IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol.005, LIIR314
DOI: https://doi.org/10.52731/liir.v005.314

Analyzing the Learning Process in a Card Operation-Based Programming Learning Support System

Natsumi Tanabe *, Shimpei Matsumoto *

Abstract

The Card Operation-based Programming Learning Support System (COPS) was developed as a tool to assist programming learning with a focus on helping learners understand the structure of programs. Previous studies have demonstrated that when this system is introduced into classes, it maintains learning outcomes equivalent to those of traditional coding-based learning while reducing learning time. However, as the difficulty of tasks increases, learners tend to systematically search for answers relying solely on feedback functions (hints) without engaging in thoughtful problem-solving, a phenomenon known as knowledge-free solutions. Therefore, this study aims to enable instructors to intuitively grasp learners engaged in inappropriate learning activities by analyzing and visualizing the learning log data from COPS. Traditionally, the Levenshtein distance has been used as a quantitative measure to evaluate learners' learning progress. In this study, we aim to enhance the analysis and visualization of the learning process by incorporating the concept of taboo choices, aiming to improve the accuracy of detecting learners engaged in inappropriate learning activities. Taboo choices are selections with heavier penalties when chosen compared to regular dummy choices. We compared the detection accuracy of inappropriate learners using three patterns of visualization results with different weights for the penalties of taboo choices. The results demonstrate the significance of introducing the taboo choices mechanism in the analysis and visualization process.

Keywords: programming learning, data analysis, understanding program structure, contraindicated options

1 Introduction

In recent years, the global proliferation and advancement of digital technology have heightened the significance of information education. Particularly emphasized through programming learning is the acquisition of "Computational Thinking (CT)," referring to the ability to comprehend problems and devise procedures and methods to solve them using computers. CT encompasses skills such as decomposing complex problems, logically constructing solution methods, identifying patterns and regularities and converting them into algorithms, and predicting and evaluating outcomes.

However, due to the multifaceted nature of programming, beginners face significant learning burdens [1][2]. Conversely, reducing the learning burden could lead to diminished apprehension towards programming among learners and foster enjoyable learning experiences. Furthermore, the potential for enhanced learning outcomes in programming, driven by increased motivation and the ability to explore more content, is considerable.

* Hiroshima Institute of Technology, Hiroshima, Japan

To mitigate the learning burden and effectively support programming learning, an information structure-oriented approach that decomposes the various technical elements required for programming and structures the entire learning process into open-ended tasks, allowing for the easy allocation of cognitive resources to essential learning, is considered effective [3]. The Card Operation-based Programming Learning Support System (COPS) was developed based on this concept. In COPS, learners divide the program code of a task into meaningful segments and present them as multiple cards, arranging them in the correct order. This approach allows learners to focus on understanding the structure of the program, thereby concentrating their cognitive resources on essential learning.

However, programming learning using the card operation method is not intended as a standalone solution; it complements traditional lectures and coding exercises, enabling more efficient learning. Due to limited classroom hours in educational institutions such as universities, conducting all programming activities within this time frame imposes a significant burden on learners. Therefore, in systems like the card operation learning support system, it is envisaged that classes could be divided into lectures focusing on grammar and exercises concentrating on algorithms, with coding exercises conducted during self-study time.

Figure 1 depicts the actual problem exercise screen of COPS, displaying the problem statement, answer field, and choice cards. Learners follow the instructions in the problem statement to insert cards containing program code from the choice field into the answer field and rearrange them in the correct sequence. Upon submission, feedback on correctness is provided, and administrators can access various log data of learners.

Figure 1: exercise screen of COPS

Previous research has demonstrated that introducing this system into classes can maintain the same learning effectiveness as traditional coding-based learning while reducing learning time. Subsequently, an analysis of learning log data revealed learners engaging in inappropriate learning activities, particularly relying on the feedback function without thoughtful consideration, as task difficulty increased.

In response, a method to quantify and visualize learner responses using the Levenshtein distance concept was proposed to identify learners engaging in inappropriate activities. Results confirmed a statistically significant lower number of card operations in the group demonstrating understanding of programming compared to those who did not, establishing a relationship between the number of card operations and learner comprehension.

