Creation of Training Data and Training for Prediction Model of Curling Scores Using Real Game Data

Tomoya Iwasaki[1], Wataru Noguchi[2][a], Yasumasa Tamura[3][b], Shimpei Aihara[4][c]
and Masahito Yamamoto[3][d]

[1] *Graduate School of Information Science and Technology, Hokkaido University, Hokkaido, Japan*
[2] *Education and Research Center for Mathematical and Data Science, Hokkaido University, Hokkaido, Japan*
[3] *Faculty of Information Science and Technology, Hokkaido University, Hokkaido, Japan*
[4] *Department of Sport Science and Research, Japan Institute of Sport Sciences, Tokyo, Japan*

Keywords: Curling, Digital Curling, Curling AI, Expectimax, Evaluation Function, Wining Probability, Game Tree Search, Results Book.

Abstract: Curling is a sport in which two teams take turns shoting stones at each other on an ice field to compete for total scores. Curling is a highly strategic sport, and the strategy of stone delivering has a significant impact on the outcome of the game. To verify strategy of curling, "digital curling" is a platform that reproduces curling on a computer. Following the previous research of curling AI using game tree search and evaluation function by Ataka et al., real game data was obtained and trained into a neural network of evaluation function. In this study, we propose a method to obtain stone position information from real game data. Also, the model was trained from the obtained data. The results show that models trained with realistic data correspond better to realistic situations than conventional models trained with data generated by algorithms. However, in situations where there are many stones on the sheet, the model was also found to be insufficiently accurate as is the case with conventional models.

1 INTRODUCTION

Curling is a sport played on a field of ice, in which two teams take turns delivering stones and compete for the final total score. In curling, not only the skill of delivering the stones to the target position, but also the strategy of where and how to deliver the stones has a great influence on the game result. Because of this highly strategic aspect, curling is also called "Chess on Ice".

To evaluate curling strategy, a computational simulation platform called *digital curling* has been utilized recently. The digital curling platform enables AI-based curling players to play against each other (Uehara and Ito, 2021). The strategies verified by this simulator are expected to be applicable to real curling, and curling AI researches are active on this platform.

Various approaches have been attempted to study curling AI by digital curling. Among them, Yamamoto et al. discretize the the field (called "sheet" in curling) into a grid and use the game tree search (Yamamoto et al., 2015). And Katoh et al. use the Expectimax search algorithm to perform game tree search in curling with uncertainty (Katoh et al., 2016). These studies have shown that the search for candidate shots in continuous and uncertain curling games is possible with game trees. In addition, Yamamoto et al. develop a neural network based state evaluation function and showed that it is more effective than hand-crafted evaluation function (Yamamoto et al., 2018). Furthermore, Ataka et al. added the game situation to the input of the neural network to allows score prediction based on the game situation such as score difference (Ataka et al., 2020).

In this study, we aimed to utilize game data from actual curling competitions for training a score prediction model. In previous research used data automatically generated based on algorithms or obtained from self-competitions between curling AIs for training data. However, such automatically generated data may be inadequate as training data in that it is far from the actual game phase and lacks diversity. Therefore, we aimed to utilize game data from actual curling competitions to train a score prediction model capable of responding to various realistic situations. In the ex-

[a] https://orcid.org/0000-0002-0250-4128
[b] https://orcid.org/0000-0001-5406-1966
[c] https://orcid.org/0000-0002-8513-0204
[d] https://orcid.org/0000-0002-7326-3691

Iwasaki, T., Noguchi, W., Tamura, Y., Aihara, S. and Yamamoto, M.
Creation of Training Data and Training for Prediction Model of Curling Scores Using Real Game Data.
DOI: 10.5220/0012941100003828

In *Proceedings of the 12th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2024)*, pages 168-179
ISBN: 978-989-758-719-1; ISSN: 2184-3201

periment, we collected two sets of training data, one from actual game data and one from algorithms used in previous research, and compared the performance difference due to the difference of training data.

2 CURLING

Curling is played on a field of ice called a “sheet”. An overview of the sheet is shown in Fig. 1. Sheet has a circular area called a “house” and kickstand called a “hack”. Delivering a stone from hack side to the opposite house side is called a “shot”. Both teams shoot stones in an attempt to leave more stones in the house. When shoting, the player must release his or her hand from the stone before reaching the “hog line” on the hack side. The stones are active if they remain stationary between the hog line and back line on the house side. Otherwise, the stones are regarded as invalid and removed from the sheet. The sheet can be swept with a broom after the stone has been shot. This allows the trajectory of the stone to be adjusted.

An inning in a game is called an “*end*”, and a game usually consists of eight or ten *ends*. The team with the highest total score at the end of final *end* wins. If both teams are tied, the game continues in extra ends until one of teams wins a score difference. It is also possible to resign the game in the middle of the game, which is called “concede”.

In each *end*, both teams take turns shoting eight stones. An *end* ends when the second team shots the last stone. This last stone is called the “hammer”. To allow more stones to be stored on the sheet, the “free guard zone rule” exists. This rule prohibits the stones in the free guard zone (area between hog line and tee line, excluding house) from leaving the playing area until the fifth shot of overall is completed. It is possible to move stones slightly as long as they are not moved out of the playing area. If this rule is violated, the shot stone will be removed and the moved stone will be returned to its original position. At the end of an *end*, only the team with the stone closest to the center of the house among the stones in the house gains scores. Scores are awarded for the number of stones inside the house that are closer to the center of the house than the opposing team’s closest stone. The team that awarded scores takes the first shot in the next *end*. If there is no stone inside the house, no scores are awarded to either team. This is called a “blank *end*,” and the first team in a blank *end* plays first in the next *end*. In other words, the playing order switches if either team gains one or more score in the previous *end*.

