Title	Factor Analyses on Positive and Negative Evaluations of Games against Go Programs
Author(s)	Kuboki, Kyota; Hsueh, Chu-Hsuan; Ikeda, Kokolo
Citation	28th International Conference on Technologies and Applications of Artificial Intelligence (TAAI 2023): 343-357
Issue Date	2024-03-28
Type	Conference Paper
Text version	author
URL	http://hdl.handle.net/10119/19704
Rights	This is the author's version of the work. Copyright (C) 2024 Kyota Kuboki, Chu-Hsuan Hsueh, and Kokolo Ikeda, under exclusive license to Springer Nature Singapore Pte Ltd. The version published by Springer, Singapore is available at https://doi.org/10.1007/978-981-97-1711-8_26
Description	28th International Conference on Technologies and Applications of Artificial Intelligence (TAAI 2023), Yunlin, Taiwan, December 1–2, 2023

Factor Analyses on Positive and Negative Evaluations of Games against Go Programs*

Kyota Kuboki[1], Chu-Hsuan Hsueh[1][0000−0001−8888−3116], and Kokolo Ikeda[1]

Japan Advanced Institute of Science and Technology, Japan
{s2250002,hsuehch,kokolo}@jaist.ac.jp

Abstract. Analyzing users' preferences is important for many platforms to further improve users' satisfaction or make recommendations. This is the same for online game platforms. In this paper, we target a classical board game, the game of Go, and investigate the factors that make human players feel enjoyable when playing against Go programs on a website. In addition, we also investigate whether different players feel enjoyable in different ways. We use game records collected from the website, where players can evaluate games as enjoyable or not. We conduct statistical analyses using basic information, such as game lengths or players' win rates, as well as advanced analyses using information extracted by a strong Go program, such as the qualities of moves. The results show that some factors are generally common among players, while some factors show completely opposite preferences. For example, players generally prefer opponents with proper playing skills; meanwhile, some prefer close games and some prefer to win by large margins.

Keywords: Evaluation factor analysis· Player preference· The game of Go.

1 Introduction

In recent years, the development of computer technology and the emergence of new methods have led to the rapid growth of AI. Researchers have employed AI in many fields. In the field of games, researchers have been devoted to making AI programs stronger than top human professionals. For the game of Go, AlphaGo [8] won against a top professional Go player in 2016, and a successor named AlphaGo Zero [9] beat AlphaGo the following year. These strong Go programs are valuable for players who aim to become stronger. In fact, many professional Go players utilize strong Go programs in their studies.

Meanwhile, some players, especially amateurs, play for fun. Many such players want to enjoy the gameplay rather than knowing the best moves or playing against stronger opponents. It is desired to have Go programs for such players, leading them to games that they find enjoyable. Some researchers created such

* This work was supported by JSPS KAKENHI Grant Numbers JP23K11381 and JP23K17021. We would also like to thank Qinoa Inc. for providing the Go game records on Qinoa Igo.

Go programs based on common assumptions such as that players feel enjoyable when the skill levels do not differ too much [7, 6]. However, it is unclear whether those assumptions really hold. To our knowledge, there are no data-driven studies of what factors human players find enjoyable when playing games.

Moreover, different players may find different factors of games enjoyable. For example, some players may prefer close games, while others may prefer games that they win by large margins. Depending on the preferences of each player, it varies whether a game is enjoyable. Therefore, we consider it necessary to conduct player-by-player analyses to investigate the factors that make games enjoyable.

In this paper, we aim to analyze the factors that make a game enjoyable for each player and to clarify whether these factors differ from player to player. In cases where the factors differ, we investigate the tendencies of players. To achieve this, first, we collect game records that have been evaluated as enjoyable or unenjoyable by human players. Next, we perform statistical analyses on the collected games. We then analyze these games using the latest Go program and extract features for each game. Based on these features, we analyze the factors that contribute to enjoyable games. While some of the factors were generally common among the players, others were completely opposite among the players. A typical example has been discussed earlier, i.e., preferring close games or preferring wins with large margins.

2 The Game of Go

This section briefly explains the game of Go [1] in terms of the basic rules and terminologies. The board size is usually 19×19, and two players take turns placing black and white stones on the intersections on the board. The players aim to surround larger areas (called *territories*) than the opponents.

A game ends when both players judge that there are no places worth playing and *pass* their turns consecutively. After that, the sizes of both players' territories are counted, i.e., the number of empty intersections within each player's territory. The player with a larger territory wins. A game also ends when one of the players *resigns*, usually when the player judges that there are no chances to win.