Based on the hypothesis that analyzing and visualizing learning log data can identify learners engaging in inappropriate activities, a method compressing and presenting this data using multidimensional scaling (MDS) was proposed. Regression analysis demonstrated the ability to classify appropriate and inappropriate learning activities. Furthermore, visualization through MDS enabled the identification of similarities in learner behavior in a single diagram. However, the accuracy of detection remains unclear, and practitioners have highlighted the need for a more prominent presentation of inappropriate learners.

Consequently, the objective of this study is to enhance the accuracy of detecting learners engaging in inappropriate activities by incorporating the taboo choice concept into a quantitative method based on the traditional Levenshtein distance and visualizing it through multidimensional scaling.

2 Method

The learning log data analyzed in this study were collected from previous research. The study included four problems: 1) variable types, 2) iteration (for loops), 3) arrays, and 4) sorting, with no time limits imposed. Participants consisted of 20 university students, including third- and fourth-year undergraduates and graduate students majoring in information science who have a basic understanding of C programming. In the analysis of COPS learning logs, the Levenshtein distance has traditionally been used to quantitatively evaluate the learning progress of students. The Levenshtein distance is an indicator that measures how much two strings differ, also known as the edit distance. Specifically, it refers to the minimum number of edits (insertions, deletions, substitutions) required to transform one string into another (Figure 2).

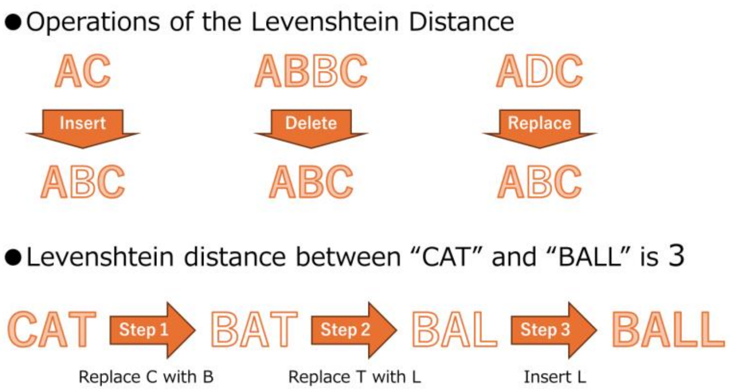

Figure 2: The concept of Levenshtein distance

In the analysis of COPS log data, the Levenshtein distance measures how much the order of cards inserted by learners in the answer field differs from the correct order. This metric allows evaluating how closely learners approach the correct answer with each card operation. Figure 3 illustrates the specific calculation method of the Levenshtein distance. The Levenshtein distance is 0 when all correct cards are inserted in their correct positions. Thus, when no cards are inserted in the answer field, the Levenshtein distance equals the number of correct cards. This is referred to as the initial distance for that problem. The Levenshtein distance at a given point is calculated by adding the items in Figure 3 to the initial distance. Additionally, the value of the Levenshtein distance when the answer field is entirely filled with dummy cards, or when all dummy cards are inserted into the answer field, is designated as the maximum distance n for that problem.

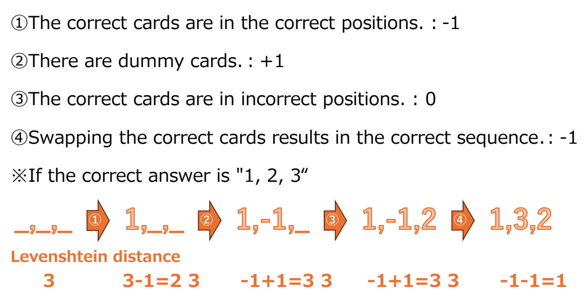

Figure 3: Calculation of Levenshtein distance in COPS learning log data analysis

The evaluation of learning data was compressed and presented intuitively through regression analysis and multidimensional scaling (MDS). Figures 4 and 5 graphically depict the temporal changes in Levenshtein distance when a learner practiced a certain problem, with linear regression lines drawn. Figure 4 illustrates an example of appropriate learning activities, while Figure 5 shows an example of inappropriate learning activities. "Appropriate learning activities" in this context refer to instances where a learner correctly answered a problem on the first attempt, with any other outcome classified as "inappropriate learning activities". Comparing the slopes of the regression lines in each graph confirms that inappropriate learning activities exhibit gentler slopes. Additionally, averaging the slopes across all data reveals a slope of -0.2084 for appropriate learning activities and -0.0074 for inappropriate learning activities, indicating a consistent trend across all data sets.