As mentioned above, in curling, the score is determined when the last shot of the second offensive team is completed. In this sense, curling is a sport in which the second team has an overwhelming advantage. Therefore, it is necessary to adopt different strategy for the first and second offensive teams. Generally, the first team aims to let the second team to score one point, so that the team will get the hammer in the next *end*. Alternatively, the first team aims to gain scores instead of the second team called “steel”. The second team, on the contrary, aims to score two or more points or, to maintain hammer by removing all stones from the house to make it a blank *end*. Both teams proceed with the game, constructing strategy to achieve the desired stone position for their team at the end of the *end*. Each team must decide on their strategy and shot stone within the allotted time limit.

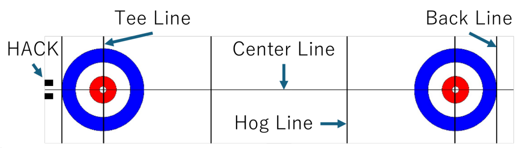

Figure 1: Overview of the sheet. The player shots a stone in the direction of the opposite house from the hack. The distance from the hack to the hog line on the hack side is 33 feet. The distance to the opposite hog line is 105 feet, to the tee line is 126 feet, and to the back line is 132 feet.

3 DIGITAL CURLING

Digital curling is a platform for reproducing curling on a computer. Digital curling enables curling AIs to play against each other. Games in digital curling are conducted through communication between the digital curling simulator and each AI. The simulation is performed by a physics engine. The game is played according to the rules of real curling.

The curling AI sends shot information to the digital curling simulator, which simulates the shot and returns the result to the AI. The AI sends shot information, including the 2-dimensional vector and direction of rotation, to the simulator. Then the simulator runs a simulation based on the current game information and the input shot information, and sends the results to each AI. The information from the simulator includes the position of each stone, the current score, and the team that will play the next shot. The positions of the stones are represented by a two-dimensional coordinate system. Simulations are performed at discretized regular intervals by a physics engine that calculate trajectories and collision detection. The results of the simulation can be viewed on the GUI as shown in Fig. 2.

The number of *ends* and AI thinking time can be set arbitrarily. The presence or absence of the free guard zone rule can also be set. These settings are disclosed in advance in curling AI competitions using Digital curling.

In real curling, due to the influence of sheet conditions and the skill of the players, it is not always possible to place the stones precisely at the target position. In real curling game, this uncertainty must be taken into account when making strategic decisions. To reproduce this, a random fluctuation is added to a given shot vector in digital curling. The setting about the scale of random fluctuation can be set for each player. In competitions, the setting is disclosed.

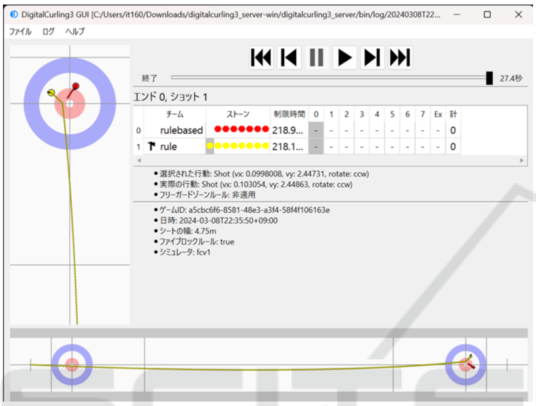

Figure 2: Simulation result on GUI. The trajectories and results of all shots during a digital curling game can be viewed on GUI.

4 CURLING AI IN DIGITAL CURLING

Previous research (Katoh et al., 2016) has shown that game tree search and evaluation functions can be used to build an AI that is superior to a curling AI built on rule based methods. In this study, we construct an evaluation function with reference to the curling AI of Ataka et al. The evaluation function consists of an expected score distribution and a winning probability table. The expected score distribution is calculated from the game situation using a score prediction model.

The score prediction model needs to be trained with curling game information such as stone positions. In the previous research, the model was trained by data generated by the algorithm. On the other hand, in this study, we aim to utilize data from real curling competitions in order to make the Curling AI to adapt diverse and realistic situations.

4.1 Search in Digital Curling

The evaluation function maps a given game situation to a quantitative value. Ataka et al. have shown that the expected winning probability can be used as the evaluation value, allowing evaluation according to the situation of the curling game. The evaluation function consists of a score prediction model and a winning probability table.

The expected score distribution is the probability distribution of the score obtained at the end of an *end*. In digital curling, the score obtained at the end of an *end* is uncertain because of the uncertainty involved in the shot. Also, in curling, the winning probability significantly depends on the number of scores obtained. Therefore, the distribution of expected score is used instead of simply relying on the value of expected score.

A winning probability table is a table showing a team's winning probability based on the number of remaining *ends* in a game and the difference in total score to the opponent team. In curling, the optimal strategy can vary greatly based on the number of remaining *ends* and the score difference. In particular, maximizing the number of scores for a single end does not always lead to maximizing the winning probability, because the blank *end* is an effective strategy for the team having hammmer. The winning probability table makes it possible to deal with such situations. By combining the winning probability table and the expected score distribution, the expected winning probability at the end of that *end* can be computed. Table. 1 shows the winning probability table used in this study. It was obtained by self-competition between curling AIs in previous research and shows the winning probability of the first team in each situation.

Table 1: Winnig probability table to be used. The table shows the winning probability of the first team for each score difference and the number of remaining *ends*.

		remaining *ends*			
		3	2	1	0
Difference of score	3	0.919	0.946	0.962	1.000
	2	0.771	0.794	0.881	1.000
	1	0.609	0.557	0.677	1.000
	0	0.340	0.279	0.260	0.220
	-1	0.162	0.122	0.042	0.000
	-2	0.034	0.021	0.011	0.000
	-3	0.015	0.014	0.011	0.000

4.2 Score Prediction Model

The score prediction model computes the expected score distribution at the end of an *end* based on the game situation without simulating all possible shots until the end of the *end*. The model is constructed by a simple fully-connected neural network. As input data for the model, positioning information of stones on the sheet and information of the game situation are used. The information for each stone, as shown in Table2, is inputted in order of proximity to the center of the house for all fifteen stones. Information of stones that do not exist on the sheet are all set to 0. Information of the game shown in Table3 is also input to the model. These are all one-hot vectors. The output of the model is an 11-dimensional distribution of expected scores. This indicates the probability of scoring from -5 to 5 scores at the end of *end*. The activation function of the output layer uses softmax.

