Proc. of the 10th Intl. Conf. on Advances in Computing, Control and Networking - ACCN 2020

ISBN No. 978-1-63248-184-9 DOI : 10.15224/978-1-63248-184-9-03

Automatic Detection of Stacking in Task using VR Space

Shota Hashimura, Fumiko Harada, Hiromitsu Shimakawa

***Abstract*— Lessons tailored to the situation of understanding have a high learning effect, but it is difficult to keep track of students' understanding. In this paper, we propose a method that use VR to estimate whether students in a task have outlooks for the answer from body movements. As a result of the experiment, we succeeded in estimating the outlook of the answer with 75% accuracy but cannot found that the prospect itself was correct or not. This makes it possible to detect a state that the student is stuck in the task and the understanding is stagnant, and it is possible to provide a lesson tailored to the student's understanding state.**

Keywords— Education, Virtual Reality

I. Introduction

Lessons with appropriate difficulty have a high learning effect.[1] If the content of the lecture is too difficult, students cannot understand, and if it is too easy, they will not be able to learn. The teacher tries to adjust the progress of the lesson to gives the students the appropriate load. To do so, teachers need to know the load on students.

In face-to-face classes, teachers try to know the load by observing the students during the class. Teachers also ask students questions during class and receive questions from students in order to know their understanding. In this case, the teacher can know the student's facial expressions and other behaviors, so that the teacher can know the load on the student.

However, in multi-person lectures, it takes too much time to check the understanding of each student in this way. For this reason, paper tests are generally performed, or e-learning tests are performed to reduce the trouble of scoring. Thereby confirming the understanding of all students. However, in such a test, it is not possible to know how the student to answer. For this reason, it is difficult to estimate the load on the student, and it is more likely that the teachers overlook when the student accidentally answers the correct answer than in the case where the student talks directly. Also, frequent tests increase the burden on both teachers and students. For this reason, it is difficult to fine-tune the progress of the class and adjust the load according to the level of understanding. Therefore, there is a need for a simple and accurate way to estimate student comprehension.

In this paper, we propose a method of estimating the lack of a prospect of the answer from the behavior of the student working on the task in the VR space and estimating the student's understanding from that. Human memory includes short-term memory and long-term memory, which have different characteristics in terms of retained elements and functions. In this study, we focus on the difference between these two memories in estimating the state of understanding. Short-term memory processes a given task for comprehension. Long-term memory stores schemas which is the pattern for comprehension frequently occur in short-term memory. Solving new tasks that have not been mastered require that short-term memory retain many factors. Short-term memory is small. Therefore, solving a task that have not been mastered make a large cognitive load. On the other hand, if students learn to solve the same type of task many times, you have a schema in long-term memory. This schema is brought from long-term memory to short-term memory and used for processing. Since a schema is treated as one chunk in short-term memory, when solving a task which a solution has been mastered, the cognitive load is reduced even if many elements appear in the task. Using this fact, if there is a lot of hesitation and mistakes when answering, then the cognitive load is imposed on the answerer because the schema has not been generated in the answerer's long-term memory. In this research, we use a VR space where gaze and hand movements can be recorded accurately to estimate that the answerer finds the schema that suits the situation from the body movement data. It means the outlook of the answer is established. In addition, we will discuss whether there is a difference in the body movement between the correct answering and wrong answering, and whether the correctness can be estimated from the body movement.

The structure of this paper is shown below. Section II describes related knowledge and related research. Section III describes how to clarify the state of understanding from behavior, and Section IV confirms the effectiveness of this method through experiments. Section V conclude this study.

