

Nghiên cứu và Ứng dụng Lọc Cộng tác trong Khai thác Mô hình Người dùng

Đỗ Thị Minh Phụng

Trường Đại học Công nghệ Thông tin – ĐHQG Tp.HCM
phungdtm@uit.edu.vn

Tóm tắt. Báo cáo này tổng quan về hướng nghiên cứu của đề tài, kết quả đạt được và dự kiến kế hoạch nghiên cứu trong thời gian tới. Mô hình hóa người dùng là xu hướng mới trong việc tăng cường khả năng thích ứng của các hệ thích nghi (adaptive system) và hệ khuyến nghị (recommender system). Bản chất của các hệ thống mô hình hóa là khai thác mô hình người dùng (user model) hay hồ sơ người dùng (user profile) để khám phá các mẫu (pattern) hay thông tin tiềm ẩn về người dùng. Lọc cộng tác (Collaborative Filtering - CF) là một trong những kỹ thuật quan trọng của khai thác mô hình người dùng và là cách tiếp cận thành công nhất để xây dựng các hệ khuyến nghị, khai thác mạng xã hội để phát hiện sở thích người dùng. Nghiên cứu tập trung vào kỹ thuật khai thác mô hình người dùng bằng CF. Tính mới của nghiên cứu là đề xuất một giải thuật tìm tập phổ biến (mẫu mua hàng) qua khai thác nhị phân (bit mining). Giải thuật này tăng tốc độ và độ hữu ích của CF khi giảm thời gian xử lý và giới thiệu cho người dùng những mặt hàng mà họ yêu thích. Đồng thời, nghiên cứu đề xuất một kiến trúc trừu tượng cho việc thi công và đánh giá các thuật toán CF, làm nền tảng cho những nhà nghiên cứu khi phát triển CF.

1 Giới thiệu

1.1 Tổng quan Vấn đề Nghiên cứu

Mô hình người dùng [25], [37] là hạt nhân của các hệ thích nghi (adaptive system) và hệ khuyến nghị (recommender system). Các nghiên cứu cơ sở bao gồm các nghiên cứu về các phương pháp xây dựng mô hình người dùng, các hệ thống thi công việc mô hình hóa người dùng, cụ thể như sau:

- Phương pháp máy học [19]: Dựa trên giao tiếp người và máy. Phân tích các dữ liệu từ sự tương tác của người dùng và hệ thống để suy luận mô hình người dùng, với nhiều chủ đề liên quan như: mạng neuron, cây quyết định, giải thuật di truyền, mạng Bayesian ... làm cơ sở cho một số phương pháp khác.
- Phương pháp Bayesian [8], [13], [23], [24], [34], [35]: Dựa trên suy luận thống kê Bayesian trong đó các quan sát hay bằng chứng học tập sẽ được dùng để suy luận xác suất về sự hiểu biết của người dùng về một phần tử tri thức nào đó.

- Phương pháp chồng (overlay) [16, 26, 34]: Mô hình người dùng là tập con của mô hình tri thức chuyên gia, trong đó mỗi phần tử trong mô hình này sẽ chứa những dữ liệu về người dùng.
- Phương pháp khuôn mẫu (stereotype) [31]: Người dùng sẽ được gom nhóm và phân lớp theo từng stereotype, mỗi stereotype là tập những đặc trưng thường xuất hiện của người dùng.
- Phương pháp dựa kế hoạch [8]: Kế hoạch là chuỗi những hoạt động nhằm đạt được mục tiêu, hệ thống theo dõi những hoạt động này để phân loại người dùng.

Các hệ thống cụ thể thi công việc mô hình hóa người dùng: Shell mô hình hóa (User Modeling Shell) [28], trình chủ mô hình hóa (User Modeling Server) [2], [9], [15], hệ mô hình hóa người dùng dựa trên tác tử (Agent-based User Model) [7], hệ mô hình hóa người dùng di động (Mobile User Model) [4], [5]. Hệ thống mô hình hóa người dùng GUMS [30] được xem là hệ độc lập ứng dụng đầu tiên. Các shell mô hình hóa đã đưa ra một nền tảng cho những trình chủ mô hình hóa phổ biến ngày nay. Những kỹ thuật mô hình hóa được áp dụng trong trình chủ cũng giống như trong shell, ví dụ, shell um được trình chủ Personis sử dụng, Personis nhúng um vào trong cấu trúc trình chủ kết hợp với lưu trữ cơ sở dữ liệu.

