Rodrigo, M. M. T. et al. (Eds.) (2021). Proceedings of the 29th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

In-process Feedback by Detecting Deadlock based on EEG Data in Exercise of Learning by Problem-posing

Sho YAMAMOTO[a*], Yuto TOBE[a], Yoshimasa TAWATSUJI[b] & Tsukasa HIRASHIMA[c]
[a]*Faculty of Engineering, Kindai University, Japan*
[b]*Global Education Center, Waseda University, Japan*
[c]*Graduate School of Advanced Science and Engineering, Hiroshima University, Japan*
*yamamoto@hiro.kindai.ac.jp

Abstract: Giving feedback to learning activities is one of the most important issues so as to realize adaptive learning. Feedback for the product of the activity (we call it "after-process feedback") has previously been implemented in many interactive and adaptive learning environments. However, feedback during the activity (we call it "in-process feedback") has been hardly implemented. When a learner gets stuck or frustrated during some stage of the process, in-process feedback is much better than after-process feedback. The difficulty in realizing in-process feedback lies in the timing and content of the feedback. To solve this, we developed and implemented affect detection based on EEG data for deciding the timing of the feedback, and knowledge state estimation based on knowledge structure for the content of the feedback. Furthermore, in this study, we realize and evaluate the in-process feedback by detecting deadlocks based on EEG data for learning through problem-posing.

Keywords: Problem-posing, in-process feedback, EEG, knowledge structure, wheel-spinning problem

1. Introduction

We have continued our research to develop and operate a learning environment for problem-posing for arithmetic word problems (Nakano et al., 2000; Yamamoto et al., 2012). The current version of the learning environment is called *Monsakun*, which covers arithmetic word problems that can be solved by one addition and subtraction, those that can be solved by one multiplication and division, and those that use 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 worked on the practical use of this system in elementary school, and based on the results of this practice, we confirmed that learning by 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).

Monsakun can be used to diagnose the posed problem based on a model of arithmetic word problems and give feedback to learners. This is feedback on the results of an exercise. We called this type of feedback "after-process feedback." However, this feedback alone is not enough for learners to grasp and modify their errors. This problem is known in intelligent tutoring system (ITS) research as the wheel-spinning problem. In the wheel-spinning problem (1) learners get stuck in the mastery learning loop without any learning occurring, and (2) learners become frustrated when they are repeatedly presented with problems that they obviously cannot solve (Beck & Gong, 2013). Therefore, even in *Monsakun*, to realize useful learning, it is important to realize not only feedback on the posed problem but also feedback on the process of problem-posing (i.e., feedback for the deadlock during learning). We call this feedback "in-process feedback." For this purpose, it is necessary to detect whether the learner is in a wheel-spinning state from two aspects: a stationary point or a cyclic transition of "the knowledge state," and the expression of negative emotions.

We previously performed a model-based analysis of the problem-posing process (Supianto et al, 2017). If the system can generate feedback based on this analysis at the time when the learner's deadlock is detected, it is critical feedback for the learner to resolve the deadlock. Detecting negative emotional states toward an exercise is useful for detecting a deadlock. For detecting (negative) emotional states, attempts have been made to use emotion estimation in learning environments owing to the spread of machine learning and inexpensive measuring devices (Ammar et al, 2010; Alqahtani et al, 2019). For example, Auto Tutor estimates the learner's emotions from the learner's facial expression and uses them as feedback (Graesser et al, 2004; Hussain et al, 2011). However, 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) and Botelho et al. (2019)) but also the detection of the knowledge states of the learner.

Therefore, in this study, to realize in-process feedback, 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 problem. In Section 2, we described the current version of *Monsakun*, and in Section 3, we describe the design of the in-process feedback. Section 4 introduces the interface of the system used for implementing the in-process feedback. In Section 5, we described simple evaluations and limitations, and in Section 6, we provide some concluding remarks.

2. Learning Environment based on Knowledge Structure "MONSAKUN"

2.1 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 by arithmetic word problems that can be solved through one addition and subtraction. This system works on tablets.

Figure 1 shows the interface of *Monsakun*, upon on which the learner poses problems. The learner is given a "calculation" and a "story" as constraints when posing a problem, as shown in the upper left part of the figure. On the right side, the learner is given multiple simple sentence cards 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 on the left, the diagnostic button below it becomes active. When the learner taps this button, *Monsakun* diagnoses the problem posed and feeds the results back. This feedback is the after-process feedback. Learners receive this result and deepen their understanding of the knowledge structure by repeating the problem-posing through trial and error.

