Deep Reinforcement Learning Compared to Human Performance in Playing Video Games

Jovana Markovska, Domen Šoberl

University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Glagoljaška 8, SI-6000 Koper, Slovenia

Abstract

Through deep reinforcement learning, a computer can learn to play simple video games under the same conditions as human players — by watching the pixels on the screen and issuing the button-press type of actions. In this paper, we investigate how quickly a computer can reach and surpass human performance in a simple two-player video game, given no background knowledge of how to play the game. We implement a Deep Q-Learning (DQN) algorithm using Python and integrate it with the Atari 2600 emulator that runs the Pong game. We train the neural network for up to 3 million training steps and evaluate its performance after every 100.000 steps. As a reference, we measure the performance of 18 human players and take their average rating as the human-level baseline. We propose an evaluation metric that considers the obtained game points and the player's endurance during the game. We find that the Deep Q-Learning algorithm can surpass beginner-level human players in playing the Pong game after about 4 hours of training.

Keywords

reinforcement learning, deep learning, Q-learning, deep neural networks

1. Introduction

With the advances of deep learning [1], it became possible to implement a deep reinforcement learning algorithm (Deep Q-Network, DQN) that can learn how to play video games under the same sensory conditions as human players [2]. Traditionally, reinforcement learning was used with high-level game states, e.g., board positions with the tic-tac-toe game [3]. On the other hand, humans observe a screen image composed of raw pixels, and it is up to them to make an abstract meaning out of it. For example, suppose a human player is given no instructions on how a game is played, the mechanics of it, who the agent is, and what the goal of the game is, except for positive or negative feedback when the score has been changed. The player would, by trial and error, first make sense of the individual objects seen on the screen, determine the mechanics of the game by trying out various actions with the game controller, and find out what the goal is supposed to be by observing under what circumstances the scores change. Deep learning has enabled computers to perform under similar conditions. Instead of feeding the algorithm with an abstract representation of the game's state, the convolutional layers within a DQN learn how to extract high-level features from raw pixels [4], while the fully connected layers learn a game's strategy.

Human-Computer Interaction Slovenia 2022, November 29, 2022, Ljubljana, Slovenia

89222055@student.upr.si (J. Markovska); domen.soberl@famnit.upr.si (D. Šoberl)

In this paper, we aim to answer the research question of how quickly DQN can reach and surpass human performance in a simple two-player video game. We present the results of a video game tournament that we organized to evaluate the human progress in playing the classic *Pong* game on the Atari 2600 platform. We propose an evaluation metric that encodes the player's endurance and the achieved score, which we compare with the values achieved by the computer at different stages of training.

2. Methodology

We implemented the DQN reinforcement learning algorithm [2] in Python 3.10. We used Tensorflow 2.10 [5] to construct the deep neural network, *OpenAI Gym* [6] in combination with the ALE (Arcade Learning Environment) [7] to simulate the Atari 2600 gaming platform, and *Petting Zoo* [8] to adapt the ALE for multi-agent playing. The input to the learning algorithm was the screen image of size 160×210 pixels. The output was an integer representing the action to be executed. ALE supports 18 distinct actions of the original Atari 2600 controller. We reduced the number of actions to the four required for playing the Pong game: 0 (no action), 1 (fire), 2 (up) in 3 (down). The same four actions were available to human players by pressing keys on the keyboard. The agent was competing against the built-in computer opponent. The agent received the positive reward +1 when hitting the opponent's goal and the negative reward -1 when receiving the goal from the opponent. The same rules applied to human players. The game ended when either side reached 21 points.

We trained the DQN for 3 million steps, which accounted for 797 played games. The game runs natively at 20 steps per second, which takes 41.67 hours of real-time gameplay to reach 3 million steps. In our case, the training process lasted about 49 hours on a desktop computer without GPU support. The training started with a high exploration factor to obtain diverse playing experiences, which means that most actions were chosen randomly. The ratio of actions chosen randomly vs. through the learned policy was linearly reduced with the number of training steps until it reached the ratio of 1:10 at 1 million steps and remained constant from that point on. We saved the weights and biases of the trained Q-network every 100,000 steps, thus obtaining 30 different AI players, each later one supposedly performing better than the preceding ones. All 30 AI players were then given to play 100 full games against the computer opponent to evaluate their average rating.

