Evaluating Adaptive Learning Model

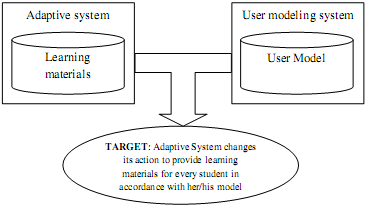
*Abstract*—Distance learning or e-learning is a trend of modern education, which brings new chance of study to everyone. Thus, everyone can study at anywhere and anytime so that they can improve and update their knowledge in lifelong time. Adaptive learning is a research branch of e-learning, which give adaptation and personalization to users in learning context. Different people receive different learning materials / teaching methods in accordance with their individual information / characteristics such as knowledge, goal, experience, interest, background, etc. Such individual information is structured in a format so-called user model. User model is the heart of adaptive learning system. User model is managed by user modeling system. There are many theories and practical methods to build up user model and adaptive learning system and each method has particular aspects but it is very difficult to determine which method or system is good because there is no evaluation standard and each method has particular strong point and drawbacks. Therefore, the goal of this research is to propose criterions to evaluate adaptive learning system and user modeling system. Moreover, research gives an evaluation scenario considered as an example for applying proposed criterion into evaluating adaptive learning system and user modeling in learning context.

Keywords— adaptive learning, user modeling, evaluation

# Introduction

Adaptive learning system is defined as the system that has ability to change its actions to provide learning content and pedagogic environment/method for every student in accordance with her/his individual information/characteristics [2] such as knowledge, goal, experience, interest, background, etc when these characteristics vary from person to person. The description of learners’ individual information/characteristics is learner model or user model. Adaptive learning system takes advantages of learner model to improve the quality of adaptation task but it does not build up or manipulate learner model. User modeling system is responsibility for gathering information to create and update learner model. In other words, user modeling system manages user model and provide necessary information about user to adaptive system. Following figure *1* describes the interaction between user modeling system and adaptive system.

1. Interaction between user modeling system and adaptive system [1]



User modeling system is the heart of adaptive learning system. There are a lot of theories and practical methods to build up adaptive system and user modeling system. Each method has strong points and drawbacks and so it is very useful to evaluate these methods in order to determine which model is appropriate to which situation because each method tailors to concrete conditions and contexts. For example, studying via internet website is very different from studying at a course with support of network. This report focuses on how to evaluate adaptive learning system with regard to user modeling system in e-learning or distance learning context when there is no separation between adaptive learning system and user modeling system. We can consider the corporation between adaptive system and user modeling system as an integrated model so-called adaptive learning model. Thus, this research has two goals:

* Firstly, research proposes criterions to evaluate adaptive learning model.
* Secondly, research gives some scenarios as an example that applies criterions above into performing evaluation task in concrete situations.

Note that all concepts relating to term “learning” in this research refers to learning via internet or distance learning or e-learning if there is no additional explanation.

# Evaluation criterions

This research proposes three criterions of evaluation:

* Criterion *α* so-called *system criterion* tells us how adaptive learning system works with/without user modeling system. For example, when modeling server applies Bayesian network into build up learner model, criterion *α* measures the performance of adaptive system with or without the support of Bayesian network. In general, this criterion answers two following questions:
  + How adaptation is performed in adaptive system with/without the support of modeling server.
  + Whether the whole user knowledge is computed more accurately with the support of user model, for example Bayesian network.
* Criterion *β* so-called *academic criterion* tells us how well modeling server helps users to study. This criterion surveys users’ study result. The higher criterion *β* is, the better study result is.
* Criterion *γ* so-called *adaptation criterion* or satisfaction criterion measures the quality of adaptation function of learning system with the support of modeling server. After every student gives feedbacks or comments on adaptive system, these feedbacks are collected and analyzed; hence, criterion *γ* is calculated based on these feedbacks in order to estimate level of students’ satisfaction from adaptive system. The higher criterion *γ* is, the better quality of adaptation is.

Now we discuss methods to determine these criterions. Note that in this research, the default user model is Bayesian network model if there is no additional explanation.