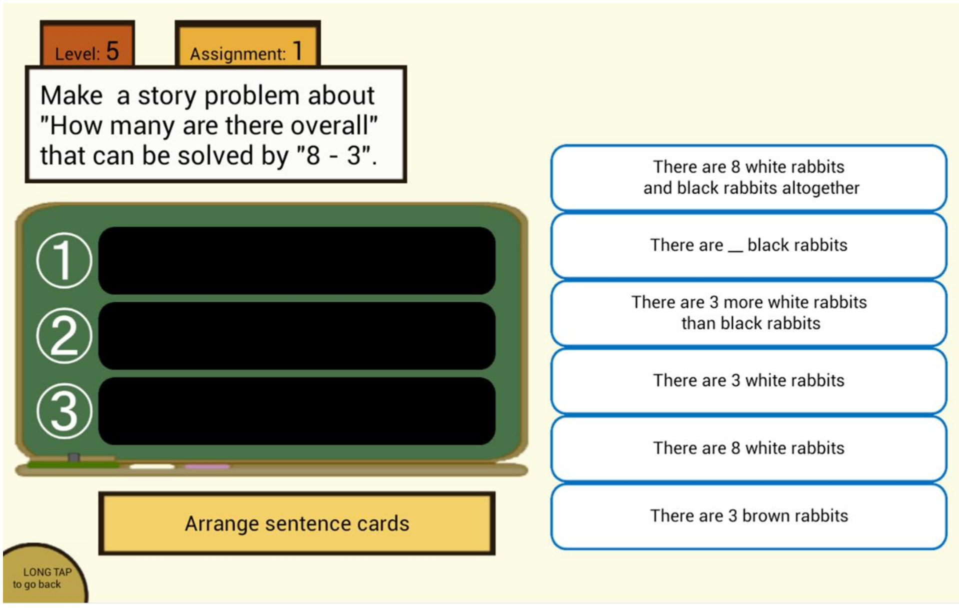

Figure 1. Interface of *Monsakun* for Problem-posing.

2.2 *Feedback Generation by Knowledge Structure (After-process feedback)*

In this section, we describe the current feedback based on a knowledge structure. Figure 2 shows a knowledge structure of 1-step addition and subtraction arithmetic word problems (Hirashima et al., 2014). Arithmetic word problems that can be solved with a single summation/difference consist of three quantitative concepts. Further, one problem is composed of two independent quantity sentences expressing the existence of the quantitative concept and one relative quantity sentence expressing the relationship between them. Each quantitative concept is expressed based on the quantity (value), what quantity the quantity is (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 existence sentence as “there is” indicates existence.

There are four story types in a problem that can be solved using 1-step addition or subtraction, i.e., combine, increase, decrease, and comparison. Furthermore, combine and increase are stories that recall addition, and decrease and comparison are stories that recall subtraction. These change the expression of the predicates of the relative quantity concept and combination of the quantity concepts.

In addition, Monsakun has level 1–3 tasks, each of which has a difficulty level set based on the knowledge structure. In level 1, the calculation that recalls by the story required in the assignment and the calculation given in the assignment are same. For example, the learner are required to pose a “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 addition. Level 2 has the same condition, although the calculation includes the given value on the left side, such as “4 +? = 9.” Level 3 is an assignment in which the calculation of the story and the calculation of the mathematical formula are different. For example, the learner is required to pose a “Increase” story problem that can be solved by calculating "9–5." See Hirashima et al. (2014) for details.

Finally, feedback using this knowledge structure is generated based on whether the problem posed by the learner satisfies the constraints of the above structure. There are a total of five errors that are fed back to the learner. The constraint violations regarding the establishment of the problem are as follows: “object correspondence,” “quantity correspondence,” and “number of independent quantity sentences and relative quantity sentence.” In addition, because “calculation” and “story” are given as assignments in *Monsakun*, there is also a constraint violation of “difference in mathematical formula” and “difference in story.” These errors are generated when the learner poses a problem and presses the diagnostic button.

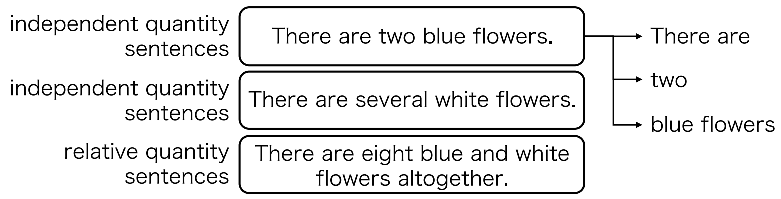

Figure 2. Triplet Structure of Arithmetic Word Problem by Solving 1-Step Summation/Difference.