Shota Hashimura
Graduate School of Information Science and Engineering
Japan

Fumiko Harada
Connect Dot Ltd.
Japan

Hiromitsu Shimakawa
College of Information Science and Engineering
Japan

II. Related knowledge

A. *Cognitive load and schema*

The load on the brain to understand things is called the cognitive load. When people try to understand things, they always store those things in their brains. In some cases, all the elementals used for understanding are kept in memory, while in other cases, several related elementals are grouped and kept. The latter case is considered a process for

ISBN No. 978-1-63248-184-9 DOI : 10.15224/978-1-63248-184-9-03

efficiently handling elements used for understanding that are performed many times. Tasks that process frequently used relevant facts are patterned and stored as grouped elements. This grouped element is called a chunk [2]. There are two types of memory, long-term memory and short-term memory, each of which has its own characteristics in what is stored on it and in the processing of the stored data. The short-term memory is also called a working memory and plays a role of temporarily storing information and processing the information. Human can understand the thing when all the elements that make it up can be kept in working memory simultaneously. [1] However, the working memory capacity is small. It can hold only about 4 chunks of information even for young adults. [3] On the other hand, long-term memory has a large capacity. The patterns of processing information are stored in long-term memory, and these patterns are called schemas. The schema has the role of combining multiple elements into one chunk, which can reduce the working memory load. Take a example when memorizing a six-character alphabet sequence called MEMORY. If a child who does not know English tries to remember this alphabet sequence. Then the child remember one character at a time like 'M', 'E', 'M', 'O', 'R', 'Y'. This six letters occupy six chunks. On the other hand, if you know the English word MEMORY, you can combine that information into one chunk, reducing the load on the working memory required for storage. The difficulty of a task is influenced by the cognitive load of the information processed to accomplish the task. The magnitude of this cognitive load is affected by the knowledge of the person solving the task. If the solver of the task has an appropriate schema, the number of chunks processed simultaneously in the working memory can be reduced, so the cognitive load also will be reduced. On the other hand, if you do not have the appropriate knowledge for the task, the cognitive load will be high because the number of chunks will not be reduced by the schema.

B. *Schema search status and schema confirmed status*

When people think about things, they search the brain for a schema that suits the situation. Sometimes a suitable schema can be found, and sometimes cannot so the thinking is stopped. Also, the found schema is not always appropriate, and sometimes a schema that leads to the wrong conclusion. When human find a schema, they confirm which schema to use, and start looking again for the next schema. Like this way, human lives by repeating the schema search state and the schema confirmed state. When a student works on an assignment in a class, if the student does not have a pattern of thinking (schema) necessary for answering, it is conceivable that the student cannot get out of the schema search state. It is possible that student's physical movements are different between the schema search state where the answer policy is not determined and the schema confirmed state where the policy is fixed. Also, there is a possibility that there is a difference in the behavior between using an appropriate schema to solve a problem and using the wrong schema.

C. *Advantages of class in VR space*

In the VR space, the user's hand movements and head movements can be recorded with high accuracy. The position and angle of the head-mounted display used to present the VR space can always be grasped by sensors, therefore the scene projected is changed according to the movement of their head in real time. When the users moves their neck, they can see what they wants to see in the VR space from the angle they wants to see, with the same sense as in ordinary life. As a result, the users can get three important factors for gaining a sense of reality in the space are: three-dimensional spatiality, real-timeness, and self-projection [4].

In recent years, VR supports a motion controller for users to operate the VR space. This motion controller is used by hand. The motion controller is also can grasp the real-time position and tilt with the sensor. As a result, the hand model is displayed at the position and tilt in the VR space corresponding to the grasped position and tilt. In addition, the pressure of grabbing is also sensed, so users can their VR hand open and close. As a result, the users feels as if his or her hand is in the VR space.

By using VR to conduct lessons and tests, it can be to sense minute movements of the user, such as changes in the field of view when moving the head in the VR space, and stray hand movements. Furthermore, the feeling of operation that the user experiences in the VR space, which is close to the real space, does not impose an extra load on the user during operation. For this reason, it is considered that the user behaves in the same way as solving a task in the real space, and that these behaviors include information indicating the student's understanding status and cognitive load.

