Estimating Learners’ Subjective Impressions of the Difficulty of Course Materials in e-Learning Environments

Kazuaki Nakamura¹, Koh Kakusho², Masayuki Murakami³, Michihiko Minoh²

¹Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University
²Academic Center for Computing and Media Studies, Kyoto University
³Research Center for Multi-Media Education, Kyoto University of Foreign Studies

¹nakamura@mm.media.kyoto-u.ac.jp, ²{kakusho, minoh}@media.kyoto-u.ac.jp
³masayuki@murakami-lab.org

Abstract
One of the problems of present e-learning compared with traditional classroom lectures is that teachers cannot assess learners’ interest in learning or willingness for learning. According to the achievement motivation theory, the learners’ interest or willingness is deeply related to the learners’ subjective impressions of the difficulty of each course material (SIDC). In this paper, we propose the method for estimating the learners’ SIDC by observing their behaviors during e-learning, in order for the teachers to be able to assess the learners’ interest or willingness. In estimating the learners’ SIDC, we use different feature processing criterion for each learner based on the correlation between the learners’ SIDC and the features extracted from their behaviors, which is differently modeled by SVM for each learner, considering that the correlation between the learners’ SIDC and the features may depend on the learners. In the result of our experiment, the learners’ SIDC were estimated with accuracy of 85.8% on an average.

Keywords
e-learning, SIDC, eye gaze, head poses, and mouse actions

1 Introductions
Recently, many companies and educational institutions have introduced e-learning to their human resource development. One of the major advantages of e-learning is that learners can learn anytime and anywhere they want. On the other hand, e-learning has a disadvantage that teachers can hardly know the learners’ learning process because the teachers cannot see the learners’ behaviors beside them. In order to cope with this disadvantage, some institutions use learning-logs stored by e-learning system. In the learning-logs, it was recorded when each learner accessed each page of course materials or which quizzes in the course materials were correctly answered by each learner. However, from the learning-logs, the teachers can know only the level of understanding of each learner. They still cannot know which part of the course materials engaged each learner’s interest or which part impressed each learner as difficult.

In classroom lectures, lecturers understand not only the level of understanding of the learners but also their interest in the lectures or willingness to learn by seeing their behaviors during the lectures. Today, many universities attempt to evaluate the lectures that they offered [1], aiming to improve the lectures. In this lecture evaluation, following two criteria are often used [2].

(1) The level of understanding of the learners for instructional contents
(2) The learners’ interest in the instructional contents or willingness to learn

The evaluation results based on above criteria are used by the lecturers to modify their course materials or instructional methods. For example, the lecturers modify the explanations or metaphors which could not engage the learners’ interest.

In contrast, the teachers can know only the level of understanding of the learners in current e-learning environments, as previously described. In other words, the teachers use only criterion (1) in current e-learning. They do not use the criterion (2). However, considering the criterion (2) is also important. Because the learners’ interest or willingness are useful for modifying the course materials, similar to the case of classroom lectures. For example, the statistical analysis for the learners’ interest can specify uninteresting parts of the course materials explicitly. The descriptions of these parts should be modified by
the teachers.

There are several factors which concern with the learners’ interest. The difficulty of the course materials is one of the factors, and it is relatively important. According to Atkinson’s theory of achievement motivation [3], when a person do a task, his/her tendency to achieve $T$ is modeled as the product of his/her subjective probability of success $PS$ and the task’s incentive for success $IS$, that is $T = PS \times IS$. The $PS$ represents how strongly the person believes that s/he can accomplish the task successfully. The $IS$ represents how strongly the person feels that accomplishing the task.

In addition to this, the $IS$ is modeled as the complement number of $PS$, that is $IS = 1 - PS$. Hence, the $T$ is definitely modeled as $T = PS \times IS = PS(1 - PS)$. This Atkinson’s model can be applied to e-learning tasks. In the e-learning case, the $PS$ represents how strongly the learners feel it difficult to understand the course materials, and the $T$ represents the learners’ interest or willingness. Thus, the learners’ subjective impressions of the difficulty of the course materials (SIDC) are considered one of the important factors which concern with the learners’ interest.