2.3 *Research Question*

We have already confirmed that *Monsakun* can provide useful feedback that improves learner understanding. This system only implements after-process feedback based on the diagnosis of the posed problem. However, 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 during the thought process. Therefore, it is crucial to give appropriate feedback at the right time when learners are confused (i.e., negative emotional state), for example, when they face a deadlock.

We previously performed a model-based analysis of the problem-posing process but were unable to generate timely feedback. Process analysis can be performed if even one card is set. Based on this analysis, the assessment can be made every time the learner inserts or removes the card from the blank. However, it is not realistic for the system to provide feedback to the learner at this frequency.

Therefore, we developed in-process feedback, which helps detect negative emotional states based on EEG data and uses its states as deadlocks to generate feedback based on a process analysis.

This is realized using a simple electroencephalograph. After detecting a deadlock, the system generates feedback based on the knowledge structure and the problem-posing situation of the learner. We believe this system can use affect detection to generate feedback that will allow learners to overcome a deadlock.

3. Feedback Design

3.1 Design of Affective Detection

A number of studies on Intelligent Tutoring Systems have considered human emotional states with multifaceted physiological data (Graesser et al, 2004; D'Mello et al, 2007), including an Affective Tutoring System (Ammar et al, 2010). The goal of these studies is to estimate the emotional states of learners based on externally acquired information (e.g., seat pressure and facial expressions from cameras). In recent years, with the advancement of measurement technology, it has become easier to obtain physiological information, and attempts have been made to estimate the emotional aspects of learners using electrocardiography (Alqahtani et al, 2019) and an EEG (Xu et al, 2018). Recent research has attempted to estimate the emotional state of learners from their multifaceted physiological information using a deep learning algorithm (Matsui et al, 2019).

In this section, we describe the development of a model that uses EEGs to detect a deadlock of a learner. The device used was MindWave Mobile 2 manufactured by NeuroSky Inc. The reason for using electroencephalography is that, in general, it allows more freedom of head movement during the measurement than functional magnetic resonance imaging (fMRI) or near-infrared spectroscopy (NIRS) (Miyauchi, 2013). In particular, learners in deadlock may cause head movements (by thinking), and the use of electroencephalography is suitable in this respect. In addition, ordinary electroencephalographs have some issues such as calibration, which imposes a burden on the learner. Wearing the device can affect the learner's affective data. Therefore, in order to minimize these effects, we used a simple electroencephalograph. The acquisition of training data by MindWave Mobile 2 and the model construction by a d-CNN are described below.

3.1.1 Learning Data

We measured the learning data using a simple electroencephalograph when applying *Monsakun* for three university students in the engineering field. Raw data can be acquired from MindWave Mobile 2 as well as values of a low α wave, low β wave, low γ wave, high α wave, high β wave, medium γ wave, θ wave, and Δ wave. We focused on the frequency spectrum, such as alpha and beta waves, assuming that they encode information in the temporal direction as relatively global information, rather than each data in a fine time interval (i.e. sampling rate). The output data included nine emotions: enjoy, hope, pride, anger, anxiety, shame, hopelessness, boredom, and the last one based on the AEQ proposed by Pekrun et al. (2011). The learner answered what kind of feelings they experienced during the exercise from among these nine emotions.

The learner wore a MindWave Mobile 2 device and worked on level 1–3 exercises of *Monsakun* in turn. The learner exercises were recorded on video. Next, the learner answered which of the nine emotions that were felt every 10 s while watching a video of the exercise. In addition, we asked them to answer whether the emotion was caused by the “exercise,” “software UI ,” or “others.” If it was caused by “software UI,” and “others,” it was deleted from the training data.

In addition, the emotional response also converted the positive emotions enjoy, 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)” Therefore, the actual output values are 1 for “the state in which the exercise is proceeding smoothly” and 0 for “deadlock” This is listed as "n/p" in the table. Therefore, the learning data are as shown in Table 1. The number of data are 572 for level 1, 557 for level 2, and 1079 for level 3.

Table 1. *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

3.1.2 Model Generation by Deep Learning (3-CNN)

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 as the machine learning because it was assumed that the activation of human emotions used is closely related to the movement shown in the EEG. The learner of machine learning was set to 3 hidden layers and 10 nodes for each layer. In addition, the dropout rate was set to 20% to prevent overfitting. Next, the activation function was set to tanh because the activation of human emotions is gradual, and the loss and evaluation functions were set to the mean square error. These settings are experimental. 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 this time.

