Research and Practice in
Technology Enhanced Learning

RESEARCH Free and Open Access

In-process feedback by detecting deadlock based on EEG data in exercise of learning by problem-posing and its evaluation

Sho Yamamoto [1,2] *, Yuto Tobe [3] , Yoshimasa Tawatsuji [4] and Tsukasa Hirashima [5]

*Correspondence:
yamamoto@info.kindai.ac.jp
Faculty of Informatics,
Kindai University,
3-chōme-4-1 Kowakae,
Higashiosaka, Osaka 577-8502,
Japan
Full list of author information is available at the end of the article

Abstract

Feedback on learning activities is one of the most important issues in achieving adaptive learning. In this study, we propose a mechanism for solving this problem by detecting the deadlock state of a learner during a learning activity and providing feedback to eliminate such a state. Feedback on the products of learning activities (we call it “after-process feedback”) has been implemented in numerous interactive and adaptive learning environments. However, feedback during an activity (we call it “in-process feedback”) has rarely been implemented. In-process feedback is considered to be much better than after-process feedback when learners have difficulty or become frustrated with the learning material during the learning process. The difficulty in implementing in-process feedback lies in the timing and content of the feedback. It has been pointed out that the detection of a deadlock must be achieved as early as possible; otherwise, it reduces the learning motivation of the learner. Therefore, we focused on electroencephalograph (EEG) data, which are difficult to cheat and can clearly detect the state of the learner. By combining EEG data with machine learning, we developed a model for detecting when a learner is stuck, allowing us to detect the timing. After that, we generate the proper feedback by estimating the knowledge state of the learner based on the knowledge structure and task response status. We implemented and evaluated the in-process feedback approach in a learning environment posing arithmetic word problems.

Keywords: Problem-posing, In-process feedback, Knowledge structure, Affective computing, EEG, Wheel-spinning

Introduction

We have continued our research on the development and operation of a learning environment posing arithmetic word problems (Hirashima et al., 2015; Nakano et al., 2000; Yamamoto et al., 2012, 2014). The current version of the learning environment is called

Monsakun, which covers arithmetic word problems that can be solved through a single addition and/or subtraction, a single multiplication and/or division, or a total of four arithmetic operations (Hirashima et al., 2014). The system diagnoses the problem posed by the learner based on the knowledge structure of the arithmetic word problem. We focused on the practical use of this system in elementary schools and, based on the results, confirmed that the system is useful for promoting the acquisition of a knowledge structure. We verified that this effect can be obtained not only by students in a regular classroom but also by students in a special-needs classroom (Yamamoto & Hirashima, 2016).

Several research efforts have focused on learning processes applied in learning environments. Beck et al. pointed out that even if a learning activity is persistent, it may not be equivalent to an activity that contributes to actual learning (Beck & Gong, 2013; Kai et al., 2018), which is a problem called wheel-spinning. Under this situation, although the learner works on a learning activity, the learner becomes stuck in the mastery learning loop without any actual learning occurring. The continuation of this situation will lead to frustration and a loss of motivation for learners (Matsuda et al., 2016; Sedek & Kofta, 1990). Although some learners may persevere in their learning, it is unrealistic to expect this of all learners. Therefore, the early detection of wheel-spinning is an important issue in promoting effective learning in a learning environment.

Monsakun can be used to diagnose the posed problem based on a model of arithmetic word problems and provide feedback to learners. This provides feedback on the results of the exercise. We called this "after-process feedback," which is useful in helping learners learning. However, it has been observed that with this feedback alone, some learners become stuck and repeat the same exercises. The more difficult the task is, the greater the number of such learners. Therefore, even with *Monsakun*, to realize useful learning, it is important to implement feedback not only on the posed problem but also on the process of the problem-posing (e.g., feedback on a deadlock during learning), which we call "in-process feedback." The difficulty in realizing in-process feedback lies in the timing and content of the feedback. The purpose of this research is to develop a function that can detect the deadlock of a learner at the proper time and provide appropriate feedback. To achieve this purpose, it is necessary to detect whether the learner is in a wheel-spinning state from two aspects: a stationary point of "the knowledge state," and the expression of negative emotions.

To detect the occurrence of wheel-spinning, a method that estimates the learner's state based on the learner's response is often used in combination with machine learning, such as Bayesian knowledge tracing (Matsuda et al., 2016; Pelánek, 2017). Various methods have also been proposed to incorporate facial expression recognition and emotion estimation into the learning environment (Graesser et al., 1999). By detecting the emotions of the learners, such approaches promote learning and support teacher intervention from an affective perspective. If wheel-spinning is a situation in which a learner is experiencing difficulty in learning, we believe it will be possible to detect wheel-spinning from this emotional perspective. Owing to the immediacy of deadlock detection and the fact that an electroencephalograph (EEG) is a primitive learner parameter, we used an EEG in building our prototype deadlock detector.

As another unique aspect of our research, we previously conducted a model-based analysis of the problem-posing process (Hasanah et al., 2015; Supianto et al., 2017). Using this mechanism, it is possible to estimate the learner's state of understanding and generate feedback for such understanding based on the knowledge structure and the exercise status of the learner. Thus, detailed feedback regarding a deadlock can be realized based on this estimation. However, only this approach, *Monsakun* cannot detect proper timing for feedback. Therefore, by combining this estimation with EEG-based deadlock detection, we believe that we can achieve our goal of in-process feedback. Thus, in terms of the wheel-spinning problem, there is limited research approaching the problem from the perspectives of not only the detection of negative emotions (e.g., Beck & Rodrigo, 2014; Botelho et al., 2019) but also the detection of the knowledge states of the learner.

To realize in-process feedback, in this study, we developed two functions: an EEG-based deadlock detection function using a simple electroencephalograph, and a feedback generating function that points out the cause of the deadlock based on the problem-posing state and the knowledge structure of arithmetic word problems. In this paper, we report on the development of the *Monsakun Affective* and a preliminary evaluation conducted by university students to examine the feasibility of in-process feedback that combines these two functions. In Section 2, we describe related research and the position taken by the present study. In Section 3, we describe the current version of *Monsakun*, and in Section 4, we describe the design of the in-process feedback. Section 5 introduces the interface of the system used for implementing such feedback. In Section 6, we provide a simple evaluation and the limitations of this study, and finally, in Section 7, we provide some concluding remarks.

Literature review

Learning environment and wheel-spinning

A learning environment provides a place for learners to acquire various skills and knowledge. For example, systems such as Cognitive Tutor, ASSISTments, SQL-Tutor, and KERMIT, which are intelligent tutoring systems (ITSs), are useful for estimating the understanding of learners while providing useful learning interventions (Heffernan et al., 2006; Mitrovic, 2003; Ritter et al., 2007; Suraweera & Mitrovic, 2002). For arithmetic and mathematics problems in particular, systems such as PAT Tutor have been proposed, their main purpose is to promote the understanding of learners through “problem-solving” (Koedinger et al., 1997). Such a system is realized as a type of mastery learning, where the learner acquires the target ability through by repeatedly solving problems. It is therefore important for learners to be able to continue their exercises in an appropriate manner.

By contrast, Kai et al. pointed out that both productive and unproductive exercise continuations occur (Kai et al., 2018). A non-productive continuation is a situation in which the learner is unable to update the target knowledge and is not making progress in learning. At this point, although the learner proceeds with the learning activity itself, the learner has not mastered anything. This situation is defined as wheel-spinning. Under this situation, learners become frustrated and are unable to maintain their motivation to learn. It is therefore important to detect this condition as soon as possible and provide appropriate support for learning (Beck & Rodrigo, 2014). Detecting the rotation of the wheel is often accomplished by using knowledge traces to derive the probability that the learner can solve the problem (Matsuda et al., 2016). In this case, various behaviors of the learner are used as parameter. With this approach, the model is built using data from several problem-solving sessions of the learner. Reducing this number of sessions is one of the challenges to the immediacy of wheel-spinning detection.