3 Related Work

3.1 Preference Analysis of Human Players in Games

One of the earliest studies to analyze player preferences and behavior within a particular game was that of Bartle's [3]. Bartle studied what players enjoyed doing on Multi-User Dungeons (MUDs). Bartle categorized players based on their preferences for objects (other players or the game world) and their corresponding behaviors (interactive or one-direction).

Tondello et al. [10] analyzed human players' preferences more thoroughly and introduced five player traits: aesthetic orientation, narrative orientation, goal orientation, social orientation, and challenge orientation. Their study demonstrated the effectiveness of the model in video games. However, since video games differ from board games in many aspects, their model is unsuitable for our study.

3.2 Analysis of Go Players

With the development of Go AI, researchers have employed the Go AI in various applications. One such application is to analyze game records to understand players' tendencies. For example, Gao et al. [4] presented a professional Go annotation dataset that includes rich in-game statistics calculated by a Go program, KataGo [11]. They showed sample tasks that could be done with the dataset, e.g., predicting mistakes during a game and predicting the outcome of a game.

Hayashita et al. [5] aimed to analyze the factors that make a game enjoyable for human Go players. First, they collected game records that have been evaluated as enjoyable or unenjoyable by human players. Next, they analyzed these games using KataGo and extracted many features for each game. Based on these features, they proposed several hypotheses about the factors that contribute to enjoyable games. However, their study did not differentiate the game records by player, possibly resulting in the strong influence of players with a large number of games. It is also possible that conflicting trends of players canceled each other out so that no trend was observed when viewed as a whole.

4 Approach

This section describes how we collect game records and analyze the games.

4.1 Collected Game Records

This study is collaboration research with Qinoa Inc., and we use game records and evaluations from Qinoa Igo [2], a website operated by Qinoa Inc. Qinoa Igo provides players with Go programs of various types and skill levels to play against. Players can evaluate games as *good* or *bad* once per game at any time point. According to Qinoa Igo's question statement, "did you find the game enjoyable, practice, etc.?", players evaluate good if they think so and evaluate bad otherwise. When a player evaluates a game as good or bad, both the game record at that time point and the evaluation are saved.

To analyze factors of enjoyment player by player, we only use game records that satisfy the following conditions, resulting in 6,911 games from 98 players.

- Games on 19×19 boards starting with empty boards
- Games with more than 50 moves
- Games played by players who played and evaluated 10 games or more[1]

[1] Qinoa Igo saved players' IP addresses, and we used the IP addresses to distinguish players. Therefore, we could not identify the same player accessed from different IP addresses and could not distinguish different players accessed from the same IP address.

4.2 Analysis Steps

We conduct the factor analyses of enjoyable games in two major steps. First, we conduct statistical analyses based on basic data such as each player's win rate and good rate (Section 5). Next, we conduct more advanced analyses using a strong Go program. In more detail, we let the Go program analyze (or, say, review) each game so that we can extract advanced features of each game for factor analyses (Section 6).

The Used Go Program We analyze the game records using KataGo [11], a powerful Go program based on AlphaGo Zero. For each state in a game, we can obtain from KataGo information such as the win rate, the territory lead by the player to move (e.g., a territory lead of 5 means that the territory size of the player to move is 5 intersections bigger than the opponent's), and the most and the second most promising moves. This study uses the following information.

- Information about a state
  - The (estimated) win rate
  - The (estimated) territory lead, etc.
- Information about a move
  - The rank in KataGo (more promising moves rank higher)
  - The (estimated) win rate after playing the move
  - The (estimated) territory lead after playing the move
  - The difference in the (estimated) territory lead from the pass move, which we call *gain*, etc.

The Extracted Features of a Game We extract features for each game, either from simple statistics or based on KataGo. Note that a game on Qinoa Igo is played by a human player and a Go program. The following presents 4 features whose effects are easier to understand while we have extracted and analyzed more features.

- The game length (i.e., the total number of moves)
- KataGo's estimated territory lead at the end of the game for the human player, abbreviated as *human's lead*
- The promising-move rate of Qinoa Igo's Go program (i.e., the ratio of the moves by Qinoa Igo's Go program that match KataGo's most and second most promising moves), abbreviated as *program's performance*
- The number of unnecessary moves of Qinoa Igo's Go program (i.e., the number of moves by Qinoa Igo's Go program where the gains estimated by KataGo are lower than 0.5), abbreviated as *program's vainness*

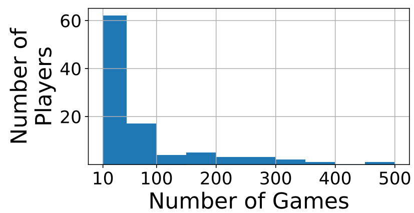

Fig. 1: The histogram of the number of games per player.