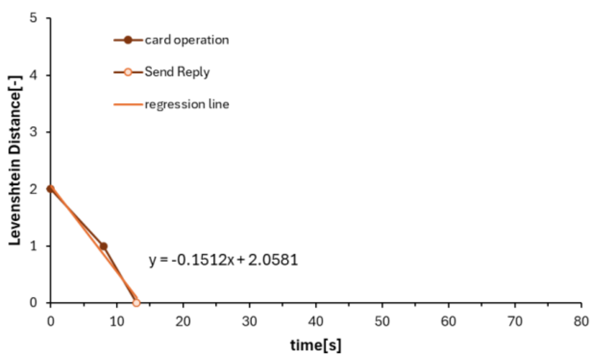

Figure 4: Temporal Change of Levenshtein Distance (Example of Proper Learning Activity)

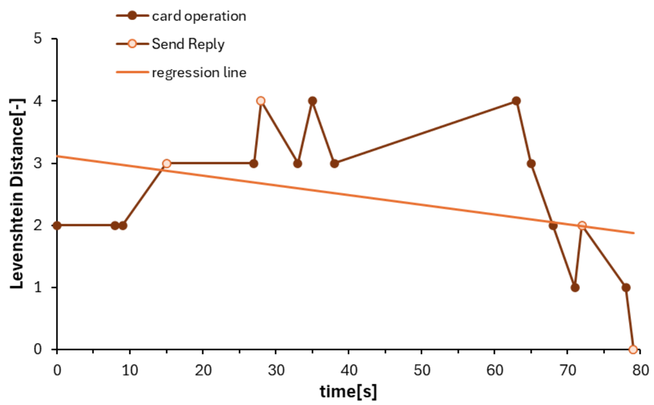

Figure 5: Temporal Change of Levenshtein Distance (Example of Improper Learning Activity)

Figure 6 illustrates a scatter plot representing the slopes and intercepts of regression lines for all learners solving a particular problem. Upon observing the graph, it is evident that appropriate and inappropriate learning activities are distinctly distributed and can be effectively classified based on these distributions.

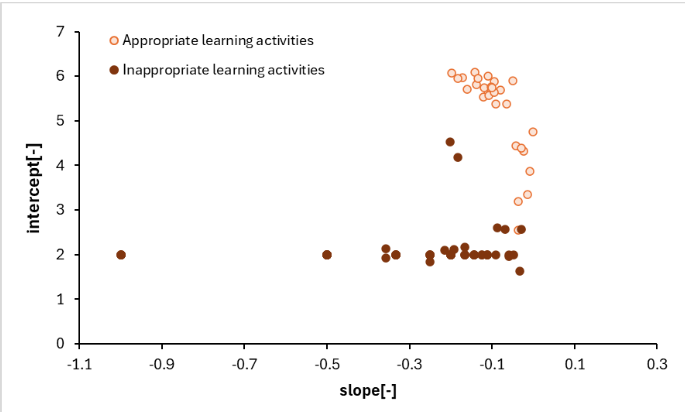

Figure 6: Slope and Intercept of Regression Lines

Next, a method was proposed to visualize using multidimensional scaling (MDS) the slopes of regression lines, intercepts, and the number of operations required to reach the correct answer as input. Figure 7 demonstrates the visualization of inappropriate learning activities in a specific problem using MDS. Multidimensional scaling is a technique for placing high-dimensional data into a lower-dimensional space, preserving the similarity and distance information of the input three data points, and aggregating them into a two-dimensional graph. The points on the graph represent learning activities, and the distances between points represent the similarity of operations between learning activities. In practice, learning activities located close to each other exhibit similar changes in Levenshtein distance, whereas those further apart exhibit different characteristics in Levenshtein distance changes. This figure demonstrates that multidimensional scaling intuitively captures the operational tendencies of learners in the problem using a single diagram.