Table 2: List of information of stones. Coordinates and distance from the center of house are represented as continuous values, while other data are represented as discrete values.

Information	Value
x coordinate	[-2.375, 2.375]
y coordinate	[32.004, 40.234]
Distance from center of house	[0, 6.78]
Is there stone in the play area	0, 1
Is there stone in the house	0, 1
Owned player	-1, 0, 1
Is there enemy stone on the inside	0, 1

Table 3: List of information of game. These information are input in one-hot vectors.

Information	Value(one-hot vector)
Score difference	-2,-1,0,1,2
Remaining *ends*	0,1,2,3

4.3 Creation of Training Data

In this study, training data was created only for the last shot of the *end*. Because the score of the *end* is determined when the last shot is completed, so multiple stages of simulation are not necessary, and the expected score distribution can be obtained with a small number of simulations.

First, the number of remaining *ends*, score difference, and stone position data are given as input data for the model. In this situation, all candidate shots are simulated on the digital curling. Each candidate shot is a vector of shots that reach each score on the sheet divided on the grid. Grid is a square and its size is the radius of the stone. No random fluctuation are added to the simulation at this time. The score is calculated from the stone position obtained after simulation. Based on the points scored, the score difference in the game situation and the number of remaining *ends*, the expected winning probability is calculated using the winning probability table. The formula for this calculation is shown in Eq. 1.

$$E(x,y) = \sum_{\Delta x=-3}^{3} \sum_{\Delta y=-5}^{5} p(x',y')w(r,d,s(x',y')) \quad (1)$$

$$\text{where} \quad x' = x + \Delta x, \quad y' = y + \Delta y$$

$p(x',y')$ is the probability of the shot stone reaching the grid around the candidate shot after the simulation when a random fluctuation is added to the candidate shot. Δx *and* Δy are the deviations from the grid of the candidate shot. We consider 3 grids in the x direction and 5 grids in the y direction centered on the grid of the candidate shot. Since the random fluctuation added to the shot are predetermined, the probability of the stones reaching the surrounding grids can be determined in advance. $w(r,d,s)$ is the expected winning probability when the fluctuation of remaining *ends* r, the score difference d, and the score at the end of the *end* is s. This can be obtained from Table. 1. $s(x',y')$ is the score at the end of the *end* when a shot is made at x' *and* y', obtained from the simulation without random fluctuation described above.

Thus, we obtain the expected winning probability for all candidate shots. The candidate shot with the highest expected winning probability obtained is considered the best shot. Finally, we simulate this best shot multiple times with the original stone arrangement. Since random fluctuations are added to this simulation, a probability distribution for each score can be obtained. This probability distribution is the expected score distribution as the target output, corresponding to the game situation as the input.

As an example, the number of remaining *ends* is 1, the score difference is 0, and the stone position is the data shown in Fig. 3. Table. 1 shows that the winning probability in this situation is about 67% if one point is gained, 74% if a blank end is assumed, and about 4% if one point is lost. In other words, blank end and scoring one point are high winning probability actions, with blank end having the highest value. In the situation, a shot as shown in the Fig. 4 is the best shot in this situation because the expected winning probability according to Eq. 1 is the highest. After simulating the best shot multiple times, the expected score distribution is as shown in Fig. 5. Fig. 5 shows an example situation that can be blank *end* about 80% of the time, but it is also a situation that a single point be scored . It also shows that there is almost no risk of steal. This means that the best shot in this situation will succeed 80% of the time, and even if it fails, there is little possibility of steal.

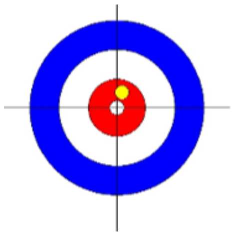

Figure 3: Example of stone position. There is only one stone of the opposing team near the center of the house. The situation of the game is that the number of remaining *ends* is one, the score difference is 0, and the next shot is the last shot of the *end*.

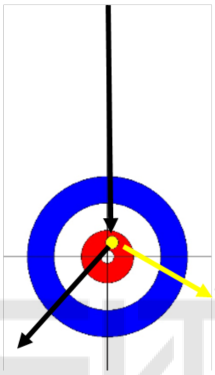

Figure 4: Example of shot. The black arrows are the shot stone trajectories. The yellow arrow is the trajectories of stones that move in collision with the shot stone.

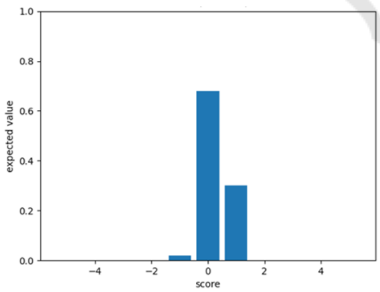

Figure 5: Expected score distribution in the example. The expected score distribution indicates that this situation is almost risk-free, but the success of the best shot (blank *end*) is not assured.

4.4 Example of Training Data

By creating training data using the method described in the previous section, it is possible to create training data that corresponds to the situation in which strategies change significantly depending on the game situation.

First, let us assume that the common situation is that the team has the red stone, next shot is the last shot and the stone position as shown in Fig. 6, With this stone position, it is easy for a red team to score one point by simply placing a stone in the center of the house, but relatively difficult for a team to score two points because it requires double takeout of the yellow stones. Blank *end* is almost impossible because team need to take out all three stones in the house and the shot stone must not be left in the house. On this case, the first situation is set with a score difference of 0, *end* 9th, and the second situation is set with a score difference of -2, *end* 10th.

