Learning to Win Games in a Few Examples: Using Game-Theory and Demonstrations to Learn the Win Conditions of a Connect Four Game

Ali Ayub[1[0000-0001-9458-477X]] and Alan Wagner[2]

[1] The Pennsylvania State University, State College PA 16802, USA
aja5755@psu.edu
[2] The Pennsylvania State University, State College PA 16802, USA
alan.r.wagner@psu.edu

Abstract. Teaching robots new skills using minimal time and effort has long been a goal of artificial intelligence. This paper investigates the use of game theoretic representations to represent interactive games and learn their win conditions by interacting with a person. Game theory provides the formal underpinnings needed to represent the structure of a game including the goal conditions. Learning by demonstration, has long sought to leverage a robot's interactions with a person to foster learning. This paper combines these two approaches allowing to a robot to learn a game-theoretic representation by demonstration. This paper demonstrates how a robot can be taught the win conditions for the game Connect Four using a single demonstration and a few trial examples with a question and answer session led by the robot. Our results demonstrate that the robot can learn any win condition for the standard rules of the Connect Four game, after demonstration by a human, irrespective of the color or size of the board and the chips. Moreover, if the human demonstrates a variation of the win conditions, we show that the robot can learn the respective changed win condition.

Keywords: Game Theory, social learning, interactive games, active learning, human-robot interaction

1 Introduction

In recent years, researchers have used interaction and demonstration to teach robots new activities [1, 2, 3]. Learning from demonstration (LfD) may offer a fast, intuitive, and relatively effort-free method for teaching a robot. Game theory provides the formal underpinnings needed to represent the structure of a social interaction. This paper combines these two approaches with the goal of allowing to a robot to learn a game-theoretic representation from demonstration with little prior knowledge.

Our research uses interactive games such as Connect Four to explore human-robot interaction. Games such as these are useful social paradigms that, we believe, could play an important role toward developing social robots. Interactive games are structured, both behaviorally and temporally, simplifying the task of organizing the robot's behavior. Moreover, the structure of the interactions in these types of games are agreed upon before the game is played and typically followed by both parties.

This paper attempts to develop a computational process allowing a robot to learn the win conditions for the game Connect Four using a single demonstration and a few robot generated examples with a question and answer session led by the robot. Learning a game's win conditions is typically one of the first steps to learning a new game. A person will often explicitly ask and receive instruction about how to win a game and learn to play by watching just a few examples and some associated instructions. Towards the development of such a computational process, this paper seeks to investigate the following questions: How can a robot be equipped with a similar ability to learn the win conditions of an interactive game by watching a few examples shown by a naïve human? How can the system learn various derivative win conditions from one specific win condition shown by a human?

The overreaching objective of this research is to develop a process that allows most non-experts to teach a robot the interactive game of their choice. We thus aim to create a general process that will eventually be used to teach a robot to play a variety of games including card games, board games, and Improv games. In this paper, we develop an approach to learn the win conditions of the Connect Four game but we believe that our approach can be applied to learn the win conditions of other interactive games. The remainder of the paper begins with a brief discussion of the related work, followed by a discussion of our approach, experiments and results.

2 Related Work

The field of artificial intelligence has made significant progress in developing systems capable of mastering games such as Chess, poker, and even Go [4, 5]. Deep reinforcement learning has recently been used to train autonomous agents to play a variety of Atari and other games [6, 7]. Although the robot does learn how to play the game with a considerable accuracy, the process requires large amounts of data, time, and accurate perception.

Learning from demonstration (LfD) offers a way to reduce time and effort to teach robot new skills. LfD has been used to learn a variety of tasks like table tennis [1], pick and place various objects [3], and drawer opening [8] but there has been less work focused on using LfD to teach interactive games. Others have investigated the potential of learning by watching just a few examples. For example, [3] presented a one-shot learning mechanism for picking up an object by watching just a single video but the process required a huge amount of preliminary data and time to train the robot on very similar types of tasks. Other researchers have employed active learning in which a robot uses a question-answer session with a human to learn about a new activity [9, 10, 11, 12]. Although using question and answer reduces the data and time required to learn a skill, it nevertheless is highly context dependent and may not be generalizable across different tasks. Some researchers have focused on the use of cognitive architectures like ACT-R and Soar to learn different goal directed actions [13, 14]. These approaches provide a formal method to reason about the actions taken to achieve a goal but there is a need to computationally represent those actions to be able to reason about them effectively. One objective of our research is to represent interactive games computationally in a way that makes knowledge transfer between known games and unknown games easy. This paper takes a step towards that objective.