In contrast to these studies, *Monsakun*, which we are continuously developing, differs in that it is targeted at the “problem-posing” of the arithmetic word problem (Hirashima et al., 2015; Nakano et al., 2000; Yamamoto et al., 2012, 2014). It has been recognized that learning to pose a problem can improve the ability to solve arithmetic word problems more than problem solving (Polya, 1945; Silver, 1994). However, because the problem space required to pose a problem is large for problem-posing learning, diagnosis is generally realized through an evaluation by peer learners or by targeting multiple-choice problems (Chang et al., 2012; Yu, 2011). By contrast, *Monsakun* requires learners to pose arithmetic word problems by integrating quantitative concepts that have a single meaning. In other words, *Monsakun* realizes problem-posing exercises in units of quantitative concepts to be understood. *Monsakun* can then diagnose the posed problems and generate feedback to the learner by comparing the posed problem to the constraints of the knowledge structure. In other words, *Monsakun* estimates the state of understanding about knowledge structure of the learner for each task and generates feedback. *Monsakun* has been used in a variety of schools and has produced results that promote the learners’ understanding of knowledge structures. However, wheel-spinning has been observed in some learners in the use of this system and therefore needs to be resolved.

Affective computing intelligent tutoring system

Affective computing improves the ability of computers to recognize human emotions and make decisions through emotional information processing (Picard, 2000). These concepts have been incorporated into an intelligent tutoring system, and Ammar et al. proposed an affective tutoring system (Ammar et al., 2010). The system uses a camera to recognize the learner's facial expressions and determine the agent's actions for learning support. Hutt et al. used eye tracking to capture the moments when learners are distracted from a task and use this information to help their learning (Hutt et al., 2016). Graesser et al. used multifaceted physiological data to estimate human emotional states and applied them to inform the ITS behavior (D'Mello et al., 2007; Graesser et al., 2004). In this way, not only the structure of the target knowledge and the training log, but also the emotional aspects are useful for ITS design.

Affective computing not only identifies the facial expressions of the learner through a camera, it also facilitates the acquisition of physiological information through recent advances in measurement technology. Therefore, attempts have been made to estimate the emotional aspects of learners using an electrocardiogram (Alqahtani et al., 2019) and an EEG (Xu et al., 2018). In addition, recent research has attempted to use deep learning algorithms to estimate the emotional state of learners using multidimensional physiological information from multiple devices (Matsui et al., 2019). Such physiological information reflects the emotions of the learners, and we believe that it is possible to extract the involuntary emotional states of the learner. In general, as mentioned in the previous section, in the field of knowledge tracing, the learner's knowledge state is estimated for each (fixed) number of questions, and feedback is provided (Corbett & Anderson, 1994). Therefore, there is a time delay in detecting when a learner has reached a deadlock. By contrast, if we can detect physiological information reflecting the emotional state of a deadlock, we can generate immediate feedback to the learner, which we believe to be important for our research.

Considering the various physiological information available, we focused on an EEG, which is widely used for emotion recognition (Li et al., 2019). In addition, an EEG can be easily acquired through recent technology, and the head is free to move during the measurement in comparison to other methods such as a Functional Magnetic Resonance Imaging (fMRI) or a near-infrared spectroscopy (NIRS) (Miyauchi, 1996). Because a learner who is stuck on a specific exercise may find learning difficult, such a learner may move their head when thinking, making the use of an EEG suitable in this respect. However, regular EEG meters have certain problems, including calibration, which can be difficult for the learner to deal with. In addition, wearing the EEG device may affect the learner's emotional data. To minimize these effects, a simple EEG was used.

Monsakun for learning by problem-posing

Learning by MONSAKUN

A brief description of learning using *Monsakun* is provided in this section. The target domain of this study is a learning environment for problem-posing of arithmetic word problems that can be solved through a single addition or subtraction. This system works on tablets. The software implements three levels based on the understanding of the required knowledge structure, which are described in the next section. The learners start with the easiest levels and work their way up to higher levels.

Figure 1 shows the interface of *Monsakun*, through which the learner poses problems. When posing the problem, the learner is given a “calculation” and a “story” as constraints, as shown in the upper-left part of the figure. As shown on the right side, the learner is given multiple simple sentence cards in order to pose a problem.

The learner can pose a problem by selecting three correct cards from the given simple sentence cards and arranging them in the proper order. When the learner finishes arranging the cards within the black blank area at the center left, the diagnostic button below it becomes active. When the learner taps this button, *Monsakun* diagnoses the problem posed and feeds the results back (Figure 2). This feedback is called after-process feedback. Learners receive the result and deepen their understanding of the knowledge structure by repeating the problem-posing through trial and error.

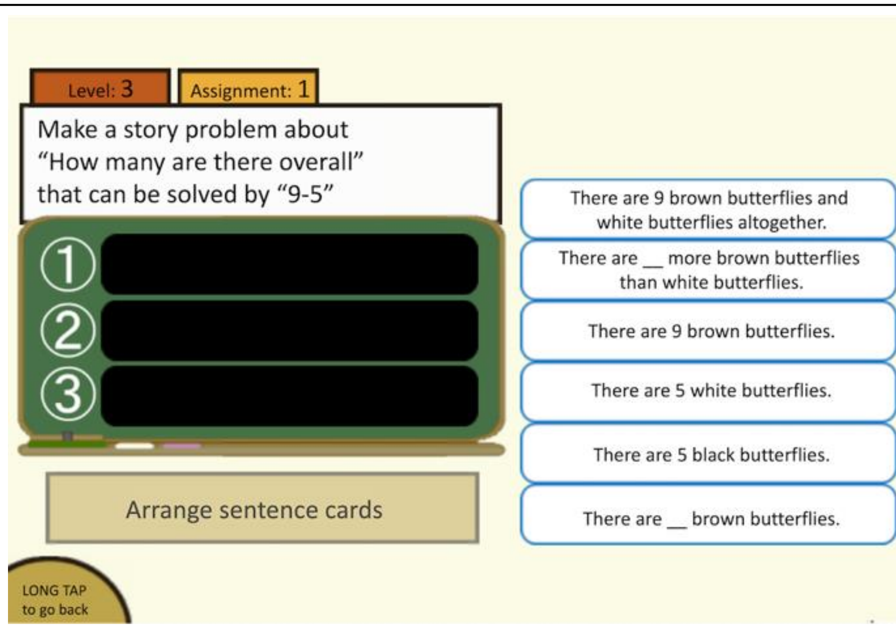

Fig. 1 Interface of *Monsakun* used for problem-posing

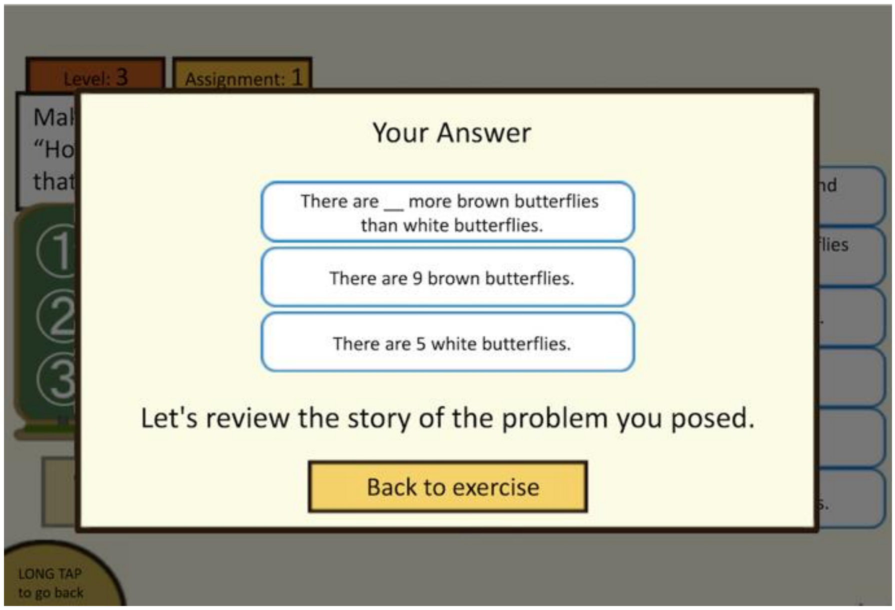

Fig. 2 Interface of feedback window

Knowledge structure of MONSAKUN

Figure 3 shows the knowledge structure of one-step addition and subtraction arithmetic word problems (Hirashima et al., 2014). Arithmetic word problems that can be solved through a single summation/difference consist of three quantitative concepts. Further, one problem is composed of two independent quantitative sentences expressing the existence of the quantitative concept and one relative quantitative sentence expressing the relationship between them. Each quantitative concept is expressed based on the quantity (value), what the object of the quantity (object), and what kind of property it has (predicate). We call a sentence that expresses this single concept of quantity a simple sentence. For example, in the case of “There are two apples,” “two” is the quantity, “apple” is the object, and “there is” is the predicate. This simple sentence is an example of an independent quantitative sentence because “there is” indicates existence.