Next, the data 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: 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 rate was set to 0.1, 0.05, and 0.01. These parameters were experimental.

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 epoch with the highest accuracy among each gradient method was adopted.

The above operations were carried out at each level of 1–3, and deadlock detection model for each level were created. Table 2 shows the data adopted for each level of 1–3.

Table 2. *Results of Machine Learning in Levels 1–3*

	Learning Rate	Epoch	Gradient Method	Accuracy	Loss
Level 1	0.1	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

3.2 Design of Feedback based on Posing Problem and Knowledge Structure

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

Monsakun can detect a constraint violation if some cards are answered by the learners. Table 3 shows the correspondence between this constraint violation and feedback. Such feedback is only generated when the EEG diagnosis detects a deadlock and when there are fewer than three 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 given in the assignment. The sentence of the feedback may change depending on this given story. Finally, whether the answered card satisfies the conditions for a feedback generation is detected and the feedback shown in the feedback sentence of Table 3 is generated. As an example, suppose there is one answered card and it is a relative quantity sentence. If the answered card is difference from given story by assignment, the learner will receive feedback that says, “Be careful about the type of story.”

Table 3. *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 quantity sentence	Irrelevant	Difference of given story	Be careful about the type of story.
			Same as given story	You’re doing good.
	Independent quantity 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 quantity sentences	Combine and Difference	Each cards’ objects are same	Be careful about the objects shown in story.
			Each cards’ objects are different	You’re doing good.
		Increase and Decrease	Each cards’ objects are same	You’re doing good.
			Each cards’ objects are different	Be careful about the objects shown in story.
	Relative and independent quantity sentences	Irrelevant	Relative quantity sentences is not correct card	Be careful about the type of story.
			Objects of relative and independent quantity sentences are different	Be careful about the objects shown in story.
			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.

4. MONSAKUN Affective

An outline of the system for implementing the in-process feedback is described in this Section. We call this system *Monsakun Affective*. As the basic operation, the level of assignment and the assignments implemented are the same as in the conventional system introduced in Section 2. However, the learner must wear an electroencephalograph before starting the exercise for receiving new feedback. The function used to detect a deadlock is being developed as a Web API. Therefore, *Monsakun Affective* sends the data from the electroencephalograph to the Web API every second, and the Web API returns the result of the deadlock judgment based on the model in Section 3.1.

Here, we describe the exercise procedure of *Monsakun Affective*. The learner wears an electroencephalograph, confirms that the data can be measured without any problems, and then logs into the system. The level selection interface is then displayed, and the learner selects the level to work on from among levels 1–3. The system then displays the practice interface shown in Figure 3. The procedure of the exercise and feedback after the problem-posing is the same as in a usual *Monsakun*, although the feedback based on an EEG is displayed in the upper-right. If the learner feels confused, the balloon in the upper right will display the feedback described in Section 3.2.

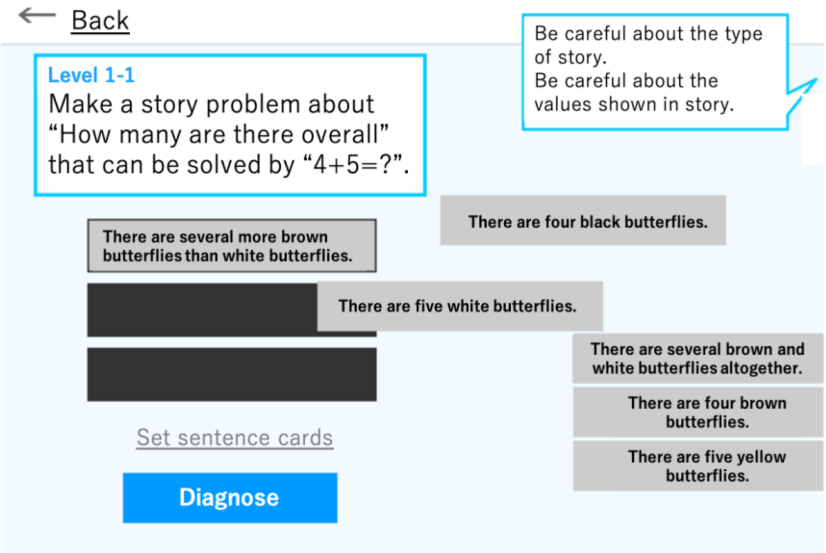

Figure 3. Interface of *Monsakun Affective* for In-Process Feedback.