In the first situation, the team should try to score one point safely, because if the team is stolen points, winning probability will significantly decrease. In the second situation, the team must attempt to score two points, even though there is a high risk, because if two points are not scored, the game is lost at that point. The expected score distributions obtained by the method in the previous section are shown in Fig. 7 for the first situation and in Fig. 8 for the second situation. The first situation shows that one point can be obtained reliably. The second situation shows that the success rate is low, but two points can be obtained. Thus, the expected score distribution in training data can take into account situations where high risk/high return is required and situation where low risk/low return is required.

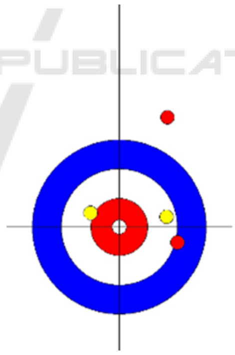

Figure 6: Position before last shot. The opposing team has the No. 1 and No. 2 stones and own team has the No. 3 stone.

5 TRANING DATA EXTRACTION FROM REAL CURLING GAMES

The data information required to compose the training data are stone position, score difference, and number of remaining *ends*. We obtain stone placement data from real games, instead of utilizing data extracted from self-playing by AI-based players or automatically generated based on algorithms.

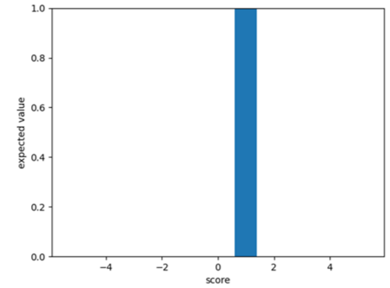

Figure 7: Expected score distribution for situation required low risk strategy. The distribution shows that the best shot is sure to succeed.

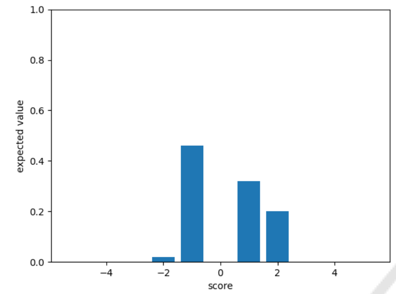

Figure 8: Expected score distribution for situation required high risk strategy. The distribution indicates that the best shot is the shot with the greatest risk.

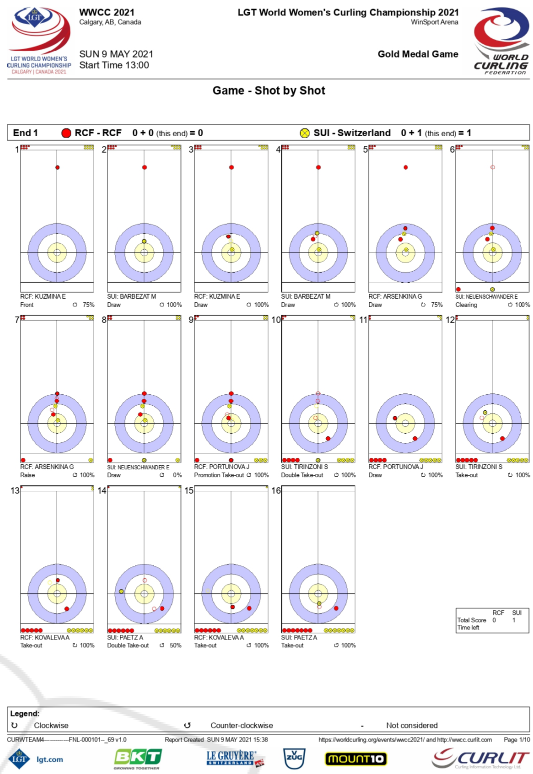

Figure 9: Example of Shot by Shot. The stone positions of the sheets are listed as images in the order of the shots.

5.1 Results Book

Game data was obtained from the "Results Book" provided by Curlit, a PDF file containing information on major curling competition in the world (Curlit, 2024). The Results Book contains pages called "Shot by Shot", which show the sheet information including the location of the stones in each shot, shown in the image (Fig. 9) (Curlit, 2022) . The coordinates of the stones are obtained from these images.

5.2 Data Extraction from Results Book

We extract three types of data: the stone positions, the first shot team in the corresponding *end*, and the shot number. These data are stored in the database. An existing work (Myslik, 2020) provided a method to extract those data from the earlier version of Result Books. Based on Myslik's work, we implemented an enhanced script compatible with the latest version of Result Books.

Data extraction is done as following. First, we split the Results Book by game. An image is taken from the shot by shot page showing stone position in shot order. The image obtained at this time is shown in Fig. 10. This image shows the number of stones remaining for each team, stones existing on the sheet, stones moved by the previous shot (represented by colored frames only) , and stones removed from the game. From this image, only the stones on the sheet necessary for the training data are extracted.

The first shot team is determined from the first shot image as shown in the Fig. 11. Since the first image shows the sheet after the first team has already made a shot, the number of stones remaining for the first team is seven. Therefore, the team with lower number of stones remaining is the first team.

Next, the contour information of the area having the same color as the stones is obtained from the image. However, areas other than stones on the sheet may be unintentionally included in the resulting contours. Therefore, it is necessary to exclude extra contour information. First, the color of the center of the contour is obtained and those that do not have the same color as the stones are excluded. The stones moved by the previous shot can be excluded by this requirement since they are represented only by a colored frame. Next, contours that are too large or too small are excluded. Since the stones on the sheet are shown at approximately the same size, it is possible to exclude contour information obtained incorrectly. Finally, contours whose center coordinates are not on the sheet are excluded. This allows for the removal of stones that are unrelated to the game, located at the top and bottom of the sheet image. The coordinates of the remaining areas, which are determined as the stones, are converted to the coordinate system of digital curling and stored in database.