D. *Relative works*

Nakamura et al. [5] study about automatic assessment of students' understanding. In this study, by acquiring the face of a student during e-learning with a camera, they succeeded in estimating the subjective difficulty that the student felt in the teaching material with an accuracy of about 75%. However, there is still a problem that it is necessary to train estimator before use because there are individual differences in face motion.

There are the researches that support lessons according to students' understanding. Philip et al. [6] have developed Codeopticon, a tool that helps teachers keep track of students' progress in programming classes. Codeopticon displays the code editing screen of 15 learners at the same time, in a tile form on the teacher's screen. If more than 15 students are in class, the screen of the student whose code has changed is displayed preferentially, making it easier for the teacher to grasp the current situation of the students. However, there is a problem that teachers have to endure a heavy load because the progress of a large number of students must be grasped manually.

Volodymyr et al. [7] created a tool to help teachers keep track of student progress during a 3D modeling workshop. In this tool, live images of the screens of all students are displayed on the teacher's PC, and the timeline showing the rough operation history of the students is also displayed. The timeline makes it easier for students to understand the progress, and the teacher's burden is reduced by automatically providing support according to the operation history. However, to automatically provide support, it is necessary to automatically estimate the student's level of

Proc. of the 10th Intl. Conf. on Advances in Computing, Control and Networking - ACCN 2020

ISBN No. 978-1-63248-184-9 DOI : 10.15224/978-1-63248-184-9-03

understanding. Since it is difficult to do so, only a simple method of support that is response to a specific action has been realized. In order to further reduce the burden on the teacher and enhance the student's learning effect, it is necessary to automatically estimate the students' understanding.

If we can obtaining more student information from the behavior during class, and finding unified features with little individual differences which are relevant to student understanding, then we can find a way to easily and accurately estimate students' understanding.

III. Estimation of understanding using VR

A. *Estimation of understanding from behavior*

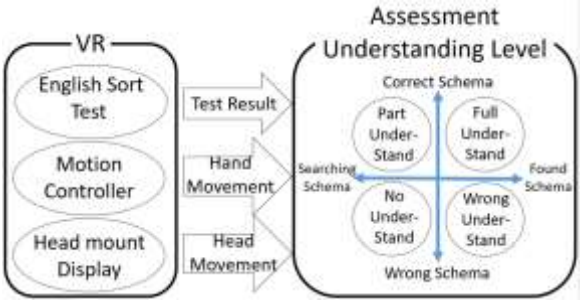

Figure 1. Method overview

In this paper, we propose a method of estimating the student's understanding from the behavior of the task answer process. Fig. 1 show the overview of the method. We take an example the English sorting problem as the task.

When people process things in their brain, they repeatedly search the brain for a schema that can be adapted and uses the found schema. In the schema search state in which the answer policy has not been determined, and in the schema determined state in which the answer policy has been determined, it is conceivable that there is a difference in physical movement. We propose a method for estimating from the body movement that the solver who is working on the task has not escaped from the schema search state. In addition, in the schema confirmed state, it is possible that there is a difference in the body movement between the state using the correct schema and using the wrong schema. Also, in schema found state, we attempt to estimate whether student have correct schema or not. For example, if the students exit the schema search state early and give the correct answer, it can be determined that the answerer has a full understanding of the knowledge required for the assignment, and they give the correct answer but stay long in schema search state, it can be assumed that they have only partial understanding. In addition, if a wrong answer is given despite leaving the schema search state at early, it is presumed that there is a fundamental misunderstanding in the understanding, and a schema search is performed for a long time and the answer is wrong, it can be estimated that learning is insufficient.

