Gradient Learning for the game of Go

Shi-Jim Yen[1], Yi-Ling Chen[1], Hsin-I Lin[1]

Abstract: Computer Go can play the role of the drosophila for the AI services of assisting human learning. For the game of Go, since human Go players have a different process of reasoning compared to Go programs today, we need to develop Go learning methods that more closely match how humans think. It is hard to do this before. Nowadays, deep learning is a powerful tool since 2012. We can use deep learning to simulate the move style of different kinds of human players. In this paper, we use several deep learning techniques and reinforcement learning to develop a gradient learning system for Go. The system contains human-like Go programs with various strengths, which allows players to learn Go skills progressively. The experimental result shows that the system can help players to improve their ranking.

Keywords: Gradient Learning, Scaffolding Theory, Go, Computer Games, Deep Learning.

1. Introduction

Computer Game is the drosophila of Artificial Intelligence. A series of AlphaGo programs created a brand new history for AI. [6][7][8] For the AI services of assisting human learning, we believe computer board games can also play the role of the drosophila. Since human Go players have a different process of reasoning compared to Go programs today, we need to develop Go learning methods that more closely match how humans think. Deep Learning takes inspiration from human cognitive processes and is similar to human intuition. As a result, Go programs developed with Deep Learning generate plays that feel more human. We use Deep Learning, Reinforcement Learning, and Monte Carlo tree serch to develop a gradient learning system for Go. The system contains human-like Go programs with various strengths, which allows players to learn the game progressively. This paper will describe the learning theories and framework of this system. Then gives the rewards from the users.

2. Learning Theory and System Framework

2.1 The Learning Theory

The basic learning theory of the gradient learning system is Scaffolding Theory. The Theory is based on the concept of Zone of Proximal Development (ZPD) in constructivist theory by Vygotsky. [1][2] For people to develop their mind they have to seek knowledge from someone wiser than you, like a mentor. In this study, we developed a concept based on Scaffolding Theory proposed by Wood et al. [3][4][5] It involves a kind of "scaffolding" process, mentions that the novice or child needs the direction by auxiliary to solve a problem, achieve their goal and finish a task which would help their unassisted efforts. As a consequence, we expanded the concept of learning gradient in this research to find out the most suitable auxiliary enable to guide learning effectively and maintain learners to focus on the direction correctly in learning. The principle of gradient learning is according to Scaffolding Theory proposed by Wood, Bruner & Ross[3] as the foundation to make learners have each auxiliary to improve personal learning skills in the correct direction.

The concept of learning gradient is that the best opponents to play against are in fact not the top experts of the field. Instead, one improves the most when playing against someone who is just one step higher on the skills gradient. If you play against an opponent who is much higher on the skills gradient, you may not have the ability yet to comprehend the opponent's actions, so the experience turns out to be less helpful for learning.

However, it is difficult to use a traditional algorithm to perform different levels in simulation. Take AI Go programs for example, despite normal programs in Go have reached the same power to defeat professional opponents, they are helpless to people learning Go. Using deep learning and reinforcement learning, we can establish a gradient learning system, allowing learners to choose a suitable from the Go program, and increase the efficiency in learning.

1 Department of Computer Science and Information Engineering, National Dong Hwa University, Shoufeng, Hulaien city 97401 Taiwan. Email: {sjyen, yi_ling10, 410521233}@mail.ndhu.edu.tw

2.2 The System Framework

The framework of our system is based on that of AlphaGo, which is constructed by deep convolutional neural network(DCNN)[9] and Monte Carlo tree search(MCTS)[6]. The strength of a Go program is based on the quality of DCNN and the simulation number of MCTS. [6][7][8]

Ikeda and Wu used MCTS to produce various strategies moves and natural position control. [11][12] However, Naturalness moves may cause a problem for MCTS based methods. Our method is to use different DCNN models trained from different Go game records. Then use different MCTS parameters setting. Finally, we could make many human-like Go programs with various strengths.