## Calculating criterion α

There are two ways to calculate criterion *α* such as using hypothesis testing and using regression model. By using hypothesis testing, suppose the amount of knowledge that user mastered over a concept *C* is quantified as a measure *kc*. Let *KU* = {*k1U, k2U,…, knU*} be knowledge vectors of user *U*, where each measure *kcU* represents the mount of knowledge *C* that user *U* mastered. Given group *A* and group *B* are groups of students learning via adaptive system with and without support of user model, respectively. Two users *i* and *j* are picked randomly in group *A* and group *B*, respectively. We have:

*Ki =* {*kiU, kiU,…, kiU*} has sample variance *si2*.

*Kj =* {*kjU, kjU,…, kjU*} has sample variance *sj2*.

Criterion *α* is represented by the statistical hypothesis testing indicates how well the Bayesian network in group *A* supports adaptive learning system. In other words, with the support of user model, adaptive learning system makes user knowledge around the average knowledge. Therefore, hypothesis test aims to variance test instead of mean test. Null hypothesis is stated that two variances are equal. Lower-tail technique is applied when the alternative hypothesis is assumed that group *A* has less variance:

*H0 : si2 = sj2*

*H1: si2 = sj2*

Suppose the significant level is *0.05*, *F*-distribution is used to test two variances.

If *F* < *f0.95, n-1, n-1* then the null hypothesis is rejected, we can include that group *A* with support of user modeling system improve adaptive learning system much more than group *B*. So, criterion *α* get Boolean value *true*, indicating the preeminence of user modeling system.

The essence of *α* is to measure the level of precision of inference methods with/without user model. By using regression technique, if an inference method is good, its predictive value, namely the whole knowledge user achieves, and all partial knowledge items user study at every stage on learning path will satisfy well a function or equation. In other words, this predictive value has small deviation/error. Suppose that partial user knowledge items like chapters, sessions, etc are represented as a set of random variable are *X1, X2, X3,…, Xn*. Let *Y* represents the total knowledge that users gain over whole course like Java course, Oracle course, etc. We try to find out the linear function of random variables *Xi* (s) so that *Y* is the expected value of such function.

*Y = a0 + a1X1 + a2X2 + a3X3 +…+anXn*

Let *ai* (s) be regression coefficients. Therefore linear function is determined, it is applied back to each user in both groups *A* and *B* so as to predict her/his knowledge so-called *estimated knowledge*. Such knowledge is compared to *real knowledge* of users. The deviation of *estimated knowledge* and *real knowledge* is called *prediction error*. The square sum of all *prediction error* reflects the measure *α*. In general, the process to calculate *α* has four steps:

1. Firstly, the regression coefficients *ai* (s) are computed by the method of mean least square. Note that because we have two linear functions for groups *A* and *B*, there are two sets of regression coefficients, each set for one group.
2. Secondly, let *ekiA* and *ekiB* be estimated knowledge of user *ith* in group *A* and *B*, respectively. Note that *ekiA* and *ekiB* are calculated by applying linear function determined in the first step.

*ekiA = YA = a0 + a1X1 + a2X2 + a3X3 +…+anXn* (in group A)

*ekiB = YB = a0 + a1X1 + a2X2 + a3X3 +…+anXn* (in group B)

1. Thirdly, let *kiA* and *kiB* be the real knowledge of user *ith* in group *A* and *B* from database, respectively. Let *erriA* and *erriB* be the prediction error of user *ith* in group *A* and *B*, respectively.

*erriA = | ekiA* – *kiA|*

*erriB = | ekiB* – *kiB|*

1. Finally, the measure *α* is simple inverse of square sum of all prediction errors.

The larger the measure *α* is, the better the level of precision of inference method is.

## Calculating criterion β

Let *KA, KB* be the knowledge vectors of group *A* and *B*, respectively where *kiA* and *kiB* is the average grades of students in group *A* and *B* over concept *i*, respectively.

*KA*= (*k1A, k2A,…, knA*) has sample variance *sA2* and sample mean .