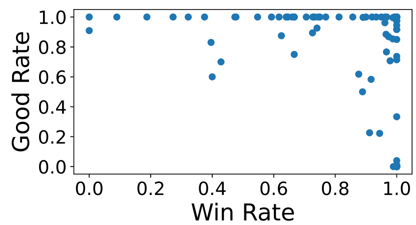

Fig. 2: The scatter plot of each player's win rate and good rate.

5 Statistical Analyses

This section presents statistical analyses based on the basic data of the collected 6,911 games from 98 players. The games were evaluated as either good or bad by the players. The overall good rate of the games was 0.664. The result showed that the players were generally enjoyed, though there was still room for improvement. In addition, the human players' win rate was 0.865, which is very high. This indicates that the human players were generally stronger than Qinoa Igo's Go programs. The following subsections will present more detailed analyses. Subsection 5.1 shows the number of games per player. Subsections 5.2, 5.3, and 5.4 show the relations between good rates and win rates, game results, and game lengths, respectively.

5.1 The Number of Games per Player

First, we investigated the number of games played by each player, and Fig. 1 shows the results. Among players who played more than 10 games, 64 players played between 10 and 50 games, and 10 players played more than 200 games. The number of games played by each player differed greatly. When averaging the games from different players, we need to consider such differences in game numbers. Take the following as an example: Player X played 300 games and player Y 10 games. If we simply calculate the average over the 310 games without considering the number of games per player, the weights of the two players become 30 : 1. Namely, player X's results dominate the analysis. On the other hand, if we calculate the average of each player and then calculate the average of these averages, i.e., players X and Y weighing 1 : 1, too big influence comes from player Y, whose data are less reliable due to a small number of games.

To reduce the influence of players with a large number of games while still giving them more influence than players with a small number of games, we weighted different players' games as follows. Specifically, for a player with n games, we set each game's weight to $\sqrt{10/n}$. For example, a player with 10 games has a weight of 1 per game, and a player with 40 games has a weight of 1/2 per game. In the following analyses, this weighting is applied.

Table 1: The weighted win rate and good rate of the target games.

Win Rate		Good Rate	
Average	Std.	Average	Std.
0.725	0.406	0.840	0.234

Table 2: The weighted good rates separated by game results.

Game Result	Good Rate				
	Overall	Player P's		Player Q's	
Human lost (resign)	0.951	-	(0/0)	1.0	(7/7)
Human lost (pass)	0.913	0.0	(0/3)	1.0	(5/5)
Human won (pass)	0.876	0.667	(24/36)	0.571	(12/21)
Human won (resign)	0.647	0.742	(75/101)	0.068	(7/103)

5.2 Relation between Good Rates and Win Rates

Table 1 lists the weighted win rate and good rate. Compared to the unweighted statistics on the whole target games, the win rate decreased from 0.865 to 0.725, and the good rate increased from 0.664 to 0.840. Note that compared to the unweighted statistics, the influence of players who played many games was reduced. We suspected that some players who played many games won almost all games and evaluated almost all games as bad. Except for such players, the general win rates were proper, and the games were generally favorable.

We further looked into each player's win rate and good rate, as plotted in Fig. 2. Note that some points overlap others in the plot. We found that many players evaluated only one of good and bad: 52 evaluated all their games as good, while 12 evaluated all their games as bad. All players with good rates of 0 had win rates higher than 95%, i.e., the players were much stronger than the Go programs. The result suggested that making Go programs stronger might help improve the good rate. In addition, among players with high win rates, some evaluated all their games as good. We considered two possible reasons: these players were satisfied with the current Go programs, or they evaluated the games irrelevantly to the gameplay.

5.3 Relation between Good Rates and Game Results

Table 2 lists the weighted good rate for four categories of game results: (i) Human lost because of resignation, (ii) human lost because of owning a smaller territory counted after two consecutive passes, (iii) human won because of owning a bigger territory, and (iv) human won because the Go program resigned. Usually, a resignation is made when the player judges there is no chance of winning. Thus, the four categories from top to bottom roughly represent human players' losses by large margins in territories, losses by small margins, wins by small

Game Length	#Games	Good Rate
50 – 100	1479	0.591
100 – 150	305	0.584
150 – 200	834	0.626
200 – 250	881	0.671
250 – 300	421	0.665
300 –	76	0.665

Table 3: Statistics of games where Go programs resigned.