The changes of the learners’ mental state or mental level including interests, willingness, or the SIDC, is considered to be expressed as the changes of the learners’ nonverbal information such as facial expressions, eye gaze, head poses, hand gestures, and so on. In classroom lectures, the teachers actually assess the learners’ interest in the lectures based on how many students raise their faces [2]. This indicates that the learners during a lecture show their mental state via their faces or head poses. Similar to this, the learners during e-learning are also considered to show their mental state via their faces, gaze or head poses.

Based on the above discussions, in this paper, we propose the method for estimating the learners’ SIDC by observing their nonverbal information such as facial expressions, eye gaze, head poses, and histories of mouse actions, aiming to add the above criterion (2) into the current e-learning environments, in which the teachers can only use the criterion (1).

In the remainder of this paper, we will first discuss about some previous works related to our purpose in section 2. In section 3, we will clarify which features of the learners’ nonverbal information are useful for estimating the learners’ SIDC through two exploratory experiments. And, based on the results of the experiments, we will propose a method for estimating the SIDC from those useful features. In section 4, we will experiment and evaluate the availability of our estimation method before giving our concluding remarks in section 5.

2 Related Works

The study of Ueno et al [4] is one of the researches which aim to assess the conditions of the learners by observing them. In this study, Ueno et al have developed the system for recording how long each learner spends time to learn each page of the course materials as time-series data. And they have proposed the method for detecting outlier from the recorded data. However, this study does not consider which mental state corresponds to the detected outlier. Therefore this method cannot estimate the learners’ mental state or mental level including the SIDC.

Affective Computing [5] is the research field whose goal is to estimate the human’s mental states in order to realize natural interaction between humans and computers by exchanging information involved with human’s mental states. Previously, much research in this field focused on only “affective” states represented by Ekman’s 6 basic emotions (“joy”, “fear”, “sadness”, “disgust”, “anger” and “surprise”) [6]. For example, in the study of Li et al [7], which have aimed to apply affective computing to e-learning, the learners’ “affective” states are classified into 8 states including Ekman’s 6 basic emotions together with “teasing” and “neutral”. They have tried to estimate the learners’ current “affective” state. However, it is not so usual that the learners express such emotion or “affective” state during learning. The learners feel “difficulty” or “boredom” rather than the basic emotion in learning situations.

From the above viewpoint, some researchers have tried to estimate other kind of mental states such as the level of interest of the learners. Nosu et al [8] have assumed the following four axes, each of which has two values (or states), as the learners’ mental state: easy/difficult, boring/interesting, confused/comprehending and tired/concentrating. Then they have tried to estimate which side of each axis the learners’ condition exists in. For this estimation, they used biometrical signals including pulse rate, breathing rate and finger temperature as well as facial expressions. However, observing biometrical signals requires the learners to wear specific devices. Wearing such devices during learning is frustrating for the learners.

Nakamura et al [9] have classified the learners’ mental states into two states: “stalled” and “not stalled”. And they have proposed the method for assessing whether each learner are “stalled” or not at each moment. In their definition, the “stalled” learner is defined as the learner who feels that they can no longer answer the given question. In this study, operation time intervals, which can be obtained without any wearable devices, are used as a feature for estimation. In order to encourage the learners to perform many operations, and in order to obtain operation time inter-
vals stably, Nakamura et al have developed the learning environment which presents the hints or explanations of the questions triggered by the learners’ operations. This environment is suitable for testing or examining the learners’ skills. However, in actual e-learning situations, there are not only test-based course materials but also the course materials consisting only of expository sentences, figures, and video clips, which do not include any tests or quizzes. This kind of course materials basically does not encourage the learners to perform any operations. Therefore, operation time interval may not be obtained stably from the learners studying this kind of course materials. In other words, it has not been clarified whether the operation time intervals are still useful for estimating the learners’ mental states. Another feature which can be obtained stably and independently of learning environments should also be considered.

