IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol.006, LIIR347
DOI: https://doi.org/10.52731/liir.v006.347

Estimating Human Difficulty for SameGame Puzzles

Motoki Miura *

Abstract

SameGame is a computer puzzle game in which players must select adjacent blocks of the same color vertically and horizontally to remove them from the screen. Generally, initial blocks are arranged randomly, but it is known that the difficulty varies depending on the initial blocks. It is also known that it can be impossible to completely remove all blocks on small boards. As a way to quantitatively estimate the difficulty of SameGame puzzles, we devised an index that quantifies the correspondence between blocks when they are removed and the discrete state of those blocks. We compared our proposed index with a random success rate, which examines the ratio of the number of answer steps divided by the number of all possible steps. By investigating the correlation with the clear rate based on human play data, we analyzed the effectiveness of several indices including the proposed method.

Keywords: Difficulty Estimation, Puzzle Game, SameGame, Clickomania

1 Introduction

SameGame is a computer puzzle game based on ”ChainShot!” released in 1985. It is also known as Clickomania or Swell-Foop. The game usually starts with blocks placed randomly on a 20×10 grid, as shown in Figure 1. When a player selects a block, adjacent blocks of the same color disappear at the same time. After these blocks are removed, remaining blocks fall down (Figure 2). When a vertical row of blocks is removed, the blocks on the right side of the row move left to fill the empty row.

1.1 SameGame rules and score calculation method

In a normal SameGame puzzle, the main objective is to get a high score. The score is calculated as $(n-2)^2$, where n is the number of blocks to be removed in one selection. Therefore, to get a high score, it is necessary to remove as many blocks as possible at once. In a normal SameGame puzzle, if you remove all the blocks you get bonus points, and you can continue the game from a randomly placed block state while keeping your score. However, since blocks are generated randomly, some stages are impossible to remove all the blocks in the end.

* Chiba Institute of Technology, Chiba, Japan

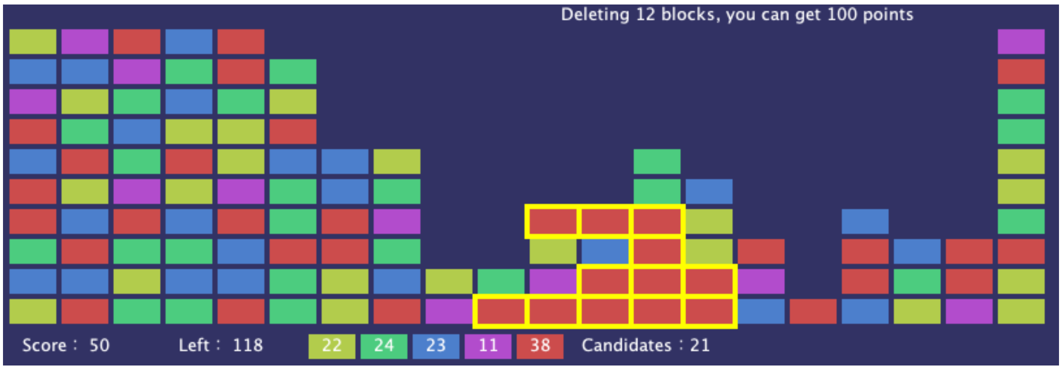

Figure 1: Standard SameGame puzzle (before removing the highlighted blocks)