In addition to the number of independent and relative quantitative sentence, there are two other constraints. The first is object correspondence. For example, “Seven green flowers

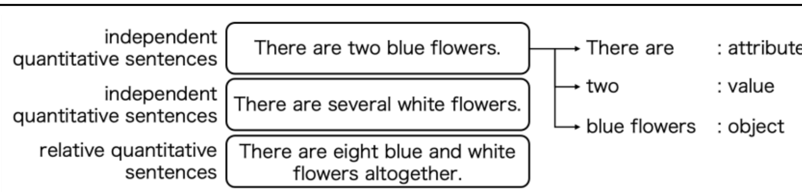

Fig. 3 Triplet structure of arithmetic word problem by solving one-step summation/difference

are blooming. Five green flowers have died. There are several green flowers in bloom.", the object correspondence in the problem is correct. However, if one of the sentences is a "red flower," the problem is not valid by the constraint of object correspondence. The second is the quantity relation. In relation to the above example, the quantity relation is "7 - 5 = ?" is established. However, when the number of the second simple sentence becomes eight, the result of the quantity relation becomes negative, "7 - 8 = ?" the result becomes negative and the quantity relation is not established. There are four story types in a problem that can be solved using one-step addition or subtraction, that is, combining, increasing, decreasing, and comparing. These stories share that the number of independent and relative quantitative sentences must be two and one, respectively. However, the relative quantitative sentence is expressed differently in each story. For example, a "decreasing" story would be "Eight apples are eaten." If it is a "combining" story, it would sound like, "There are eight apples and oranges altogether." Moreover, the conditions for the problem to be valid the object correspondence and quantity relations are different in each story. Therefore, relative quantitative sentences, quantity relationships and object correspondences are defined for each story. Therefore, for each story, relative quantitative sentences, as well as quantity relationships and object correspondences, are defined. These change the expression of the predicates of the relative quantitative concept and the combination of the quantitative concepts.

Level of assignment by knowledge structure

In addition, *Monsakun* has tasks with a difficulty level of 1–3 based on the knowledge structure. Table 1 shows this list. In each assignment, a story and a computation are presented as constraints to be satisfied by the problem to be posed by the learners. The story also has recalled operations; comparing and decreasing evoke subtraction, while combining and increasing are stories that evoke addition. Level design is based on these features. In Level 1, the calculation that recalls the story required in the assignment and

Table 1 Examples of each level design

	Formula for assignment	The Story of the problem	Correct answer example
Level 1	4 + 5 = ?	Increase	There are four apples. Five apples have fallen. There are several apples.
Level 2	4 + ? = 9	Increase	There are four apples. Several apples have fallen. There are nine apples.
Level 3	9 - 5	Increase	There are several apples. Five apples have fallen. There are nine apples.

the calculation given in the assignment are the same. For example, the learner is required to pose an “Increase” story problem that can be solved by calculating “4 + 5 = ?.” This is a story in which the “Increase” story recalls an addition, and the given calculation is also added. Level 2 has the same conditions, but the calculation presented is “4 +? = 9,” where the quantity to be sought is the left-hand side. Level 3 is an assignment in which the calculation of the story and calculation of the mathematical formula are different. For example, the learner is required to pose an “Increase” story problem that can be solved by calculating “9 - 5.” These assignments require a proper understanding of the structure because the calculations recalled by the story and the calculations presented in the assignments are different. Therefore, the difficulty of each task is defined based on the knowledge structure of the arithmetic word problems to be understood. See Hirashima et al. (2014) for details.

After-process feedback generation by knowledge structure

In this section, we describe the current feedback (after-process feedback) based on a knowledge structure. After-process feedback using this knowledge structure is generated based on whether the problem posed by the learner satisfies the constraints of the above structure. For learners, the learning goal is to acquire these constraints in arithmetic word problems that can be solved in a single addition or subtraction. There are a total of five errors that are fed back to the learner. The constraint violations regarding the establishment of the problem, also described in *Monsakun*’s knowledge structure, as follows: “object correspondence,” “quantity relation,” and “number of independent quantitative sentences and relative quantitative sentence.” In addition, because “calculation” and “story” are given as assignments in *Monsakun* as previous research, there is also a constraint violation of “difference in mathematical formula” and “difference in story.” *Monsakun* can diagnose these constraint violations by identifying the quantity relation and story of the posed problems based on *Monsakun*’s knowledge structure. These errors are generated when the learner poses a problem and presses a diagnostic button.

Research question

Through practical use in a variety of schools, we have already confirmed that *Monsakun* can provide useful feedback promoting learner understanding. *Monsakun*, the previous study, is only implements after-process feedback based on the diagnosis of the posed problem. However, we observed that some learners tend to get stuck, particularly at Level 3, at which they stop learning. These learners are stuck in an exercise loop of the same problem through superficial trial and error. For example, they might try to randomly choose different cards for the same problem. Moreover, some learners with a slow learning progress might face difficulty with problem-posing and thus be unable to continue to pose the problem. This type of confusion often occurs not the result of thinking, but the process of thinking. Therefore, it is crucial to provide appropriate feedback at the right time when learners become confused (i.e., are in a negative emotional state), for example, when they face a deadlock. Appropriate feedback provides learners with "feedback based on the identification and resolution of the cause" at "the timing when learners feel stuck during the exercise." Therefore, the following two points that fall under these categories are the research questions for this study.

RQ1. How can *Monsakun* detect when the learner is stuck?

RQ2. How to identify the cause of learner's deadlock on *Monsakun*?

To resolve this issue, we begin with RQ2, which is relevant to the previous research. We previously conducted a model-based analysis of the process of problem-posing, but were unable to generate timely feedback (Supianto et al., 2017). If one card is set up to blank area, a process analysis can be applied. Based on this analysis, each time a learner pulls a card out of a blank area, the possible reason of error at that point can be identified. However, it is unrealistic for the system to provide feedback to the learner at this frequency.

Next, to approach RQ1, we attempted to detect negative emotional states from the EEG data to detect the appropriate timing described above as deadlock. The instrument used to acquire the EEG data is a simple electroencephalograph. Machine learning can estimate outputs from complex inputs; however, it is difficult to interpret. Specifically, it can extract deadlocks; however, it cannot estimate their causes. Therefore, the relationship between identifying the causes of deadlocks and generating feedback to resolve them is using the knowledge structures described in "Knowledge Structure of *Monsakun*" section. Thus, the disadvantages of each method are compensated for by the advantages of each. We call such feedback in-process feedback. This study proposes an in-process feedback mechanism that combines these methods.

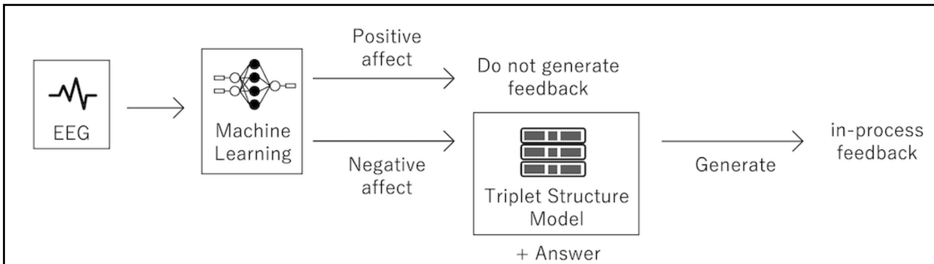

Fig. 4 Overview of in-process feedback

Methodology (design of in-process feedback)

Objective

The developed function is shown in Figure 4. The machine learning model we want to create is a classifier that divides the data obtained from a simple EEG into two states: “the exercise is progressing effectively” and “the exercise is not progressing well (deadlock).” Then, if the exercise is “not progressing well,” the system will provide feedback on the probable cause of the deadlock based on the knowledge structure and the current answer status. In this case, the cause of the deadlock is the constraints that the learner needs to understand. If we can estimate these constraints and help the learner to think about them, we believe that we can realize in-process feedback to solve the deadlock during learning. Therefore, in this section, we describe the design of a deadlock detector and an estimator of its cause.