Bản chất của các hệ mô hình hóa là khai thác tri thức từ mô hình người dùng để phục vụ cho các hệ thích nghi và hệ khuyến nghị. Có nhiều phương pháp khai thác mô hình người dùng như: khai thác mẫu tuần tự (sequential pattern) [27], [36], phân lớp tài liệu [3], xây dựng nhóm người dùng [3] và lọc cộng tác (Collaborative Filtering - CF) [1]. CF là xu hướng của việc khai thác mô hình người dùng nhằm phục vụ cho các hệ khuyến nghị. Giả định cơ bản của CF là nếu người dùng X và Y đánh giá n mặt hàng tương tự nhau, hoặc có những hành động tương tự (ví dụ, mua, xem, nghe) sẽ đánh giá hoặc có hành động trên các sản phẩm khác cũng tương tự như vậy [20].

Kỹ thuật CF sử dụng dữ liệu các mặt hàng ưa thích của nhiều người dùng trước đó để khuyến nghị cho người dùng mới những mặt hàng mà họ có thể thích. Trong kỹ thuật CF, dữ liệu đánh giá là một ma trận gồm một danh sách m người dùng $\{u_1, u_2, \dots, u_m\}$ và một danh sách n mặt hàng $\{i_1, i_2, \dots, i_n\}$ và đánh giá của mỗi người dùng u trên mặt hàng i được đặc tả bằng thang giá trị, thông thường từ 1 đến 5. Một số giá trị bị thiếu trong ma trận do người dùng không cung cấp sở thích của họ trên một mặt hàng cụ thể.

Bảng 1. Ma trận đánh giá

| | i_1 | i_2 | i_3 |
|-------|-------|-------|-------|
| u_1 | 1 | 2 | 1 |
| u_2 | 2 | 1 | 2 |
| u_3 | 4 | 1 | 5 |
| u_4 | 1 | 2 | ? |

Có nhiều vấn đề trong lọc cộng tác như dữ liệu thừa, số lượng người dùng và mặt hàng ngày càng tăng, đáp ứng tức thời yêu cầu tư vấn và một số vấn đề khác như tính đồng nghĩa (các mặt hàng tương tự hay giống nhau nhưng có tên khác nhau), dữ liệu

bị nhiễu và các vấn đề bảo vệ sự riêng tư, sự hữu ích của các mặt hàng khuyến nghị. Nghiên cứu tập trung giải quyết một số vấn đề trong lọc cộng tác như tốc độ và tính hữu ích.

1.2 Nghiên cứu Liên quan

Các hệ thống lọc cộng tác thể hệ đầu như GroupLens [29] dùng dữ liệu người dùng đánh giá để tính độ tương tự giữa các người dùng hoặc giữa các mặt hàng và đưa ra khuyến nghị theo các giá trị độ tương tự đã được tính toán. Các phương pháp lọc cộng tác dựa trên bộ nhớ (memory-based *CF*) được phát triển trong các hệ thương mại như Amazon, Barnes và Noble vì các phương pháp này dễ thực hiện và có hiệu quả cao [1], [11], [18].

Tuy nhiên, các kỹ thuật memory-based *CF* có một vài hạn chế như độ tương tự giữa các mặt hàng không đáng tin cậy trong trường hợp dữ liệu thưa. Để đạt được hiệu suất dự đoán tốt hơn và khắc phục hạn chế của các thuật toán memory-based *CF*, cách tiếp cận lọc cộng tác dựa trên mô hình (model-based *CF*) được đưa ra. Cách tiếp cận này dùng dữ liệu đánh giá “thô” để ước lượng hoặc xây dựng mô hình cho việc dự đoán. Mô hình có thể dạng khai thác dữ liệu hoặc máy học. Các kỹ thuật model-based *CF* nổi tiếng như mô hình BNs (Bayesian belief nets) *CF* [22], [33], mô hình cụm *CF* (clustering *CF* model) [3], mô hình ngữ nghĩa tiềm ẩn *CF* (latent semantic *CF* model) [10], mô hình MDP *CF* (Markov decision process-based *CF*) [6] và kỹ thuật giảm chiều ma trận (Singular Value Decomposition - SVD, Principle Component Analysis - PCA) [20].