B. *The environment that highlight behaviors*

In order to estimate student's schema search activity from behavior, we propose a VR task answering environment that can record behavior in detail and make the behavior highlight. By answering the task in the VR space, the hand and head movements in the answering process can be recorded with high accuracy. Fig. 2 and 3 show screenshots of the VR space created in this study. The movie of this VR space is located at www.de.is.ritsumei.ac.jp/publication/englishtest.mp4. We use Unity5.6.2p2 to create this VR space. We use the English sorting problem for the example of classroom tasks. When the task begins, English words cards appear on the desk in front of the student in the VR space. The solver can grasp these cards by placing his hand on the cards in the VR space and holding the controller. By sorting these English words, students can make correct English sentences. The task of students is to move these English words to the blue spaces in front of them to complete the English sentence. The hint which translated the correct sentence into Japanese is displayed above the student's head. Students can check the hint by looking up. The position of the hint is placed far from the table purposely. The student moves his head to look alternately at the hint and card. Since students wear the headset on the head, the movement of the student's gaze can be recorded from the movement.

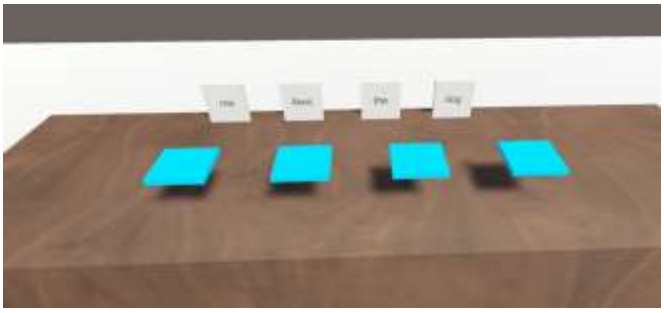

Figure 2. Screenshot of the VR space

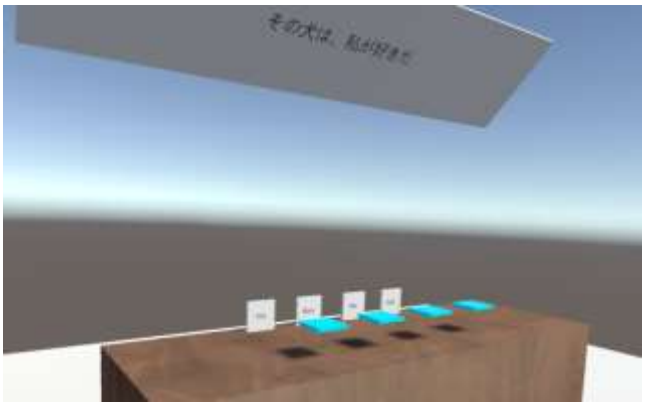

Figure 3. Overhead view of experimental environment

C. *Physical movements and how to record them in the VR space*

The Oculus Rift, a head-mounted display used in this study, provides the users with three-dimensional spatiality, real-timeness, and self-projection to obtain a sense of reality

Proc. of the 10th Intl. Conf. on Advances in Computing, Control and Networking - ACCN 2020

ISBN No. 978-1-63248-184-9 DOI : 10.15224/978-1-63248-184-9-03

in a VR space. Thus, unlike e-learning, even students who are not good at operating a personal computer etc. can work on tasks by using ordinary movements as real world. Therefore, students are able to answer the task without any extra load. In addition, by recording the movement of the head during the answer with high accuracy, it becomes possible to grasp the rough gaze movement.

Oculus Touch, the motion controller is used in this study, displays the user's hand movements in the VR space. Since the movement of the finger is also sensed, intuitive operations such as holding the English word by holding the hand and placing the English word by opening the hand are enabled. Therefore, like the operation of the head-mounted display with the head, the operation with the motion controller is excellent for students unfamiliar with electronic devices to engage tasks without extra burden. There is no haptic feedback in VR space. Since the student cannot rely on haptic feedback, when the student tries to grab an object in the VR space, it is necessary to confirm the position by looking at what he wants to grab. Therefore, when grabbing each word, the student eventually sees the word to grab. In the answering process, the student's gaze may shift to another object. If student already convinced of the answer policy, he do not need to move his gaze to another object, but if student do not have an answer policy, he moves his gaze to another object such as hints. The extra movement of the gaze is considered to be related to the hesitation. Therefore, it is possible to estimate where the student is lost in their task.