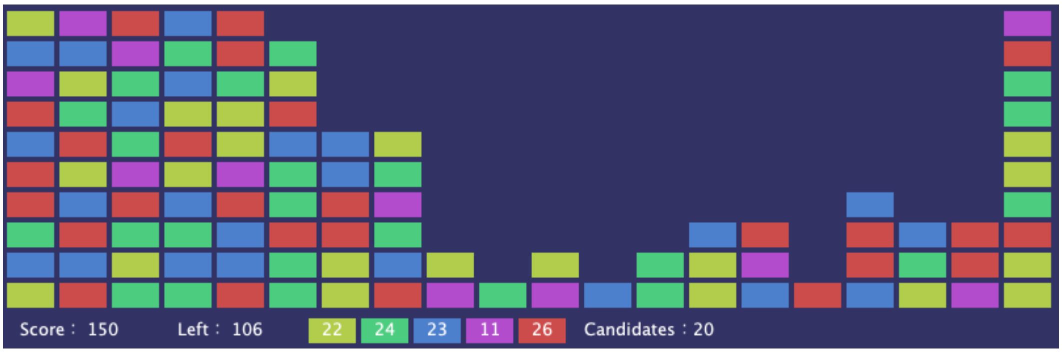

Figure 2: Stage after clearing highlighted blocks of Figure 1

1.2 Proposal for a SameGame puzzle with a guaranteed solution to be removed all of the blocks

Therefore, as a variation of SameGame, we improved a SameGame puzzle with a guaranteed solution to be removed all of the blocks. By providing stages with a guaranteed solution, we can improve the game experience for players.

We also focus on and discuss its human difficulty as a fundamental characteristic. We believe that by presenting the stages in order from easier to harder, based on the idea of small steps, players can gradually become more adept at the game strategy. To achieve this, we considered how to estimate the difficulty of the prepared stages.

A method has been proposed to efficiently generate problem sets for improvement of the Puyo Puyo puzzle[1], which is a type of puzzle in which pieces fall like Samegame. However, no similar research had been conducted on Samegame.

Here, we consider a puzzle stage in which blocks of four colors are placed on a 5×5 board. The reason for considering a smaller board is that it takes longer to search the solution space, and we want to ensure that the game is not too difficult for humans to solve.

As mentioned above, the stages are generated randomly in a typical SameGame puzzle. However, to guarantee solution to be removed all of the blocks, we prepare stages in advance. Previous research [2] has shown that when two types (two colors) of blocks are randomly generated, the probability of clearing all blocks increases as the length and width of the blocks increase; however, when the number of colors increases or the length and width of the board decreases, it is not guaranteed that all blocks can be cleared in the end. Regarding the difficulty of the problem, [3] has shown that even a stage with only two colors and two rows is NP-complete.

2 Exploring the Solution Space of 5×5 Sized SameGame

First, we randomly placed blocks of four colors on a 5×5 board, then searched the solution space to extract stages that had one or more solutions to be removed all blocks. Note that we excluded stages that were identical due to color substitution, as they were duplicates.

As a result, we obtained 3,027 guaranteed stages[1] in the above process. For each stage, we enumerated all possible paths, then calculated a random success rate and a path depth. The random success rate (RSR) is a proportion which is the number of paths that clear all blocks divided by the number of all possible paths a player can take. Statistically, the higher the RSR, the higher the clear rate when randomly selecting blocks to be removed. The path depth means the number of selections of deleting blocks until all of the blocks removed and cleared. We calculated a path depth average for each 3,027 stage, as a representative value of the stage.

Figure 3 shows the histogram of the path depth average. The mean value of the path depth average was 8.94.

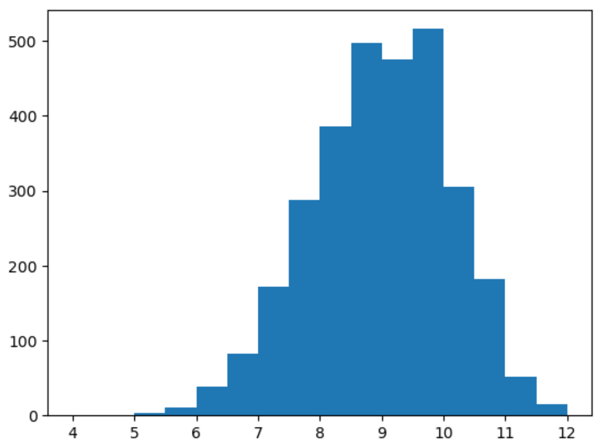

Figure 3: Frequency distribution of the path depth average

Next, we analyzed the RSR. Figure 4 shows the histogram of the RSR. We found that stages with a low probability (it may be difficult for human) were more likely to be generated. Regarding the two factors, we confirmed the relationship between the RSR and the PDA. The scatterplot of the RSA and the PDA (Figure 9 in appendix) revealed no notable bias.

Initially, we used the RSR as a provisional measure of difficulty, but when we actually played the game, we found that some stages were easy to solve even if the RSR was low. Therefore, in this paper, we devised a method that takes into account block movement and adjacent relationships to define a more appropriate level of difficulty. We also verified the validity of this method based on actual user play data.