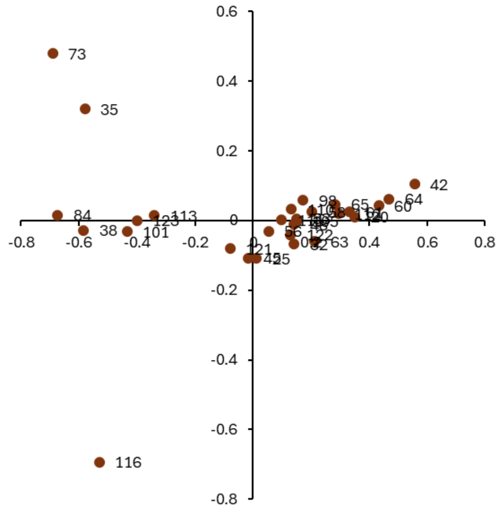

Figure 7: Example of Visualization using Multidimensional Scaling

Furthermore, to enhance the accuracy of detecting inappropriate learning activities, this study incorporated the concept of taboo options and evaluated their effectiveness through experiments. Taboo options in this study refer to choices that incur a heavy penalty compared to typical dummy choices when selected. Choosing a taboo option is considered indicative of insufficient understanding of programming or of answering without thoughtful consideration. In the same learning log data as previous studies, one of the dummy options was set as a taboo option. Taboo options are designed to measure understanding, and the criteria for setting them include the following: (1) adopting errors related to important programming concepts, which are the main topics of the problem, and (2) adopting errors that can measure whether the problem statement is understood. Simple errors such as typographical errors are not adopted. For example, Figure 8 shows a problem statement that asks to input 8 into n using scanf and output it. Three choices are given. Among these choices, the dummy choices are 1 and 3. The most appropriate dummy choice among these dummy choices is 3 because, despite the problem statement requiring the use of scanf, choice 3 does not use scanf, indicating a lack of understanding of the problem.

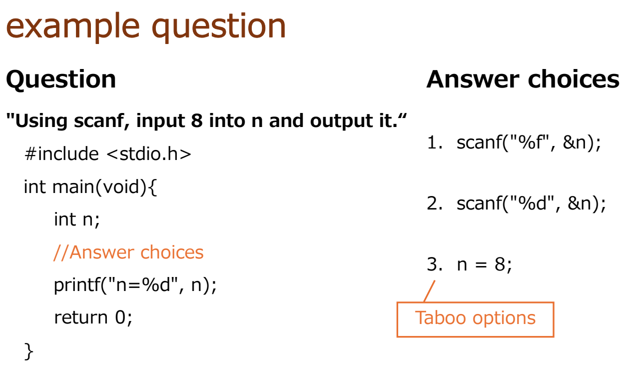

Figure 8: Criteria for Setting Forbidden Choices

Penalties for choosing taboo options need to be significantly increased to differentiate them from dummy options, but the optimal penalty level remains uncertain. In this study, comparisons were made between the conventional method (without taboo options) and two proposed methods, denoted as n×1 (Proposed Method 1) and n×2 (Proposed Method 2). When all answer slots are filled with dummy cards or when all dummy cards are inserted into the answer slots, the Levenshtein distance value is the maximum distance n for that problem. For the learning log data of the four problems presented earlier, a total of twelve multidimensional scaling graphs were prepared for the conventional method, Proposed Method 1, and Proposed Method 2. In the experiment, these graphs were randomly shown to 17 subjects, who were undergraduate and graduate students studying information science. The appropriate learning activities are also reflected in the results of this multidimensional scaling, as learners engaged in appropriate activities also selected taboo options. Since learners engaging in inappropriate activities alone might skew the similarity in operational tendencies depending on the problem, it was deemed necessary to have a benchmark for appropriate learning activities. Before conducting the detection task, the purpose of this study, the multidimensional scaling method, and the experimental procedures were explained, and it was ensured that the subjects had a sufficient understanding before proceeding with the responses. With three methods per question and a total of four questions evaluated, a total of twelve visualization results were randomly presented to the subjects. The time limit was set at 20 seconds per question, with no limit on the number of learners to be detected. Furthermore, the visualized results presented to the subjects allow them to judge appropriate and inappropriate activities, but the display of learners who chose taboo options has been omitted. From the data obtained in the experiment, four evaluation indices of the confusion matrix were calculated for each problem, and a comparison of detection accuracy was made through F-tests and t-tests for the F-value.

3 Results

Figures 9, 10, and 11 depict the visualization results for a specific problem using the conventional method, Proposal 1, and Proposal 2. In this problem, n is 6, and the values of taboo options are "1," "6," and "12," respectively. Orange points represent learners engaging in appropriate learning activities, while blue points represent those engaging in inappropriate learning activities. The red numbers indicate learners who chose taboo options. Upon observing learners represented by points distant from the dense cluster, it is evident that they often choose taboo options. However, in the conventional method, learners who chose taboo options are also present near the center of the dense cluster. In contrast, in Proposal 1 and Proposal 2, learners who chose taboo options tend to be located farther from the dense cluster as the weight of the taboo options increases.