*KB*= (*k1B, k2B,…, knB*) has sample variance *sB2* and sample mean .

The measure *β* for each group is computed as accumulative probability of assumption user in such group has mastered over course.

Let Φ is cumulative function for standard normal distribution. Note that Φ should be accumulative function for *t-*distribution with *n – 1* degree of freedom for more accurate. Here we use standard normal distribution as example for convenience. The higher criterion *β* is, the better study result is because the academy criterion *α* is measured as the probability of event that student’s grade is higher than or equal to *0.5*.

## Calculating criterion γ

Suppose a questionnaire is built up by expert and it is composed of *n* questions *Q =* (*q1, q2,…, qn*). Users in each group rate on each question where rating value may be binary satisfied and unsatisfied. By the simplest way, criterion *γ* is defined as the ratio of the number of satisfied users to the whole number of users.

In enhance method, the rating values range in an interval, for example [*0…5*], where value *0* and *5* indicates least and most satisfied. Therefore, we have two rating matrices *A* and *B* for two groups. Each row in rating matrix is composed of rating values of a user; in other words, each cell represents a rating that a user gives to a question. Each matrix is “*compressed*” into a mean vector. Let *μA* and *μB*be the mean vectors of groups *A* and *B,* respectively. The measure *γ* is computed as the module (length) of mean vector. Which group has higher measure *γ* will satisfy users much more.

*γA = |μA|* and *γB = |μB|*

There are three steps to compress rating matrix and to calculate *γ*:

1. Firstly, rating matrix is “shrunk” by projecting it onto its eigenvectors. The number of columns *k* is much smaller than *n*, *k << n*. These columns represent essential questions. Tables I and II shows rating matrix and “shrunk” rating matrix, respectively.
2. Rating matrix as collection of users’ feedbacks

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *a11* | *…* | *a1j* | *…* | *a1n* |
| *…* | *…* | *…* | *…* | *…* |
| *ai1* | *…* | *aij* | *…* | *ain* |
| *…* | *…* | *…* | *….* | *…* |
| *am1* | *…* | *amj* | *…* | *amn* |

1. Rating matrix is shrunk by projecting it onto its eigenvectors

|  |  |  |
| --- | --- | --- |
| *a11* | *…* | *a1k* |
| *…* | *…* | *…* |
| *ai1* | *…* | *aik* |
| *…* | *…* | *…* |
| *am1* | *…* | *amk* |

1. Secondly, each column of matrix corresponding to each question is assumed as a statistical distribution *F*i. Thus the mean of *F*i is estimated by *μi*.
2. Finally, the mean vector of this matrix is composed of all estimates *μi* and the criterion γ is the module of such mean vector.

In general, the higher criterion *γ* is, the better quality of adaptation is.

# An evaluation scenario

Evaluation scenario is the example for demonstrating how to calculate and apply aforementioned criterions into evaluating the quality of adaptive learning model. E-learning cannot replace face-to-face teaching and it should exist parallel and support traditional education. Thus, this scenario makes the comparison between face-to-face learning manner and distance learning manner. This evaluation scenario is divided into three main acts in which students and teacher play the roles of actors.

1. *Study act*: Teacher teaches and students learn in both face-to-face manner and e-learning manner via website. Suppose students are classified into three groups *A, B* and *C*. Groups *A* and *B* represent face-to-face manner and e-learning manner via website, respectively. Especially, group *C* represents e-learning manner with support of user model, namely Bayesian network.
2. *Feedback act*: Students give feedbacks to teacher and teacher collects and analyzes them.
3. *Evaluation act* is done by teacher; thus, criterions *α, β* and *γ* are calculated according to data collected from two above acts. The quality of adaptive learning in groups *A*, *B* and *C* are determined based on such criterions.

Study act has *5* scenes:

1. Teacher builds up school’s curriculums and set up adaptive e-learning website with/without the support of user modeling system.
2. Teacher teaches and students in groups *A, B* and *C* learn by face-to-face manner.
3. Students in groups *B* and *C* go on website and study by themselves. Teacher monitors them and put up important notice.
4. Students in groups *A, B* and *C* do tests and exercises via website.
5. Teacher evaluates students based on their test results.

