega_{Ratings}$ bằng cách sử dụng những dữ liệu đơn giản đã có trong profile mà không cần yêu cầu người này cung cấp những feedbacks ?". Nếu trả lời được câu hỏi này thì sẽ góp phần giải quyết vấn đề đã nêu trong 2.2.

Về mặt lý thuyết, chúng tôi sử dụng các khái niệm cơ bản của lý thuyết tập thô (Rough Sets Theory) (Pawlak, 1982, 2004) và đề xuất một số mở rộng để giải quyết bài toán suy diễn cộng đồng.

3.3.1. Khái niệm cơ bản của lý thuyết tập thô

Trước hết, tập thuộc tính A được chia thành hai phần rời nhau: $A = C \cup D$, với D chứa duy nhất một thuộc tính quyết định (decision attribute) d, và C chứa các thuộc tính điều kiện (condition attributes) còn lại. Bảng cộng đồng T sẽ được gọi là bảng quyết định (decision table), và mỗi dòng có thể được xem như một luật (decision rule). Ví dụ, trong Bảng 1 với D = {Ratings}, ta có luật u7:

(Occupation = "Công chức", Location = "TPHCM", Topic = "WEB") → (Ratings = "Nhóm R5")

Trong một luật, vế trái chứa các điều kiện (conditions) và vế phải được gọi là quyết định (decision). Từ đây về sau, một luật có thể được biểu diễn dưới các dạng như: C → D hay C → d. Xét về tính hiệu quả, người ta thường cố gắng tìm kiếm tập con thuộc tính điều kiện $P \subseteq C$ sao cho: P → D

Một luật u : P → D được gọi là nhất quán hay chắc chắn (certain decision rule) nếu:

$$[u]_P \subseteq [u]_D$$

trong đó, $[u_P]$ và $[u_D]$ là các lớp tương đương của u theo các quan hệ R_P và R_D (xem [2] trong 3.1).

Trong Bảng 1, ta có u_3, u_7, u_8 và u_9 là các luật nhất quán.

Cho trước P và D, ta định nghĩa miền khẳng định hay miền dương (positive region) tương ứng với R_P, ký hiệu $POS_P(D)$, là tập hợp chứa tất cả các luật nhất quán, và P là một tập rút gọn (reduction) của C nếu:

$$POS_P(D) = POS_C(D) \quad [3]$$

3.3.2. Khái niệm cơ bản của lý thuyết tập thô

Giả sử vectơ P_u chứa một cộng đồng

U	α = Occupation	α=Location	α=Topic	α=Rating
u_1	Doanh nhân	TPHCM	CNPM	Nhóm R1
u_2	Công chức	TPHCM	CNPM	Nhóm R4
u_3	Công chức	TPHCM	CSDL	Nhóm R2
u_4	Công chức	TPHCM	CNTT	Nhóm R1
u_5	Công chức	TPHCM	CNTT	Nhóm R4
u_6	Công chức	TPHCM	CNTT	Nhóm R3
u_7	Công chức	TPHCM	WEB	Nhóm R5
u_8	Công chức	Hà Nội	CSDL	Nhóm R5
u_9	Doanh nhân	Hà Nội	CSDL	Nhóm R5
u_{10}	Doanh nhân	Huế	CSDL	Nhóm R3
u_{11}	Doanh nhân	Huế	CSDL	Nhóm R2
u_{12}	Doanh nhân	Huế	CSDL	Nhóm R3

Bảng 1. Ví dụ bảng quyết định với D = {Ratings}.

không hoàn chỉnh (thiếu hoặc sai) thuộc Ω_α. Khi đó, hệ thống sẽ chọn D = {α} và C = A \ D, và tìm một luật nhất quán trong $POS_C(D)$ để suy diễn ra cộng đồng phải tìm trong P_u. Cách tiếp cận này sẽ làm giảm nhẹ rất nhiều công sức của người sử dụng trong quá trình định vị.