[1]Note: We did not list all the 5×5 stages that can be solved.

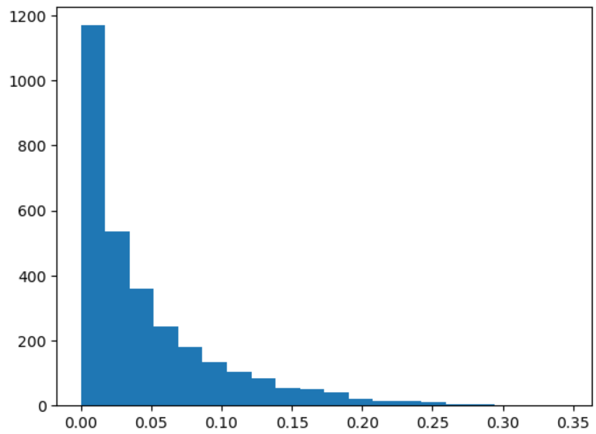

Figure 4: Frequency distribution of the random success rate (RSR)

3 Proposed Difficulty Index

When considering the difficulty of the SameGame puzzle, in which the objective is to erase all blocks, we focused on the discrete state of adjacent blocks when they are erased. In easy stages, many blocks of the same color are adjacent to each other in the initial state, and there is often no need to move the blocks to be erased closer together. Conversely, in difficult stages, blocks that need to be erased in correspondence with each other are scattered throughout the stage, and even if adjacent blocks are erased early on, players may get stuck. In the initial stage, leaving adjacent blocks unerased is generally considered to be more difficult as the number of the adjacent blocks increases, as this is counterintuitive. Therefore, we thought that by quantifying this "discrete state of corresponding blocks," we could more accurately estimate the difficulty of the SameGame.

We will explain the calculation method of the proposed index using Figure 5. This figure shows an answer check viewer we created to confirm correct answers for each stage of SameGame. The viewer shows one of the solutions that requires the fewest number of deletions among all solutions that can erase all the blocks.

Now, we show the answer of stage 2, which has high RSR (0.317) but relatively difficult for human to solve (see Figure 5). The figure contains eight phases of the stage. The upper left shows the initial phase of the stage, and the area highlighted in yellow shows the blocks to be removed at this phase. Each block has a number and a coordinate displayed on it. The number $(0,1,2,\cdots)$ in the upper left of each block indicate in which step (at which stage) the block will be erased in the currently displayed solution. The coordinate in the lower right shows the initial position of the block in (x,y).

For example, in the initial state in the upper left of Figure 5, the blocks to be erased in the fourth step (they are painted cyan and are labeled 4) are discrete as $(2,0)$, $(3,2)$, and four-connected blocks ($(2,4)$, $(3,4)$, $(4,4)$, and $(4,3)$). Additionally, the blocks painted green that are eliminated in the seventh step are also discrete ($(1,0)$, $(1,4)$, $(5,1)$, and $(5,4)$) and can be seen to be distributed throughout the stage.

The double circle (◎) on (or near) the block indicates the coordinate position of the corresponding block(s). If it is a single block (non-adjacent or different number labels), the coordinate position is a center of the block. If multiple blocks are adjacent, the double circle indicates their "center of gravity." The index of "dispersion of corresponding blocks"