Interactive two-player games can be formally represented using Game theory as stochastics games [15]. A game formally represents a strategic interaction among a set of players and a solution to a game is the set of strategies that can be used to best play the game. Previously, limited research has been done to investigate the application of game theory to control the interactive behavior of a robot with humans. Researchers in [16], attempted to formulate an abstracted link from game theory to control the interactive behavior of a social robot. In our previous work [17], we formulated the game-theoretic underpinnings needed to represent the interactive games and showed some preliminary results using simulations on how game theory can be used to learn different games by interaction with humans.

3 Representing Connect Four Using Game Theory

This paper focuses on the use of the techniques, mentioned in the previous section, on the game Connect Four. Connect Four requires player to place game chips in a 7x6 vertical matrix. Players win by creating a row, column, or diagonal of four continuous chips. The game is computationally simplistic which benefits our research because a broad age range of people can be easily taught to play the game.

Connect Four can be represented computationally as an extensive-form game (fig. 1). It is a perfect information extensive-form game because at each stage both players have complete information about the state of the game, actions taken by the other player and the actions available to the other player in the next stage. At each turn, a player selects a column to place their respective colored chips, hence at each stage a player has a maximum of seven actions available. Fig. 1 shows one stage of the extensive-form representation of the game.

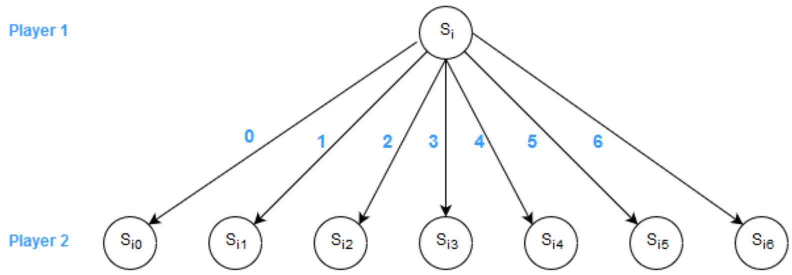

Fig. 1. Extensive-form game representation for one stage of the Connect Four game is depicted above. The lower nodes represent the game state after one of the seven actions (0-6) is chosen by player 2, the upper node depicts the current game state when player 1 chooses an action.

Images of the Connect Four game (Fig. 2 left) can be directly translated into an intermediate matrix format (Fig. 2 middle) indicating which player has pieces occupying specific positions in the matrix. This matrix can then be used to generate possible extensive-form games (Fig. 2 right) that can be checked against the game's win conditions. More importantly, the extensive-form game can be translated back into matrices and used to predict what different game states should look like or, as described later, presented to a person as potential win conditions for verification.

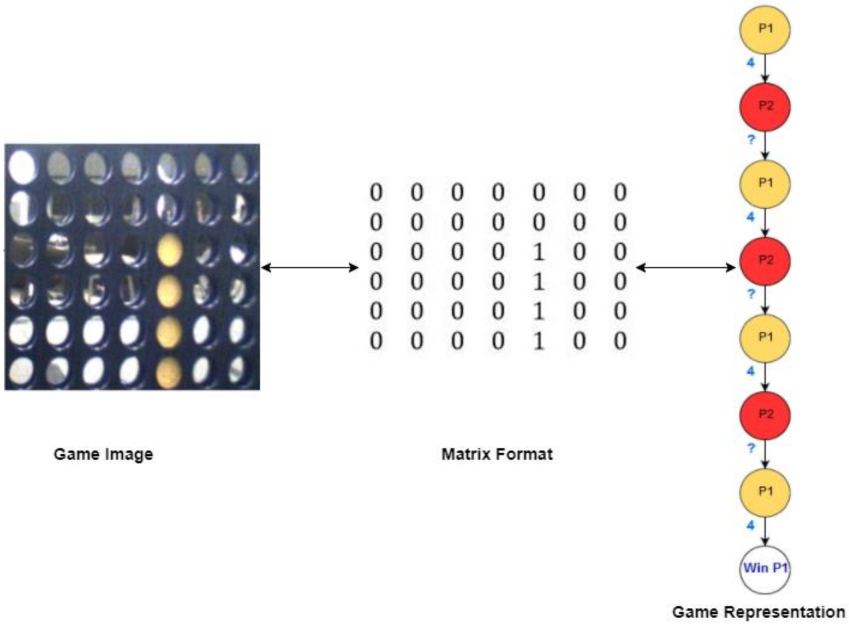

Fig. 2. A column win condition for the Connect Four game seen from the robot's perspective is shown above (left). The associated extensive-form representation is shown on the right, only the actions taken by the robot are depicted without representing all actions available at each stage. The numbers along with the arrows show the action number chosen by both players (4 by the human and ? by the robot since robot's actions are unknown). Best viewed in color.