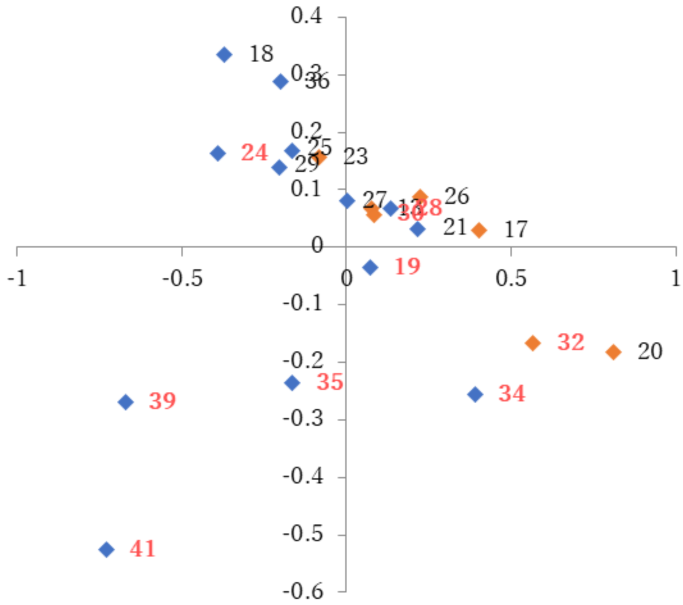

Figure 9: Visualization using multidimensional scaling method (conventional method)

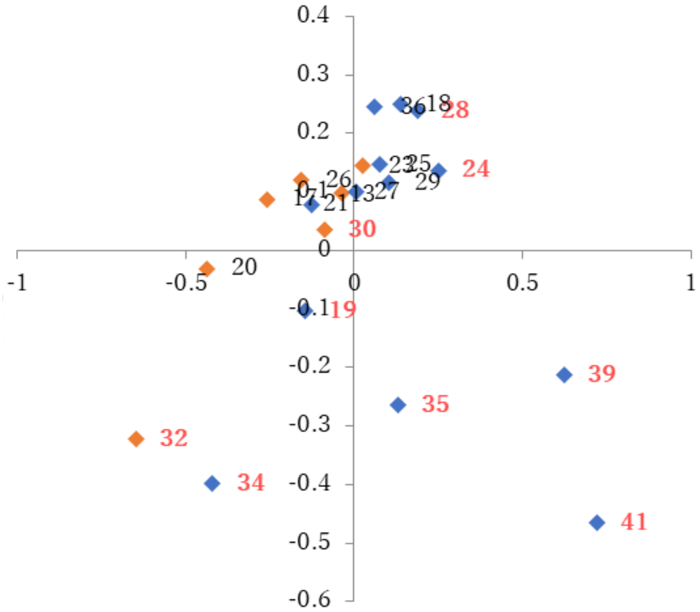

Figure 10: Visualization using multidimensional scaling method (proposed method 1 [n×1])

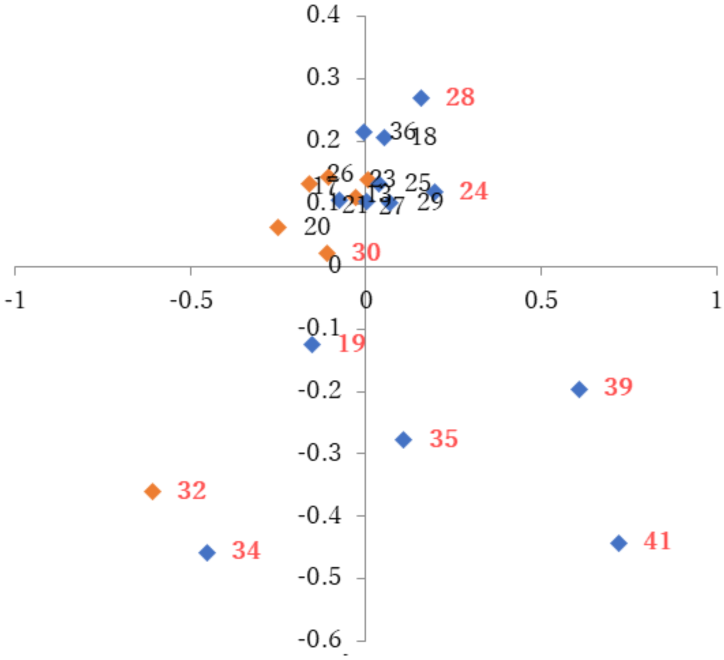

Figure 11: Visualization using multidimensional scaling method (proposed method 2[n×2])