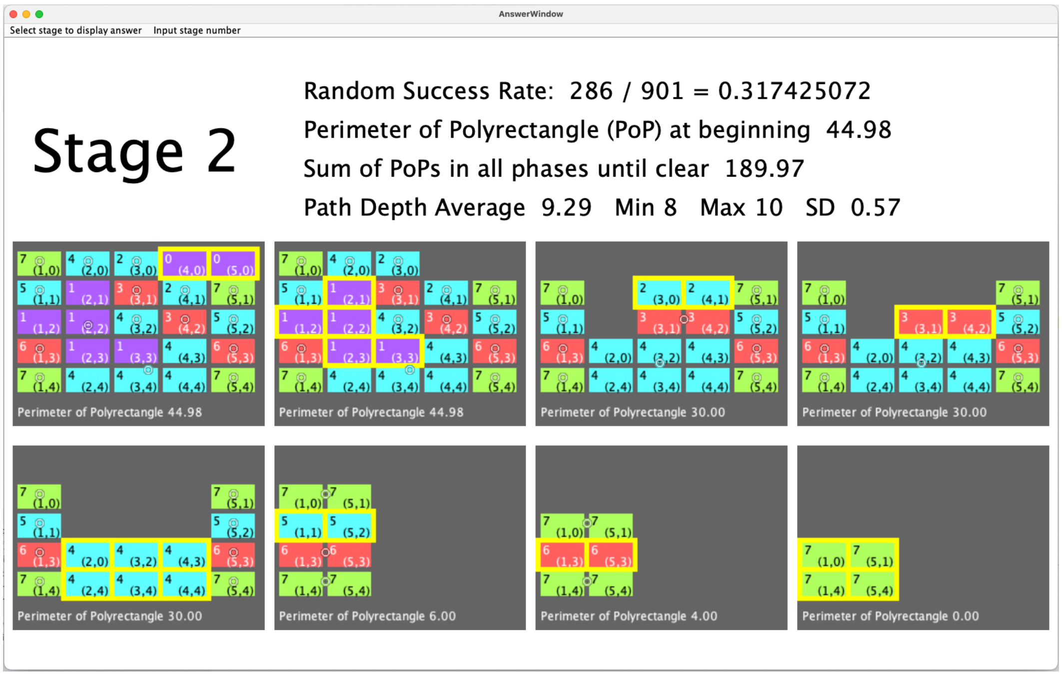

Figure 5: An answer checker that displays one of the correct answer with shortest path depth. This view also shows the calculation process of the proposed discrete index. The top of the view shows the random success rate, the perimeter of polyrectangle (PoP) at beginning, and sum of PoPs until clear, and the average, minimum, maximum, and standard deviation of the path depth of the answers (in this stage, 286 solutions).

proposed in this paper is calculated by the perimeter distance of the convex hull, which is the degree of dispersion of the block coordinates to be deleted at each stage (or their centers of gravity, if the blocks are adjacent rather than single). However, if the coordinates are only one (all corresponding blocks are adjacent and deleted in one operation), the perimeter distance is 0.

In this study, we calculated two types of candidates: the perimeter distance in the initial state and the total value of the perimeter distance at each stage of the solution, and decided to consider the appropriateness of the difficulty. The reason for calculating the total value as a candidate is that the discrete state of corresponding blocks changes from time to time as the stages progress, which is thought to easily affect the difficulty felt by the player.

The top of the Figure 5 also shows the number of solutions that can be completely cleared in this stage, the number of all solutions, the random success rate, as well as the average, minimum, maximum, and standard deviation of the path depth (number of deletion until clear) in all solvable solutions (in this example, 286 solutions).

4 Verification

In order to collect the difficulty perceived by human players, we created a web application of SameGame (see Figure 6) that collects play logs. The SameGame application has 375 stages, which are arranged in order of the random success rate. Players basically play in this order. The percentage of times the stage was cleared out of the number of times the stage was started was defined as the “clear rate,” and this was used as an estimate of the difficulty perceived by human players. We collected the play data from November 28, 2023 to Octover 15, 2024. The data had 4,281 entries. However, stages that were cleared less than twice were excluded. Finally, we analyzed remaining 172 stages were analyzed. When a browser accessed the site for the first time, a random string ID was issued in the cookie. The number of clients (types of unique IDs logged) that played at least once was 298.

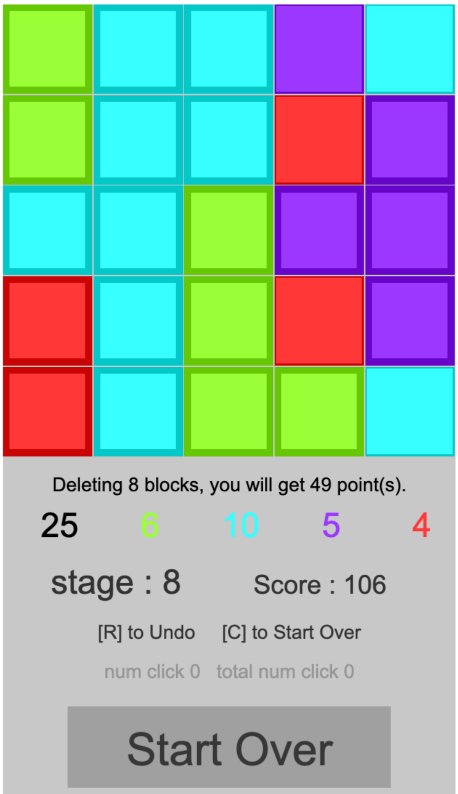

Figure 6: Screenshot of the 5×5 sized Web SameGame for the Experiment.
`https://p5.istlab.info/zenkeshi_en`