Teacher’s role in study act:

* Teaching face-to-face in traditional manner.
* Building up knowledge domain and creating web resources for this domain such as defining html lesions, tests, exercises, etc.
* Creating user model, for example, creating Bayesian network and its weights for knowledge domain.
* Setting up user modeling system and e-learning adaptive website.
* Monitoring students’ learning process.
* Sending test results and school report to students.

Students’ role in study act:

* Students in groups *A*, *B* and *C* go to class to study in face-to-face manner.
* Students in groups *B* and *C* learn themselves on adaptive learning web sites. Note that website / learning materials are adapted to each student based on their knowledge and characteristics.
* Students in groups *A*, *B* and *C* do tests / exercises via website.

Feedback act has *3* scenes:

1. Teacher creates the questionnaire to survey students’ feeling about both adaptive learning website and curriculum such as very satisfied, satisfied and not satisfied.
2. Students answer or rate on such questions online.
3. Teacher collects students’ feedbacks and analyzes them.

Evaluation act has *2* scenes:

1. Teacher calculates three criterions based on students’ feedback and test results.
2. Teacher makes the decision about the quality of face-to-face teaching manner and e-learning manner with/without support of user modeling system.

For example, there are two classes *A* and *B*. Class *A* is only taught in face-to-face manner, otherwise students in class *B* study both in face-to-face manner and adaptive learning environment. Study results and students feedback are collected from both two classes. Each class has the same number of students, namely *10*. Let *GA* and *GB* be the average study results of classes *A* and *B*, respectively.

*GA =* {*4, 5, 3, 6, 2, 8, 3, 5, 8, 6*}

*GB =* {*6, 8, 9, 10, 6, 7, 9, 6, 9*}

Suppose there are *4* students in class *A* and *2* students in class *B* who don’t satisfy teaching curriculum. Sample mean and deviation of class *A* are =5.0 and *sA* = = *2.05*. Sample mean and deviation of class *B* are = 7.0 and *sB* = = *1.66*. Let *αA* and *αB* be academy criterions of class *A* and class *B*, respectively.

Let *βA* and *βB* be satisfaction criterions of class *A* and class *B*, respectively, we have:

Suppose the weights of criterion *α* and criterion *β* are *0.7* and *0.3*, respectively. Let *evalA* and *evalB* be the total evaluations on group A and group B, respectively.

*evalA = 0.7\*0.5 + 0.3\*0.6 = 0.53*

*evalB = 0.7\*0.99 + 0.3\*0.8 = 0.93*

When *evalA* is greater than *evalB*, it is possible to conclude that the teaching method in class *B* is more effective than the one in class *A* because class *B* takes advantages of adaptive learning method.

# Conclusion

This conclusion gives some comments on evaluation criterions. As aforementioned, there are three criterions such as system criterion *α*, academy criterion *β* and adaptation criterion *γ*. That two of three criterions, concretely *α* and *β*, assessing user knowledge implicates that evaluation of adaptive learning model focuses on the effect of education which is ability to help students to improve their knowledge although adaptation and personalization are significant topics in adaptive learning. You can recognize that the education never goes beyond the main goal that increases amount of human knowledge. Evaluation scenario, an example for demonstrating how to determine these criterions, indicates that study is lifelong process for everyone and so, classes and courses are short movies in this lifelong process. Both students and teachers are actors and their roles can mutually interchange, for example, teaching is the best way to learn and student is the best teacher of teacher.

##### References

1. Alfred Kobsa. Generic User Modeling Systems. User Modeling and User-Adapted Interaction 2006 (UMUAI-2006).
2. Peter Brusilovsky and Eva Millán. User Models for Adaptive Hypermedia and Adaptive Educational Systems. The Adaptive Web – Methods and Strategies of Web Personalization. Lecture Notes in Computer Science, Springer-Verlag Berlin Heidelberg 2007, ISBN-13: 978-3-540-72078-2.