In addition, since unconscious human behaviors depend on individual learners, different behaviors need to be employed for estimating the mental state of each learner. In previous studies above, the same features extracted from same behaviors are used for every learner.

In this paper, we consider the learners’ nonverbal information which can be observed independently of learning environments without any wearable devices. It includes facial expressions, eye gaze, and head poses, which we call facial information. Such facial information can be easily obtained by observing the face by cameras. Since facial information is deeply related on the mental state or mental level of each learner, we focus on the facial information as well as operation time intervals. In order to specify useful behaviors of the learners’ facial information, we conduct two examinations. Firstly we examine which behaviors of the learners’ facial information are considered when we humans estimate the learners’ SIDC (in section 3.1). Secondly we examine the correlations between the learners’ SIDC and each of the behaviors in order to clarify the behavior actually correlates to the learners’ SIDC and how the correlations are different by individual learners (in section 3.2). Based on the results of these examinations, we develop the estimator for the learners’ SIDC using machine learning methods (in section 3.3).

3 Estimating the learners’ SIDC from their behaviors

3.1 Which behaviors are useful for estimating the SIDC?

Which behaviors of the learners are considered when we humans estimate the learners’ SIDC? One of the ways for answering this question is to give a questionnaire to the participants who are observing the learners during e-learning. In our previous study [10], we have developed the system for recording an e-learner’s facial image sequence seen through his/her computer screen with the viewpoint behind the screen. By this facial image sequence, we are observe both the e-learner’s facial information, including facial expressions, eye gaze and head poses, and course materials displayed on the computer screen (see Figure 1). Using this system, we perform the following exploratory experiment.

First, we recorded videos of two Japanese e-learners answering English word test. The videos are presented to 7 participants different from the learners, for answering a questionnaire. In this exploratory experiment, we made two kinds of English word tests: One consists of easy words, and the other consists of difficult words. The difficulty level of each English word is evaluated by the SVL12000, which is the vocabulary lists of various levels created for Japanese students of English. We asked the two learners to answer both tests, so that we recorded two kinds of videos by each learner. On the other hand, we asked the 7 participants to estimate SIDC of each learner. We also asked the participants why they estimated so. In the result of this experiment, all participants correctly selected the videos of the learners answering the difficult test. The reason for the answer is shown in Figure 2.
Since the reasons “Moving the face frequently” and the “Tilting the head to the side frequently” are considered two different descriptions of the same behavior of head movement, these two reasons are merged. The “Mumbling something” may be the behavior that appears only in limited situations. Since we used English word test, some of the learners mumble the words included the quiz sentences in this experiment. However, this is not the case of general e-learning situations. The number of subjects who pointed out the “Showing confused look on the face” is only one. This indicates that the “Showing confused look on the face” is less important behavior than the other behaviors listed in the Figure 2, at least in this experiment. From above discussion, it turns out that the “Tilting the head to the side frequently”, the “Gazing at the same object for a long time”, and the “Taking long time to answer the tests” are important behaviors for humans to estimate the learners’ SIDC. We use following values as the features for estimating the learners’ SIDC, each of which corresponds to one of the three important behaviors above.

(1) “Tilting the head to the side frequently”

While the learner is tilting his/her head, the angle of the head tilting from the vertical direction \( \omega \) (see Figure 3) is also changing with time. We use the variance of the \( \omega(t) \) within a certain interval of time, which is described as \( V[\omega(t)] \), as the feature corresponding to the “Tilting the head to the side frequently”.

(2) “Gazing at a same object for a long time”

While the learner is gazing at a single object, his/her point of gaze \( g(t) \) is a point in \( \mathbb{R}^2 \) does not move widely on the display. For this case, the average amount of the variation of the \( g(t) \) within a certain interval of time, which is described as \( E\left[\frac{d}{dt}g(t)\right] \), becomes small. Similarly, the total variance of the \( g(t) \), which is described as \( V[g(t)] \), also becomes small. We use the \( E\left[\frac{d}{dt}g(t)\right] \) and the \( V[g(t)] \) as the features corresponding to the “Gazing at a same object for a long time”.