Figure 7 shows the scatterplot matrix of the random success rate (RSR), the perimeter of polyrectangle (PoP) at start, the sum of the PoPs, and the clear rate in actual play data, as well as the correlations and p values in the test for no correlation. The p value is rounded to the fourth decimal place. The number of * to the right of the p value indicates the number of equations that are satisfied out of the three inequalities $p < 0.05$, $p < 0.01$, and $p < 0.001$.

In our previous study [4], the RSR was used as a provisional measure of difficulty. The upper right of Figure 7 (correlation coefficients $r = 0.23$, $p = 0.003$) show that there is weak correlation between the RSR and the clear rate. For the PoP at start, no correlation ($r = -0.10$, $p = 0.21$) was observed with the actual clear rate. For the sum of PoPs, a weak correlation ($r = -0.20$, $p = 0.010$) was observed with the actual clear rate. From these results, it can be said that the sum of PoPs is effective than the PoP at start, and the effectiveness is similar to the RSR for estimating the clear rate as a measure of difficulty.

Figure 8 shows the scatterplot matrix of the path depth indices (min, average, and max) and the clear rate, as well as the correlations and p values in the test for no correlation. From this matrix, we can see that the correlation coefficients of minimum path depth with

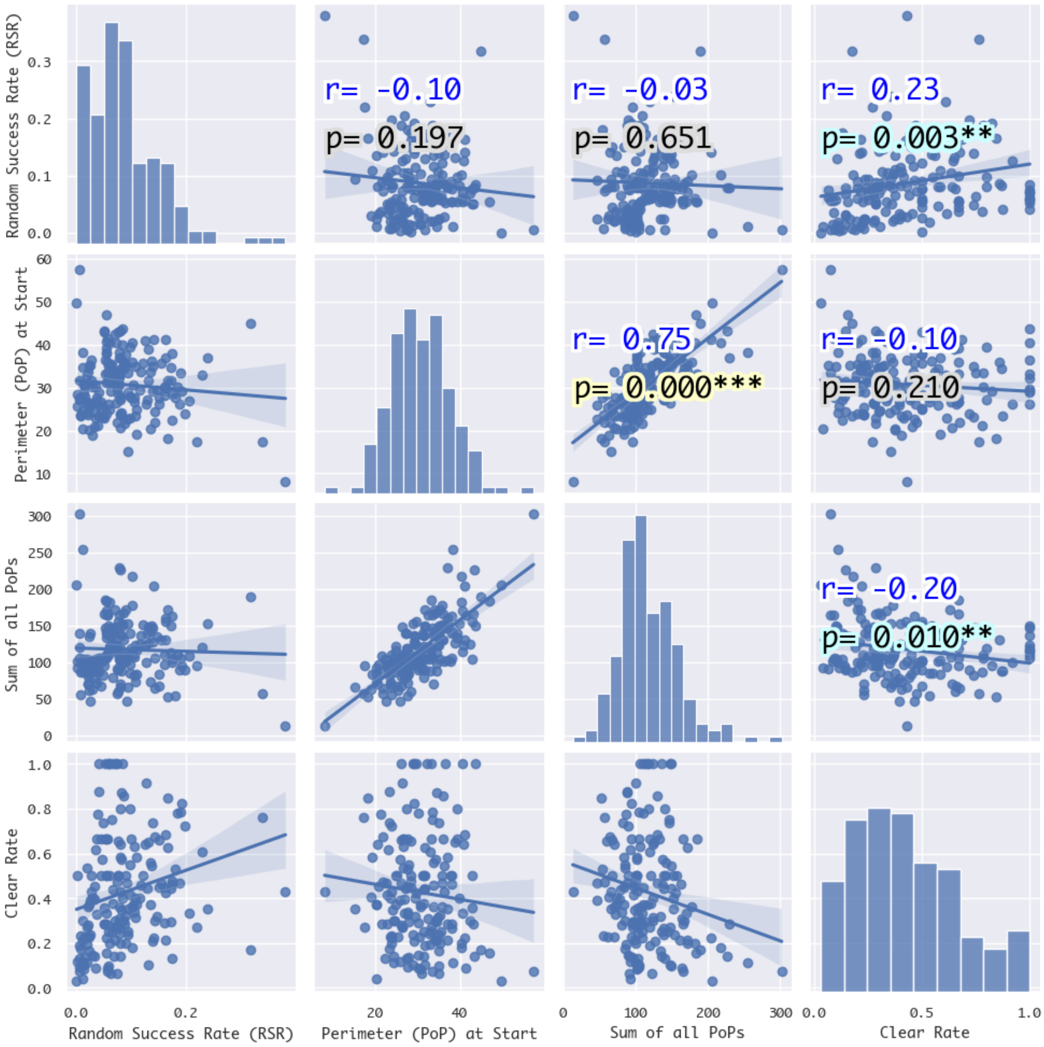

Figure 7: Scatterplot matrix of the random success rate (RSR), the perimeter of polyrectangle (PoP) at start, the sum of all PoPs, and the clear rate

the clear rate ($r = -0.26, p = 0.000$) shows a weak correlation. However, the average and maximum of path depth did not show significant correlations.

In particular, the correlation between the minimum path depth and the clear rate is larger than the correlation of the sum of PoPs discussed earlier, so this is a useful indicator for estimating the clear rate as a measure of difficulty. However, since the minimum path depth is a discrete value, it is possible to improve the accuracy of the estimation by referring to the sum of PoPs and the RSR when considering the difficulty within a group of problems with the same minimum path depth.

5 Discussion and Summary

We investigated the correlation between the clear rate when humans solve the four-color 5×5 sized SameGame puzzle and several indices that can be calculated in advance. As a result, we were able to confirm the effectiveness of the indices including our proposed

