Recommendation Algorithm
for Learning Materials
That Maximizes Expected Test Scores

Tomoharu Iwata[1], Tomoko Kojiri[2],
Takeshi Yamada[1], and Toyohide Watanabe[2]

[1] NTT Communication Science Laboratories,
2-4, Hikaridai, Seika-cho, "Keihanna Science City," Kyoto, 619-0237, Japan
{iwata,yamada}@cslab.kecl.ntt.co.jp
[2] Graduate School of Information Science, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya, 464-0803, Japan
{kojiri,watanabe}@watanabe.ss.is.nagoya-u.ac.jp

Abstract. We propose a recommendation algorithm for learning materials that enhances learning efficiency. Conventional recommendation methods consider user preferences and/or levels, but they do not directly consider the learning efficiency. With our method, the learning efficiency is quantified by the expected improvement in the test score, and materials are recommended so as to maximize this expected improvement. The expected improvement is calculated with logistic regression models that employ the user's test result obtained before learning as input. Experimental results using fill-in-the-blank exercises for English learning show that our method yields major improvements in performance compared with random material recommendation.

Keywords: e-learning, personalization, adaptive tutoring systems, logistic regression.

1 Introduction

With the rapid growth of network and database technologies, e-learning systems have become widely used in various domains. With e-learning systems, the recommendation of suitable materials or exercises for each user is important because users have different learning levels, prior knowledge, and goals. In a textbook, the sequence of materials and exercises is fixed. This is very inefficient because some users find themselves undertaking exercises that are too easy or reading materials that are not examined in the target test.

In this paper, we propose a recommendation algorithm for learning materials that directly enhances learning efficiency. Although many personalized learning material recommendation algorithms have been proposed, they do not directly maximize learning efficiency. Instead, they consider user preferences or/and levels [1,2,3,4]. With our method, the learning efficiency is quantified by the expected

improvement in the test score, and materials are recommended so as to maximize the expected improvement. Intuitively speaking, our method recommends materials so that users provide correct answers to questions that they answered incorrectly prior to learning.

The expected improvement is calculated by logistic regression models [5] using the questions that the user answered incorrectly before learning as input. By training the logistic regression models using the learning log data and user test results, we can automatically extract information about which learning materials contribute to improvements in the test score. Since our method does not require meta data of the materials, it is applicable to any course.

2   Proposed Method

2.1   Preliminaries

Let x_i and y_i be variables that represent whether question i is correctly or incorrectly answered before and after the learning phases, respectively, as follows:

$$x_i = \begin{cases} 1/-1 & \text{if question } i \text{ is correctly/incorrectly answered before learning,} \\ 0 & \text{otherwise,} \end{cases}$$

(1)

$$y_i = \begin{cases} 1/-1 & \text{if question } i \text{ is correctly/incorrectly answered after learning,} \\ 0 & \text{otherwise,} \end{cases}$$

(2)

where $x_i = 0/y_i = 0$ means question i has not been answered before/after the learning phase. The results of the set of test questions $\boldsymbol{V}$ before and after the learning phase are represented by vectors $\boldsymbol{x} = (x_i)_{i \in \boldsymbol{V}}$ and $\boldsymbol{y} = (y_i)_{i \in \boldsymbol{V}}$, respectively.

Let z_j be a variable that represents whether material j is recommended in the learning phase as follows:

$$z_j = \begin{cases} 1 & \text{if material } j \text{ is recommended in learning,} \\ 0 & \text{otherwise.} \end{cases}$$

(3)

The recommended materials are represented by a vector $\boldsymbol{z} = (z_j)_{j \in \boldsymbol{M}}$, where $\boldsymbol{M}$ represents a set of learning materials.

2.2   Recommendation Algorithm

The goal of our method is to select materials to be used in the learning phase from a set of materials in order to enhance learning efficiency. Learning efficiency is quantified by the expected improvement in the test score.

The expected improvement in the test score given recommended materials $\boldsymbol{z}$ is written as follows:

$$E(\boldsymbol{z}) = \sum_{i \in \boldsymbol{V}} P(i)P(x_i = -1)P(y_i = 1|x_i = -1, \boldsymbol{z}),$$

(4)

Table 1. The material recommendation procedure with our method

---
1. Input the test result before learning $\boldsymbol{x}$,
2. Initialize the recommended material vector: $\boldsymbol{z} = (0, \cdots, 0)$,
3. Select a material to be recommended $\hat{j}$ by (6),
4. Update the recommended material vector: $\boldsymbol{z} = \boldsymbol{z}^{+\hat{j}}$,
5. Return to step 3 unless an end condition is satisfied.