Based on the results obtained from the experiment, confusion matrices were computed for the four evaluation metrics (accuracy, recall, precision, F-score) for each pattern (Tables 1, 2, 3, 4). When comparing each problem, Proposal 1 and 2 consistently outperformed the conventional method across all metrics. However, in the comparison between Proposal 1 and Proposal 2, significant variations were not observed, and in some metrics, Proposal 2 exhibited lower values than Proposal 1. Two factors could explain these results. Firstly, the recall rate is affected. Since the recall rate represents the proportion of correctly predicted instances out of all actual positive instances, if there are many incorrect predictions among the instances predicted as positive, the recall rate will decrease. Secondly, there is the influence of the values of the taboo options on the distribution of visualization results in multidimensional scaling. From the visualization results of multidimensional scaling, it is observed that there are learners scattered both near dense clusters and in distant positions. However, as the value of the taboo options increases, this tendency becomes more pronounced, leading to the inference that only learners excessively distant from dense clusters were detected. Furthermore, upon examining the average evaluation indices for all participants across all problems, it was observed that as the values of the taboo options increased, the numerical values of the evaluation metrics also increased (refer to Table 5). Additionally, the results of the F-test did not reveal any significant differences ($p < .05$) between Proposal n×1 and n×2.

Table 1: Four evaluation indicators (questions on variable types)

evaluation metrics	Accuracy	Recall	Precision	F-measure
conventional method	64%	31%	77%	42%
n×1	74%	48%	91%	61%
n×2	74%	47%	92%	58%

Table 2: Four evaluation indicators (questions on repetition)

evaluation metrics	Accuracy	Recall	Precision	F-measure
conventional method	56%	20%	38%	24%
n×1	73%	40%	74%	48%
n×2	71%	29%	82%	40%

Table 3: Four evaluation indicators (questions on repetition)

evaluation metrics	Accuracy	Recall	Precision	F-measure
conventional method	85%	43%	64%	47%
n×1	88%	59%	71%	59%
n×2	92%	76%	87%	77%

Table 4: Four evaluation indicators (questions on sorting)

evaluation metrics	Accuracy	Recall	Precision	F-measure
conventional method	80%	55%	83%	62%
n×1	83%	58%	90%	66%
n×2	85%	59%	92%	67%

Table 5: Mean value for all questions for all subjects

evaluation metrics	Accuracy	Recall	Precision	F-measure
conventional method	71%	37%	66%	44%
n×1	80%	51%	82%	58%
n×2	81%	53%	88%	61%

When Welch's t-test (two-tailed) was conducted for the F-score, significant differences ($p < .05$) between the conventional method and the proposed method were observed in several problems (see Figure 12). However, no significant differences ($p < .05$) were found between the proposed methods in all problems. This trend was consistent when examining the average values across all problems for all participants (see Figure 13).