To evaluate the performance of human players, we organized a tournament in playing the Pong game, which was attended by 18 students of the University of Primorska. None of the participants were proficient in playing this game before; all could be considered complete beginners. Each participant played three full games against the computer opponent in an Atari 2600 emulator on a laptop computer. Players used a standard computer keyboard and played the game on the same system. They were not allowed to practice the game beforehand. A single game typically lasted about 15 minutes. Compared to today's standard of computer gaming, Pong is not a particularly engaging game. We observed that the participants started feeling bored and demotivated after finishing three games, which is why we limited the experiment to playing only three games. Our experiment, therefore, measures the performance of beginner players.

We argue that the final game score alone is not a good indicator of the player's performance. For example, suppose that two players both lose the game without hitting a single goal, but one player manages to play the game for a longer time than the other. Although they both received the same final score, the player with more extended gameplay demonstrates more proficiency. We, therefore, define the player's rating measured in one gameplay as:

$$\text{rating} = 10^3 \cdot \frac{\sum_{i=1}^{n} r_i \cdot d_i^{-1}}{n}, \tag{1}$$

where n is the number of episodes (either side scores a goal, and a reward is collected), $r_i = \pm 1$ is the award received by the agent, and d_i is the duration of the episode in steps. The constant 10^3 is chosen to scale the rating into an easily readable interval. This rating is used the same for AI players and humans.

3. Results

The ratings (1) achieved by human players in all 54 played games are shown in Table 1. A negative rating indicates that the human player performs worse than the built-in computer opponent, and a positive value indicates that the human player performs better. As seen from the table, participants lost all the games but were improving in performance with each game. The average rating over all 54 games was -8.03, with a standard deviation of 2.84. Therefore, we conclude that human player rating at the beginner's level is typically between -10.87 and -5.19. This interval is depicted on the plot in Figure 1 by a green strip.

Table 1
Ratings achieved by the participants in all 54 tournament games.

Player	Game 1	Game 2	Game 3	Player	Game 1	Game 2	Game 3
1	-7.72	-9.92	-5.78	10	-14.46	-9.39	-6.37
2	-12.94	-8.74	-7.73	11	-4.96	-2.39	-2.02
3	-7.71	-6.00	-5.67	12	-7.30	-8.49	3.92
4	-12.45	-10.01	-3.73	13	-9.66	-7.06	-5.92
5	-16.42	-15.00	-13.2	14	-14.58	-14.44	-10.06
6	-12.9	-7.36	-5.64	15	-5.54	-4.69	-6.60
7	-6.93	-7.49	-9.59	16	-9.90	-4.72	-4.41
8	-14.72	-10.23	-5.31	17	-5.28	-3.96	-6.49
9	-7.41	-5.67	-3.03	18	-11.55	-5.84	-4.28

We evaluated the 30 trained AI players separately from the organized tournament, but the ratings are comparable because we used the same evaluation metric. The red plot in Figure 1 shows the average rating over 100 played games for each level of AI training, with the yellow strip representing the standard deviation. The rating converges relatively quickly towards zero, which means that the trained AI player reaches the performance of the computer opponent but does not surpass it. By observing the hard-coded strategy of the computer opponent, it becomes clear that it follows a near-optimal defense strategy by always keeping the paddle aligned with the vertical position of the ball. This makes the opponent hard to defeat but not difficult to