Giả sử, người sử dụng mới là một công chức ở "TPHCM", thường quan tâm đến chủ đề "CSDL" và cộng đồng của người này theo tiêu chuẩn Ratings vẫn chưa được xác định:

P_u=("Công chức", "TPHCM", "CSDL", _)

Khi đó, ta có thể áp dụng luật nhất quán u_3 trong Bảng 1:

(Occupation = "Công chức"», Location = "TPHCM", Topic = "CSDL") → (Ratings = "Nhóm R2")

và thu được: P_u = ("Công chức", "TPHCM", "CSDL", "Nhóm R2")

Trong trường hợp hệ thống không tìm thấy bất kỳ luật nhất quán nào mà thay vào đó là những luật mâu thuẫn với nhau (cùng các điều kiện ở vế trái nhưng khác quyết định bên vế phải) thì phương pháp thường dùng để giải quyết là chọn quyết định phổ biến nhất trong các luật này (Clark & Niblett, 1989). Một số phương pháp suy diễn giá trị khác cũng đã được nghiên cứu và đề xuất (Grzymala-Busse & Hu, 2000).

3.3.3. Suy diễn nhiều cộng đồng không hoàn chỉnh trong vectơ định vị

Trong một hệ thống thích nghi cộng đồng đa tiêu chuẩn, vectơ định vị P_u của một người sử dụng thường chứa cùng lúc nhiều giá trị không hoàn chỉnh. Trong khi đó, theo lý thuyết tập thô cũng như hầu hết các phương pháp suy diễn theo luật thì thuộc tính quyết định d phải được xác định trước, và hệ thống chỉ có thể suy diễn được giá trị của thuộc tính này. Khi có nhiều cộng đồng cần suy diễn trong vectơ P_u thì cách giải quyết tự nhiên nhất là lần lượt chọn các thuộc tính có vấn đề làm thuộc tính quyết định trong các quá trình suy diễn và các thuộc tính sẽ được chọn lựa để suy diễn theo một thứ tự hoàn toàn ngẫu nhiên. Như vậy, chất lượng suy diễn các cộng đồng có vấn đề trong P_u không phải lúc nào cũng được bảo đảm. Vấn đề đặt ra là: "Có phải tất cả các cộng đồng theo những tiêu chuẩn khác nhau đều có thể được suy diễn ?" và "Có hay không một thứ tự tối ưu (cục bộ) trong việc suy diễn các giá trị tiêu chuẩn α_j ?". Chúng ta cần phải xây dựng một chiến lược suy diễn hay thay thế các giá trị có vấn

U	a = Occupation	α = Location	α = Topic	α = Rating	POS_C (Topic)	POS_C (Ratings)
u_1	Doanh nhân	TPHCM	CNPM	Nhóm R1	x	x
u_2	Công chức	TPHCM	CNPM	Nhóm R4		x
u_3	Công chức	TPHCM	CSDL	Nhóm R2	x	x
u_4	Công chức	TPHCM	CNTT	Nhóm R1	x	
u_5	Công chức	TPHCM	CNTT	Nhóm R4		
u_6	Công chức	TPHCM	CNTT	Nhóm R3	x	
u_7	Công chức	TPHCM	WEB	Nhóm R5	x	x
u_8	Công chức	Hà Nội	CSDL	Nhóm R5	x	x
u_9	Doanh nhân	Hà Nội	CSDL	Nhóm R5	x	x
u_{10}	Doanh nhân	Huế	CSDL	Nhóm R3	x	
u_{11}	Doanh nhân	Huế	CSDL	Nhóm R2		
u_{12}	Doanh nhân	Huế	CSDL	Nhóm R3		

Bảng 2. Ví dụ tính toán độ đo M_1

đề trong vectơ định vị.

Tóm lại, mục tiêu của chúng tôi trong phần này là định nghĩa một số độ đo giúp cho việc đánh giá chất lượng của một tiêu chuẩn trong quá trình suy diễn cộng đồng. Trước hết, các độ đo sẽ cho phép hệ thống xác định tập hợp những tiêu chuẩn α_j nào có thể được suy diễn và hơn nữa, có thể xác định một độ ưu tiên (suy diễn) trên tập hợp này.