---

where $P(i)$ represents the probability that question i is asked in the test, $P(i) + \bar{P}(i) = 1$, in which $\bar{P}(i)$ represents the probability that question i is not asked in the test, $P(x_i = -1)$ represents the probability that question i is incorrectly answered before learning, and $P(y_i = 1|x_i = -1, \boldsymbol{z})$ represents the probability that question i is correctly answered after the learning phase when the question i is incorrectly answered before the learning phase and materials $\boldsymbol{z}$ are recommended. In (4), $E(\boldsymbol{z})$ is regarded as the expected number of questions that are incorrectly answered before the learning phase and correctly answered after the learning phase given recommended materials.

When the probabilities that questions are asked are uniform, and questions that are incorrectly answered before learning are known, the expected improvement in the test score can be simplified as follows:

$$E(\boldsymbol{z}|\boldsymbol{x}) \propto \sum_{i \in \boldsymbol{V}} I(x_i = -1)P(y_i = 1|x_i = -1, \boldsymbol{z}), \qquad (5)$$

where $I(A)$ represents an indicator function, i.e. $I(A) = 1$ if A is true, $I(A) = 0$ otherwise. We use (5) as the expected improvement in the test score in the following sections for the simplicity.

Our method sequentially selects a material that maximizes the expected improvement from materials that have not yet been recommended as follows:

$$\hat{j} = \arg \max_{j:z_j=0} E(\boldsymbol{z}^{+j}|\boldsymbol{x}), \qquad (6)$$

where $\boldsymbol{z} = (z_j)_{j \in \boldsymbol{M}}$ represents currently recommended materials, and $\boldsymbol{z}^{+j}$ represents recommended materials when material j is newly recommended, or $z_{j'}^{+j} = 1$ if $j = j'$ and $z_{j'}^{+j} = z_{j'}$ if $j \neq j'$. Table 1 shows the material recommendation procedure with our method. Examples of end conditions include those where the number of recommended materials, the expected improvement, or the time period of the learning phase exceeds a certain threshold.

2.3 Improvement Model

When recommending materials, our method requires improvement model $P(y_i = 1|x_i = -1, \boldsymbol{z})$, which is the probability of the improvement of question i given recommended materials $\boldsymbol{z}$. The improvement is modeled based on logistic regression [5] as follows:

$$P(y_i = 1|x_i = -1, \boldsymbol{z}) = \frac{1}{1 + \exp(-(\mu_i + \boldsymbol{\theta}_i^\top \boldsymbol{z}))}, \qquad (7)$$

where μ_i and $\boldsymbol{\theta}_i = (\theta_{ij})_{j \in \boldsymbol{M}}$ are unknown parameters. Intuitively speaking, μ_i represents the ease with which question i is improved, and θ_{ij} represents the influence of material j on the improvement of question i.

The unknown parameters $\boldsymbol{\Theta} = \{\mu_i, \boldsymbol{\theta}_i\}_{i \in \boldsymbol{V}}$ can be estimated by maximizing the following log likelihood using the learning log data and test results for a set of users $\boldsymbol{N}$:

$$L(\boldsymbol{\Theta}) = \sum_{n \in \boldsymbol{N}} \sum_{i \in \boldsymbol{V}} \Big(I(x_{ni} = -1 \wedge y_{ni} = 1) \log P(y_{ni} = 1 | x_{ni} = -1, \boldsymbol{z}_n)$$
$$+ I(x_{ni} = -1 \wedge y_{ni} = -1) \log P(y_{ni} = -1 | x_{ni} = -1, \boldsymbol{z}_n) \Big), \qquad (8)$$

where x_{ni} and y_{ni} indicate whether question i is correctly or incorrectly answered by user n before and after the learning phases, respectively, and $P(y_{ni} = -1 | x_{ni} = -1, \boldsymbol{z}_n)$ represents the probability that question i is incorrectly answered by user n after the learning phase when question i is incorrectly answered before the learning phase and materials $\boldsymbol{z}$ are recommended. The global optimum solution is guaranteed because the above log likelihood based on logistic regression models is a convex function.

3   Experiments

3.1   Setting

We evaluated our method using fill-in-the-blank questions for English learning, which are employed in many tests designed to evaluate grammatical knowledge, such as TOEIC and TOEFL. In the fill-in-the-blank questions used here, users select appropriate words with the correct grammar for the blank in the sentence from four options.

We implemented a web-based e-learning system for the evaluation. In the experiment, a user takes a test before and after the learning phase (pre-test and post-test) to measure the effect of learning on improving the test score. One question is presented to a user on one web page, and the user answers each question in series. The questions in the pre- and post-tests are the same, and there are 40 questions, $|\boldsymbol{V}| = 40$.