Bên cạnh lọc cộng tác, lọc nội dung (Content-based Filtering - CBF) là một hướng tiếp cận khác của hệ khuyến nghị. *CBF* thực hiện khuyến nghị bằng cách phân tích nội dung của thông tin văn bản và tìm các quy tắc trong nội dung. Điểm khác chính của *CF* và *CBF* là *CF* chỉ dùng dữ liệu đánh giá người dùng-mặt hàng để thực hiện dự đoán và khuyến nghị, trong khi *CBF* thực hiện dự đoán dựa trên các đặc trưng của người dùng và các mặt hàng [12]. Cả hai hệ khuyến nghị *CBF* và *CF* đều có những hạn chế. Trong khi các hệ *CF* không khai thác thông tin đặc trưng của người dùng hay mặt hàng, các hệ *CBF* không tận dụng những tương tự về sở thích tiềm ẩn trong cộng đồng [21].

Kỹ thuật lai *CF* (Hybrid *CF*) như thuật toán content-boosted *CF* [17] và Personality Diagnosis [14], kết hợp kỹ thuật *CF* và kỹ thuật *CBF* với hy vọng tránh được những hạn chế của các cách tiếp cận trên và do đó kỹ thuật này cải thiện được hiệu suất tư vấn.

CF là kỹ thuật mạnh trong khai thác mô hình người dùng khi nó có khả năng phát hiện những thông tin ẩn trong mạng xã hội, đồng thời không đòi hỏi nhiều thông tin về người dùng như các hệ mô hình hóa hiện có. Tuy nhiên việc cài đặt chúng gặp bốn vấn đề sau, cần được giải quyết:

1. Quá nhiều giải thuật *CF* và không có chuẩn chung để xây dựng chúng. Điều này gây trở ngại cho những nhà nghiên cứu.
2. Các độ đo đánh giá *CF* cần được chuẩn hóa.
3. Tốc độ và thời gian đáp ứng cần được cải thiện.

4. Một số giải thuật *CF* tuy có độ chính xác cao khi ước lượng giá trị đánh giá nhưng các mặt hàng chúng giới thiệu chưa thật sự hữu ích cho người dùng.

Nghiên cứu tập trung vào khai thác mô hình người dùng bằng lọc cộng tác và giải quyết các vấn đề trên.

Bảng 2. Tổng quan các kỹ thuật *CF*

| Loại <i>CF</i> | Kỹ thuật thi công | Ưu điểm chính | Hạn chế chính |
|-------------------------|--|--|--|
| Memory-based <i>CF</i> | <ul style="list-style-type: none"> * <i>CF</i> dựa trên láng giềng (giải thuật <i>CF</i> dựa trên mặt hàng /dựa trên người dùng với độ đo Pearson, cosine) * dựa trên mặt hàng /dựa trên người dùng khuyến nghị N – xếp hạng cao nhất | <ul style="list-style-type: none"> * dễ thi công * dễ dàng thêm dữ liệu mới * không cần quan tâm nội dung mặt hàng * phối hợp tốt với những mặt hàng đồng đánh giá | <ul style="list-style-type: none"> * phụ thuộc đánh giá người dùng * hiệu suất giảm khi dữ liệu thưa * không thể khuyến nghị cho người dùng những mặt hàng mới * bị giới hạn bởi tập dữ liệu lớn |
| Model-based <i>CF</i> | <ul style="list-style-type: none"> * mạng Bayesian * gom cụm * tiến trình ra quyết định Markov * ngữ nghĩa tiềm ẩn * phân tích dữ liệu thưa * kỹ thuật giảm chiều ma trận (SVD, PCA) | <ul style="list-style-type: none"> * giải quyết tốt vấn đề dữ liệu thưa và dữ liệu lớn * cải thiện hiệu suất dự đoán * theo sát cảm nhận trực quan của người dùng về mặt hàng | <ul style="list-style-type: none"> * chi phí xây dựng mô hình lớn * ước lượng thiếu chính xác * độ hữu ích của mặt hàng khuyến nghị chưa cao * mất thông tin hữu ích khi giảm chiều ma trận |
| Lai (Hybrid <i>CF</i>) | <ul style="list-style-type: none"> * kết hợp <i>CF</i> và <i>CBF</i> (Fab) * content-boosted <i>CF</i> * kết hợp memory-based và model-based <i>CF</i> (Personality Diagnosis) | <ul style="list-style-type: none"> * khắc phục hạn chế của <i>CF</i> và <i>CBF</i> * cải thiện hiệu suất dự đoán * khắc phục vấn đề ma trận thưa và “cửu xác” trong <i>CF</i> | <ul style="list-style-type: none"> * độ phức tạp và chi phí tăng khi thi công * cần những thông tin bên ngoài và thường không sẵn có những thông tin này |