(3) “Taking a long time to answer the tests”

In general, logging in/out each page of the course materials or answering quizzes are conducted by the use of mouse actions in e-learning. In the case that the learner takes a long time to answer the tests, the number of mouse actions per a unit time becomes small. We use the inverse of average number of clicking left button \( IC_L \) and wheeling \( IW \) within a certain interval of time, which are described as \( E[IC_L] \) and \( E[IW] \), as the features corresponding to “Taking a long time to answer the tests”.

The values \( \omega(t) \) and \( g(t) \) defined in above (1) and (2) can be obtained using our system shown Figure 1. In order to synthesize the e-learner’s facial image sequence seen with the viewpoint behind the screen, our system detect the central point of each eye and the mouth from the images captured by the stereo-cameras, and calculate the 3D-positions of the central points based on their binocular parallax. In this paper, we define the learner’s head orientation as the normal vector of the plane spanned over the central point of the left eye, the right eye and the mouth, in order to calculate it using the system shown in Figure 1. The inner product between the head orientation \( \omega(t) \) and the vertical vector gives the angle \( \omega(t) \) with each time \( t \). On the other hand, the point of gaze \( g(t) \) is defined by the intersection of the screen and the learner’s lines of sight. The learner’s lines of sight are decided by the learner’s head orientation and his/her ocular motions. The change of the ocular motions is represented as the change of the pattern on the eye-region of the facial images. We obtain the \( g(t) \) using the eye-region of the facial images and head orientation \( \omega(t) \) with each time \( t \).

We regard the five values defined in (1), (2) and (3) as the candidates of the features for estimating the learners’ SIDC (see Table 1). We call the candidate 1,
2 and 3 of Table 1 “facial features”, and call candidate 4 and 5 “terminal features”.

Table 1: The candidates of the features for estimating the learners’ SIDC

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The variance of the angle of the head tilting</td>
<td>$V[\omega(t)]$</td>
</tr>
<tr>
<td>2</td>
<td>The average of the variation for points of gaze</td>
<td>$E\left[\frac{d}{dt}g(t)\right]$</td>
</tr>
<tr>
<td>3</td>
<td>The total variance of points of gaze</td>
<td>$V[g_x(t)] + V[g_y(t)]$</td>
</tr>
<tr>
<td>4</td>
<td>The inverse of average number of clicking left button</td>
<td>$E[ICL]$</td>
</tr>
<tr>
<td>5</td>
<td>The inverse of average number of wheeling</td>
<td>$E[IW]$</td>
</tr>
</tbody>
</table>

3.2 Analyzing the correlation between the feature candidate and the SIDC

The five feature candidates described in section 3.1 correspond to the behaviors which are useful for humans to estimate the learners’ SIDC. However, the features actually correlated to the SIDC could depend on individual learners, as described in section 2. In this section, we examine which candidate listed in Table 1 is actually correlated to each learner’s SIDC by following exploratory experiment.

(1) Overview of the exploratory experiment

First we prepared 18 English word tests with various difficulties, based on the SVL12000. Each of the tests includes 15 quizzes, and each quiz consists of one English word and four Japanese words. Only one of the Japanese words has the same meaning of the English word, and the others are not related to the English word. The answerers of these quizzes are expected to select the Japanese word with the same meaning of the English word from the four alternatives by clicking the left button of their mouse.

Four learners are given these tests as well as questionnaires on the difficulty of each test after they finished answering the test. In each questionnaire, the learners evaluated the difficulty of each test in 4 levels based on their own judgment. (The learners wrote ‘1’ to the questionnaire when they feel that the test is very easy, and they wrote ‘4’ when they feel the test is very difficult.) While each learner was answering the tests, we recorded the video of his/her face using the system shown in Figure 1 and obtained the facial features 1, 2 and 3. On the same time, we obtained the terminal features 4 and 5 using an application which can record the history of mouse actions performed by each learner. Then we calculate the Spearman’s rank correlation coefficients (SRCC) between the each feature and the resultant values of the questionnaire.