4 Teaching a Robot the Win Conditions of Connect Four

Learning by demonstration from a human can be considered an inference problem in which the goal is to watch the actions taken by the human to accomplish the task. In the context of playing Connect Four, the person demonstrates a win condition for the game to the robot. Using the approach described above, the win condition is represented as an extensive-form game and we attempt to surmise what the rule underlying the win condition is by presenting the person with internally generated example board states and asking them whether or not the board depicts a winning game. Intuitively, the robot's behavior resembles the action of a person trying to learn a new game by generating fictional situations and asking whether or not these situations would result in a win. In order to create the extended-form game structure, the robot asks a series of questions, beginning with basic questions. The robot first asks two questions: "How many players can play this game?" And "Is this a type of game in which players take alternative turns?" Once these questions are answered by the person, the robot knows how the player's actions will iterate and a largely empty extended-form game structure can be created. Next, the robot needs to match the preprogrammed basic components of the game such as what the board looks like, what are chips and their associated colors, how to physically perform the actions related to the game to the game itself. It does that by asking the person the name of the game. We used online code for the Connect Four game which includes tools for creating the requisite robot behavior and identifying the Connect Four game pieces [18]. In the future we hope to have the robot learn these items as well.

Next the robot learns the game's win conditions. First, the robot asks the human for a demonstration of a win condition (e.g. Fig. 2 left). The robot converts the visual information into an extended-form game (Fig. 2 right). Figure 2 depicts the process of a column win. Clearly there are many other arrangements of the game pieces that will lead to other columns wins. Even though the robot does not have access to all of the game situations it can leverage the human to generate rules for winning. The robot asks the person about potential win conditions by creating different extended-form games and then converting these games to images representing the game situation that are presented to the person. The construction and use of images representing different possible game situations fosters *common ground* between the robot and the person [19]. Common ground describes the shared context that allows two individuals to fluidly communicate about a common topic. The image of the board obtained from the camera on the robot is altered to reflect possible extended-form games representing example game situations. Along with the image of the game situation a simple yes/no question is asked to confirm if that game situation will be a win condition or not. Fig. 3 depicts all the stages in learning the win condition of the connect four game.

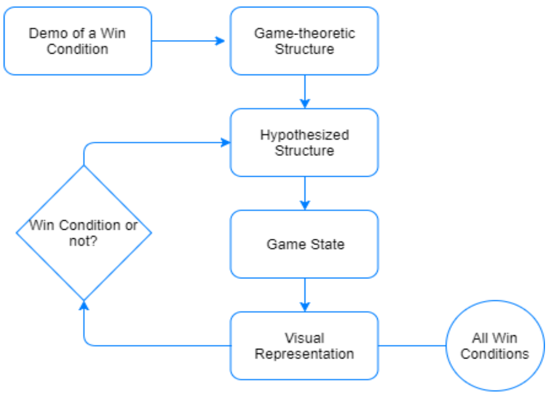

Fig. 3. A block diagram of our approach to learn the win conditions of the connect four game

Our process uses the extensive-form representation to generate potential game situations. To create these game situations, the robot first attempts to evaluate the unknowns (depicted as ? in Fig 2). To do this, the robot generates a random action that differs from the robot's action for each of the unknowns in the demonstrated win condition and presents the situation to the person. Next the robot generates game situations in which the person takes the same action as the robot (resulting in mixed column situations). Next, the robot considers moves made prior, after, or in between the column action moves. The robot then tests if the total number of actions have to be equal to the number of actions shown for the winning player, and can these actions be different than the one showed in the demonstration. An aspect of our approach is that the extensive-form formulation restricts the possible game situations that need to be asked about. Once a couple of samples of different types of extensive-form game situations are characterized by the person as containing win conditions or not, we generalize across related extensive-form games. Hence, we leverage the extensive-form game not only to create game situations for presentation to the person, but also as a means for creating subcategories of win-conditions that, collectively, describe a win rule.

