A Personalized E-Learning System Based on User Profile Constructed Using Information Fusion

Xin Li and Shi-Kuo Chang
Department of Computer Science,
University of Pittsburgh, USA,
{flying, chang}@cs.pitt.edu

Abstract

In this paper, we describe a personalized e-learning system which can automatically adapt to the interests and levels of learners. The system is designed based on the IEEE Learning Technology Systems Architecture (IEEE LTSA) to achieve high scalability and reusability. A feedback extractor with fusion capability is proposed to combine multiple feedback measures to infer user preferences. User profile, which stores user preferences and levels of expertise, is collected by user profiler to deliver personalized information using the collaborative filtering algorithm.

1. Introduction

Comparing with the traditional face-to-face style teaching and learning, e-learning is indeed a revolutionary way to provide education in life long term. Nowadays more and more people have benefited from various e-learning programs. However, high diversity of the learners on the Internet poses new challenges to the traditional "one-size-fit-all" learning model, in which a single set of learning resource is provided to all learners. In fact, the learners could have various interests; even sharing with the common interests, they may have different levels of expertise, and hence they can not be treated in a uniform way. It is of great importance to provide a personalized system which can automatically adapt to the interests and levels of learners.

User profiling is a promising approach towards the personalized e-learning systems where user profile including interests, levels and learning patterns can be assessed during the learning process. Based upon the profile, personalized learning resource could be generated to match the individual preferences and levels. Furthermore, learners with the common interests and levels can be grouped, and feedbacks of one person can serve as the guideline for information delivery to the other members within the same group.

In fact, user profiling is also the key process of many other applications; for example, the recommendation systems [1, 4, 8, 9] mainly depend on user profiles in terms of similarity and differences to provide particular suggestions. The personalized web search engine [11] can construct user profiles from browsing history and consequently provide personalized results to match the information needs of individuals. Comparing with these applications, user profiling is more feasible and important in e-learning system because learning is a much more continuous process than other activities such as online news reading and web searching.

Most approaches of user profiling are heavily depending on the user feedbacks to construct user profiles. The feedback can be assessed explicitly by rating, or implicitly by the user behaviors such as print and save. In this paper, we are not advocating either of these two approaches since both of them have significant advantage and disadvantages [11]. Instead we propose a system which can combine multiple feedback measures to get more complete and accurate profiles using the information fusion techniques.

Our e-learning system is designed based upon the IEEE Learning Technology Systems Architecture (LTSA), where multiple means of information delivery are provided including a chatting room, a customized web browser and whiteboard. A *feedback extractor* with fusion capability is designed to combine multiple feedback measures such as reading time, the number of scroll, print/save and relational index on chatting history. *User profile*, which stores user preferences and levels of expertise, is collected by *user profiler* to deliver personalized information using the collaborative filtering algorithm [8].

The rest of the paper is organized as follows:

Section 2 is for the related research. The system architecture based on the IEEE LTSA is described in Section 3, and then main components of our system are discussed in detail: *learning resources* and *user profile* in Section 4, *feedback extractor* in Section 5 and *user profiler* in Section 6. The experimental system and result analysis are described in Section 7, followed by the discussion and future research in Section 8.

2. Related Research

As the related research, the personalized e-learning system and recommendation system are discussed respectively. Information fusion, which is a relatively new concept in this domain, is also briefly reviewed.

2.1. Personalized E-Learning System

Blochl et al. [2] proposed an adaptive learning system which can incorporate psychological aspects of learning process into the user profile to deliver individualized learning resource. The user profile is placed in multi-dimensional space with three stages of the semantic decisions: cognitive style, skills and user type. However, both the means to acquire user's feedback and the algorithms to update user profile have not been addressed in the presentation.

SPERO [10] is a personalized e-learning system based on the IEEE Learning Technology Systems Architecture (LTSA). It could provide different contents for the foreign language learners according their interests and levels. The problem of SPERO system is that it is largely using questionnaires and e-surveys to build user profiles, which costs the users too much extra work.

2.2. Recommendation Systems

User profiling is the key process of recommendation systems, which collect user feedback for items in a given domain and assess user profiles in terms of similarities and differences to determine what to recommend. Depending on underline technique, recommendation systems can be divided into collaborative filtering-based [8] content-based [4] and hybrid [1, 9] approaches. Classified by means to acquire feedback, they can be categorized as explicit rating [1, 8, 9], implicit rating [8] and no rating needed [4] systems.