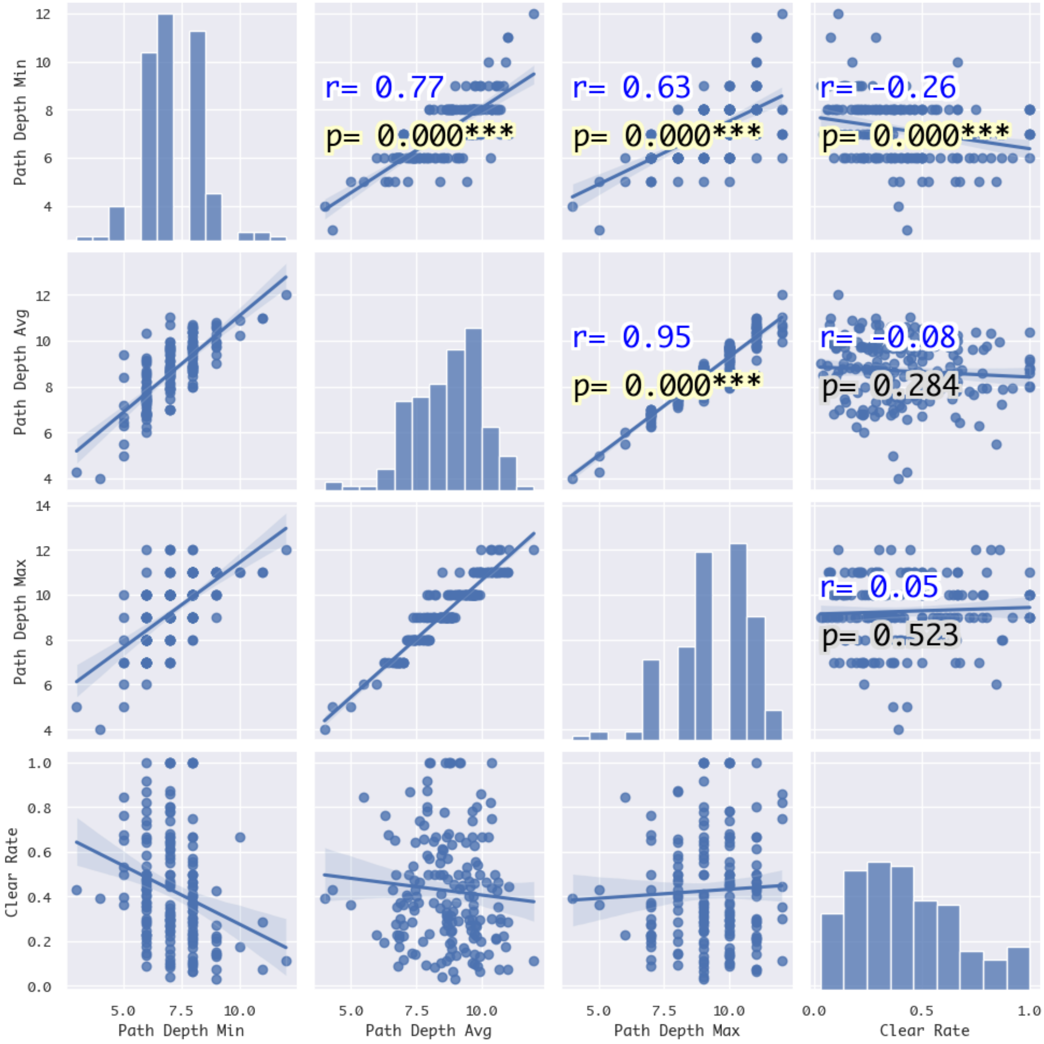

Figure 8: Scatterplot matrix of the path depth indices (min, max, and average) and the clear rate

method. From the evaluation, the minimum path depth was the best index for estimation. However, the correlation value was not high. We will continue to look at better metrics and gather more data, and will continue to evaluate based on that data.

In this study, the sum of PoPs was calculated using only one of the solutions with the minimum number of operations. Therefore, we would like to consider indicators that take into account all possible solutions.

Acknowledgments

The part of this research was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research JP22K12319.

References

[1] Ryutaro Takahashi and Kokolo Ikeda. Mating Problem Generation of Puyo-Puyo for Training. In *IPSJ SIG Technical Report, Vol.2018-GI-39, No.10*, pages 1–7, March 2018. (in Japanese).