Table 1 shows the ranking system for Go. Our system has 6 DCNN models for 6 kinds of Go programs as in Table 2. There are 30 Go programs with 30 different levels, respectively. Each is made with different model and different MCTS parameters setting for the 30 levels, respectively.

Table 1. The ranking system for Go.

	Beginner			Kyu amateur		
Rank	25K	…	10K	9K	…	1K
#Srial	0	…	15	16	…	24

Dan amateur			Pro		
1D	…	9D	1P	…	9P
23	…	33	34	…	42

Table 1 The setting for different strength Go programs. The unit of the number is thousand.

levels	#trained games	#trained Boards	accuracy	#MCTS simulations
>10K	100	21,940	43.30%	0.5-1
7K-9K	120	27,563	45.80%	1-3
4K-6K	120	27,170	46.30%	1-3
1K-3K	120	27,069	47.50%	3-5
1D-3D	120	26,905	48.40%	6-12
4D-7D	120	26,494	50.22%	12-100

3. Rewards from the Users

The different strength Go programs are operating in U-gen no ma of Nihon Ki-in (日本棋院幽玄之室) with the bot account shows in Figure 1. The name of the bot is GoTrend. There are Dan GoTrend and Kyu GoTrend, and Beginner GoTrends bots as in Figure 2, Figure 3, and Figure 4, respectively.

Figure 1. U-gen no ma of Nihon Ki-in

Figure 2. The Dan GoTrend bots.

Figure 3. The Kyu GoTrend bots.

Figure 4. The Beginner GoTrend bots.

So far, the system has provided about one million AI Go moves per month. U-gen no ma of Nihon Ki-in conducted a questionnaire survey for the users. The results of the questionnaire show that the system is quite popular and with positive comments among users. Figure 5 shows why users like this system. It can be seen that many people started to contact AI services out of curiosity, then prefer to use this system. As can be seen in Figure 6, users believe that the power of Go in the system is appropriate and there is no sense of naturalness moves.

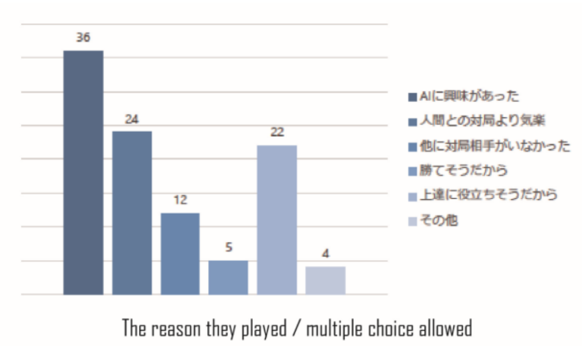

Figure 5. The reason they played with our system/ multiple choice allowed.

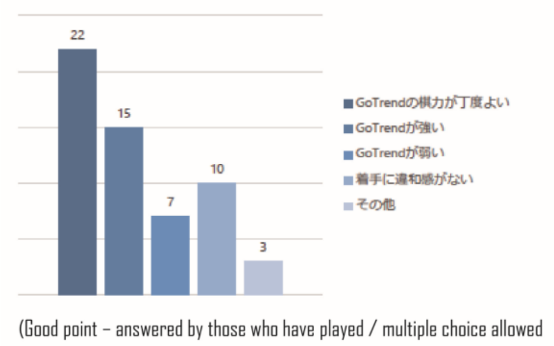

Figure 6. Good point of our system.

Figure 7 and Figure 8 show the user's power of Go distribution and the user's Go age. The user's age of Go is mostly more than 10 years, and the power of Go is mainly amateur. Thus, it can be seen that most of the users are elderly people. Because of attracting the use of senior citizens, this system has great potential in the development of the long-term care systems. In addition, the users in this system are part of the beginner. A big problem for beginners in the past is that it is not easy to find an opponent. Because the system provided a variety of opponents with different power of Go, the number of beginners has increased a lot, which is one of the contributions of this system.