Game Length	#Games	Good Rate
50 – 100	0	-
100 – 150	0	-
150 – 200	0	-
200 – 250	98	0.968
250 – 300	933	0.922
300 –	556	0.818

Table 4: Statistics of games that ended by pass where human players won.

margins, and wins by large margins. The general tendency was that the good rate decreased from the top to the bottom. We interpreted that players tended to evaluate a game as good (or instructive) when they lost and as bad (or boring) when they won by a large margin.

However, we also observed that different players, even with similar skill levels, evaluated differently. Table 2 shows concrete examples from players P and Q, whose skill levels were estimated to be close. Player Q had a similar tendency to the overall one, though with a significantly lower good rate of 0.068 for the games that the Go programs resigned. On the other hand, player P had the opposite tendency, with the highest good rate of 0.742 when the Go programs resigned. The results showed an example that players had very different preferences.

5.4 Relation between Good Rates and Game Lengths

From the investigation of game results and good rates, we found that the good rate of games in which the Go programs resigned was low. We considered two possible reasons: human players being much stronger than the Go programs and bad timings of resignation. In Go, the number of moves required to end a game (i.e., game length) differs depending on the game results. In particular, a game that ends by resignation may end early if the difference in playing skill is too large.

To find the cause of the low good rate in games that Go programs resigned, we further grouped games according to the lengths. Table 3 shows the results. For comparison, we did similarly for games that ended by pass where the human players won. Table 4 shows the results. For games that Go programs resigned, the good rates got lower as the games were shorter. Short games in this case mean that human players were much stronger than the Go programs. Big differences in territories had occurred at the early stages of the games. Namely, human players won easily and could not enjoy playing, making them dissatisfied. On the contrary, for games that ended by pass where human players won, the good rates got lower as the games were longer. We considered the reason to be that the Go programs did not resign nor pass but continued to play even when the game results were clear, which was annoying and made human players unenjoyable.

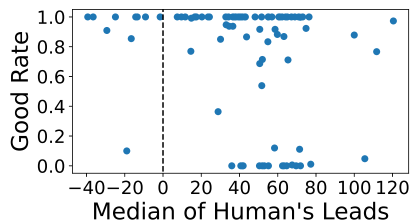

Fig. 3: The scatter plot of each player's good rate and human's lead median.

6 Advanced Analyses Using A Strong Go Program

This section conducts various analyses based on advanced features of games obtained from KataGo.

6.1 Relation between Differences in Playing Skills and Good Rates

From the results in Table 2 and Fig. 2, we observed that the good rates were low when the human players were much stronger than the Go programs or when the Go programs had to resign because human players already led a lot in the territory size. In this subsection, we look at the win-loss results in more detail and use KataGo to analyze how much human players led in territories (i.e., the feature of human's lead) at the ends of the games. Since win-loss results were required, we excluded the games where players evaluated good/bad when the games had not ended, resulting 5,373 games from 82 players.

Even when the same player plays against the same Go program, the results (i.e., the territory lead) usually differ each time. We consider the median of the territory leads for a given player to be able to indicate approximately how much the player and the Go program differ in playing skills. Meanwhile, it is interesting to investigate the relation between the good rate and human's lead of each game, where the details will be presented in Section 6.3.

Fig. 3 shows the scatter plots of each player's human's lead median (differences in playing skills) and good rate. We observed that some players evaluated all games only as good (46 players) or bad (10 players), similar to Fig. 2. For those who evaluated all bad, the medians of human's leads concentrated at 40–80. Such territory leads usually indicate big gaps in playing skills (roughly 4–7 handicap stones). On the other hand, when we looked at the players with good rates close to 1, we found a wide range of players, and many of them evaluated games to be good, even when the territory leads were between 40 and 80.

In order to clarify how each player evaluates good and bad differently, we conducted the following analyses that excluded players with extremely high or low good rates. Specifically, we targeted players with good rates between 0.1 and 0.9 and obtained 997 games from 21 players.