D. *Edit distance*

The solver's schema search activity cannot be directly observed. In order to use the activity that cannot be quantified as the objective function in this method, we propose to treat the change of the edit distance in the English sorting problem as an alternative to the schema search activity. The edit distance is an measure of similarity between character strings. Fig. 4 shows a calculation image of the edit distance in this method.

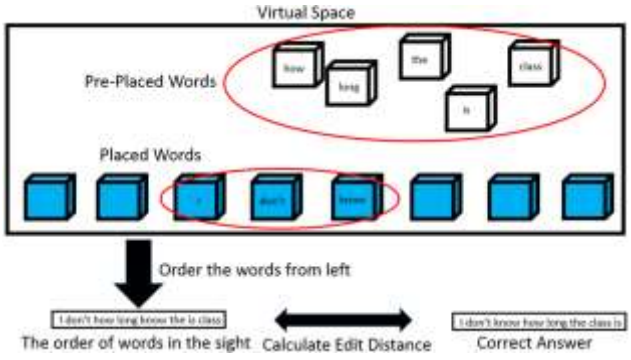

Figure 4. Calculate edit distance in VR

By sorting all the words in order from the one with the coordinates on the left side, the words are arranged from the viewpoint of the answerer. The similarity between the student's word sequence and the correct word sequence is determined by the edit distance (Jaro–Winkler distance [8]). By differentiating the edit distance, the change in the edit distance was calculated, and make this change smooth by taking a moving average over a width of 3 seconds. Since human working memory can only store about four chunks at a same time, it is difficult to solve problems without rearranging words cards. In other words, the time which the change of the edit distance is 0 is the time that the student has not found a suitable schema for the answer. In other words, the student cannot move forward or backward toward the answer in that time. It is assumed that a time zone which the smoothed change is 0 is the schema search state, and a non-zero time is the schema found state. Further, of the non-zero time zones, a time zone having a positive value is a state using a correct schema, and a negative time zone is a state using an incorrect schema.

E. *Estimation of activity about schema using neural network*

In this study, we train neural network estimator of the schema search activity by giving time series data of body movements. We use the GRU layer and the Dense layer for the neural network. The optimizer is Adam. A GRU layer with 512 nodes was used as the layer next to the input layer, and Relu was used as a GRU activation function. The GRU layer is connected to the next Dense layer with 2 nodes via the DropOut layer with a dropout rate of 0.6, and the estimated probability of the objective function binary value is transmitted to the output layer by the SoftMax activation function. The learning rate was 0.0001 and the batch size was 1024. As explanatory variables, the speed of the X, Y, Z coordinates of the hand in the answer process obtained from the motion controller (Oculus Touch) is used. Also, the values obtained by measuring the horizontal and vertical angular velocities of the head obtained from a headset (Oculus Rift) were used. These values are recorded every 20 milliseconds. Each horizontal direction is converted absolute values. Before the training, these explanatory variables were sliced in 1/50 second intervals with a window width of 20 seconds, and used as explanatory variables. As the objective variable, we used the label for whether the schema was being searched, which was created by the method described in Chapter 3.4. Training was also performed with the binary value of whether the change in the edit distance was positive or negative as the objective variable.

IV. Experiment

A. *Content of the experiment*

We asked nine college students to answer English word sorting tasks in the VR space, and collected answer data for 90 questions. The position and tilt of the hand were recorded from the signal of the motion controller, and the position and tilt of the head were recorded from the signal of the head-mounted display. The coordinates of the English word card in the VR space were also recorded. The recording frequency of each data is every 20 milliseconds. The English word sorting tasks were set with different difficulty, from simple tasks at the Japanese junior high school level to difficult tasks used for Japanese university entrance examinations.