[2] Frank W Takes and Walter A Kosters. Solving SameGame and its Chessboard Variant. In *Proceedings of the 21st Benelux Conference on Artificial Intelligence, BNAIC*, pages 249–256, 2009.

[3] Aviv Adler, Erik D Demaine, Adam Hesterberg, Quanquan Liu, and Mikhail Rudoy. Clickomania is Hard, Even with Two Colors and Columns. *The Mathematics of Various Entertaining Subjects: Research in Games, Graphs, Counting, and Complexity*, 2:325, 2017.

[4] Motoki Miura. A basic study of game characteristics of the SameGame puzzle in which all the blocks can be erased. In *WISS2023 Web Proceedings [3-C05]*, *2 pages*, 2023. (in Japanese).

A Relationship between random success rate and path depth average

Figure 9 shows a scatterplot of random success rate (Figure 4) and path depth average (Figure 3).

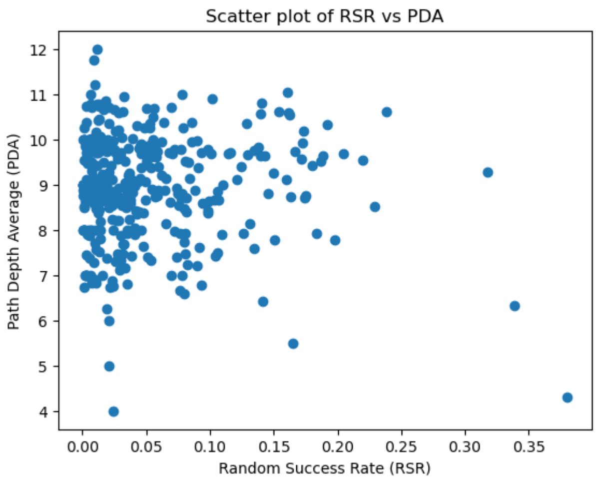

Figure 9: Scatterplot of random success rate and path depth average