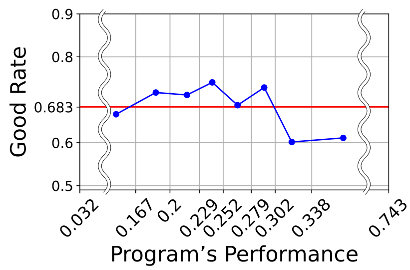

Fig. 4: The plot of program's performance and good rates.

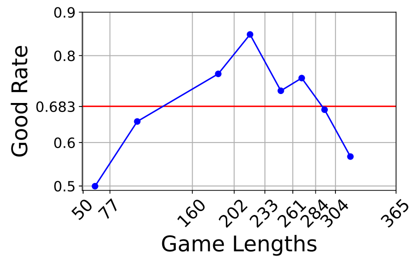

Fig. 5: The plot of game lengths and good rates.

6.2 Relation between Single Features and Good Rates

According to the analyses so far, we confirmed that game results and playing skills of Go programs had big impacts on the evaluations of good and bad. Therefore, we further investigated three features, introduced in Section 4.2, that related to game results and playing skills: Program's performance, the game length, and human's lead.

We analyzed the overall tendency of players using single features. We plotted the features and the good rates following the steps below to show the tendencies.

1. For the target feature to investigate, we sorted the games according to the feature in ascending order.
2. We separated the games into several groups so that the *weighted* number of games in each group was the same. The weight of a game has been discussed in Section 5.1 (i.e., $\sqrt{10/n}$). The number of groups to separate was decided by the Sturges' rule based on the total weights of the games.
3. We calculated the weighted good rate and the weighted average of the target feature. We plotted the former as the y-axis and the latter as the x-axis.

The Relation between the Program's Performance and Good Rates Fig. 4 plots program's performance and good rates. The red line shows the overall weighted good rate. The values of program's performance were widely distributed, indicating two possibilities: (i) the provided Go programs had various skill levels, or (ii) even the same Go program might do well or poorly in each game. The good rate became low when the Go programs did too well (right end) or too poorly (left end). We considered it natural for human players to be dissatisfied when the opponent in that game was too strong or too weak.

The Relation between the Game Lengths and Good Rates Fig. 5 plots the game lengths and good rates. The curve had a clear mountain shape, where the good rates were low when the games were too short or too long. We suspected it to be a combination of two tendencies shown in Section 5.4: too quick

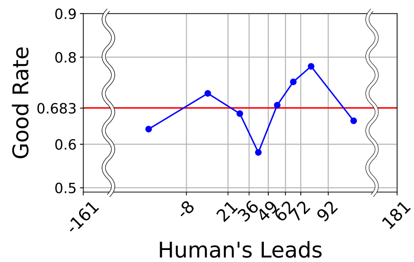

Fig. 6: The plot of human's leads and good rates.

resignation and too many unnecessary moves. More specifically, for cases the Go programs resigned too quickly (left end), players might not enjoy because the opponent was too weak. On the other hand, for cases the games were long (right end), players might be dissatisfied because the opponent did not resign despite a big difference in territory or because the opponent played many unnecessary moves in endgames instead of playing pass to finish the games.

The Relation between Human's Leads and Good Rates Fig. 6 plots the human's leads and good rates. The good rates were low at the two ends and in the center (around $x = 40$). For the two ends, we suspected that human players did not enjoy when winning or losing by large margins in territories. Regarding the center where the good rate dropped drastically, we considered it unnatural and suspected that the result was a combination of different tendencies from different players.

6.3 Analyses of Individual Players Based on Two Features

Section 6.2 has shown the relations between single features and good rates, where the effects were summed up from several players. We found that the shape of the curve in Fig. 6 (human's lead and good rate) was hard to explain. We considered two possible causes of the unnatural shape: (1) different players had different preferences, which should not be summed up, and (2) the feature human's lead might influence human players' enjoyment in conjunction with other features, which should not be ignored. Therefore, in this subsection, we analyze player by player based on two features to see each player's tendencies or preferences.

Human's Lead and Program's Performance We first analyzed human's lead and program's performance together. We supposed the two features to have a relation because it is natural that human players' territory leads are smaller if more moves of the Go programs are promising moves.