In fact, user's feedbacks are so important that only very few content-based recommendation systems require neither explicit rating nor implicit rating. For example, Surflen [4] is a recommendation system using data mining techniques to assess the association rules on web pages through user’s browsing history without the feedbacks. However, it's hard to find user's exact interests just based on the browsing history, since it always happens that users open a page they don't like or just by mistake. This problem becomes even more severe in the situation that the system is sparsely used.

GroupLens system [8], which filters Usenet news, is a collaborative filtering system using *n-nearest neighbor-based* algorithm. In this algorithm, user profile is assessed based on a subset of appropriate *n* users similar to this user. The early version of GroupLens gathers user's feedback only by explicit rate. However, observing the extra costs of the explicit rating, in the latest version it also uses reading time as an implicit indicator.

Fab system [1] is also using the collaborative filtering model, meanwhile introducing the content analysis by a “topic” filtering. Web pages are initially ranked by the topic filter and then sent to user's personal filters. Users are required to give an explicit rate, and this feedback is used to modify both the personal filter and the original topic filter.

2.3. Information Fusion

The key commonality underlying applications which require information fusion is that they need retrieve information on the same object from multiple data sources[5]. For example, in our approach, the multiple indicators are available to assess user preference; it is a fusion problem to combine them to get more complete and accurate results.

Information fusion is intensively investigated in sensor-based data processing systems such as intelligent surveillance systems, robotics vision and medical diagnoses systems, where multiple levels of fusion process are formulated and many algorithms have been developed[5, 6]. The problem we are tackling in this paper can be categorized as a decision-level identity fusion: the goal is a joint combined declaration from individual indicators. The techniques used on this level include voting, Bayesian inference, Dempster Shafer's method, and so on.

3. System Architecture

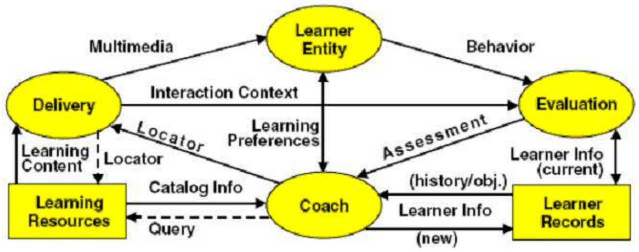

Figure 1 The IEEE Learning Technology Systems Architecture (IEEE LTSA)

The IEEE Learning Technology Systems Architecture (IEEE LTSA) [7] is a component-based framework for general learning system with high scalability and reusability. It includes three types of components shown in Figure 1:

- *Processes*: learner entity, evaluation, coach and delivery;
- *Stores*: learning records and learning resources;
- *Flows*: learning preference, behavior, assessment information, learner information, query catalog info, locator, learning content, multimedia and interaction context.

where the *processes* and *stores* exchange information through the *flows*.

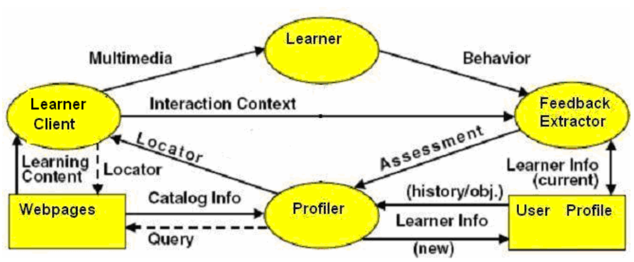

Figure 2 The Architecture of Our System

In our system, we instantiate the abstract conceptual models in IEEE LTSA by the real components shown in Figure 2:

- *Learning Resources* are represented as *Webpages* including multimedia resources such as video and audio clips.
- *Learner's Record* is implemented by *User Profile*, which stores performance, preference, etc.
- *Evaluation Entity* is implemented by *Feedback Extractor*, which can infer learner preference by fusing the multiple feedback measures.
- *Delivery Entity* is implemented by *Learner Client*.
- *Coach Entity* is implemented by *User Profiler* which interacts with other components:
  1. Receiving the preference assessment from *Feedback Extractor*.
  2. Assessing/Updating *User Profile*.
  3. Providing a guideline for information delivery for *Learner Client*.

The link between *Coach* and *Learner Entity* is removed since the feedbacks are collected implicitly.