Design of machine learning for affective detection

Learning data

Applying *Monsakun* to three university students in the engineering field, we measured the learning data using a simple EEG, i.e., a MindWave Mobile 2 by NuroSky, Inc., which provides an SDK for system development. EEG data can be acquired from the MindWave Mobile 2, as well as values of a low alpha wave, low beta wave, low gamma wave, high alpha wave, high beta wave, medium gamma wave, theta wave, and delta wave. We focused on the frequency spectrum, such as alpha, beta waves and so on, assuming that they encode information in the temporal direction as relatively global information, rather than each data point in a fine time interval (i.e., sampling rate).

For the output data, we used the Achievement Emotions Questionnaire (AEQ) proposed by Pekrun et al. (2011). The AEQ is a classification of nine basic emotions during learning. These emotions include “enjoyment,” “hope,” “pride,” “anger,” “anxiety,” “shame,” “hopelessness,” “boredom” and “other emotions.” In addition, the emotional response also converted positive emotions, such as enjoyment, hope, and pride, into “a state in which the exercise is proceeding smoothly.” We also transformed the negative emotions anger, anxiety, shame, hopelessness, and boredom into “a state in which the exercise is not proceeding smoothly (deadlock).” The other emotions are none of the above, and thus they have been removed from the learning data. The actual output value for “The exercise is progressing” is 1, and the actual output value for “feeling stuck” is 0. Therefore, the training data are shown in Table 2 (output values are written as n/p; negative values, as 0; and positive values, as 1).

Table 2 Example of training data for affective estimation on *Monsakun*

Delta	High alpha	High beta	Low alpha	Low beta	Low gamma	Mid gamma	Theta	n/p
207,877	6,426	7,024	12,333	14,276	1,040	1,094	56,468	0
74,278	15,553	2,553	8,419	7,101	3,227	580	18,357	1

In this experiment, the following were explained to the subjects and their consent was obtained before the start of the experiment: (1) EEG data will be obtained, (2) this data will be used to estimate the learner’s deadlock state (or to construct a system for this purpose), and (3) sufficient care will be taken to ensure that EEG data will not be leaked outside the system for use in this system only. In addition, the report of this emotional state was based on the learner’s subjective evaluation, and extreme caution was exercised to avoid forced exposure to the emotional state by the experimenter.

The next section describes the procedure for acquiring training data. The learner wore a MindWave Mobile 2 device and worked on Level 1–3 exercises of *Monsakun* in turn. Learner exercises were recorded on a video. Next, the learner answered which of the nine emotions the learner felt every 10 seconds while watching a video of the exercise. In addition, we asked them to answer whether the emotion was caused by the “exercise,” “software UI (User Interface),” or “others.” If it was caused by “software UI” or “others,” it was deleted from the learning data. Therefore, the above output data are only those caused by the “exercise.”

The data acquired were from the EEG data taken every 1 second, and the response data of the experimental participants were recorded every 10 seconds. In other words, the granularity of the measurement time of the EEG data used for input and the response data used as output data do not match. We therefore interpolated the data for the response data obtained every 10 seconds, assuming that the emotional state of the learner persists until the next time the data are obtained, and matched it with the temporal granularity of the EEG data.

As a result, the number of data points is 572 for Level 1, 557 for Level 2, and 1079 for Level 3. The number of data points for each emotion acquired is also shown in Table 3 for each level. At all levels, the number of positive and negative emotions was approximately the same. In Level 3, there were slightly more negative emotions and more dead ends.

Model generation by deep learning

In this section, we describe the construction of a model for deadlock estimation based on an EEG, created using the learning data described in the previous section. Deep learning was used for machine learning because it was assumed that the activation of human emotions is closely related to the movement shown in the EEG. The learner of machine

Table 3 Number of each emotion in the adopted data

	Positive			Negative				
	Enjoy	Hope	Pride	Anger	Anxiety	Shame	Hopelessness	Boredom
Level 1	100	129	109	30	104	20	80	0
Level 2	89	99	150	20	159	0	40	0
Level 3	106	280	113	130	240	50	160	0

learning was set to 3 hidden layers and 10 nodes for each layer. Thus, it is a three-layer deep neural network. In addition, the dropout rate was set to 20% to prevent an overfitting. Next, the activation function was set to tanh because the activation of human emotions was gradual, and the loss and evaluation functions were set to the mean square error. These settings were used in the experiment. The batch size is a standard value of 32, and the data were divided into 95% training data and 5% test data because few learning data were prepared in this study.

Next, parameters examined for the optimal model construction are described. In this study, we verified the model accuracy by changing the gradient method, number of epochs, and learning rate. Seven gradient methods were examined, i.e., SGD, Adadelta, Adagrad, Adam, Adamax, RMSprop, and Nadam. When the epoch was examined experimentally using each gradient method, an overfitting occurred at 1000 epochs or more; thus, we decided to examine four numbers of epochs: 100, 300, 500, and 800. The learning rates were select from 0.1, 0.05, and 0.01, which were determined experimentally.

The procedure used for building the model is as follows: After using the above learner, the learning rate was first examined by fixing the number of epochs to 300 using each gradient method. Next, using the most accurate learning rate, we examined the model with 100–800 epochs. At this time, the epoch with the highest accuracy in each gradient method was used as the representative value. Finally, the gradient method with the highest accuracy using decided learning rate and epoch was adopted.

The above operations were carried out at each level of 1–3, and a deadlock detection model for each level was created. Table 4 lists the data adopted for each level of 1–3.

Table 4 Results of machine learning in levels 1–3

	Learning rate	Epoch	Gradient method	Accuracy	Loss
Level 1	0.10	500	Adadelta	0.778	0.134
Level 2	0.05	100	RMSprop	0.778	0.183
Level 3	0.01	300	Nadam	0.704	0.152

Design of feedback based on posing problem and knowledge structure

The design of the feedback applied during an exercise using the knowledge structure is described. The system generates feedback when a learning deadlock is detected using the deadlock detection model based on the EEG and machine learning described in the previous section. Therefore, the system should assess the in-process problem rather than the after-process problem. This is based on a model-based analysis of the problem-posing process that has already been implemented (Supianto et al., 2017).

Monsakun can detect a constraint violation based on the triplet structure model if some of the cards are answered by the learners. Table 5 shows the correspondence between this

Table 5 Correspondence between type of answered cards, constraint violations and feedback sentences

Number of answered cards	Kind of card	Given story	Condition	Feedback sentence
1	Relative quantitative sentence	Irrelevant	Difference from given story	Be careful about the type of story.
			Same as given story	You're doing good.
	Independent quantitative sentence	Irrelevant	Set cards not used for assignments	Be careful about the objects shown in story.
			Set cards used for assignments	You're doing good.
	Irrelevant	Irrelevant	Set cards not contained correct value	Be careful about the values shown in story.
			Set cards contained correct value	You're doing good.
2	Two independent quantitative sentences	Combine and Difference	Each card's objects are same	Be careful about the objects shown in story.
			Each card's objects are different	You're doing good.
		Increase and Decrease	Each card's objects are same	You're doing good.
			Each card's objects are different	Be careful about the objects shown in story.
	Relative and independent quantitative sentences	Irrelevant	Relative quantitative sentence is not correct card	Be careful about the type of story.
			Objects of relative and independent quantitative sentences are different	Be careful about the objects shown in story.
			Not applicable (n/a)	You're doing good.
	Irrelevant	Irrelevant	Set cards contained incorrect value	Be careful about the values shown in story.
			Set cards contained correct value	You're doing good.

constraint violation and feedback. Such feedback is only generated when the EEG diagnosis detects a deadlock, and when there are fewer than two cards answered. First, the system checks the number of cards answered by the learner when the EEG detects a deadlock. The type of simple sentence card answered is detected. The system then confirms the story provided in the assignment because the feedback sentence may change depending on the given story. Finally, whether the answered card satisfies the conditions for feedback generation is detected, and the feedback shown in the feedback sentence of Table 5 is generated. For example, suppose one card, which is a relative quantitative sentence, has been answered. If the answered card is different from the given story based on the assignment, the learner will receive feedback that states, "Be careful about the type of story." Multiple feedback may be generated as a result of these estimations.