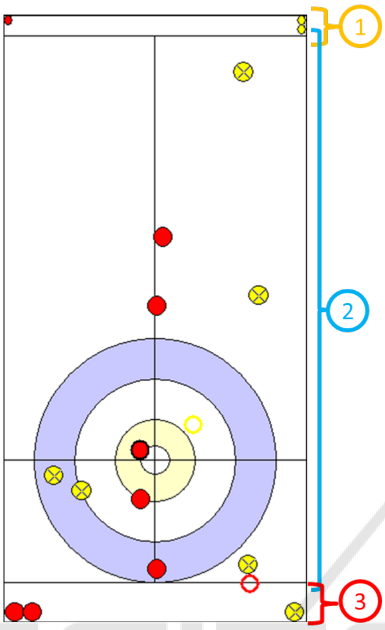

Figure 10: Image showing stones on sheet obtained from Shot by Shot. Area 1 shows the number of stones before the shot for both teams. Area 2 shows the position of the stones on the sheet and various information representing the progress of the game. Area 3 shows the number of stones from both teams already removed from the game.

5.3 Data Extraction Results

In this work, game data was obtained from the Results Book of eight competitions after the "No tick rule" was applied. The competitions used were European curling Championships 2022, European curling Championships 2023, Pan Continental curling Championships 2022, Pan Continental curling Championships 2023, World Junior curling Championships 2022, World Junior curling Championships 2023, World Women's curling Championships 2022 and World Men's curling Championships 2022. In total, we were able to obtain data for 532 competitions with 74,857 games. The number of data extracted per shot is shown in Table 4. More than 4600 data were extracted for each shot.

A comparison of the original image and the extracted data illustrated in the coordinate system of digital curling is shown in Fig. 12. In Fig. 12, the positions of stones are almost the same as the original image in the coordinate system of digital curling. This shows that this method can obtain accurate stone

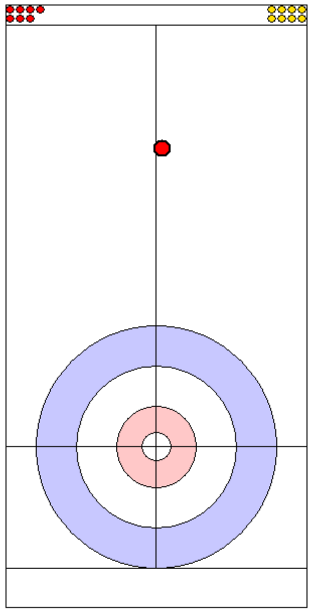

Figure 11: Sheet image after the first shot. This is the first sheet image on the page for each end. In this case, the red team made the first shot, so there is one less red stone remaining at the top of the image.

position data even when there are many stones and other information on the sheet.

Fig. 13 compares the distribution of the number of stones each team holds: the number of first team's stones on x-axis, and that of second team's stones on y-axis. According to Fig. 13, the data extracted from Results Book (on the right) obviously contains more situations with considerably many stones than the data generated by the algorithm (on the left). Also, the first team tends to hold more stones than the second team in the data extracted from Results Book. This comparison shows that the data extracted from Results Book includes diverse realistic situations.

5.4 Creation of Training Data from Extracted Data

Following the procedure outlined in Section 4.3, we created the training data to be used in training from the game information extracted in the previous section. The training data consists of pairs of game information, including extracted stone positions, remaining *ends*, and score difference, along with the corresponding expected score distribution.

Below, we evaluate the quality of the created data. By comparing the actual game results with the expected score distribution generated by the method in Section 4.3, we will confirm that the training data is somewhat realistic, i.e., not unnatural compared to the actual player's strategy.

Table 4: Number of data per shot. About 4600 data were obtained per each shot. The reason for the smaller number of data in the late game is that the game may end before last *end* due to concede.

Shot number	1-9	10	11	12	13	14	15	16
Number of data	4690	4687	4684	4682	4663	4663	4650	4590

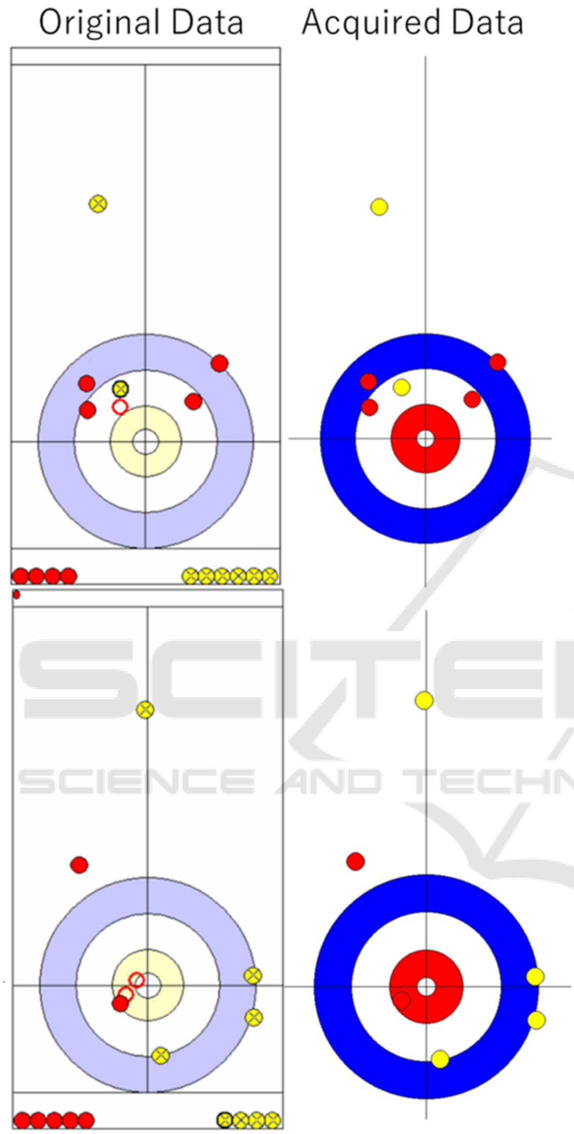

Figure 12: Comparison of extracted data with the original image. The original image is on the left and the extracted data is illustrated again by coordinate of digital curling on the right.