M_1. Độ đo dựa trên chất lượng suy diễn. Ta định nghĩa tiêu chuẩn D_2 ưu tiên hơn D_1 như sau:

$D1 \lhd D2 \Leftrightarrow | POSA\backslash D1(D1) | \leq | POSA\backslash D2(D2) |$ [4]

Giả sử vectơ của một người sử dụng mới như sau: P_u = ("Doanh nhân", "Huế", _, _). Bảng 2 cho chúng ta thấy Topic có độ ưu tiên cao hơn Ratings, và để hiệu chỉnh P_u, hệ thống sẽ bắt đầu bằng cách chọn D = {Topic}. Các luật u_{10}, u_{11} và u_{12} cho phép suy diễn giá trị "CSDL" cho Topic:

P_u = ("Doanh nhân", "Huế", "CSDL", _)

Sau cùng, các luật u_1, u_2, u_3, u_7, u_8 và u_9 trong POS_C(Ratings) không thể dùng cho P_u, và "Nhóm R3", là giá trị thống trị trong các luật u_{10}, u_{11} và u_{12}, có thể được gán cho tiêu chuẩn Ratings.

Ví dụ trên cũng cho thấy rằng việc suy diễn không phải lúc nào cũng có thể thực hiện được do kích thước hạn chế của các POS_C(D). Vì vậy, trên thực tế, việc vận dụng thêm các phương pháp suy diễn khác dựa trên luật là hết sức cần thiết. Thông thường, trong các phương pháp này, chất lượng của một luật suy diễn r : C → D được thể hiện bằng hai yếu tố certain và support như sau:

$$Certain(r) = \frac{\#occurences(r)}{\#occurences(C)}$$

$$Support(r) = \frac{\#occurences(r)}{|U|}$$

Như vậy, độ đo M_1 sẽ được cải biên như sau.

Cho ζ_D là một tập hợp các luật suy diễn. Chất lượng của ζ_D được định nghĩa bởi:

$$\varphi(\zeta_D) \equiv \varphi(D) = \sum_{r \in \zeta_D} Support(r).Certain(r)$$

Kế đến, ta định nghĩa độ ưu tiên giữa D_1 và $D_2 \in A$ như sau:

$D_1 \lhd D_2 \Leftrightarrow | \varphi(D_1) | \leq | \varphi(D_2) |$ [5]

Đây là sự tổng quát hóa của công thức [4] vì với các luật nhất quán, ta có:

$$\varphi(D) = \frac{|POS_C(D)|}{|U|}$$

M_2. Độ đo hướng đến sự thuận tiện cho người sử dụng. Về phía người sử dụng, chúng tôi đề xuất một độ đo hướng đến sự thuận tiện và dễ dàng trong việc cung cấp dữ liệu cần thiết cho sự định vị trong các cộng đồng. Trước tiên, định nghĩa về tập rút gọn của C (xem [3]) là khá ngặt và trên thực tế thì các tập rút gọn P thường gần bằng C trong khi hệ thống không thể yêu cầu người sử dụng cung cấp quá nhiều thông tin. Do đó, quá trình suy diễn các cộng đồng có thể không thực hiện được nếu thông tin do người sử dụng cung cấp không phủ hết những thuộc tính điều kiện trong các tập rút gọn của C. Vì vậy, chúng tôi sử dụng các tập rút gọn xấp xỉ (approximative reduction) của C nhằm mục đích thu hẹp tối đa vế trái P của các luật r : P → D mà vẫn bảo đảm được chất lượng suy diễn ở một mức độ nhất định.