5. Experimental Use

5.1 Procedure

The purpose of this experiment is to confirm whether in-process feedback is properly generated using the developed prototype system. The subjects were five engineering college students, who differed from the students who acquired the learning data.

First, the subject was instructed on how to use *Monsakun Affective* and the experiment procedure. Next, the subject used the developed prototype system to work on an assignment of levels 1–3. At this time, 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 subject answered the usage questionnaire.

5.2 Evaluation of Timing and Contents

Table 4 shows various exercise logs of the system. All values are averages for all subjects by level. If we check the exercise time, level 3 is overwhelmingly large. The number of feedbacks is divided into after-process feedback and in-process feedback. The number of all feedbacks are only for mistakes and do not include the number of feedbacks for correct answers. There are more the number of feedback for level 3 than each number of feedback for level 1 and 2 as well. The number of assignments for each level 1-3 is 10 questions. Therefore, the number of posed problems for Levels 1 and 2 is almost the same as the number of assignments. However, the number of posed problems for level 3 is about three times the number of assignments. The number of steps is the number of times the card is put in and out of the blank. It takes a minimum of 3 steps to pose the correct problem. Therefore, the minimum number of steps for each level is 30. More than this number of steps is the steps in which the learner performed various thinking activities other than giving the correct answer, and this is described as the number of search steps. The number of steps was lowest at level 2 and highest at level 3.

Moreover, there were more number of in-process feedbacks than the number of after-process feedbacks. *Monsakun* assessed that many feedbacks are required at the in-process timing in addition to the after-process feedback.

Table 5 provides answers regarding the adequacy of the in-process feedback timing and content. The subjects judged that the timing of in-process feedback was appropriate for about 70%.

Although there is a possibility of an overfitting with machine learning, the possibility of an overfitting is low because the subject is not a student who acquired the learning data and logs of multiple learners were combined into such data. Although 10% to 20% of the feedback was deemed “unnecessary,” this level decreased to less than 10% as the difficulty increased. We suspect that this result is because the learner was received feedback when he/she was thinking in the same way as the content of the feedback. Currently, the system generates feedback the moment it detects a deadlock; however, it may be better to generate feedback later.

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* needed it, but that the learner also determined it to be meaningful.

Table 4. *Logs of Exercise using Monsakun Affective (N=5)*

	Exercise time	Number of feedbacks		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 5. *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

5.3 Questionnaire

The contents and results of the questionnaire are shown in Figure 4. There are six answers: “strongly agree,” “agree,” “somewhat agree,” “somewhat disagree,” “disagree,” and “strongly disagree.” However, in Q2 to Q4, there are 6 answers: "Very difficult", "Difficult", "Somewhat difficult", "Somewhat easy", "Easy", and "Very easy". In the Figure 4, the graph is drawn by replacing "very difficult" with "strongly agree." First, all learners were able to concentrate on the exercises, and the simple electroencephalograph was not in the way. Most of the respondents answered that levels 1 and 2 were easy, but level 3 was difficult, even for college students. In addition, as an important result, many of the subjects pointed out that it was difficult to notice new feedback. This led to the negative answers to Q6 and 7.

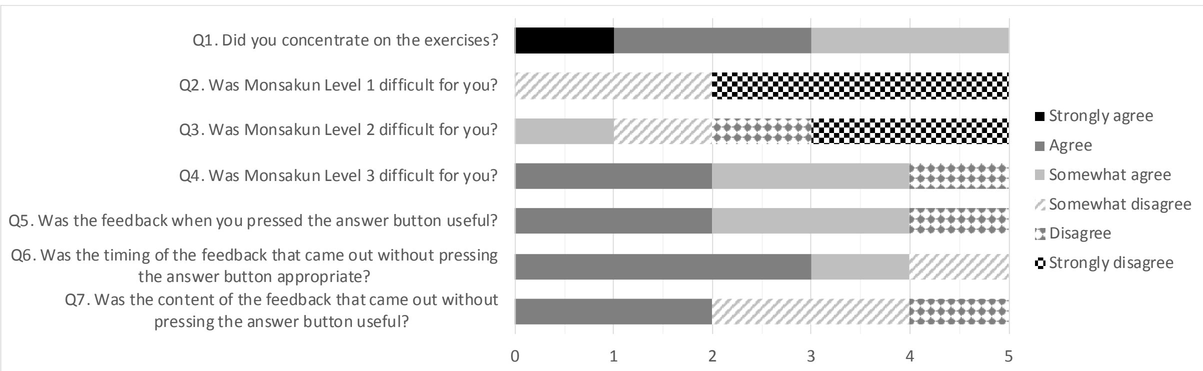

Figure 4. Contents and Results of Questionnaire (N = 5).