Note that we divided above 18 tests into two groups, each of which includes 9 tests. The difficulties of the two groups are same. We perform the exploratory experiment twice using one group of test per once. This is because it is better to catch more general correlations.

(2) Correlations between the SIDC and the feature candidates

The SRCC between the each feature candidate and the SIDC, which is the resultant values of the questionnaire in this experiment, are shown in Table 2. We perform a two-sided t-test with a significant level equal to 0.05. In Table 2, the items with ‘*’ mean that the correlation is significant under the t-test. The result shown in Table 2 indicates two facts. One is that some of the feature candidates listed in Table 1 does not have a significant correlation with the SIDC for some learners. The other is that at least one feature candidate have a significant correlation with the SIDC for each learner.

Table 2: Spearman’s rank-correlation coefficient between each feature candidates and the SIDC of each learner

<table>
<thead>
<tr>
<th>Learner</th>
<th>Facial features</th>
<th>Terminal features</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.775* -0.564* 0.086</td>
<td>0.861* 0.477</td>
</tr>
<tr>
<td>B</td>
<td>0.575* -0.760* -0.253</td>
<td>0.692* 0.733*</td>
</tr>
<tr>
<td>C</td>
<td>0.844* 0.170 0.258</td>
<td>0.907* 0.873*</td>
</tr>
<tr>
<td>D</td>
<td>0.673* -0.727* -0.538*</td>
<td>0.742* 0.752*</td>
</tr>
</tbody>
</table>

Note: ‘*’ means that the correlation is significant.

(3) Are the correlations different by individual learners?

The result shown in Table 2 also indicates that the correlations between each feature candidate and the SIDC are different by individual learners. For example, the SIDC of subject A has a significant correlation with the facial feature 2 but not with the terminal feature 5. In contrast, the SIDC of subject C has a significant correlation with the terminal feature 5 but not with the facial feature 2. The facial feature 3 has a significant correlation with the SIDC only for the subject D. They indicate that the features which are qualitatively correlated to the SIDC depend on individual learners.
The facial feature 1 and the terminal feature 4 have a significant correlation with the SIDC for all the learners. However, these correlations are also different by individual learners under a quantitative analysis. Figure 4 and 5 are the scatter graphs for the quantitative relations between the SIDC and the feature 1 and 4 respectively. The horizontal axis of the graphs represents the values of feature 1 and 4, and the vertical axis represents the level of the SIDC. Figure 4 shows that the values of facial feature 1 for subject A exceed $1.0 \times 10^{-4}$ even when his SIDC is 1, which is the lowest value. In contrast, most of the values of facial feature 1 for subject D are no more than $1.0 \times 10^{-4}$ even when his SIDC is 3. Figure 5 shows that the maximum value of terminal feature 4 for the subject A is around 6 when his SIDC is 4. But for the subject C, most of the values of terminal feature 4 exceed 6 when his SIDC is 4. These facts indicate that the quantitative correlations between the SIDC and the features are different by individual learners, even if the features are qualitatively correlated to the SIDC for every learner.

### 3.3 Developing different estimators for each learner using SVM

From the discussion in section 3.2 (3), it becomes clear that the qualitative and quantitative properties of the correlations between the SIDC and the features are different by individual learners. It means that each learner’s SIDC cannot be estimated by only a single criterion or a single estimator for every learner. We need to develop different estimators for individual learners, by modeling the properties of the correlations between each learner’s SIDC and the features using machine learning methods. We first train the different 2-class classifiers for individual learners, each of which can estimate whether each learner’s SIDC are high or low based on the five features listed in Table 1, using some pattern recognition method. Then we estimate each learner’s SIDC using the trained classifiers.

There are many pattern recognition methods for training a classifier: k-Nearest Neighbors, Support Vector Machines (SVM), Neural Networks, Decision Trees, and so on. We use SVM [11] for training the classifiers by considering that it can realize non-linear discrimination boundaries with high generalization capability.