Cho trước ngưỡng chất lượng θ. Các tập rút gọn xấp xỉ P của C được định nghĩa như sau:

$$R_D^{(\theta)} = \{P \subseteq C \mid \sigma(P) \geq \theta\}$$

với $$\sigma_{C,D}(P) = 1 - \frac{POS_{C\backslash P}(D)}{POS_C(D)}$$

Một tập rút gọn xấp xỉ $P \in R_D^{(\theta)}$ chính là tập hợp những tiêu chuẩn có thể "chấp nhận được" và hữu dụng trong trường hợp

hệ thống quan tâm nhiều đến tính hiện thực của các điều kiện ở vế trái của các luật hơn là sự chính xác tuyệt đối của chúng. Ví dụ, nếu $\theta = 0, 8$, thì $R_{Ratings}^{(\theta)}$ sẽ chứa 2 tập rút gọn xấp xỉ của C là P_1 = {Location, Topic} và P_2 = {Occupation, Location}, trong khi nếu $\theta = 1$ thì sẽ không có tập rút gọn nào thật sự bé hơn C.

Xuất phát từ $R_D^{(\theta)}$, ta có thể định nghĩa các độ ưu tiên trên các thuộc tính quyết định, hoặc bằng cách a) giới hạn kích thước các tập rút gọn xấp xỉ, hoặc b) xác định danh sách các thuộc tính điều kiện C_0 phải chứa trong các tập rút gọn xấp xỉ:

$$D_1 \lhd D_2 \Leftrightarrow |\{P \in R_{D1}^{(\theta)} \mid |P| \leq \lambda\}| \leq |\{Q \in R_{D2}^{(\theta)} \mid |Q| \leq \lambda\}| \quad [6]$$

$$D_1 \lhd D_2 \Leftrightarrow |\{P \in R_{D1}^{(\theta)} \mid C_0 \subseteq P\}| \leq |\{Q \in R_{D2}^{(\theta)} \mid C_0 \subseteq Q\}| \quad [7]$$

Công thức [6] ưu tiên cho tiêu chuẩn nào tạo ra nhiều tập rút gọn xấp xỉ có kích thước nhỏ (cần ít thông tin của người sử dụng cho quá trình suy diễn), trong khi công thức [7] quan tâm nhiều đến nội dung của các tập rút gọn xấp xỉ hơn là kích thước của chúng (ưu tiên những tiêu chuẩn có thể được suy diễn bằng những điều kiện đơn giản như nơi cư trú, tuổi tác, nghề nghiệp, …). Trong mọi trường hợp, đây là những độ đo hướng về sự thuận tiện cho người sử dụng.

Trên thực tế, vì bài toán xác định tất cả các tập rút gọn của C, cần thiết cho việc tính toán các công thức [6] và [7], là một bài toán NP-khó (Skowron & Rauszer, 1992), cho nên ta có thể giới hạn trước kích thước của $R_D^{(\theta)}$ nhằm làm giảm độ phức tạp của tính toán hoặc tìm cách xác định ngưỡng θ sao cho trong phần lớn trường hợp thì các tập rút gọn xấp xỉ đều chứa C_0.

Tóm lại, về nguyên tắc thì cấu trúc (các cột) của bảng T phụ thuộc vào một hệ thống ứng dụng cụ thể vì thuộc tính quyết định phải được xác định ngay từ ban đầu, trong khi cách tiếp cận suy diễn cộng đồng đa tiêu chuẩn của chúng tôi cho phép bỏ qua sự xác định này và như vậy đã tạo ra sự đa dạng trong góc nhìn "cấu trúc" đối với bảng T mà chúng tôi đã đề cập trong phần 3.1.

IV. KẾT LUẬN

Trong bài báo này, chúng tôi đã đề xuất một mô hình quản lý cộng đồng đa tiêu chuẩn nhằm khai thác triệt để các thuộc tính trong các profiles của một ITS (thông tin cá nhân, mục tiêu, trình độ, nền tảng kiến thức, chủ đề quan tâm, …), và qua đó có thể cung cấp những tài nguyên thích nghi một cách đa dạng cho người học. Nhìn chung, mô hình đề xuất đã góp phần giải quyết những vấn đề trong sự thích nghi được đề cập trong những phần đầu.