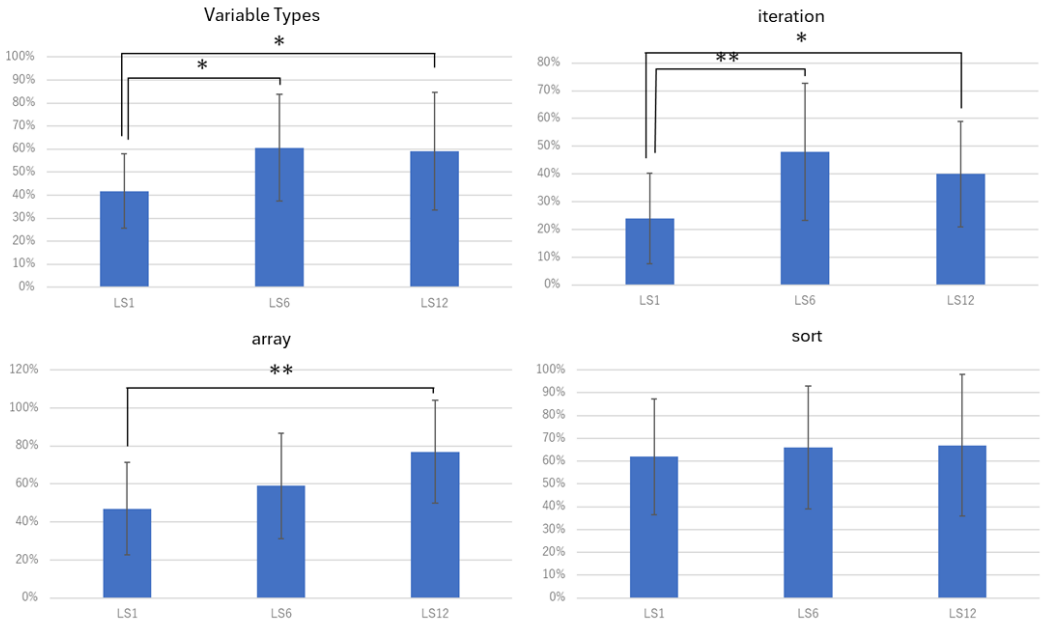

Figure 12: F-value t-test (by question)

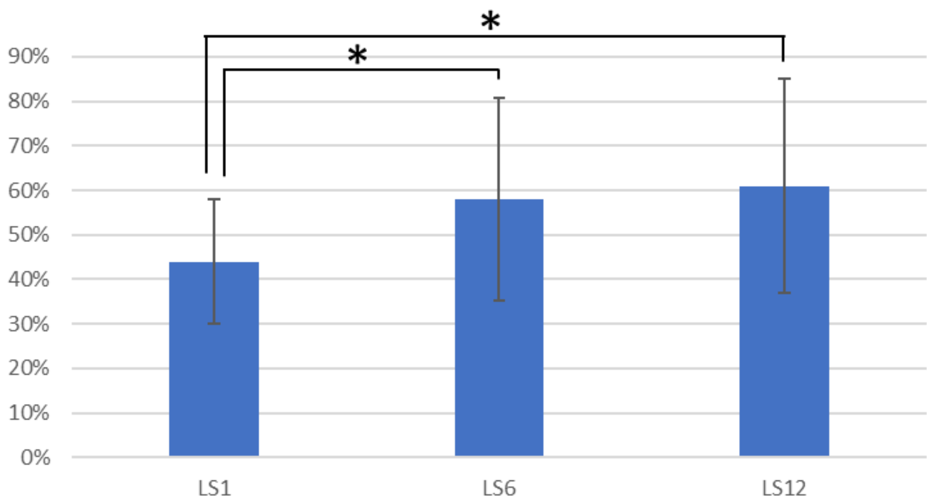

Figure 13: Mean value for all questions for all subjects

4 Conclusion

In this study, we verified the effectiveness of introducing the concept of taboo choices in the evaluation and visualization of the learning process in programming learning using COPS. We set one of the dummy choices as a taboo choice and evaluated and analyzed it by imposing a heavier penalty on the Levenshtein distance when it was selected. We presented three patterns of visualization results with different weights for the penalty of taboo choices to participants simulating teachers and listed up learners suspected of selecting taboo choices. The detection accuracy was evaluated using F-tests and Welch's t-tests (two-sided). As a result, the detection accuracy of proposed methods 1 and 2, which set taboo choices, significantly exceeded that of the conventional method without taboo choices. This indicates that the introduction of the concept of taboo choices in evaluation and visualization improves the detection accuracy of learners engaged in inappropriate learning activities. However, no significant difference in detection accuracy was observed depending on the weight of the penalty for taboo choices, suggesting that excessively large penalties may not lead to improved detection.

Acknowledgement

This work was partly supported by Grant-in-Aid for Scientific Research (C) No. 22K02815 and No.23K02697 from the Japan Society for the Promotion of Science (JSPS).

References

[1] S.Matsumoto, Y.Hayashi, and T.Hirashima, Development of a program-ming learning system through card manipulation focusing on thinking about relationships between parts, Transactions of the Institute of Electrical Engineers of Japan C, Vol.138, No.8, pp.999-1010 (2018).

[2] M.Makino, Data Mining in e-Learning, Transactions of the Japan Society for Educational Technology, vol.31, no.3, pp.271-283 (2007).

[3] S.Garner, A Tool to Support the Use of Part-Complete Solutions in the Learning of Programming, Proceeding deconference, pp.222-228(2001).