Fig. 7 shows players A and B's scatter plots of human's lead (x-axis) and program's performance (y-axis) in the games evaluated as good (blue circles) and bad (red crosses). Player A evaluated the games he/she lost ($x < 0$) as

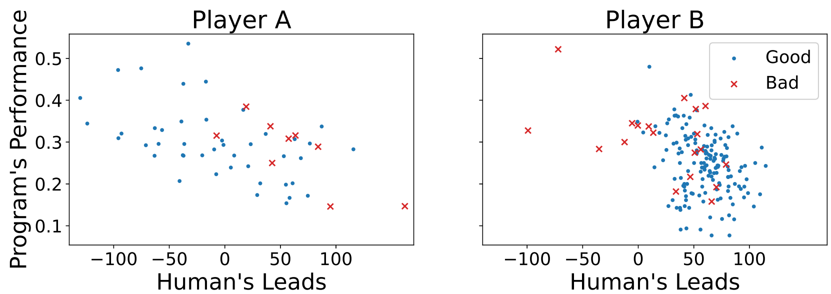

Fig. 7: Scatter plots of human's lead and program's performance for Players A and B.

good more often than the games he/she won ($x > 0$), which might indicate that losing made him/her learn something and was favorable. Among the games that Player A won, good was more often evaluated when the Go programs played fewer promising moves (lower y). For different games with the same value of human's lead (i.e., the same x), lower y values were likely to indicate that Player A also played fewer promising moves, or say, Player A played more bad moves. Similarly, we interpreted that Player A learned something from actually playing the bad moves and was satisfied with this. In contrast, Player B had more games evaluated as bad on the left side (losses or close games). We interpreted this to indicate that player B enjoyed winning games by large margins. In the games that he/she won, he/she evaluated many games as bad where the program's performance was high. Although the tendency to have more bad games in the upper parts of the figures was the same as Player A, we interpreted Player B to feel more comfortable winning when the opponent was weaker.

Scatter plots are suitable for seeing the tendency of each player in detail, but when there are many players, it is hard to analyze the tendencies at once. Therefore, we represented each player with two points (and the dotted line connecting the two points), where the circle point was the centroid of the player's good games and the cross point was the centroid of the player's bad games.

Fig. 8 shows the 21 players whose good rate was between 0.1 and 0.9. We observed that circle points (good games) concentrated more on the center of the figure, while the cross points (bad games) were widely distributed. We interpreted that players generally preferred games with moderate territory differences and opponents with moderate playing skills. Despite the general tendency, we found differences in the relation between the circles and the crosses. Thus, we further grouped the players as follows. The red group contained players who preferred close games (circle points closer to $x = 0$ than cross points). The orange group contained players who were sensitive to the programs' performance (the dotted line being close to vertical). The blue group contained players who preferred to win by large margins rather than close games. To sum up, we concluded that

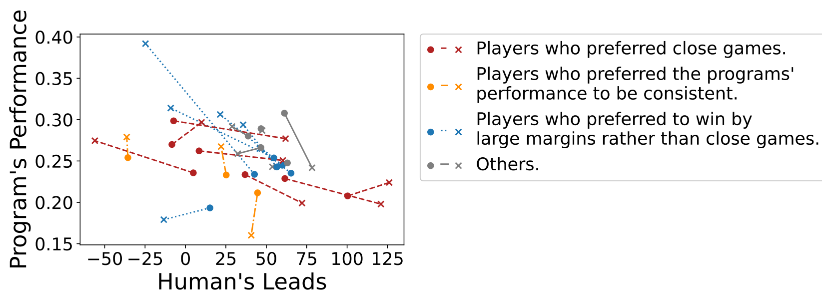

Fig. 8: Players' centroids of good and bad games plotted based on human's lead and program's performance.

players had general tendencies (preferring moderate territory differences with proper-level opponents), while players did differ from others.

Human's Lead and Game Length We also analyzed human's lead and game length together. The reason was that we supposed even with the same territory lead (say 50), it was not strange for players to feel differently when the games ended at different timings (say the 100th, 200th, or 300th moves).

Fig. 9 shows 4 relatively characteristic players' scatter plots of human's lead (x-axis) and game length (y-axis).[2] Player B, who was discussed earlier (Fig. 7) to prefer to win by large margins, was likely to evaluate close games ($x \approx 0$) or long games (higher y) as bad. Player C, in contrast, seemed not to prefer to win by large margins or by Go programs' resignation early in the games. Instead, we suspected Player C to prefer close games or losses, though close games that were too long were likely to be evaluated as bad (this tendency was especially clear for Player E who will be discussed later). Player D was similar to Player C in that they tended to evaluate games as good when they lost while as bad when they won. But they were different in that Player D only evaluated the games he/she won by large margins in the middle games as bad. Player E seemed not to prefer to win by Go programs' resignation early in the games as Player C, but Player E more often evaluated games as good when the game lengths were between 100 to 300, regardless of the values of human's lead. Games with more than 300 moves were almost evaluated as bad. We suspected the reason to be that the Go programs played many unnecessary moves and did not resign even when the winners were clear or did not pass even when the territories were fixed.