The game information used is from World Women's curling Championships 2022. The situation is that the team is the yellow stone, score difference of -2 and the last shot in the 7th *end*. The stone positions is as shown in Fig. 14. The expected score distribution in this situation is shown in Fig. 16. In the actual game, the team's shot was successful, the stone position was as shown in Fig. 15, and the team scored three points. From Fig. 16, in the expected score distribution we created, the probability of obtaining three points is nearly 80%. In other words, it was shown that it is possible to obtain an expected score distribution in line with real players' strategies as training data for the model from the positions of stones in actual games.

6 MODEL TRAINING AND RESULTS

The obtained training data will be used to train and validate the score prediction model. Validation is performed by comparing models trained by real data with models trained by data generated by conventional algorithms.

6.1 Training Model

The number of stone positions used to train the model is 4650 for both real and generated data. For each of these stone positions, there are 5 different cases in terms of score difference and 4 cases in terms of the number of remaining *ends*, so the training data is 93,000. The x-coordinates of the stone position were inverted to increase the number of data, since the effect of inversion on the x-coordinates is negligible. The final number of training data is 186,000. One hundred out of 4650 stone positions were used as validation data. This means that the model has 182,000 training data and 4,000 validation data.

The model has three hidden layers, each with 200 neurons. The activation function of the hidden layer uses ReLU. The hidden layer includes batch normalization and dropout. The optimization function used during training was Adam, the learning rate was 0.001, the batch size was 1024, the dropout rate was 0.3, and the loss function was the mean squared error. The training was conducted for 100 epochs with these parameters. The loss during training was as shown in Fig. 17. Mean squared error is used for the loss function. The training and test loss are decreasing, but the test loss is higher.

6.2 Results

We validated whether the trained models could predict accurate expected score distribution in multiple

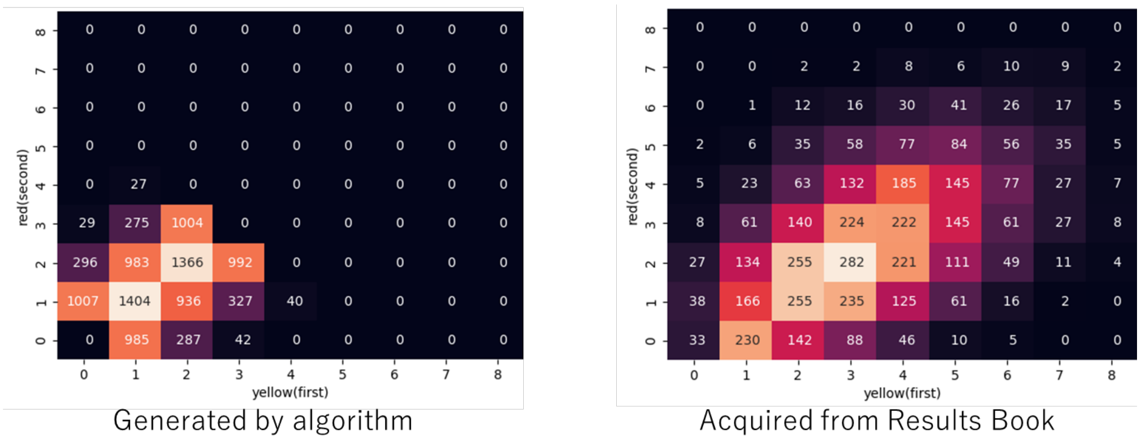

Figure 13: Distribution of number of stones for each team in train data. In each figure, the horizontal axis shows the number of stones for the first team and the vertical axis shows the number of stones for the second team.

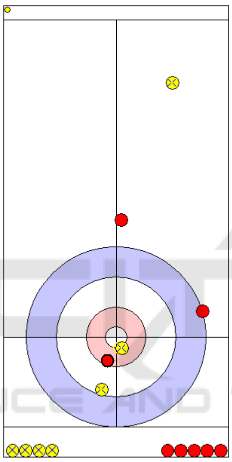

Figure 14: Position before last shot. Own team (yellow stone) has the No. 1 stone, but own team must take out the opposing team's No. 2 stone in order to obtain multiple scores.

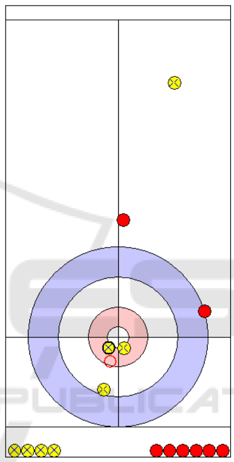

Figure 15: Stone position after Fig. 14 shot in the actual competition. The shot was successful and own team scored 3 points.

situations. In which figure of stone positions in each situation, the stones of the own team are shown in red and those of the opposing team in yellow. The prediction results for these situations are show in Fig. 18, 19 and 20. In the figure, from left to right, the correct score distribution, the prediction results from the real model, and the prediction results from the conventional model. The correct score distribution is the expected score distribution created according to the training data creation method described in Section 4.3.

Situation 1 is a situation in which the stone position is Fig. 18, the score difference is 0, and the number of remaining *ends* is 0. The stone position in Fig. 18 is a simple situation with a small number of stones. The prediction result in this case is shown in Fig. 18. In this situation, the team can win the game

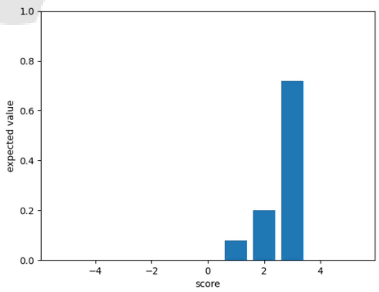

Figure 16: Expected score distribution created from the situation. The probability of three scores, which is the score obtained in the actual game, is nearly 80%. This result is in line with the actual strategy of the players.

if team's stone becomes the No. 1 stone, so the team have to take out the opponent's stone in the house and leave shot stone in the house. Since this shot is of low difficulty, the correct score distribution shows that a

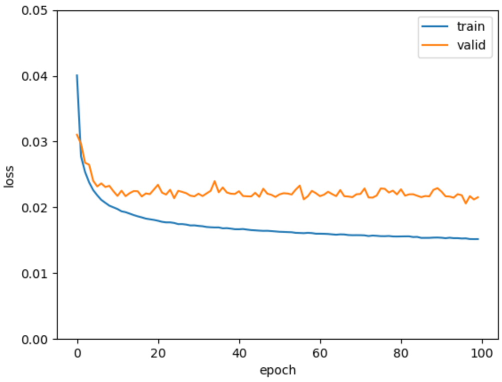

Figure 17: The learning curve. Mean squared error is used for the loss function. The orange curve is for validation data and the blue curve is for train data.

point will almost certainly be scored. In this case, both models are accurate in their predictions.