As an example of how these predicted question answer sessions are performed, we show two different predicted questions related to the game-theoretic structure shown in Fig. 2. Given the game depicted in the figure, the robot first recognizes that player 2's actions are not shown, total number of actions shown by the human are four and there is only one type of action shown for player 1, and player 1 won the game. The robot generates two game situations related to the actions not shown for player 2, one focusing on whether or not the robot can take actions other than the ones shown by player1 (moves to column 4) (Fig. 4 Left) and another examining if the robot can take actions that are the same as the one shown by player 1 (moves to column 4) (Fig. 4 Right). For a column win, the answer to the later situation is no.

Fig. 4. Two examples of the different game situations presented to the human on the robot's monitor (located where a head would be). Best viewed in color.

We have used this process to generate win rules for column wins, row wins, and diagonal wins. The general process for creating the game situations for presentation to the person is the same regardless of the type of win rule. For some win rules, however, additional game situations need to be presented to the person. For example, diagonal wins require a specific pattern of moves by the other player. The robot therefore uses the diagonal win example to generate situations that break this pattern in different ways. The number of game situations required to learn a game rule depends on the rule. For a column win, seven game situations are generated. For a row win, eight game situations are created. And for a diagonal win, 12 different situations must be presented to the human.

5 Experiments

To evaluate this system, we used the Baxter robot manufactured by Rethink robotics. Google's text-to-speech API was used to communicate the questions in natural language to the person. The person answered the questions by typing inputs into a computer to avoid errors induced by the speech-to-text conversion process. The experimenter served as the robot's interactive partner for all of the experiments. To evaluate our approach, we performed several experiments.

We hypothesized that the process described above would allow the robot to learn the three Connect Four win rules (four games pieces in a row, column, or diagonal). We tested the process by providing the robot with a single correct demonstration of one type of win rule (e.g. a column win) and a human then correctly answered the robot's questions about the self-generated game situations ("Is this a win for yellow?"). We repeated this process for the other types of win rules (row and diagonal).

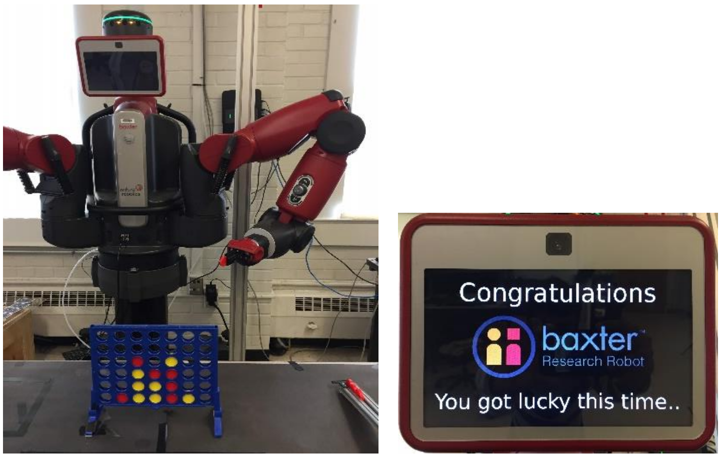

Fig. 5. An example of robot playing the connect four game after learning the rules (Left) and the message shown by the robot after the human wins the game (Right).

Next we tested the robot's ability to use the win rules to play a real game against a human opponent. We verified that the robot could correctly use the rules it had learned by playing ten games against the experimenter. Fig. 5 (Left) shows an example of robot playing the game after learning the win conditions. For all of these games the robot correctly applied the rules and demonstrated its ability to correctly identify if it or the person had won the game. Fig. 5 (Right) depicts a screen the robot displays when it recognizes a winner. These experiments demonstrated that the robot could learn the win rules from a single demonstration and by using question and answer to present the person with different game situations, ultimately arriving at a set of extensive-form games constituting a win.

To further check the ability of our approach to learn different win rules we created variants of the Connect Four game for testing. The first variant which we called Connect Five, is the same as Connect Four but requires the players to create patterns of five in a row. Applying the same process, the robot was able to correctly learn the win rules for the Connect Five game and play the game against an opponent. We then created a Connect Three game. Once again, applying the same process, the robot was able to correctly learn the win rules for the Connect Three game and play the game against an opponent.