2 Mục đích Nghiên cứu

Nghiên cứu chính của đề tài là khai thác mô hình người dùng bằng lọc cộng tác với hai mục tiêu:

- Nghiên cứu và phát triển một kiến trúc trừu tượng cho lọc cộng tác. Kiến trúc này chuẩn hóa các giải thuật *CF* và làm nền tảng cho những nhà nghiên cứu khi phát triển giải thuật *CF* mới. Mục tiêu này giải quyết vấn đề 1 và 2.
- Đề xuất một giải thuật *CF* để khai thác hiệu quả mô hình người dùng. Giải thuật được thi công theo kiến trúc đề xuất. Giải thuật này tăng tốc độ và độ hữu ích của

lọc cộng tác khi giảm thời gian xử lý và giới thiệu những mặt hàng mà người dùng yêu thích. Mục tiêu này giải quyết vấn đề 3 và 4.

3 Nội dung Nghiên cứu

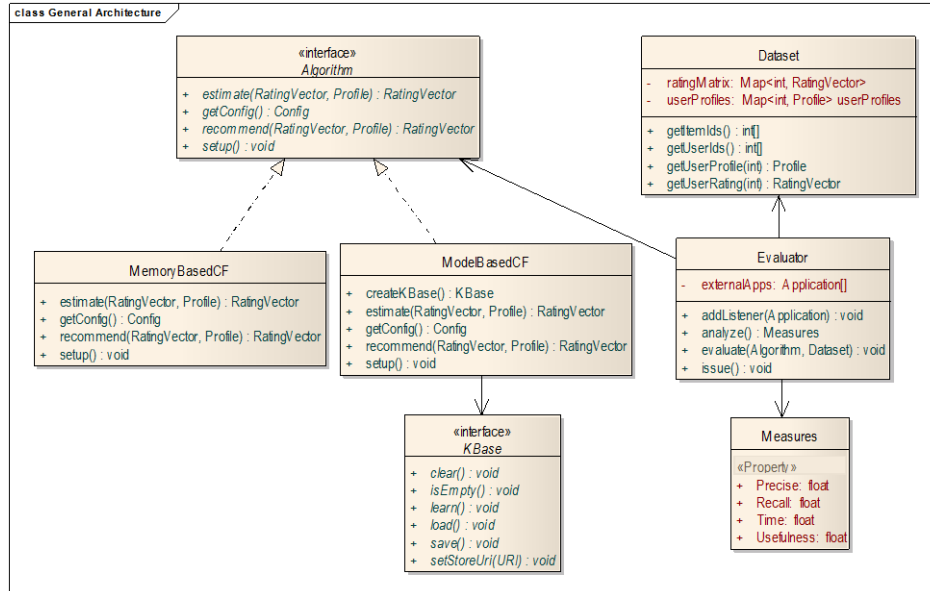
3.1 Đề xuất Kiến trúc Trừu tượng Để Thi công và Đánh giá Các Thuật toán CF