Situation 2 is a situation in which the stone position is Fig. 19, the score difference is -1, and the number of remaining *ends* is 0. The stone position in Fig. 19 is more complicated than situation 1 because there are more stones. In this situation, the team can win the game by scoring 2 points, so it is necessary to leave the shot stone in the house while taking out the opposing team's stones without touching the our team's stones in the house. Since this shot is a highly difficult shot, the correct score distribution shows that although there is a high possibility of gaining two points, there is a risk of gaining only one point or, even loosing the game by being steeled by the opponent. From Fig. 19, we can see that the model learned with real data is able to predict the similar distribution as the correct score distribution. On the other hand, the conventional model has the highest probability of gaining 2 points, but there is also a probability of gaining 3 points, which is impossible, indicating that the conventional model does not accurately understand the situation in prediction.

Situation 3 is a situation in which the stone position is Fig. 20, the score difference is 0, and the number of remaining *ends* is 0. The stone position in Fig. 20 is extremely complicated with a large number of stones. In this situation, the team can win by scoring one point, but because of the large number of stones around the house, even a slightly off shot will result in a failure. Therefore, the correct score distribution shows that the probability of winning, i.e., pointing one, is about 20% and the game is lost in most cases. Fig. 20 shows that the model trained on real data predicts approximately similar trends to the correct score distribution, but the accuracy is low and prediction is not accurate. On the other hand, the conventional model shows completely different results from the correct score distribution, indicating that it cannot handle this situation.

These results indicate that the model trained by real data is more accurate in predicting more difficult situations than the conventional model. However, it was also found that accurate prediction is still difficult in situations with a large number of stones. Complex situations such as Situation 3 are difficult to learn in the current situation due to their low frequency of occurrence in the acquired training data. Also, in some results there are outputs of scores that could never happen due to the number of stones on the sheet. One reason for these problems may be that the model consists of only simple layers. With only simple layers, learning the relative positions of stones is more difficult than with models such as CNN or transformer. As a result, the model's prediction accuracy may be low in situations with low frequency of occurrence, or it may produce outputs that are not realistic. Therefore, to solve this problem, model modification may be effective in addition to simply increasing the training data. An effective model would be one that can better take into account interrelationships between objects, such as a transformer.

7 CONCLUSIONS

In this study, as part of the creation of curling AI in digital curling, real game data was used as training data for the score prediction model that follows previous research. The real game data was obtained from the Results Book. As a result, we were able to accurately obtain the stone position data necessary for the training data. The training data generated by this data contains more diverse and realistic aspects compared to those of previous studies. The results of the training showed that the model trained by real game data improved prediction accuracy in realistic aspects compared to the model trained by data created by conventional algorithms.

One challenge with this approach is that real game data is finite, limited by the amount of data provided by the Results Book . In addition, the real data model is still not accurate in situations where there are many stones on the sheet because there are few similar situations.

To overcome these challenges, we plan to investigate the data augmentation methods suitable for the real game data, introduce a new prediction model architecture, like transformers, including input features and preprocessing. Data increase in the Results Book by future competitions could also improve the model performance. We also plan to create training

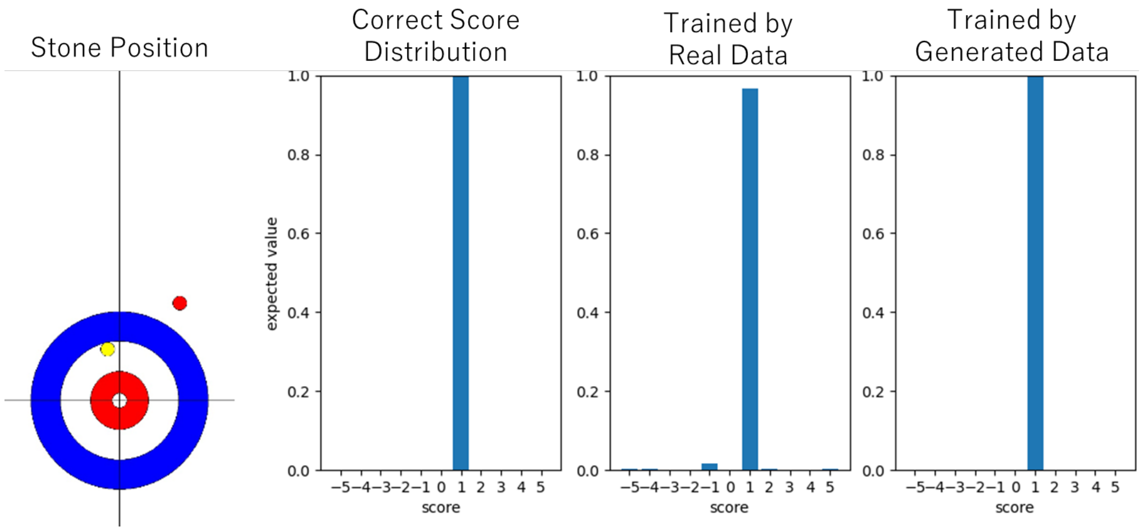

Figure 18: Stone position, correct score distribution and prediction results for each model in Situation 1. In this situation, the score difference is 0, and the number of remaining *ends* is 0.