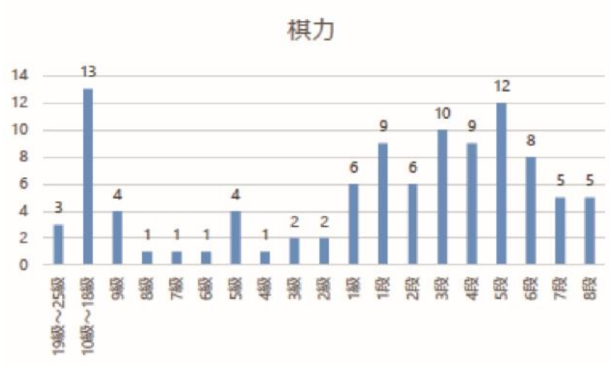

Figure 7. Go skill of the users.

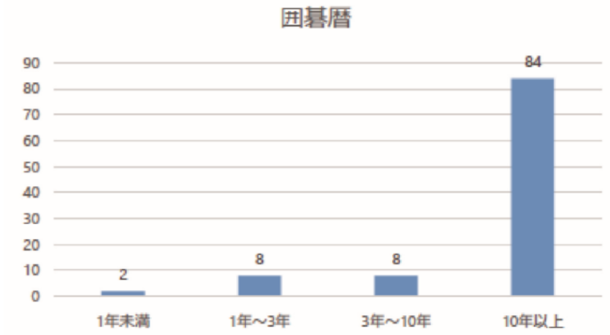

Figure 8. Go history of the users.

4. Concluding Remarks

We use Deep Learning, Reinforcement Learning, and Monte Carlo tree search to develop a gradient learning system for Go. The system contains human-like Go programs GoTrend with various strengths, which allows players to learn the game progressively. Go players can choose the suitable Go program to compete and learn its skills. The questionnaire shows that our programs GoTrend is quite popular and with positive comments among users.

Acknowledgments The authors would like to thank anonymous referees for their valuable comments in improving the overall quality of this paper. This work was supported in part by the Ministry of Science and Technology of Taiwan under the contract 108-2634-F-259-001- through Pervasive Artificial Intelligence Research (PAIR) Labs, Taiwan.

References

[1] Vygotsky, L.S.: Mind in society: The development of higher psychological process (M. Cole, V. John-Steiner, S. Scribner and E. Souberman, eds.). Cambridge, MA: Harvard University Press (1978).

[2] Gindis Kozulin, G., Ageyev, M.: Vygotsky's Educational Theory in Cultural Context, Cambridge: Cambridge University Press. ISBN 0521528836 (2003).

[3] Wood, D., Bruner, J.S., Ross, G.: The role of tutoring in problem solving. J. Child Psychol. Psychiatry 17, 89–100 (1976).
[4] Hogan, K., Pressley, M.: *Advances in learning & teaching. Scaffolding student learning: Instructional approaches and issues.* Cambridge, MA, US: Brookline Books (1997).
[5] Cazden, C. (1979). Peekaboo as an instructional model. Papers and Reports on Child Language Development, 17, 1-19.
[6] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. (2016).
[7] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, "Mastering the game of go without human knowledge," *Nature*, vol. 550, pp. 354–, Oct. 2017.
[8] Tian, Y., Zhu, Y.: Better computer go player with neural network and long-term prediction. International Conference on Learning Representations (2016).
[9] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. *CoRR*, vol. 1409.4842 (2014).
[10] Krizhevsky, A., Sutskever, I., Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems 25, 1097–1105 (2012).
[11] Wu, I-C., Wu, T.-R., Liu, A.-J., Guei, H., Wei, T.: On Strength Adjustment for MCTS-Based Programs. In: Thirty-Third AAAI Conference on Artificial Intelligence (2019).
[12] Ikeda, K., Viennot, S, "Production of various strategies and position control for monte-carlo goentertaining human players." In: 2013 IEEE Conference on Compu-tational Inteligence in Games (CIG). pp. 1-8, (2013).