Feedback based on constraint violations can be considered directive feedback because it encourages the learner to become aware of the constraints being violated. We are currently generating feedback for all constraint violations, which the learner caused. In addition, if the learner is feeling stuck on an exercise but is satisfying the constraints that need to be satisfied, feedback such as "You're doing good" can be generated. This type is called encouraging feedback. These two types of feedback help learners who feel stuck make smooth progress in their learning. Because both types of feedback are generated based on the knowledge structure and its constraints, we believe that they are effective in promoting learning.

Monsakun Affective

System overview

Figure 5 shows an overview of the system. We refer to this system as *Monsakun Affective*. The system is divided into an EEG-based deadlock estimation part and an exercise part. The EEG-based deadlock estimation was developed as a Web API (Application Programming Interface) using Python. Here, a model generated by machine learning is implemented. The exercise part was also developed in Python, with Kivy as the Graphical

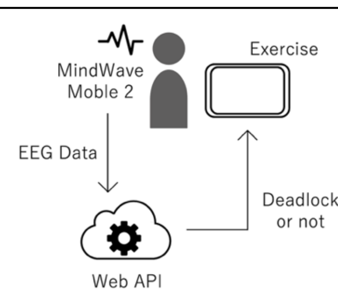

Fig. 5 Framework of *Monsakun Affective*

User Interface (GUI). As the reason for separating the EEG-based deadlock estimation part from the exercise part, MindWave library (NeuroPy) is stopped by an old version of Python, and it is difficult to combine with machine learning models.

Monsakun Affective monitors the brain wave data sent from MindWave Mobile 2 every second and sends the data to the Web API sequentially. The Web API uses the data to provide feedback to *Monsakun Affective* on whether the learner has become stuck. If the learner is diagnosed as such, *Monsakun Affective* will then generate feedback from the learner's answers. If the learner is diagnosed as not being stuck, an empty string is generated. These will be displayed on *Monsakun Affective*.

Several interface and exercise

Here, we describe the exercise procedure for *Monsakun Affective*. First, learners select their grade, pair, and number, and if their username is correct, they enter their password to log in. After logging in, they are presented with the level selection interface shown in Figure 6. In the level selection interface, there is an icon in the lower-right corner of the screen confirming the connection of MindWave Mobile 2. Here, the learner can confirm the connection of a simple EEG. The connection can be checked by connecting the device to the Personal Computer (PC), attaching it, and selecting the appropriate port on *Monsakun Affective*. Once the exercise is ready, learners will work on the assignments in order starting from Level 1. The implemented levels and assignments, which are the same as those introduced in Section 3, are those used to generate models for the machine learning.

After selecting a level, the main interface shown in Figure 7 is displayed. This screen is essentially the same as *Monsakun* described in the above sections. The difference is the

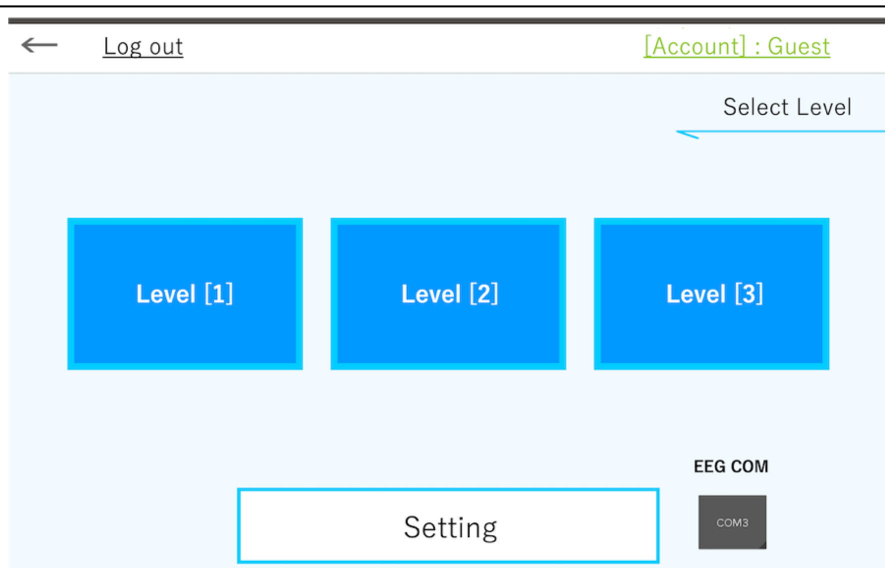

Fig. 6 Interface of level selection

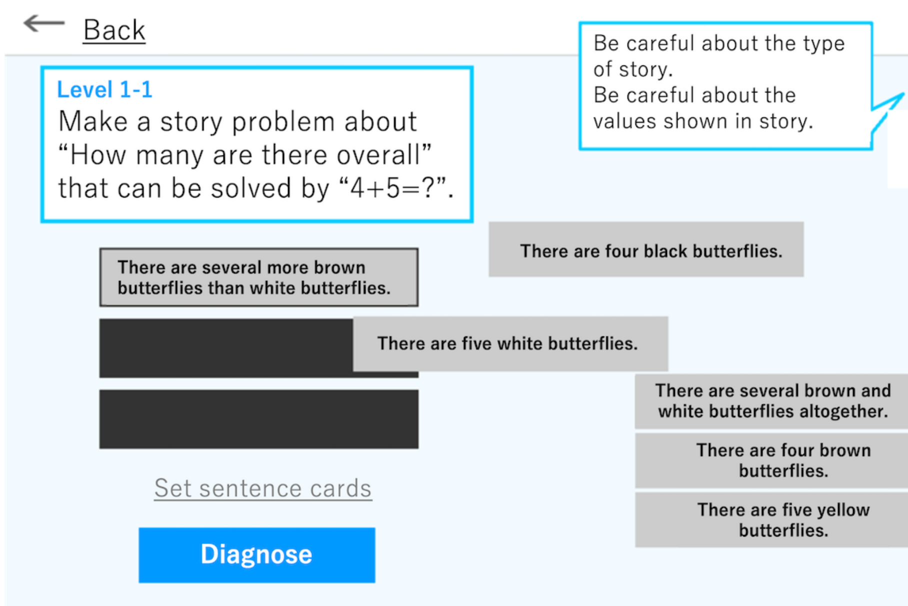

Fig. 7 Interface of *Monsakun Affective* for in-process feedback

callout displayed in the upper-right corner, which is always visible. When the *Monsakun Affective* detects a deadlock, in-process feedback is displayed here. When the learner has solved all assignments of the selected level, the system will automatically return to the level selection interface. In this way, the learner solves the assignments for all levels.

Preliminary evaluation

Procedure

The purpose of this experiment was to confirm whether the in-process feedback was properly generated using the developed prototype system. Specifically, about each research question, the timing of the feedback and the content generated as feedback (i.e., pointers based on the reason for the impasse) require to be evaluated for appropriateness. The subjects were five engineering college students who differed from students who acquired the learning data. The reason we selected university students were chosen as the subjects for this study is that they must have metacognitive abilities to evaluate the content and timing of the feedback, which is the purpose of the evaluation. We have also confirmed that there is no significant difference in the analysis of processes based on knowledge structures between the university students and elementary school students (Hasanah et al., 2015). Therefore, the intended evaluation can be conducted even with university students as subjects.

First, each subject was instructed on how to use *Monsakun Affective* and the experiment procedure. Next, the subject used the developed prototype system to work on the assignment of Levels 1–3. The exercise was recorded on video. Next, the subject was asked to answer whether the timing and content of the feedback implemented were appropriate while watching the video. There were four answers regarding the timing: “appropriate,” “early,” “slow,” and “not necessary.” There are two types of content answers, “appropriate” and “inappropriate.” Finally, the subjects answered the questionnaire.

Evaluation of timing and contents

Table 6 shows various exercise logs of the system. All values are the averages for all subjects by level. If we check the exercise time, Level 3 is overwhelmingly large. The feedback is divided into after-process feedback and in-process feedback. The feedback amount is only for mistakes and does not include the feedback for correct answers. More feedback is given for Level 3 than for Levels 1 and 2. Ten questions are assigned for each level of 1–3. The number of problems posed for Levels 1 and 2 is almost the same as the number of assignments. However, the number of posed problems for Level 3 is approximately three times the number of assignments. The number of steps is the number of times the card is placed in and taken out of the blank area. It takes a minimum of three steps to pose the correct problem. Therefore, the minimum number of steps for each level is 30. The number of steps in which the learner conducts various thinking activities other than giving the correct answer, which is described as the number of search steps. The number of steps was the lowest at Level 2 and highest at Level 3. Therefore, because the subjects were university students, Levels 1 and 2 were relatively easy for them, and the difficulty of Level 3 was appropriate for applying in-process feedback.