Như đã đề cập trong phần nghiên cứu liên quan (Phần 1.2), có rất nhiều giải thuật CF và mỗi giải thuật có những đặc trưng khác nhau và không có kiến trúc hoặc chuẩn chung để thi công chúng. Ngoài ra, mỗi nhà nghiên cứu phát triển thuật toán theo phương thức riêng của họ. Đây là những nguyên nhân của sự phức tạp khi xây dựng và đánh giá các thuật toán CF. Vì vậy đề tài đề xuất một kiến trúc trừu tượng cho việc thi công và đánh giá các thuật toán CF. Nền tảng (framework) là cơ sở hạ tầng hoặc phần mềm thi công kiến trúc trừu tượng này. Kiến trúc tổng thể được biểu diễn bằng ngôn ngữ UML, gồm 4 lớp (class) và giao diện (interface) cơ bản: *Algorithm*, *Kbase*, *Dataset* và *Evaluator*. Mỗi giao diện và lớp được xem như các chuẩn công nghệ phần mềm cho thuật toán CF. Nhà nghiên cứu sẽ tuân theo chuẩn này khi họ áp dụng framework để viết thuật toán mới.

- Giao diện *Algorithm* biểu diễn thuật toán trừu tượng. Nhiệm vụ chính của nhà nghiên cứu là thi công giao diện này theo mục tiêu thuật toán mới của họ. Trong hầu hết các trường hợp, họ kế thừa trực tiếp hai lớp *MemoryBasedCF* và *ModelBasedCF*, hai lớp này dẫn xuất trực tiếp từ *Algorithm*. *MemoryBasedCF* và *ModelBasedCF* lần lượt đại diện cho thuật toán memory-based CF và model-based CF.
- Giao diện *KBase* biểu diễn cơ sở tri thức kết hợp với giải thuật *ModelBasedCF*. Cấu trúc của *KBase* rất linh hoạt và nó phụ thuộc vào ý tưởng và mục đích của thuật toán.
- Lớp *Dataset* gồm ma trận đánh giá (rating matrix) và hồ sơ người dùng (user profile). Mỗi dòng của ma trận đánh giá được biểu diễn bởi lớp *RatingVector*. Hồ sơ người dùng được biểu diễn bởi lớp *Profile*. Framework quản trị *Dataset*.
- Lớp *Evaluator* để đánh giá thuật toán theo bốn độ đo (*Measures*): thời gian (Time - T), độ chính xác (Precise - P), độ bao phủ (Recall - R) và độ hữu ích (Usefulness - U). Các độ đo này được định nghĩa bên trong *Evaluator*. Hai lớp dẫn xuất từ *Evaluator* là *EstimateEvaluator* và *RecommendEvaluator*. *Evaluator* đọc và cung cấp *dataset* cho *Algorithm*. Cuối cùng, nó đánh giá thuật toán bằng cách tính toán *Measures* dựa trên kết quả thi hành thuật toán.

Khi các nhà nghiên cứu phát minh ra một thuật toán mới, họ cần thực hiện hai việc sau:

- Viết lớp thuật toán dẫn xuất trực tiếp từ *MemoryBasedCF* hoặc *ModelBasedCF*. Trong trường hợp kế thừa từ *ModelBasedCF*, họ phải thi công *KBase* tương ứng với mục tiêu thuật toán.
- Nén lớp thuật toán, lớp *KBase* cùng với tất cả các file liên quan vào trong một file nén được gọi là *plug-in* và chếp *plug-in* này vào trong framework.



Hình 1. Kiến trúc tổng quan của framework

3.2 Đề xuất Giải thuật mới Thi công Framework

Các thuật toán *CF* tập trung vào sự chính xác và yếu tố đáp ứng thời gian thực (real-time) khi giới thiệu một món hàng cho người mua. Như đề cập trong phần nghiên cứu liên quan (Phần 1.2), có ba hướng tiếp cận khi xây dựng các thuật toán cho các hệ thống tư vấn:

- Memory-based *CF*: Nạp toàn bộ dữ liệu đánh giá vào bộ nhớ rồi tính xu hướng, sở thích người mua dựa vào các phép toán ma trận.
- Model-based *CF*: Ý tưởng chính của phương pháp là xây dựng một mô hình suy diễn từ dữ liệu đánh giá và từ mô hình này sẽ tìm ra xu hướng, sở thích người mua để giới thiệu những món hàng phù hợp.
- Hybrid *CF*: Kết hợp hai phương pháp trên.