Regarding Players C and E, their scatter plots helped to explain the seemingly unnatural M-shape in Fig. 6. Strong players like Player C often forced

[2] We suspected the reason why the data points for Players B, C, and E look like arcs to be that (1) The Go programs resigned early in the games when the human players already led a lot in territory. (2) After around 150 moves, if the human players led a lot, the Go programs resigned. (3) If the differences in territory were not large, the games proceeded to the ends.

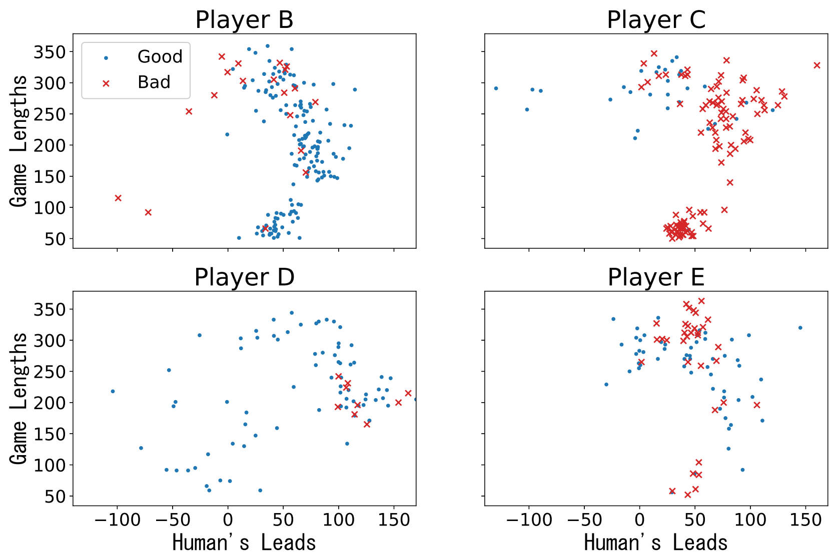

Fig. 9: Scatter plots of human's lead and game length for Players B, C, D, and E.

the Go programs to resign quickly and evaluated such short games as bad. In such cases, the human's leads were around 30–50. Also, some players like Player E evaluated very long games as bad, where the Go programs hesitated to resign/pass and the human's leads were around 40–60 [3]. We suspected these tendencies to be a possible reason for lower good rates around human's leads of 40–50 in Fig. 6.

Fig. 10 shows the analyzed 21 players' centroids of good and bad games depicted in a similar way to Fig. 8, but the y-axis here is game length instead of program's performance. Similar to Fig. 8, circle points (good games) could be more often found around the center of the figure than the cross points (bad games). The results suggested that players generally preferred games with moderate territory differences and lengths. Nevertheless, according to the relation between the circles and crosses, we found several groups of players. The blue group contained players who preferred to win by large margins. The red group contained players who preferred close games. The orange group contained players who were sensitive to game lengths, dissatisfied either with Go programs' resignation in early games or with too-long games that might contain unnecessary moves. Again, we found both general tendencies among players (preferring games with moderate territory differences and lengths) and differences in players' preferences.

[3] When the human's lead became larger than this range, probably the Go programs resigned earlier.

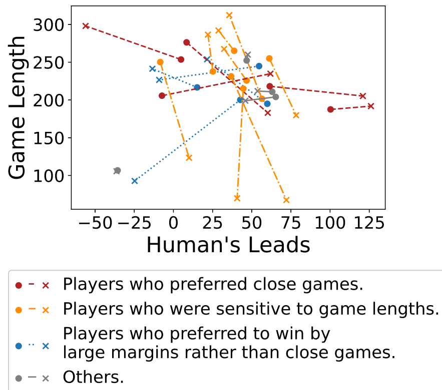

Fig. 10: 21 players' centroids of good and bad games plotted based on human's lead and game length.

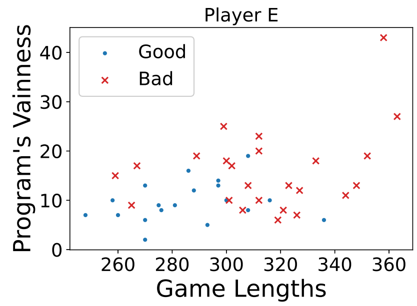

Fig. 11: Scatter plot of game length and program's vainness for games ended by pass where Player E won.