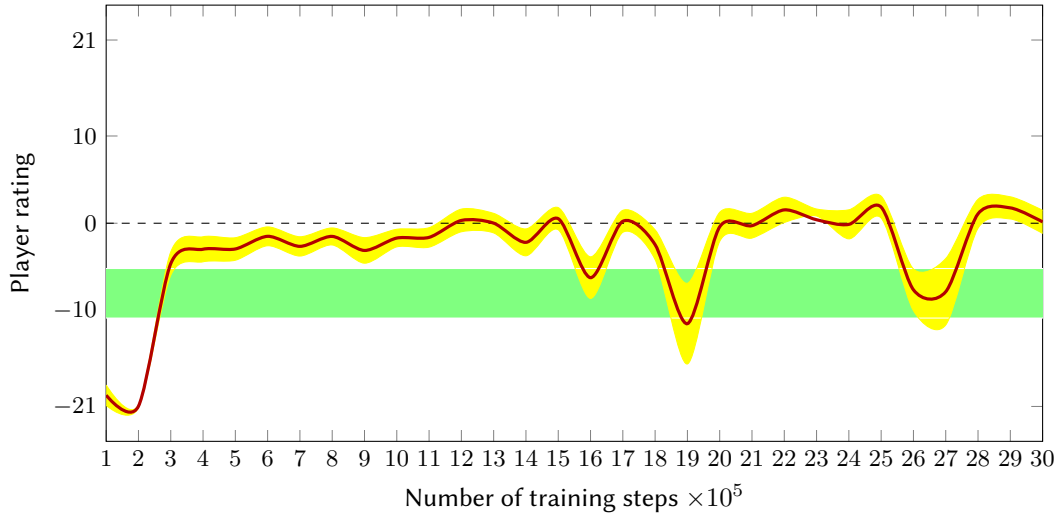

Figure 1: Comparison of the ratings achieved by deep reinforcement learning at different levels of training with average human rating.

match by inventing a similar defensive strategy. The training is also not stable, which we can see on the plot as occasional drops in performance along the training timeline. However, these instabilities are quickly compensated with further training. Human performance was surpassed in about 300,000 training steps, which is roughly 4 hours of real-time gameplay.

4. Discussion

We researched the performance of deep learning against humans in playing a simple video game. Our research question was how quickly computers could reach and outperform human players in action gaming when playing under the same sensory conditions as humans. We chose to conduct our experiment on the classical Atari 2600 Pong game and measured the performance of 18 people who played 54 games altogether. The number of games was determined after a discussion with the participants, who declared that after the third game, they were demotivated to carry out the experiment longer. We argued that the final score alone is an insufficient measurement of performance and proposed a rating that also incorporates the player's endurance. The comparison between the measured human ratings and the ratings achieved through deep reinforcement learning showed that the beginner's human level of performance in playing Pong could be surpassed by AI if trained for about 4 hours or longer. However, this may not be the case with more proficient players. Neither the AI player nor any of our participants ever managed to surpass the performance of the built-in computer opponent, but an experienced human Pong player might.

The participants were not versed in playing this game, so if they were allowed to play more games or to practice before attending the experiment, the measured human rating would probably be higher. We observed that the average ratings improved with every game played. The biggest advancement was made between the first and the second gameplay, mainly because

the participants were getting used to controlling the paddle during the first game, and only in the later games focused more on the strategy of playing. A significant drawback was also the use of a keyboard instead of the original analog controller for which the Pong game was originally made.

In this research, we only considered playing against the computer opponent. The idea of a human playing against a trained AI player is an interesting one, but it is not easy to come up with a feasible approach. Expecting the participants to play for hours with a reinforcement learning algorithm is not practical. It is also likely that the training would take a long time because humans tend to experiment with more complex playing strategies and can be inconsistent in their decisions. If trained in advance against the computer opponent, the AI player overfits to the hard-coded strategy and fails to play against humans. Another possibility is to implement adversarial training of two AI players that play against each other. However, the strategy space the two agents would have to explore is vast, and there is no guarantee that the training would be stable. Moreover, the learned adversarial strategies may not work well against human players.

References

[1] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–44.

[2] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. A. Riedmiller, A. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, Human-level control through deep reinforcement learning, Nature 518 (2015) 529–533.

[3] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, second ed., The MIT Press, 2018.

[4] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (2017) 84—-90.

[5] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL: https://www.tensorflow.org/, software available from tensorflow.org.

[6] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, W. Zaremba, OpenAI Gym, arXiv:1606.01540, 2016.

[7] M. Bellemare, Y. Naddaf, J. Veness, M. Bowling, The arcade learning environment: An evaluation platform for general agents, Journal of Artificial Intelligence Research 47 (2012).

[8] J. K. Terry, B. Black, L. Santos, Multiplayer support for the arcade learning environment, 10.48550/arXiv.2009.09341, 2020.