For estimating the learners’ SIDC, input vector $a$ is first defined as five-dimensional vector whose elements are the facial feature 1, 2 and 3 and the terminal feature 4 and 5. At the same time, the class label of the input vector $a$, which is represented as $b$, is defined as follows.

$$b = \begin{cases} 
1 & \text{if SIDC was low when a was obtained} \\
1 & \text{otherwise}
\end{cases}$$

Under these definitions, the discrimination function for estimating each learner’s SIDC, which is represented as $f$, is trained for individual learners. In this training step, the supervised data set for training consists of the pair $(a, b)$ which was obtained when the $b$ was known. In the estimation step, the class label $b$ of given input vector $a$ is estimated as $b = f(a)$.

### 4 Experiment

#### 4.1 Experimental Result

To examine the performance of the proposed method, we conduct an experiment by using the data sets which were gathered for the exploratory experiment described in section 3.2.
In the exploratory experiment described in section 3.2, we divided 18 English word tests into 2 groups, G1 and G2, each of which includes 9 tests. As the result, we have two data sets, each of which is obtained when each learner answered the tests included in group G1 and G2 respectively. Note that each data set consists of the features (the facial feature 1, 2 and 3, and the terminal feature 4 and 5) and the resultant values of the questionnaire. In the experiment described in this section, we trained a classifier using one data set as a training data set, and examined the performance of the trained classifier using the other data set.

Input vector $a$ was defined as five-dimensional vector whose elements are the facial feature 1, 2 and 3, and the terminal feature 4 and 5, as described in section 3.3. Class label $b$ of each input vector $a$ was decided according to the resultant value of the questionnaire for the test which was answered when the $a$ was obtained. In more detail, $b$ was decided $=-1$ if the resultant value of the questionnaire was 1 or 2, and was decided $=1$ if the resultant value was 3 or 4.

We show the result of this experiment in Table 3. This result shows that the learners’ SIDC were estimated with accuracy of 85.8% on an average by the proposed method.

<table>
<thead>
<tr>
<th>Subject</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>97.6%</td>
<td>81.1%</td>
<td>90.8%</td>
<td>73.3%</td>
<td>85.8%</td>
</tr>
</tbody>
</table>

4.2 Discussion

In classroom lectures, lecturers predict the learners’ learning conditions or mental conditions by observing the behaviors of raising hand or facial expressions of the learners for reference how to design the instructions [12]. Ikuta has examined how accurately the lecturers can predict the learner’s learning condition defined as binary variable meaning good or not good in his research [12]. According to the result of the examination, the hitting rate of the lecturers’ prediction is around 75%. It indicates that the lecturers can conduct their lectures smoothly even if they can predict the learners learning conditions or mental conditions with accuracy of only around 75%.

In comparison with this, our proposed method could estimate the learners’ SIDC with at least 73.3% in the experiment described in this section, which is almost same as the accuracy of the lecturers’ prediction. Hence, our proposed method is judged to be an enough useful approach from a practical point of view.

5 Conclusions

In this paper, we proposed the method for estimating e-learners’ SIDC in order for teachers to assess the e-learners’ interest in or willingness for learning. Since the information for estimating SIDC should be obtained independently of learning environments without any wearable devices, we used facial information as well as mouse actions of the learners as the information for estimation. However it is unclear which behaviors on the facial information or mouse actions are useful for estimating the learners’ SIDC.

In this paper, we first assumed that the learners’ behaviors which we humans consider for estimating the learners’ SIDC are also useful for auto-estimation, and specified such behaviors by an exploratory experiment. Based on the result of this experiment, we adopted tilting head, variation of gaze and the frequency of mouse actions as the useful behaviors, and extracted several features from the behaviors. Then we examined how each of the extracted features is correlated with each learner’s SIDC by another exploratory experiment. In the result of this experiment, it turned out that the properties of the correlations between the features and the SIDC are different by individual learners. We therefore developed different estimators for individual learners, differently modeling the properties of the correlations between the features and each learner’s SIDC using SVM. In the result of an experiment, the learners’ SIDC were estimated with accuracy of 85.8% on an average by the estimators developed by the proposed method.

One of the future works is to estimate not only the SIDC but also other mental states of the learners. It is one example how strongly the learners concentrate on learning.

References