Moreover, more in-process feedback was given than after-process feedback. *Monsakun Affective* assessed that a large amount of feedback is required for the in-process timing, in addition to the after-process feedback.

Table 7 provides answers regarding the adequacy of the in-process feedback timing and content. The subjects judged that the timing of the in-process feedback was appropriate approximately 70% of the time. Although there is a possibility of overfitting with machine

Table 6 Logs of exercise using *Monsakun Affective* (N=5)

	Exercise time	Amount of feedback		Number of posed problems	Number of steps	
		After-process	In-process		Total	Exploring steps
Level 1	4m48s	1.4	15.4	11.4	55.8	25.8
Level 2	4m43s	1.0	16.8	11.0	44.6	14.6
Level 3	16m28s	17.4	40.8	27.4	156.6	126.6

Table 7 Result of timing and content of in-process feedback in Levels 1–3

	Timing				Content	
	Appropriate	Early	Slow	Not necessary	Appropriate	Inappropriate
Level 1	0.66	0.14	0.01	0.18	0.87	0.13
Level 2	0.70	0.19	0.00	0.11	0.87	0.13
Level 3	0.77	0.15	0.01	0.07	0.91	0.09

learning, the possibility of an overfitting is low because the subjects are not students who have acquired the learning data, and logs of multiple learners are combined into learning data. Although 10% to 20% of the feedback was deemed "unnecessary," this level decreased to less than 10% as the difficulty increased.

Next, we considered the answers to the feedback content. Approximately 90% of the feedback content was considered appropriate. The feedback deemed the most inappropriate was "Think about the values and objects shown in the story." It is possible that the subject did not consider this to be an error because they merely overlooked the objects and values. It was shown that in-process feedback not only determined that *Monsakun Affective* required it, but that the learner also determined it to be meaningful.

Next, we consider the answers to the feedback content. Approximately 90% of the feedback content was considered appropriate. The feedback that was deemed the most inappropriate was "think about the values and objects shown in the story." It is possible that the subject did not consider it an error because they merely overlooked objects and values. It was shown that in-process feedback not only determined that *Monsakun Affective* needed it, but that the learner also determined it to be meaningful.

Questionnaire

The contents and results of the questionnaire are presented in Figure 8. There are six answers: "strongly agree," "agree," "somewhat agree," "somewhat disagree," "disagree," and "strongly disagree." However, in Q2 to Q4, there are six answers: "Very difficult," "Difficult," "Somewhat difficult," "Somewhat easy," "Easy," and "Very easy." In

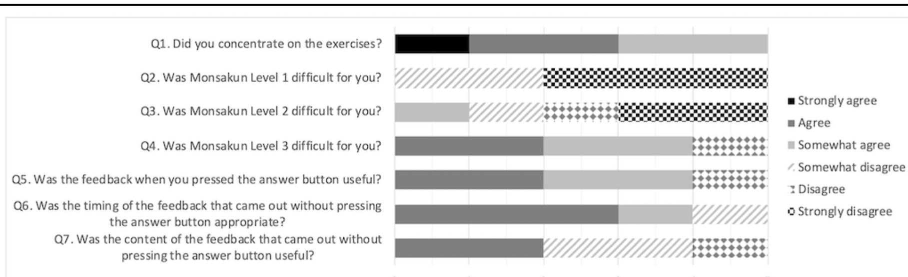

Fig. 8 Contents and results of questionnaire (N = 5)

Figure 8, the graph is drawn by replacing "very difficult" with "strongly agree." Questions Q5 through Q7 also asked the subjects to give reasons for their answers.

First, from the answer to Q1, all learners were able to concentrate on the exercises, and the simple EEG device was not a distraction. For Q2 to Q4, most of the participants answered that Levels 1 and 2 were easy, but Level 3 was difficult, even for college students. This is consistent with the results of the exercise log. In addition, Q5 is a question regarding the usefulness of the after-process feedback, and the participants gave positive feedback. Moreover, Q6 and Q7 are questions regarding the usefulness of in-process feedback. The responses to this question included "I was too focused on the assignment and didn't notice (the feedback) much" and "I did not read (the feedback) much and answered the assignment." These answers indicate that it is difficult to notice in-process feedback. This led to negative answers for Q6 and Q7. We received positive comments regarding the learning process in the system, such as "I was able to gradually approach the answers" and "The information I wanted at the time was concisely displayed." They also suggested the usefulness of encouraging feedback, such as "It gave me confidence because it came up early when I was worried for a moment whether (the learning activity) was actually right." These statements indicate the usefulness of both directive and encouraging feedback.

Discussion

We were able to develop a system that can detect a deadlock in a learner and return feedback by using the knowledge structure and affective detection. Detecting a deadlock based on EEG data using a simple electroencephalograph is considered to be sufficiently practical, with approximately 70% of the results showing an appropriate timing. An overfitting may have occurred as a result of machine learning. However, the learning data used was a combination of data from multiple learners, and the test subjects were different from the learners who collected the learning data. Therefore, we believe that the possibility of an overfitting is low. This model should be verified in future studies. In addition, we considered that the simplification of the model by making the objective variable binary also contributed to the outcome. The objective variable could be simplified because the system has a sufficient knowledge structure and can estimate the error of the learner during an exercise. Thus, even if the system cannot detect detailed emotions, meaningful feedback can be generated that will resolve the deadlock of the learner.

From the exercise log, it was found that the subjects were conducting problem-posing activities for arithmetic word problems through trial and error. In particular, Level 3 is remarkable. In *Monsakun*, a number of steps greater than the shortest number of steps possible indicates exploratory thinking rather than answering correctly. Therefore, all steps can be the target of feedback. Although it is possible to provide feedback through all steps, however, doing so hinders the exercise and is unrealistic. In this experiment, an average of

167 instances of feedback were generated per person for the 30 questions. By contrast, in-process feedback was generated an average of 73 times per person, which is a reduction of approximately 40%. In the other hands, the number of instances of after-process feedback is too small. In this experiment, the average is 19.8 per person. Considering that approximately 70% of the in-process feedback was valid, this means that *Monsakun Affective* generated approximately 30 additional feedback instances needed for smooth learning. Therefore, it is considered that sufficient feedback for smoothly learning is realized.

The above results confirm the feasibility of in-process feedback, which can be realized using machine learning based on simple EEG data to estimate the timing of deadlock in RQ1 and knowledge structure-based estimation of the cause of deadlock in RQ2. Therefore, to contribute to the field, we proposed one of the solutions to wheel-spinning, also mentioned as an issue of the study. This study also uses machine learning and knowledge structures to solve the timing and content of in-process feedback. This method generates feedback for each learner's action in the exercise using knowledge structures. Because this is too much feedback, the feedback generation is narrowed down to an appropriate number of times (timing) using a biometric-based machine learning model. Therefore, we believe that the proposed method of combining symbolic and statistical models to solve learners' deadlock during exercises is different from many wheel-spinning solutions and provides novel findings. This study was a preliminary evaluation of the specific domain of arithmetic word problems. However, since the in-process feedback framework itself is a general framework, similar effects can be expected if the system to which it is applied has a significant knowledge structure.

From a practical standpoint, we also believe that such a framework would contribute to solving the deadlock in online learning that has become popular due to COVID-19. It is sometimes difficult to ask questions promptly to a teacher online, and teachers also take time to grasp the learner's learning situation. Under these circumstances, in-process feedback can care for more learners than after-process feedback alone. This method can be used in actual classes to provide effective learning for learners who can learn to some extent but do not fully understand. However, since it requires a simple electroencephalograph, further consideration of a device is required. This point is also described in the limitations.

Limitation

The limitations of this study are described from two aspects: the limitations of the proposed method and those of its preliminary evaluation. Because this study proposes a method called the in-process feedback, the following limitations should be addressed in the future, and we believe that the significance of this study didn't detract from.