4. Learning Resources and User Profile

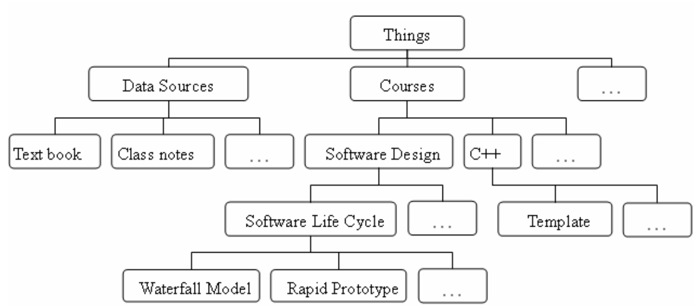

Figure 3 An Example of Ontology Knowledge Base

Learning Resources (e.g. webpages) are organized by the topics which are structured in ontology knowledge base (OKB). An example of OKB is shown in 3. The each topic is attached with several keywords.

Definition 4.1 a webpage *pg* is a 4-tuple:

$$pg = <id, co, tp, l>$$

where

- *id* is a unique identification number.
- *co* is the content of *pg*.
- *tp* is the topic.
- *l* is the expertise level of *pg* such as beginning, intermediate and advanced.

The user profile is defined as follows:

Definition 4.2 A user profile *upf* is a 4-tuple:

$$upf = <id, bh, ch, ls>$$

where

- *id* is a unique identification number.
- *bh* is the browsing history represented as $\{<pg_1,r_1>, <pg_2,r_2>, \ldots, <pg_n, r_n>\}$, where pg_i is a webpage read by the user; r_i is user preference on pg_i in [0..1] (i=1, 2, …, n).
- *ch* is the chatting history in natural languages.
- *ls* is the levels of expertise represented as $\{<tp_1,l_1>,$

<tp_2,l_2>, …, <tp_n, l_n>}, where tp_i is a topics; l_i is the levels on tp_i in terms of beginning, intermediate and advanced (i=1, 2, …, n).

5. Feedback Extractor

Feedback Extractor collects feedbacks to make a final assessment of user preference. In this section, firstly the feedback measures are described. Next, a fusion model is proposed to fuse these measures.

5.1. Feedback Indicators

Definition 5.1 Feedback indicator for a webpage *pg* is a function which returns 0 or 1, where 0/1 means the negative/positive correlation with user preference.

Four implicit feedback indicators are employed in our system including: reading time, scroll, print/save and relational index.

- Reading Time: return 1 if user read *pg* longer than φ_t, where φ_t is a predefined threshold; 0 otherwise.
- Scroll: return 1 if the number of user scrolls (either mouse or keyboard Pageup/Pagedown) on *pg* is greater than φ_s, where φ_s is a predefined threshold; 0 otherwise.
- Print/Save: return 1 if user prints/saves *pg*; 0 otherwise.
- Relational Index: return 1 if keywords of *pg* appear in user's chatting history *ch* more than φ_r times, where φ_r is a predefined threshold; 0 otherwise.

Although these indicators could be pretty much context-related, we can make them objective measures by normalizing the learning resources. For example, the webpages are trimmed to have roughly the same length and layout, so reading time could effectively indicates the interest of users.

5.2. Fusion Model

To compile all these feedbacks for a final assessment of user preference, we map the problem into an information fusion process. Define ***H*** is a hypothesis that user has positive preference, given independent indicators I_1, I_2, ..., I_n (n>1), the posteriori probability of P(**H**| I_1, I_2, ..., I_n) is the joint declaration, which can be assessed using Bayesian method:

$$P(H \mid I_1, I_2, ..., I_n) = \frac{P(I_1, I_2, ..., I_n \mid H)\cdot P(H)}{P(I_1, I_2, ..., I_n)}$$

$$= \frac{\prod_{i=1}^{n} P(I_i \mid H)\cdot P(H)}{\sum_{i=1}^{n}(P(I_i \mid H)\cdot P(H) + P(I_i \mid \neg H)P(\neg H))}$$

where

- $P(\neg H) = 1 - P(H)$
- $P(I_i \mid H)$ and $P(I_i \mid \neg H)$ are probability of the observing indicator I_i given *H*.
- $P(H)$ is the priori probability of the hypothesis *H* (without having observed the evidence)

$P(I_i \mid H)$, $P(I_i \mid \neg H)$ and $P(H)$ are called *model parameters* which can be assessed through the statistical analysis on training data.

6. User Profiler

User Profiler is a core component in our system. In this section, we firstly present a brief overview on it in terms of the input and output. Next, the two main functions are discussed in detail.

6.1. Overview

Briefly *User Profiler* has two tasks:

1. Assessment of Expertise Level
   - Input: user's browsing history with the preference assessed by *Feedback Extractor*.
   - Output: user's levels of expertise.
2. Providing Guideline for Delivery
   - Input: user's browsing history and levels of expertise.
   - Output: a list of webpages, which are potentially interesting to users.