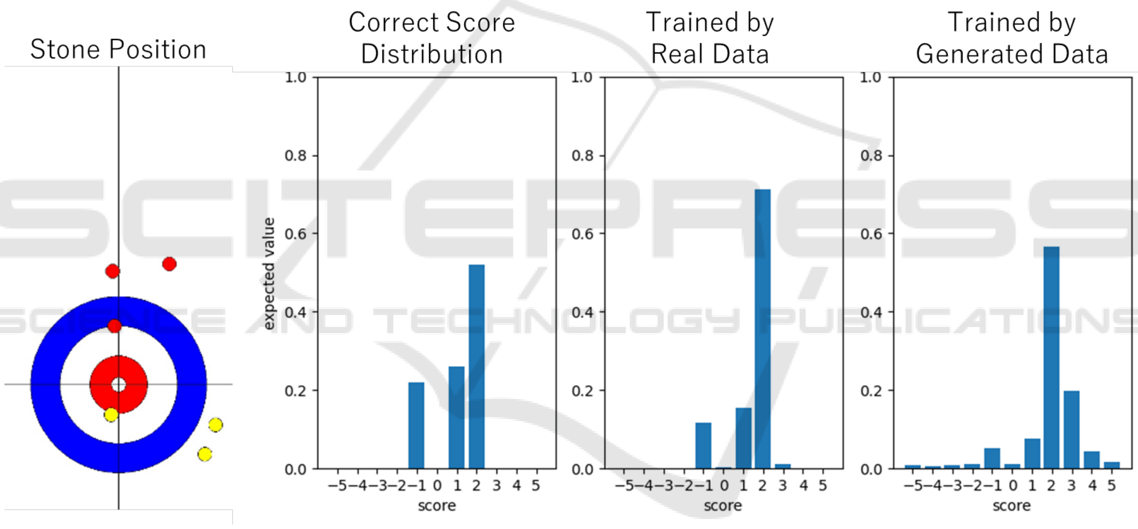

Figure 19: Stone position, correct score distribution and prediction results for each model in Situation 2. In this situation, the score difference is -1, and the number of remaining *ends* is 0.

data and conduct training for cases other than the last shot. Eventually, we would like to incorporate the completed model into a curling AI and evaluate the method through games on digital curling.

ACKNOWLEDGEMENTS

This work was supported by the “The Enhancement of HPSC Infrastructure through Technology Innovation Project” of Japan Sports Agency.

REFERENCES

Ataka, K., Noguchi, W., Iizuka, H., and Yamamoto, M. (2020). Learning of evaluation function in digital curling considering the probability of scores at each end. *Proceedings of AAAI20 Workshop, Artificial Intelligence in Team Sports*, pages 1–6.

Curlit (2022). WWCC2022 resultsbook. https://curlit.com/PDF/WWCC2022_ResultsBook.pdf.

Curlit (2024). Curling results. https://curlit.com/results.

Katoh, S., Iizuka, H., and Yamamoto, M. (2016). A method of game tree search in digital curling including uncertainty. *Journal of Information Processing*, 57(11):2354–2364.

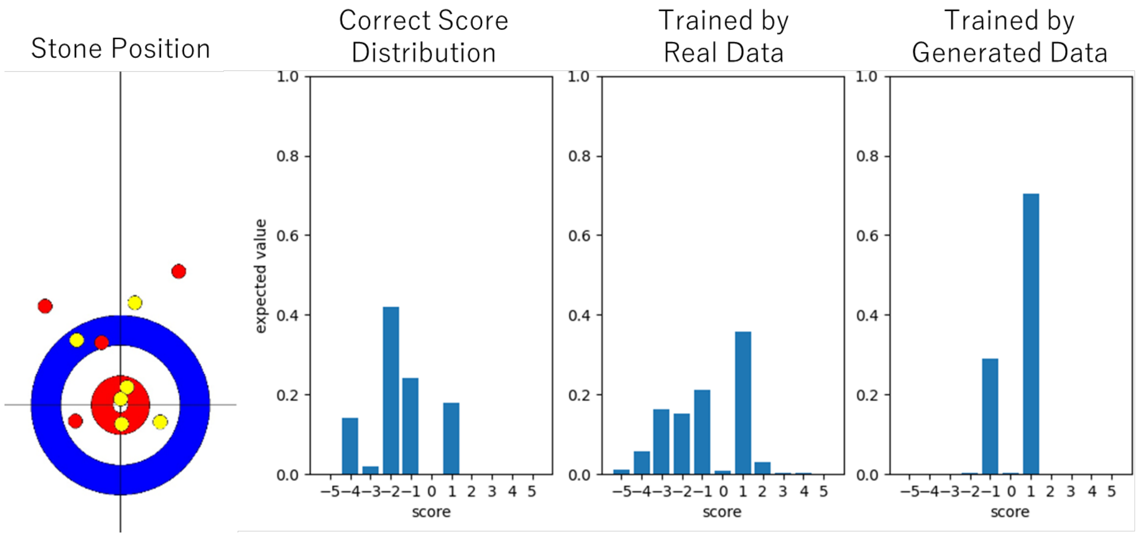

Figure 20: Stone position, correct score distribution and prediction results for each model in Situation 3. In this situation, the score difference is 0, and the number of remaining *ends* is 0.

Myslik, J. (2020). Curling analytics. https://www.jordanmyslik.com/portfolio/curling-analytics/.

Uehara, K. and Ito, T. (2021). Improving the simulator of digital curling to get closer to actual measurement data. Technical Report 18, The University of Electro-Communications,The University of Electro-Communications.

Yamamoto, M., Katoh, S., and Iizuka, H. (2015). Digital curling strategy based on game tree search. *Proceedings of 2015 IEEE Conference on Computational Intelligence and Games*, pages 474–480.

Yamamoto, M., Katoh, S., and Iizuka, H. (2018). Learning of expected scores distribution for positions of digital curling. *Proceedings of Workshop on Curling Informatics (WCI 2018)*, pages 8–9.