As for the limitation of the proposed method, this method assumes a learning environment that generates feedback based on the knowledge structure of the learning target. Therefore, for a learning environment such as a simple multiple-choice question, this method may not be entirely successful because the causes of the learner’s errors cannot be extracted. However, we believe that the in-process feedback framework, i.e., estimation of deadlock from non-verbal information by machine learning and estimation of its error causes using knowledge structures, has generality.

Examining the detailed quality of feedback is also an important issue. Although we have confirmed that useful feedback can be generated to facilitate learning, the breakdown of its effectiveness is unknown. There are two types of in-process feedback: those pointing out violations of the constraints and those pointing out the satisfaction of the constraints. The former points out errors, whereas the latter points out the correct answer, which encourages the learner. Therefore, there may be differences in their effects. It is also important to consider the difference in learning gains between the after-process and in-process feedback.

Another limitation is the accuracy of the deadlock detection model using machine learning. We must consider the small number of training data and their representativeness and comparability. In addition, a more rigorous examination of the distribution of the acquired EEG data and of the distribution of each mental state is needed in order to more scientifically examine the relationship between physiological data and mental states. About the collection of training data, it is difficult to verify that the introspection report at the time of acquisition of the learning data was correct (D’Mello et al., 2007) because, in the case of this study, it takes time from the end of the exercise to the introspection report to be created. Learners must recall their feelings during the exercise. Moreover, it is also necessary to understand that learners occasionally do not report honestly about their affective states.

Furthermore, in this study, we experimentally adjusted the hyperparameters of machine learning. By contrast, Young et al. proposed a method using a genetic algorithm to adjust the hyperparameters of a convolutional neural network consisting of three layers. We plan to use this method to determine the appropriate hyperparameters in the future (Young et al., 2015). Comparison with other methods such as long short-term memory (LSTM) may be required.

The interface presenting the in-process feedback is inappropriate. It is therefore necessary to consider how to present new feedback.

Thereafter, we discuss the limitations of the preliminary evaluation results. Because this study examines the realization of in-process feedback, we believe that the results described above were achieved among university students who could evaluate their activities, indicating this method’s validity. However, the small number of subjects did not guarantee the generality of the results, and future studies with more subjects are needed. In addition, the application to elementary school students, based on previous research, needs to be addressed in future studies, considering ethical issues because the machine learning model (i.e., obtaining training data from students) must be constructed again.

Finally, the in-process feedback proposed in this study has a variety of possible applications. It may be possible to combine it with Bayesian knowledge tracing and deep knowledge tracing using the parameters of the various response situations of learners to achieve more flexible and rich support. In addition, a simple electroencephalograph could also be considered to use camera images can also be considered in place of a simple EEG device. This is because there is a precedent for estimating emotional states by facial expression estimation using camera images (Calvo & D'Mello, 2010). Such an approach can lead to immediate use in the field, as chromebooks and other devices have become widely available due to COVID-19. It is also conceivable that a framework similar to this system could be realized in various domains by using different knowledge structures.

Conclusion and future works

We studied the realization and verification of in-process feedback used to resolve deadlocks during the learning process through problem-posing. This feedback function detects a deadlock in a learner based on emotion detection using brain wave data and machine learning and identifies the cause of the deadlock based on the knowledge structure and state of the problem-posing.

The existing system realizes the assessment and feedback of a posed problem based on the knowledge structure. This is feedback on the posed problem and can be said to be a post-process feedback, which is useful for learners to modify their errors. However, when using such a system, if the learner becomes confused while working on an exercise, this feedback will not work properly.

Therefore, we aimed to develop in-process feedback for learners who find themselves in a deadlock during an exercise. Previously, we were able to analysis the problem-posing process, and thus generating feedback at the right time is a significant challenge. We used EEG-based affect detection to address this challenge. The accuracy of the developed deadlock detection system was approximately 70%, and the appropriateness of the feedback sentence was approximately 90%. Overall, the accuracy was approximately 70%. This result shows that in-process feedback can be effectively realized to resolve the deadlock of a learner. Hence, we believe that the feasibility of in-process feedback has been demonstrated by estimating when deadlock occurs by machine learning based on simple EEG data in RQ1 and by estimating the cause of deadlock based on knowledge structures in RQ2. We believe that this will also provide insight into solving the wheel-spinning problem in an ITS.

However, we must consider how to provide such feedback. A learner pointed out that it was difficult to notice the suggested feedback during the exercise. Furthermore, development of machine learning models should also be considered. For example, it is necessary to consider the accuracy of the introspection report of learners when acquiring learning data. We also plan to improve the values of the hyperparameters. In future work, we plan to verify (1) the difference in learning gain between in-process and after-process feedback and (2) the difference in the effect of each feedback when the EEG is replaced with another device.

Abbreviations

EEG: Electroencephalograph; ITS: Intelligent Tutoring System; fMRI: Functional Magnetic Resonance Imaging; NIRS: Near-infrared Spectroscopy; AEQ: Achievement Emotions Questionnaire; UI: User Interface; n/a: Not Applicable; GUI: Graphical User Interface; PC: Personal Computer; API: Application Programming Interface; LSTM: Long Short-Term Memory.

Authors' contributions

Sho Yamamoto and Yoshimasa Tawatsuji drafted the manuscript. Sho Yamamoto and Yuto Tobe developed *Monsakun Affective* and conducted case study. Tsukasa Hirashima provided insights and reviewed the manuscript. Sho Yamamoto acquired funding for the research. The authors read and approved the final manuscript.

Authors' information

Sho Yamamoto, Ph.D, is a full Lecturer of Kindai University, Japan. He is also affiliated with the Cyber Informatics Research Institute at Kindai University. His research interests include learning/learner experience design, knowledge-based systems, learning engineering, and intelligent learning support systems.
Yuto Tobe is an undergraduate student of Kindai University, Japan. He trained in learning engineering methodology at Kindai University, at Yamamoto Laboratory.
Yoshimasa Tawatsuji, Ph.D, is a Lecturer of Waseda University, Japan. His research interests include human-agent interaction, model of others, neural mechanism of uncanny valley, and emotions.
Tsukasa Hirashima, Ph.D, is a full professor of Hiroshima University, Japan. His research interests include knowledge-based modeling of learning activities and design & development of learning support environments.

Funding

The work is supported in part by JSPS KAKENHI Grant Numbers 20K03192.

Availability of data and materials

Not applicable.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Faculty of Informatics, Kindai University, Osaka, Japan.
[2] Cyber Informatics Research Institute, Kindai University, Osaka, Japan.
[3] Faculty of Engineering, Kindai University, Hiroshima, Japan.
[4] Center for Data Science, Waseda University, Tokyo, Japan.
[5] Graduate School of Advanced Science and Engineering, Hiroshima University, Hiroshima, Japan.

Received: 23 January 2022 Accepted: 18 October 2022
Published: 28 February 2023 (Online First: 28 Deccember 2022)

References

Alqahtani, F., Katsigiannis, S., & Ramzan, N. (2019, August). ECG-based affective computing for difficulty level prediction in intelligent tutoring systems. In *2019 UK/China Emerging Technologies (UCET)* (pp. 1–4). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/UCET.2019.8881872

Ammar, M. B., Neji, M., Alimi, A. M., & Gouardères, G. (2010). The affective tutoring system. *Expert Systems with Applications*, *37*(4), 3013–3023. https://doi.org/10.1016/j.eswa.2009.09.031

Beck, J., & Rodrigo, M. M. T. (2014, June). Understanding wheel spinning in the context of affective factors. In S. Trausan-Matu, K. E. Boyer, M. Crosby & K. Panourgia (Eds.), *Intelligent Tutoring Systems. ITS 2014. Lecture Notes in Computer Science, vol 8474* (pp. 162–167). Springer, Cham. https://doi.org/10.1007/978-3-319-07221-0_20

Beck, J. E., & Gong, Y. (2013, July). Wheel-spinning: Students who fail to master a skill. In H. C. Lane, K. Yacef, J. Mostow & P. Pavlik (Eds.), *Artificial Intelligence in Education. AIED 2013. Lecture Notes in Computer Science, vol 7926* (pp. 431–440). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39112-5_44

Botelho, A. F., Varatharaj, A., Patikorn, T., Doherty, D., Adjei, S. A., & Beck, J. E. (2019). Developing early detectors of student attrition and wheel spinning using deep learning. *IEEE Transactions on Learning Technologies*, *12*(2), 158–170. https://doi.org/10.1109/TLT.2019.2912162