6.2. Assessment of Expertise Level

Basically expertise levels are determined by the average preferences. The webpages user has read on any topic *tp* could have different levels in terms of beginning, intermediate and advanced. The level of the user on *tp* is the one which has the highest average preference.

6.3. Guideline for Delivery

The information delivery is based on the collaborative filtering algorithm [8]. Given two users U_1 and U_2, pg_1, pg_2, …, pg_n are the common pages they both read, with the feedback $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ respectively. Assume the average feedbacks of the page $pg_1, pg_2, \ldots, pg_n$ are $\varpi_1, \varpi_2, \ldots, \varpi_n$, a similarity function S on U_1 and U_2 is defined using Pearson correlation coefficient:

$$S(U_1, U_2) = \frac{\sum_{i=1}^{n} (x_i - \varpi_i) \times (y_i - \varpi_i)}{\sqrt{\sum_{i=1}^{n} (x_i - \varpi_i)^2 \times \sum_{i=1}^{n} (y_i - \varpi_i)^2}} \qquad (1)$$

Given any active user U_x, using (1) could find the n users who have the highest similarity, named n neighbors $\{U_1, U_2, \ldots, U_n\}$ of U_x, the preference of U_x on page pg — p_x can be predicted by the preferences of the neighbors which is already known, denoted as $p_1, p_2, \ldots, p_n$. Given ϖ is the average rating on page pg,

$$p_x = \varpi + \frac{\sum_{i=1}^{n} p_i \times S(U_x, U_i)}{\sum_{i=1}^{n} S(U_x, U_i)} \qquad (2)$$

The webpages with the highest interest predictions are the potential interesting pages for U_x, which will be delivered without requests.

7. Prototype System and Experiments

In this section, we briefly introduce our prototype system, followed by the experiments and some preliminary results.

7.1. Prototype System

A prototype system has been implemented. The system diagram is shown in Figure 4. Figure 5 shows the main interface of learner client. Several means are provided for the information delivery including a chatting room (right bottom), a customized web browser (left) and whiteboard (right top). With the help of communication server, multiple feedback measures are recorded for feedback extractor; user profiler can push the web pages without requests.

7.2. Experiments

For system training purpose, we ask a group of students to do the following experiments:

- Step 1: select a topic such as "E-R diagram" and "C++", let the students indicate their levels on it in terms of beginning, intermediate, or advanced.
- Step 2: ask the students to use leaner client (Figure 5) reading the prepared webpages.
- Step 3: require the students seriously rating the interest on every article they have read from 1 (the least) and 5 (the most).

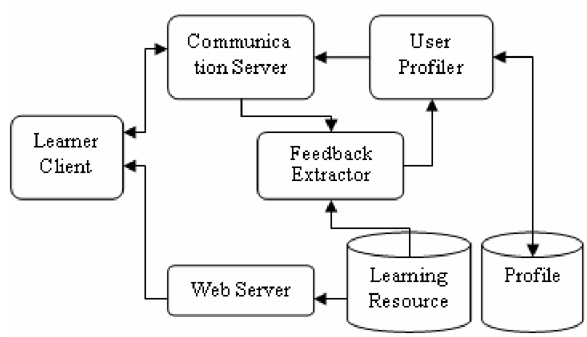

Figure 4 System Diagram

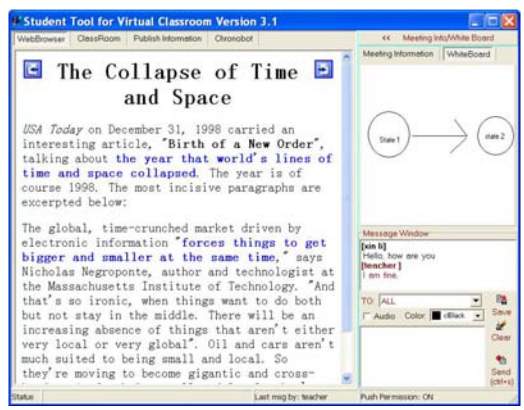

Figure 5 Learner Client

7.3. Preliminary Results and Analysis

Some preliminary results of the experiments have been collected. Figure 6 shows three indicators with explicit rating (the relational index is not employed in the experiments). The thresholds of the reading time φ_t and scroll φ_s are simply determined by the average value of medium evaluation (rating 3). The *model parameters* of the fusion model discussed in Section 5.2 can be assessed based upon these thresholds.

Compared with the provided levels by users, the level assessment algorithm described in Section 6.2 has 83.2% accuracy. However, the evaluation means of information delivery are still under development.