B. *Estimation of schema search activity*

We try to estimate whether the student is searching the schema or not from the physical movement. To create an objective variable, the edit distance at each time in the

Proc. of the 10th Intl. Conf. on Advances in Computing, Control and Networking - ACCN 2020

ISBN No. 978-1-63248-184-9 DOI : 10.15224/978-1-63248-184-9-03

answering process was calculated by the method described in chapter C of section III, and the derivative of the value was used to calculate the change in the edit distance. This change in edit distance was smoothed by taking a moving average every three seconds in the past. The time when the change of the smoothed edit distance is 0 is regarded as the schema searching state, and the remaining time is schema found state. Using this labeling as the objective variable, and using body movement as an explanatory variable, we trained neural network. These variables were down-sampled and used 70297 data during schema search state and the same number of data of schema found state. We did cross-validation. Four-fifths of the entire data was used as training data and the remaining one-fifth as validation data. The accuracy of validation data was highest when only hand acceleration was used as an explanatory variable. Since the maximum correct answer rate was 0.74856 on average, it can be said that the schema search activity of the students affected the physical movement in the answering process.

C. *Estimation of using correct schema or wrong schema*

The schema found state contain two states, using correct schema state and using incorrect schema. We tried to estimate whether the state using correct schema and the state using incorrect schema from the body motion. We calculated the moving average of the change in the edit distance, and determined the data having a positive value is the state using a correct schema, and having a negative value was the state using an incorrect schema. We did down-sampling of 70297 of schemas found state data to 4978 using correct schema data and 4978 using incorrect schema states. We used these down-sampled data for training. We did cross-validation. The accuracy rate of five times cross-validation tests was around 50%, so the estimation was not successful.

V. Conclusion

In this paper, we proposed a method for estimating the lack of outlook of answers using VR. As a result of the experiment, we confirmed that it is possible to estimate the state where the answer policy is decided and the state where the answer is not decided with 75% accuracy from body movement. On the other hand, it was also revealed that whether the answer itself was a correct answer or a wrong answer did not affect the student's physical movement, so that correctness could not be estimated from the physical movement.

References

[1] Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. Educational Psychology Review, 19(4), 469-508.

[2] Fred Paas , Alexander Renkl & John Sweller (2003) Cognitive Load Theory and Instructional Design: Recent Developments, Educational Psychologist, 38:1, 1-4, DOI: 10.1207/S15326985EP3801_1

[3] Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences, 24, 87-185

[4] S.Tachi, M.Sato, M.Hirose(2010), “Science of Virtual Reality(バーチャルリアリティ学)”, The virtual reality society of japan

[5] K. Nakamura, K. Kakusho, M. Murakami, and M. Minoh(2010). "Estimating Learners’ Subjective Impressions of the Difficulty of Course Materials by Observing Their Faces in e-Learning" The IEICE Transactions on Information and Systems(Japanese Edition) Vol.J93-D No.5 pp.568-578

[6] Guo, Philip J. "Codeopticon: Real-time, one-to-many human tutoring for computer programming." Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology. ACM, 2015.

[7] Dziubak, Volodymyr, et al. "Maestro: Designing a System for Real-Time Orchestration of 3D Modeling Workshops." The 31st Annual ACM Symposium on User Interface Software and Technology. ACM, 2018.

[8] Winkler, William. (1990). String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. Proceedings of the Section on Survey Research Methods.

Shota Hashimura

received Bachelor degree of Information Science and Engineering in Ritsumeikan Univ. in 2018. He currently study at Graduate School of Information Science and Engineering, and study about Data Engineering.

Fumiko Harada

received B.E., M.E, and Ph.D degrees from Osaka Univ. in 2003, 2004, and 2007, respectively. She joined Ritsumeikan Univ. as an assistant professor in 2007, and is currently a lecturer. She engages in the research on real-time systems and data engineering. She is a member of IEEE.

Hiromitsu Shimakawa

received Ph.D degree from Kyoto Univ. in 1999. He joined Ritsumeikan Univ. in 2002. Currently, he is a professor in Ritsumeikan Univ. His research interests in include data engineering, usability, and integration of psychology with IT. He is a member of IEEE and ACM.