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, *1*(1), 18–37. https://doi.org/10.1109/T-AFFC.2010.1

Chang, K. E., Wu, L. J., Weng, S. E., & Sung, Y. T. (2012). Embedding game-based problem-solving phase into problem-posing system for mathematics learning. *Computers & Education*, *58*(2), 775–786. https://doi.org/10.1016/j.compedu.2011.10.002

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, *4*(4), 253–278. https://doi.org/10.1007/BF01099821

D'Mello, S., Picard, R. W., & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, *22*(4), 53–61. https://doi.org/10.1109/MIS.2007.79

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 180–192. https://doi.org/10.3758/BF03195563

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & Tutoring Research Group. (1999). AutoTutor: A simulation of a human tutor. *Cognitive Systems Research*, *1*(1), 35–51. https://doi.org/10.1016/S1389-0417(99)00005-4

Hasanah, N., Hayashi, Y., & Hirashima, T. (2015). Analysis of problem-posing process of arithmetical word problem as sentence integration: Viewpoint of first selected sentence. In G. Chen, V. Kumar, Kinshuk, R. Huang & S. C. Kong (Eds.), *Emerging Issues in Smart Learning. Lecture Notes in Educational Technology* (pp. 85–88). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44188-6_11

Heffernan, N. T., Turner, T. E., Lourenco, A. L., Macasek, M. A., Nuzzo-Jones, G., & Koedinger, K. R. (2006). The ASSISTment Builder: Towards an analysis of cost effectiveness of ITS creation. In G. Sutcliffe & R. Goebel (Eds.), *Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference* (pp. 515–520). The AAAI Press.

Hirashima, T., Hayashi, Y., Yamamoto, S., & Maeda, K. (2015, November). Bridging model between problem and solution representations in arithmetic/mathematics word problem. In H. Ogata, W. Q. Chen, S. C. Kong & F. Qiu (Eds.), *Proceedings of the 23rd International Conference on Computers in Education ICCE 2015* (pp. 9–18). Asia-Pacific Society for Computers in Education.

Hirashima, T., Yamamoto, S., & Hayashi, Y. (2014, June). Triplet structure model of arithmetical word problems for learning by problem-posing. In S. Yamamoto (Ed.), *Human Interface and the Management of Information. Information and Knowledge in Applications and Services. HIMI 2014. Lecture Notes in Computer Science, vol 8522* (pp. 42–50). Springer, Cham. https://doi.org/10.1007/978-3-319-07863-2_5

Hutt, S., Mills, C., White, S., Donnelly, P. J., & D'Mello, S. K. (2016). The eyes have it: Gaze-based detection of mind wandering during learning with an intelligent tutoring system. In T. Barnes, M. Chi & M. Feng (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 86–93). International Educational Data Mining Society.

Kai, S., Almeda, M. V., Baker, R. S., Heffernan, C., & Heffernan, N. (2018). Decision tree modeling of wheel-spinning and productive persistence in skill builders. *Journal of Educational Data Mining*, *10*(1), 36–71. https://doi.org/10.5281/zenodo.3344810

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8*(1), 30–43.

Li, T. M., Chao, H. C., & Zhang, J. (2019). Emotion classification based on brain wave: A survey. *Human-centric Computing and Information Sciences*, *9*(1), 1–17. https://doi.org/10.1186/s13673-019-0201-x

Matsuda, N., Chandrasekaran, S., & Stamper, J. C. (2016). How quickly can wheel spinning be detected?. In T. Barnes, M. Chi & M. Feng (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 607–608). International Educational Data Mining Society.

Matsui, T., Tawatsuji, Y., Fang, S., & Uno, T. (2019, June). Conceptualization of IMS that estimates learners' mental states from learners' physiological information using deep neural network algorithm. In A. Coy, Y. Hayashi & M. Chang (Eds.), Intelligent Tutoring Systems. ITS 2019. Lecture Notes in Computer Science, vol 11528 (pp. 63–71). Springer, Cham. https://doi.org/10.1007/978-3-030-22244-4_9

Mitrovic, A. (2003). An intelligent SQL tutor on the web. *International Journal of Artificial Intelligence in Education*, *13*(2-4), 173–197.

Miyauchi, S. (1996). Non-invasive study of human brain function and psychophysiology. *Japanese Psychological Review* (in Japanese), *56*(3), 414–454. https://doi.org/10.5674/jjppp1983.15.11

Nakano, A., Hirashima, T., & Takeuchi, A. (2000). A learning environment for problem posing in simple arithmetical word problem. In *Proceedings of the 8th International Conference on Computers in Education* (pp. 91–98). Asia-Pacific Society for Computers in Education.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pelánek, R. (2017). Bayesian knowledge tracing, logistic models, and beyond: An overview of learner modeling techniques. *User Modeling and User-Adapted Interaction*, *27*(3), 313–350. https://doi.org/10.1007/s11257-017-9193-2

Picard, R. W. (2000). *Affective computing*. MIT press. https://doi.org/10.7551/mitpress/1140.001.0001

Polya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton University Press. https://doi.org/10.2307/j.ctvc773pk

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, *14*(2), 249–255. https://doi.org/10.3758/BF03194060

Sedek, G., & Kofta, M. (1990). When cognitive exertion does not yield cognitive gain: Toward an informational explanation of learned helplessness. *Journal of Personality and Social Psychology*, *58*(4), 729–743. https://doi.org/10.1037/0022-3514.58.4.729

Silver, E. A. (1994). On mathematical problem posing. *For the Learning of Mathematics*, *14*(1), 19–28.

Supianto, A. A., Hayashi, Y., & Hirashima, T. (2017). Model-based analysis of thinking in problem posing as sentence integration focused on violation of the constraints. *Research and Practice in Technology Enhanced Learning*, *12*(1), 1–21. https://doi.org/10.1186/s41039-017-0057-5

Suraweera, P., & Mitrovic, A. (2002, June). KERMIT: A constraint-based tutor for database modeling. In S. A. Cerri, G. Gouardères & F. Paraguaçu (Eds.), *Intelligent Tutoring Systems. ITS 2002. Lecture Notes in Computer Science, vol 2363* (pp. 377–387). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-47987-2_41

Xu, T., Zhou, Y., Wang, Z., & Peng, Y. (2018). Learning emotions EEG-based recognition and brain activity: A survey study on BCI for intelligent tutoring system. *Procedia Computer Science*, *130*, 376–382. https://doi.org/10.1016/j.procs.2018.04.056

Yamamoto, S., & Hirashima, T. (2016). A case study of interactive environment for learning by problem-posing in special classroom at junior high school. In W. Chen, J.-C. Yang, S. Murthy, S. L. Wong & S. Iyer (Eds.), *Proceedings of the 24th International Conference on Computers in Education ICCE 2016* (pp. 282–287). Asia-Pacific Society for Computers in Education.

Yamamoto, S., Akao, Y., Murotsu, M., Kanbe, T., Yoshida, Y., Maeda, K., Hayashi, Y., & Hirashima, T. (2014). Interactive environment for learning by problem-posing of arithmetic word problems solved by one-step multiplication and division. In C.-C. Liu, H. Ogata, S.C. Kong & A. Kashihara (Eds.), *Proceedings of the 22nd International Conference on Computers in Education ICCE 2014* (pp. 89–94). Asia-Pacific Society for Computers in Education.

Yamamoto, S., Kanbe, T., Yoshida, Y., Maeda, K., & Hirashima, T. (2012, November). A case study of learning by problem-posing in introductory phase of arithmetic word problems. In G. Biswas et al. (Eds.), *Proceedings of the 20th International Conference on Computers in Education* (pp. 25–32). Asia-Pacific Society for Computers in Education.

Young, S. R., Rose, D. C., Karnowski, T. P., Lim, S. H., & Patton, R. M. (2015, November). Optimizing deep learning hyper-parameters through an evolutionary algorithm. In *Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments MLHPC '15* (pp. 1–5). Association for Computing Machinery. https://doi.org/10.1145/2834892.2834896

Yu, F. Y. (2011). Multiple peer-assessment modes to augment online student question-generation processes. *Computers & Education*, *56*(2), 484–494. https://doi.org/10.1016/j.compedu.2010.08.025

Publisher's Note