Mặc dù các thuật toán model-based có nhiều ưu điểm nhưng chúng gặp phải vấn đề về tính hữu ích của tiến trình khuyến nghị “các mặt hàng giới thiệu cho người dùng có thể không phải là mặt hàng mà họ yêu thích”. Vì vậy mục tiêu thứ hai của đề tài là đề xuất một giải thuật theo hướng tiếp cận model-based *CF* nhằm tăng độ hữu ích của tiến trình khuyến nghị. Giải thuật này tìm tập phổ biến (mẫu mua hàng) qua khai thác nhị phân (bit mining) và giới thiệu những mẫu được yêu thích này cho người dùng. Ngoài tăng độ hữu ích, giải thuật còn tăng tốc độ khi giảm thời gian xử lý.

Giải thuật đề xuất gồm hai bước:

- Bước 1: Dùng kỹ thuật khai thác để tìm tập phổ biến cực đại là các món hàng mà người mua thường chọn nhất.
- Bước 2: Dựa trên tập phổ biến cực đại, bất cứ khi nào người mua cần, hệ thống sẽ giới thiệu những món hàng trong tập phổ biến cực đại sao cho phù hợp với họ nhất.

Bước 1 thực hiện ngoại tuyến (off-line) nên không ảnh hưởng tốc độ, bước 2 thực hiện trực tuyến đáp ứng thời gian thực yêu cầu người dùng.

Ở bước 1, chúng tôi đề xuất một giải thuật mới khác dựa trên giả định “*các mặt hàng có tần suất mua càng cao thì càng có khả năng xuất hiện trong tập phổ biến cực đại*”.

Giải thuật mới được thi công theo chuẩn mà framework đề xuất (Phần 3.1). Nói cách khác các khái niệm *Algorithm*, *Kbase*, *Dataset*, *Evaluator* sẽ được hiện thực hóa và áp dụng cho giải thuật mới này.

4 Phương pháp Nghiên cứu

Cơ sở lý thuyết: khai thác dữ liệu (data mining) và cách tiếp cận công nghệ phần mềm (ngôn ngữ mô hình hóa UML).

Kỹ thuật thi công và thử nghiệm:

- Môi trường phát triển là nền tảng J2EE (Java-based & J2EE): Java hiện rất phổ biến, các công nghệ mở hỗ trợ Java rất nhiều, mặt khác tính đa hệ của Java cùng với khuynh hướng mở là lý do chính khiến hệ thống được thi công bằng Java.
- Dữ liệu thử nghiệm cho thuật giải *CF* đề nghị là MovieLens [32]. MovieLens được phát triển bởi GroupLens. MovieLens là bộ dữ liệu chuẩn cho các thuật toán tư vấn, được sử dụng rộng rãi nhất. Thông qua giao diện website, GroupLens đóng gói những đánh giá của người dùng đối với các bộ phim thành bộ dữ liệu MovieLens. Sử dụng bốn độ đo để đánh giá giải thuật: thời gian, độ chính xác, độ bao phủ và độ hữu ích.
- Hệ hỗ trợ phát triển và đánh giá các thuật toán *CF*, hệ này thi công kiến trúc trừu tượng.

5 Dự kiến Kế hoạch Nghiên cứu

- Xây dựng kiến trúc trừu tượng cho giải thuật *CF*.
- Nghiên cứu và cài đặt giải thuật *CF* mới dựa trên kiến trúc trừu tượng đề xuất.