Trước hết, cách tiếp cận của chúng tôi hỗ trợ sự tạo lập cộng đồng theo mọi tiêu chuẩn trong profiles, khi cần thiết, mà không bị giới hạn ở tiêu chuẩn Ratings như trong kỹ thuật CF cổ điển. Hơn nữa, các cộng đồng trong cách tiếp cận của chúng tôi không còn chỉ là những kết quả trung gian tạm thời trong quá trình tìm kiếm tài nguyên thích nghi mà thật sự trở thành một thực thể hiện hữu bền vững trong hệ thống thích nghi và là một phương thức giao tiếp mới bên cạnh feedbacks. Ngoài ra, cách tiếp cận suy diễn cộng đồng cũng đã góp phần làm giảm nhẹ công sức của người sử dụng trong quá trình định vị, đặc biệt đối với những tiêu chuẩn khó tạo lập cộng đồng như nền tảng kiến thức, ...

Mặt khác, trên thực tế, các hệ thống sử dụng CF phải đối mặt với vấn đề về mật độ thưa thớt của ma trận R (sparsity problem) dẫn đến chất lượng kém trong tính toán hệ số tương quan Pearson giữa hai người sử dụng. Ngược lại, cách tiếp cận suy diễn cộng đồng theo tiêu chuẩn Ratings hoàn toàn không gặp phải vấn đề quan trọng này cũng như không dựa vào giả thiết không vững chắc về tính đầy đủ của feedbacks (xem 1.1.b).

Trước mắt, chúng tôi mong muốn nghiên cứu khai thác mở rộng các cộng đồng như một phương thức giao tiếp mới giữa người sử dụng và hệ thống, ngoài những

feedbacks. Trên cơ sở này, ngoài quá trình suy diễn cộng đồng của hệ thống, người sử dụng cũng có thể quan sát, nhận diện và tự chọn cho mình những cộng đồng thích hợp, dựa trên một mô hình giao tiếp trực quan với hệ thống qua các cộng đồng (Nguyen và cộng sự, 2005).

Xa hơn nữa, chúng tôi mong muốn nghiên cứu khai thác các cộng đồng để cập nhật ngược lại profile của người sử dụng. Ví dụ, vào thời điểm tiến hóa đột biến của profile mà quá trình cập nhật dựa trên feedbacks lại diễn ra quá chậm và không phải luôn đầy đủ (xem 1.1.b), người sử dụng có thể được định vị vào một cộng đồng mới (bằng sự suy diễn hay tự chọn qua giao tiếp với hệ thống), và sau đó sẽ được kế thừa profile đặc trưng của cộng đồng mới này. Điều này đòi hỏi phải tiến hành những nghiên cứu về khái niệm profile của cộng đồng và sự tích hợp những profiles lại với nhau.

TÀI LIỆU THAM KHẢO

[1] Amato, G., & Straccia, U. (1999). User Profile Modeling and Applications to Digital Libraries. In Proceedings of the 3rd European Conference on Research and Advanced Technology for Digital Libraries (ECDL'99), LNCS, 1696, France (pp. 184-197).

[2] Breese, J.-S., Heckerman, D., & Kadie C. (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In Proceedings of the 14th Conference on Uncertainty In Artificial Intelligence (UAI'98), Wisconsin, USA (pp. 43-52).

[3] Brooks, C., Greer, J., Melis, E., & Ullrich, C. (2006). Combining ITS and eLearning Technologies: Opportunities and Challenges. In Proceedings of the 8th International Conference on Intelligent Tutoring Systems (ITS 2006), Jhongli, Taiwan (pp. 278-287).

[4] Brusilovsky, P. (2001). Adaptive Hypermedia. User Modeling and User-Adapted Interaction, 11, 87-110.

[5] Clark, P., & Niblett, T. (1989). The CN2 induction algorithm. Machine Learning, 3 (4), 261-284.