5.4 *Discussion and Limitation*

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 about 70% of the results showing an appropriate timing. Overfitting may have occurred as a result of machine learning. However, the learning data used is 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 overfitting is low. Of course, this model should be verified in the future. In addition, we considered a simplification of the model by making the objective variable binary also contribute to the system. 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 clear 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. Especially level 3 is remarkable. In *Monsakun,* the number of steps greater than the shortest number of steps indicates a process of exploratory thinking rather than answering the correct answer. Therefore, all these steps can be the target of feedback. It is possible to give feedback with all these steps, but it hinders the exercise and is not realistic. In-process feedback reduces this feedback to around 40% on average. In contrast, the number of after-process feedback is too small. Considering that about 70% of in-process feedback was effective, it is considered that sufficiently useful feedback was realized. Based on the above, we confirmed the feasibility of in-process feedback. This research is aimed at arithmetic word problems; however, if the system has a significant knowledge structure, the same effect can be expected.

However, the interface of the system is inappropriate. It is necessary to consider how to present new feedback. It is also 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. (2012) proposed a method using a genetic algorithm for adjusting the hyperparameters of a convolutional neural network consisting of three layers. We are going to plan that this method will be used to determine the hyperparameters in the future.

6. Concluding Remarks and Future Work

We worked on the realization and verification of in-process feedback to resolve deadlocks in learning by 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 errors.

However, when using such a system, if the learner becomes confused while thinking about an exercise, this feedback will not work properly. Therefore, we aimed to develop in-process feedback for learners who are 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 solve this challenge. The accuracy of the developed deadlock detection system was approximately 70%, and the appropriateness of the feedback sentence was approximately 90%. This result shows that in-process feedback can be realized effectively in resolving the deadlock of a learner. We believe this will provide insight into solving the wheel-spinning problem in an ITS.

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

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)* (UK), Glasgow (pp. 1-4). IEEE.

Ammar, M. B., Neji, M., Alimi, A. M., & Gouardères, G. (2010). The affective tutoring system. *Expert Systems with Applications*, 37(4), 3013-3023.

Beck, J. E., & Gong, Y. (2013, July). Wheel-spinning: Students who fail to master a skill. *In International conference on artificial intelligence in education* (USA), Memphis, Tennessee (pp. 431-440). Springer, Berlin, Heidelberg.

Beck, J., & Rodrigo, M. M. T. (2014, June). Understanding wheel spinning in the context of affective factors. In *International conference on intelligent tutoring systems* (USA), Honolulu, HI (pp. 162-167). Springer, Cham.

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.

D'Mello, S., Picard, R. W., & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, 22(4), 53-61.

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.

Hirashima, T., Yamamoto, S., & Hayashi, Y. (2014, June). Triplet structure model of arithmetical word problems for learning by problem-posing. In *International Conference on Human Interface and the Management of Information* (Greece), Heraklion, Crete (pp. 42-50). Springer, Cham.

Hussain, M. S., AlZoubi, O., Calvo, R. A., & D’Mello, S. K. (2011, June). Affect detection from multichannel physiology during learning sessions with AutoTutor. In *International Conference on Artificial Intelligence in Education* (New Zealand), Auckland (pp. 131-138). Springer, Berlin, Heidelberg.

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 *International Conference on Intelligent Tutoring Systems* (Jamaica), Kingston (pp. 63-71). Springer, Cham.

Miyauchi, S. (2013). Non-invasive study of human brain function and psychophysiology (2nd edition), *Japanese Psychological Review*, 56(3), 414-454. (in Japanese)

Nakano, A., Hirashima, T., & Takeuchi, A. (2000). A Learning environment for problem posing in simple arithmetical word problem, In *Proceedings of the 16th International Conference on Computers in Education,* (pp. 91-98).

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.

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.

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.

Yamamoto, S., & Hirashima, T. (2016). A case study of interactive environment for learning by problem-posing in special classroom at junior high school. In *Proceedings of the 24th International Conference on Computers in Education* (India), Mumbai (pp. 282-287).

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 *Proceedings of the 20th International Conference on Computers in Education* (Singapore), (pp. 25-32).

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* (USA), Austin, Texas (pp. 1-5). Association for Computing Machinery.