The materials recommended in the learning phase are fill-in-the-blank questions with solutions and explanations. One question is recommended to each user on one web page for learning, and the solution and explanation are presented on one web page after the user has answered the question. Please note that the users are not supplied with solutions and explanations in the pre- and post-tests. There are 80 learning materials, or questions, $|\boldsymbol{M}| = 80$, and 40 of these are recommended to each user in the learning phase. The learning materials are different from the questions in the pre- and post-tests. However, about half of the materials are related to test questions, for example they involve questions about the same idioms and grammatical rules.

Table 2. AUC of improvement models based on the Bernoulli distribution and the logistic regression

Bernoulli distribution	Logistic regression
0.556	**0.592**

3.2   Evaluation of Improvement Models

Our method requires improvement model $P(y_i = 1|x_i = -1, \boldsymbol{z})$. We constructed and evaluated improvement models using the log data of 52 users with random material recommendations, $|\boldsymbol{N}| = 52$. We compared improvement models based on the logistic regression in (7) with the Bernoulli distribution. The Bernoulli model assumes that the improvement does not depend on the recommended materials $\boldsymbol{z}$. The parameters can be estimated based on the maximum likelihood.

For the evaluation measurement, we used the area under the ROC curve (AUC) of the problem to predict whether or not questions that were incorrectly answered in the pre-test are correctly answered in the post-test. A higher AUC represents better predictive performance. We computed AUC using leave-one-out cross-validation, which means that we used 52 evaluation data sets, in each of which one user's data are used for the evaluation and the data of the other 51 users are used for the training. Table 2 shows the AUC. The AUC of the improvement model based on the logistic regression is higher than that of the Bernoulli model, which implies that the recommended materials are important in terms of predicting the improvement in the score, and we can predict the improvement in the test score with the logistic regression model.

The highest and second highest θ_{ij} were $\theta_{i_1 j_1} = 0.388$ and $\theta_{i_2 j_2} = 0.208$, respectively, where question i_1 and material j_1 are about the idiom 'stop by', and question i_2 and material j_2 are about the idiom 'across the street'. This result is natural because the recommendation of materials about the same idioms can improve the test score. Even though our method does not use information related to question or material content, it automatically extracts the relationship between questions and materials using the learning log data and test results.

We analyzed the relationship between questions and materials using question i_1 and material j_1, which are about the idiom 'stop by', as an example. In the pre-test, 32 users answered question i_1 incorrectly. The probability of the user answering question i_1 correctly in the post-test when material j_1 was recommended was $\hat{P}(y_{i_1} = 1|x_{i_1} = -1, z_{j_1} = 1) = 15/17$. In contrast, the probability when material j_1 was not recommended was $\hat{P}(y_{i_1} = 1|x_{i_1} = -1, z_{j_1} = 0) = 2/15$. This result indicates that the recommendation of material j_1 is effective in improving the response to question i_1.

3.3   Evaluation of Recommendation Algorithms

We evaluated the learning efficiency of the proposed recommendation algorithm by comparing it with a random recommendation algorithm. 38 users studied materials recommended at random, and 49 users studied materials recommended

Table 3. Average improvements in the test scores with random recommendation and our method

Random	Our method
4.474	**8.125**

by our method. Table 3 shows the average improvements in the test scores on a 100-point scale. Our method provided statistically significant increases compared with the random recommendation (one-tailed t-test, $p < 0.04$). In the questionnaires, 86% of users answered that there were helpful materials in the learning phase with our method.

4 Conclusion

We proposed a method for recommending learning materials that maximizes learning efficiency, or the expected improvement in the test score. The experimental results encourage us to believe that our learning material recommendation approach is promising and will become a useful tool for e-learning. Although we modeled the expected improvement in the test score with logistic regression using only the learning log data and test results, we could also use content information about the learning materials and test questions such as difficulties, and user attributes such as levels. We plan a further verification of our method by applying it to other courses of learning.

References

1. Chen, C.M., Hong, C.M., Chang, M.H.: Personalized learning path generation scheme utilizing genetic algorithm for web-based learning. WSEAS Transactions on Information Science and Applications 3(1), 88–95 (2006)
2. Li, X., Chang, S.K.: A personalized e-learning system based on user profile constructed using information fusion. In: Proceedings of the 11th International Conference on Distributed Multimedia Systems, Banff, Canada, pp. 109–114 (September 2005)
3. Stern, M., Woolff, B.P.: Curriculum sequencing in a web-based tutor. In: Goettl, B.P., Halff, H.M., Redfield, C.L., Shute, V.J. (eds.) ITS 1998. LNCS, vol. 1452, pp. 574–583. Springer, Heidelberg (1998)
4. Tang, T.Y., Mccalla, G.: Smart recommendation for evolving e-learning system. In: Proceedings of the 11th International Conference on Artificial Intelligence in Education, Workshop on Technologies for Electronic Documents for Supporting Learning, pp. 699–710 (2003)
5. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, New York (2001)