[6] Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., & Sartin M. (1999). Combining Content-Based and Collaborative Filters in an Online Newspaper. In Proceedings of the 22nd International ACM Conference on Research and Development in Information Retrieval (SIGIR'99), USA.

[7] Grzymala-Busse, J.-W., & Hu, M. (2000). A Comparison of Several Approaches to Missing Attribute Values in Data Mining. In Proceedings of the 2nd Conference on RS and Current Trends in Computing (RSCTC'00), Canada.

[8] Herlocker, J.-L. (2002). Understanding and Improving Automated Collaborative Filtering Systems. Ph.D Dissertation, University of Minnesota, USA.

[9] Jain, A.-K., Murty, M.-N., & Flynn P.-J. (1999). Data Clustering: A Review. ACM Computing Surveys, 31 (3), 264-323.

[10] Kiared, S.-A., Razek, M.-A., & Frasson, C. (2006). The Pyramid Collaborative Filtering Method: Toward an Efficient E-Course. In Proceedings of the Intelligent Tutoring Systems (IASTED'06), Puerto Vallarta, Mexico (pp. 248-257).

[11] Kumar, V.-S. (1996). Computer-Supported Collaborative Learning: Issues for Research. In Proceedings of the 8th Annual Graduate Symposium on Computer Science, Canada (Retrieved August 2004).

[12] Madhour, H., & Forte, M.-W. (2006). Semantic learning and extended student model: Towards an AHAM-based adaptive system. In Proceedings of the 5th IASTED International Conference on Web-based Education (IASTED'06), Puerto Vallarta, Mexico (pp. 349-354).

[13] Miyahara, K., & Pazzani, M.-J.

(2000). Collaborative Filtering with the Simple Bayesian Classifier. In Proceedings of the 6th Pacific Rim International Conference on Artificial Intelligence (PRICAI'00), Australia (pp. 679-689).

[14] Montaner, M., López, B., & De La Rosa, J.-L. (2003). A Taxonomy of Recommender Agents on the Internet. Artificial Intelligence Review, 19, 285-330. Kluwer Publishers.

[15] Nguyen, A.-T., Denos, N., & Berrut, C. (2005). Cartes de communautés pour l'adaptation interactive de profils dans un système de filtrage d'information. Actes du 23ème Congrès annuel de l'Informatique des Organisations và Systèmes d'Information và de Décision (INFORSID'05), France (pp. 253-268).

[16] Nguyen, A.-T., Denos, N., & Berrut, C. (2007). Improving New User Recommendations with Rule-based Induction on Cold User Data. In Proceedings of the ACM Conference on Recommender Systems 2007 (RecSys 2007), Minnesota, USA (to appear).

[17] Pawlak, Z. (1982). Rough Sets. International Journal of Computer and Information Sciences, 11 (5), 341-356. Plenum Publishing Corporation.

[18] Pawlak, Z. (2004). Some Issues on Rough Sets. Transaction on Rough Sets I, LNCS 3100.

[19] Pintwart, N., Aleven, V., Ashley, K., & Lynch, C. (2006). Using Collaborative Filtering in an Intelligent Tutoring System for Legal Argumentation. In Proceedings of Workshops held at the 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Ireland (pp. 542-551).

[20] Rashid, A., Albert, I., Cosley, D., Lam, S.-K., McNee, S.-M., Konstan, J.-A., & Riedl, J. (2002). Getting to Know You: Learning New User Preferences in Recommender Systems. In Proceedings of the 7th International Conference on Intelligent User Interfaces (IUI'02), California, USA (pp. 127-134).

[21] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In Proceedings of the Conference on Computer Supported Cooperative Work (CSCW'94), NC, USA.

[22] Skowron, A., & Rauszer, C. (1992). The Discernibility Matrices and Functions in Information Systems. Intelligent Decision Support: Handbook of Applications and Advances of Rough Set Theory, Series: Theory and Decision Library, 11, 331-362. Kluwer "Công chức" Publishers, Dordrecht.

[23] Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. Cambridge Handbook of the Learning Sciences. Cambridge, UK: Cambridge University Press.