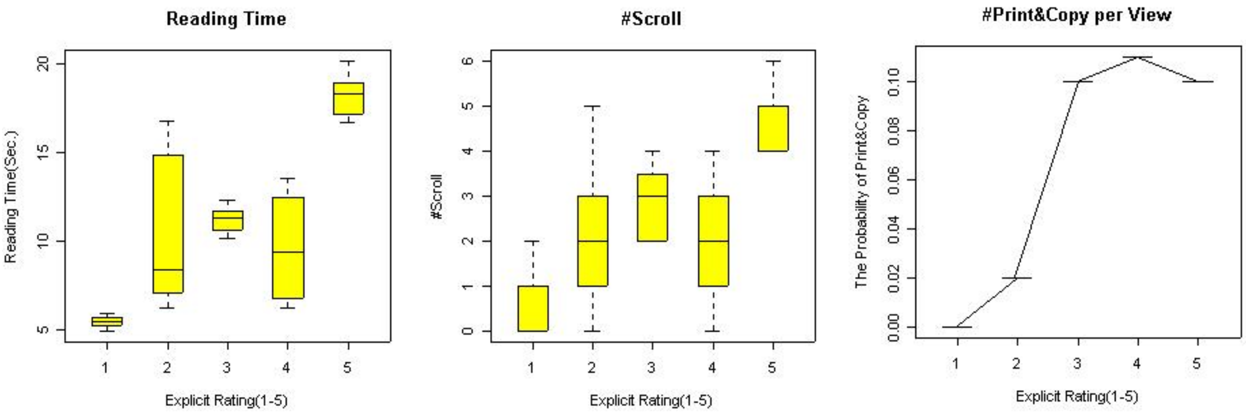

Figure 6 Feedback Measures vs. Explicit Rating

8. Discussion and Future Research

In this paper, we described our current ongoing research on the personalized e-learning system: the system is designed based upon the IEEE LTSA architecture; a *feedback extractor* with fusion capability is introduced to combine multiple feedback measures; user profile is collected by *user profiler* to deliver personalized information; the prototype system and preliminary results are presented.

However, the fourth indicator – relational index is not testified in the experiments yet. Furthermore, the usability of the system has not been fully verified by the end users, especially for the quality of information delivery. On the other hand, although ontology knowledge is used for the content classification, the structure of it could be much more complicated and the usage of it can be extended to the feedback extracting and user profiling. All these could lead to some very interesting topics and will be the subjects of our future research.

9. Acknowledgement

This research is a part of Chronobot project [3], which is supported in part by the Industry Technology Research Institute (ITRI) and the Institute for Information Industry (III) of Taiwan. We would like to thank Chieh-Chih Chang and Jui-Hsin Huang for the valuable discussion and comments.

References

[1] M. Balabanovi and Y. Shoham. Fab: content-based, collaborative recommendation. *Communication of the ACM*, 40(3):66–72, 1997.

[2] M. Blochl, H. Rumershofer, and W. Wob. Individualized e-learning systems enabled by a semantically determined adaptation of learning fragments. In *Proceeding of the 14th international workshop on Database and Expert Systems Applications*, pages 640–645, 2003.

[3] S. K. Chang. A chronobot for time and knowlege exchange. In *Tenth International Conference on Distributed Multimedia Systems*, pages 3–6, San Francisco Bay, CA, Sep. 2004.

[4] X. Fu, J. Budzik, and K. J. Hammond. Mining navigation history for recommendation. In *Proceedings of the 5th international conference on Intelligent user interfaces*, pages 106–112, 2000.

[5] D. L. Hall and J. Llinas, editors. *Handbook of Multisensor Data Fusion*. CRC Press, New York, 2001.

[6] D. L. Hall and S. A. H. McMullen, editors. *Mathematical Techniques in Multisensor Data Fusion*. Artech House Publishers, 2004.

[7] IEEE LTSC P1484.1. http://ltsc.ieee.org/wg1/index.html.

[8] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. Grouplens: applying collaborative filtering to usenet news. *Communication of the ACM*, 40(3):77–87, 1997.

[9] S. E. Middleton, N. R. Shadbolt, and D. C. D. Roure. Ontological user profiling in recommender systems. *ACM Trans. Inf. Syst.*, 22(1):54–88, 2004.

[10] SPERO: Tele-Informatics System for Continuous Collection, Processing, Diffusion of Material for Teacher Training in Special Education. http://www.image.ntua.gr/spero.

[11] K. Sugiyama, K. Hatano, and M. Yoshikawa. Adaptive web search based on user profile constructed without any effort from users. In *Proceedings of the 13th international conference on World Wide Web*, pages 675–684, 2004.