Tài liệu tham khảo

1. Xiaoyuan Su, Taghi M. Khoshgoftaar: A Survey of Collaborative Filtering Techniques. In: Hindawi Publishing Corporation, *Advances in Artificial Intelligence*, Volume 2009, Article ID 421425, 19 pages, doi:10.1155/2009/421425
2. Alfred Kobsa, Josef Fink: An LDAP-Based User Modeling Server and its Evaluation. *User Modeling and User-Adapted Interaction 2006 (UMUAI-2006)*
3. Jiawei Han, Micheline Kamber: *Data Mining: Concepts and Techniques*. Second Edition. © 2006 by Elsevier Inc.
4. Alfred Kobsa: *Generic User Modeling Systems*. *User Modeling and User-Adapted Interaction 2006 (UMUAI-2006)*.
5. Dominikus Heckmann: *Ubiquitous User Modeling*. PhD Thesis at Universität des Saarlandes. Volume 297 *Dissertationen zur Künstlichen Intelligenz* (2005)
6. G. Shani, D. Heckerman, R. I. Brafman: An MDP-based Recommender System. *Journal of Machine Learning Research*, vol. 6, pp. 1265-1295 (2005)
7. Gustavo González, Cecilio Angulo, Beatriz López, Josep Lluís de la Rosa: Smart User Models: Modelling the Humans in Ambient Recommender Systems. In: *Proceedings of the Workshop on Decentralized, Agent Based and Social Approaches to User Modelling (DASUM 2005)*, pp. 11–20 (2005)
8. Xiangyang Li, Qiang Ji: Active Affective State Detection and Assistance with Dynamic Bayesian Networks. In: *IEEE Transactions on Systems, Man, and Cybernetics: Special Issue on Ambient Intelligence*, vol. 35, no. 1, pp. 93–105 (2005)
9. Peter Brusilovsky: KnowledgeTree: A Distributed Architecture for Adaptive E-Learning. In: *Thirteen International World Wide Web Conference, WWW 2004 (Alternate track papers and posters)*, New York, NY, pp. 104-113, 10.1145/1013367.1013386 (2004)
10. Thomas Hofmann: Latent Semantic Models for Collaborative Filtering. In: *ACM Transactions on Information Systems*, Vol.22, No.1, pp. 89-115 (January 2004)
11. G. Linden, B. Smith, and J. York: Amazon.com recommendations: item-to-item collaborative filtering. In: *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80 (2003)
12. L. Si and R. Jin: Flexible mixture model for collaborative filtering. In: *Proceedings of the 20th International Conference on Machine Learning (ICML '03)*, vol. 2, pp. 704–711, Washington, DC, US (August 2003)
13. Cristina Conati, Abigail Gertner and Kurt Vanlehn: Using Bayesian Networks to Manage Uncertainty in Student Modeling. In *Journal of User Modeling and User-Adapted Interaction*, Volume 12, Issue 4, pp. 371 - 417. ISSN:0924-1868(November 2002)
14. D. Y. Pavlov and D. M. Pennock: A maximum entropy approach to collaborative filtering in dynamic, sparse, highdimensional domains. In: *Advances in Neural Information Processing Systems*, pp. 1441–1448, MIT Press, Cambridge, Mass, USA (2002)
15. Judy Kay, Bob Kummerfeld, Piers Lauder: Personis: A server for user models. In: *Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2002)*, pp. 201-212 (2002)
16. Owen Conlan, Declan Dagger, and Vincent Wade: Towards a Standards-based Approach to e-Learning Personalization using Reusable Learning Objects. In: *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare and Higher Education (E-Learn 2002)*, pp. 210–217 (September 2002) http://www.cs.tcd.ie/~oconlan/publications/eLearn2002_v1.24_Conlan.pdf.
17. P. Melville, R. J. Mooney, R. Nagarajan: Content-boosted collaborative filtering for improved recommendations. In *Proceedings of the 18th National Conference on Artificial Intelligence (AAAI '02)*, pp. 187–192, Edmonton, Canada (2002)