Game Length and Program's Vainness For Player E's results in Fig. 9, we considered a possible reason for the bad games to be that the Go programs made many unnecessary moves and hesitated to resign or pass in endgames. To confirm this, we conducted a further analysis of games that ended by pass where Player E won.

To quantize how often the Go programs played unnecessary moves, we used the gains of moves, i.e., the differences in territory leads compared to the pass move, as explained in Section 4.2. A small gain means that the move has no different effects on the game compared to passes. We counted the number of moves with gains lower than 0.5 and defined it as program's vainness. We plotted this metric with the game length to see how these two metrics influenced Player E's evaluations, as shown in Fig. 11. In games evaluated as good, the Go programs played fewer unnecessary moves in general, which supported our assumption. When the game lengths exceeded 300, the games were evaluated as bad more often, regardless of the number of unnecessary moves (program's vainness). Even when the game lengths were lower than 300, the games were evaluated as bad more often when there were more unnecessary moves. We considered two possible explanations for the dissatisfaction. First, it took longer to end the games when many unnecessary moves were made despite the winner being clear, which was a waste of time. Second, longer games, even when unnecessary moves were few, might make Player E exhausted and give him/her bad impressions.

7 Conclusions

In this paper, we analyzed the factors that make a Go game enjoyable and investigated how these factors differ from player to player. First, we collected game records from a website providing human players with intermediate-level

Go programs, where human players can evaluate games as good or bad. After performing statistical analyses using basic information, such as game results and game lengths, we employed KataGo, a strong Go program, to extract advanced features for analyses, such as expected territory leads or qualities of moves.

As a general tendency, games that Go programs won or played well were more likely evaluated as good, probably due to the fact that many players were stronger than the prepared Go programs. Conversely, very short games or games that Go programs lost by large margins were likely evaluated as bad. In addition, games were evaluated as bad when Go programs did not resign nor pass but continued to play even when the game results were clear. All of these tendencies were understandable to some extent.

Nevertheless, we also found that not all players had the same evaluation tendencies. For example, some players preferred to win by large margins, and some players preferred longer games. It is challenging to satisfy all of these players with non-adaptive Go programs since players sometimes have almost opposite preferences. One solution is to offer a variety of program options to suit different preferences. Another solution is to develop adaptive Go programs that refer to individual players' evaluation histories and game records.

In the future, we plan to conduct analyses using more advanced features. For example, in the game of Go, it is well known that there are players who are aggressive and like a lot of battles, while others may prefer to surround territories in peaceful ways. We believe that by inferring such advanced preferences of each player, the Go programs will be able to adapt to these preferences and improve the player's satisfaction.

References

1. https://en.wikipedia.org/wiki/Go_(game)
2. https://igo.qinoa.com/ja/
3. Bartle, R.: Hearts, clubs, diamonds, spades: Players who suit muds. Journal of MUD research **1**(1), 19 (1996)
4. Gao, Y., Zhang, D., Li, H.: The professional go annotation dataset. IEEE Transactions on Games pp. 1–10 (2023). https://doi.org/10.1109/TG.2023.3275183
5. Hayashita, M., Ikeda, K., Hsueh, C.H.: Factor analysis for Go AI to produce good games. Tech. rep., JAIST (2023), the 49th Meeting of the Game Informatics Research Group
6. Hsueh, C.H., Ikeda, K.: Playing good-quality games with weak players by combining programs with different roles. In: IEEE CoG 2022. pp. 612–615 (2022)
7. Liu, A.J., Wu, T.R., Wu, I.C., Guei, H., Wei, T.H.: Strength adjustment and assessment for mcts-based programs [research frontier]. IEEE Computational Intelligence Magazine **15**(3), 60–73 (2020). https://doi.org/10.1109/MCI.2020.2998315
8. Silver, D., Huang, A., et al.: Mastering the game of go with deep neural networks and tree search. nature **529**(7587), 484–489 (2016)
9. Silver, D., Schrittwieser, J., et al.: Mastering the game of go without human knowledge. nature **550**(7676), 354–359 (2017)
10. Tondello, G.F., Arrambide, K., et al.: "i don't fit into a single type": A trait model and scale of game playing preferences. In: INTERACT 2019. pp. 375–395 (2019)
11. Wu, D.J.: Accelerating self-play learning in Go. arXiv, abs/1902.10565 (2020)