18. B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl: Itembased collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on World Wide Web (WWW '01), pp. 285–295 (May 2001)
19. Geoffrey I. Webb, Michael J. Pazzani, Daniel Billsus: Machine Learning for User Modeling. In User Models User-Adapted Interaction, vol.11, no. 1-2, pp. 19–29 (2001)
20. K. Goldberg, T. Roeder, D. Gupta, and C. Perkins: Eigentaste: a constant time collaborative filtering algorithm. In Information Retrieval, vol. 4, no. 2, pp. 133–151 (2001)
21. A. Ansari, S. Essegiaier, and R. Kohli: Internet recommendation systems. In: Journal of Marketing Research, vol. 37, no. 3, pp. 363–375 (2000)
22. David Heckerman, David Maxwell Chickering, Christopher Meek, Robert Rounthwaite, Carl Kadie: Dependency Networks for Inference, Collaborative Filtering and Data Visualization. In Journal of Machine Learning Research 1, pp. 49-75 (2000)
23. Michael Mayo, Antonija Mitrovic: Using a Probabilistic Student Model to Control Problem Difficulty. In Gauthier G., Frasson C., and VanLehn K. (Eds.), Proc. of 5th International Conference on Intelligent Tutoring Systems, Springer-Verlag, pp. 524-533 (2000)
24. Nicola Henze: Adaptive Hyperbooks: Adaptation for Project-Based Learning Resources. PhD Thesis at University of Hannover (Supervisors: Prof. Dr. W. Nejdl, Prof. Dr. U. Lipeck) (2000)
25. Nora Koch: Software Engineering for Adaptive Hypermedia Systems. PhD thesis, Ludwig-Maximilians-University Munich/Germany (2000)
<http://www.pst.informatik.unimuenchen.de/personen/kochn/PhDThesisNoraKoch.pdf>.
26. Peter Brusilovsky: Methods and Techniques of Adaptive Hypermedia. In User Modeling and User-Adapted Interaction, vol. 6, no. 2–3, pp. 87–129 (1996)
<http://www2.sis.pitt.edu/~peterb/papers/UMUI96.pdf>.
27. Rakesh Agrawal, Ramakrishnan Srikant: Mining Sequential Patterns. In: Proceedings of the Eleventh International Conference on Data Engineering, pp. 3-14, March 06-10 (1995)
28. Brajnik, G., Tasso, C: A Shell for Developing Non-monotonic User Modeling Systems. In: International Journal of Human-Computer Studies 40, pp. 31-62, DOI: 10.1006/ijhc.1994.1003(1994)
29. P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl: Grouplens: an open architecture for collaborative filtering of netnews. In Proceedings of the ACM Conference on Computer Supported Cooperative Work, pp. 175–186, New York, NY, USA (1994)
30. Tim Finin, David Drager: GUMS: A General User Modeling System. In: Proceedings of the Canadian Society for Computational Studies of Intelligence (CSCSI-86) (1986)
31. Elaine Rich: User Modeling via Stereotypes. In: COGNITIVE SCIENCE 3, 329-354 (1979).
32. Movielens dataset 2011. Home page is <http://www.movielens.org>. Download dataset from <http://www.grouplens.org/node/12>.
33. Phung Do, Vu Thanh Nguyen, Tran Nam Dung, Loc Nguyen: Model-based approach for Collaborative Filtering. In: Proceeding of 6th International Conference on Information Technology for Education (IT@EDU2010), pp. 19-20 (August 2010), Ho Chi Minh city and Phan Thiet, VietNam, pp. 217-225
34. Loc Nguyen, Phung Do: Combination of Bayesian Network and Overlay Model in User Modeling. In: International Journal of Emerging Technologies in Learning (iJET), Vol. 4 No. 4. ISSN: 1863-0383. Archive: <http://online-journals.org/i-jet/article/view/684> (December 2009)
35. Loc Nguyen, Phung Do: Evolution of parameters in Bayesian Overlay Model. In: Proceedings of The 2009 International Conference on Artificial Intelligence (ICAI'09) 13-16, USA, pp. 324-329. ISBN: 1-60132-107-4, 1-60132-108-2 (1-60132-109-0) (July 2009)

36. Loc Nguyen, Phung Do: Learning Concept Recommendation based on Sequential Pattern Mining. In: Proceedings of The 2009 Third International Digital Ecosystems and Technologies Conference (IEEE-DEST 2009), Istanbul, Turkey, pp. 66-71. ISBN: 978-1-4244-2346-0 (31 May-3 June 2009)
37. Loc Nguyen, Phung Do: Learner Model in Adaptive Learning. In: Proceedings of World Academy of Science, Engineering and Technology, volume 35, November 2008, ISSN: 2070-3740. WCSET 2008: World Congress on Science, Engineering and Technology, Paris, France. Archive: <http://www.waset.org/pwaset/v35/v35-70.pdf> (November 21-23 2008)