These experiments tested the robot's ability to learn the same win conditions under different circumstances. We varied the characteristics of the demonstration, the type of win conditions (row, column, diagonal), and the nature of the win conditions (connect 3, 4, and 5). The robot learned all of the win conditions in all of these experiments. We considered conducting formal human subject experiments but pilot testing indicated that the robot's performance was likely to remain perfect so long as the person accurately answered the questions. We are currently also testing this approach on variants (different board colors, sizes, textures) of the same game. Here again we expect that the robot will learn all of the conditions because method is independent of variations in the game board or pieces. Given that the natural source of variance in this paradigm is primarily the human subject and/or the game, future work will focus on testing this approach on games with a variety of rule versions (such as card games).

6 Conclusion

This paper has illustrated the use of game-theoretic representations of interactive games as a means of learning the win conditions of these games by interacting with a human. Our experiments show that a single demonstration accompanied with a few directed examples can be used to learn the win conditions of for the Connect Four game. The extended-form game is used represent the game state, to create realistic images for communication with the person, and to devise potential win conditions. Our experiments demonstrate the generality of the approach across win conditions, demonstrations, and game variants.

This paper makes several important contributions. First, a method that connects the robot's perception of the game state to a computational and generally applicable representation is developed. These connections allow to the robot to generate visual predictions of future game states for grounded communication with the human. Second, our approach leverages the extended-form representation to make predictions about different game states. This limits the space of games states to manageable number that the robot can ask the person about. Finally, we present a method that, we believe, will be generally applicable to a variety of different games (and possibly other HRI environments). Although we have yet to show it, we see no reason that prevents this method from generalizing across a wide variety of games and contexts. Ultimately we hope that this avenue of research will not only offer a means for a robot to structure its interactions with a person, but also allow the robot to bootstrap an interactive exchange by using similar experiences represented in extended-form as a model for an upcoming interaction.

Naturally this research is not without limitations. We assume that the person demonstrates a valid win condition and that they correctly answer the questions about the other game situations. Moreover, this work does not address how the robot learns about the game's actions or the game components (board, tokens). Still, we strive to make as few assumptions as possible about the nature or structure of the game. Our hope is that the system will allow the robot to learn the structure of complete games.

The next step of our research will be to examine how the rules learned in this game can be transferred to less similar games. Considering, for example, card games one might use the process to look at different variants of poker or other games. Here learning by demonstration could perhaps be used to bootstrap the learning of new games from previously learned ones. Ultimately, we believe that these techniques take us one step closer to robots that can learn to interact across a wide variety of situations.

Acknowledgement

Support for this research was provided by Penn State's Teaching and Learning with Technology (TLT) Fellowship.

References

1. M¨ulling, K., Kober, J., Kroemer, O. and Peters, J., "Learning to select and generalize striking movements in robot table tennis.," *The International Journal of Robotics Research (IJRR),* 2013.
2. Pastor, P., Hoffmann, H., Asfour, T. and Schaal, S., "Learning and generalization of motor skills by learning from demonstration," *International Conference on Robotics and Automation (ICRA),* 2009.
3. Yu*, T., Finn*, C., Xie, A., Dasari, S., Zhang, T., Abbeel, P. and Levine, S., "One-Shot Imitation from Observing Humans via Domain-Adaptive Meta-Learning.," in *Robotics: Science and Systems (RSS),* Pittsburgh, PA, USA, June, 2018.
4. Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., L. Sifre, Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K. and Hassabis, D., "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm," in *arXiv Reprint arXiv:1712.01815,* 2017 .
5. Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Driessche, G. v. d., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N. and Sutskever, I., "Mastering the game of Go with deep neural networks and tree search," *Nature,* vol. 529, pp. 484-489, 2016.

6. Kamei, K. and Kakizoe, Y., "An Approach to the Development of a Game Agent Based on SOM and Reinforcement Learning," in *5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, Kumamoto, Japan, 2016.
7. Dobrovsky, A., Borghoff, U. M. and Hofmann, M., "An approach to interactive deep reinforcement learning for serious games," in *7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Wroclaw, Poland, 2016.
8. Rana, M. A., Mukadam, M., Ahmadzadeh, S. R., Chernova, S. and Boots, B., "Towards robust skill generalization: Unifying learning from demonstration and motion planning," in *2017 Conference on Robot Learning (CoRL)*, 2017.
9. Whitney, D., Rosen, E., MacGlashan, J., Wong, L. L. and Tellex, S., "Reducing Errors in Object-Fetching Interactions through Social Feedback," in *IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, 2017.
10. Racca, M. and Kyrki, V., "Active Robot Learning for Temporal Task Models," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, Chicago, IL, USA, 2018.
11. Cakmak, M. and Thomaz, A. L., "Designing robot learners that ask good questions," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* , Boston, Massachusetts, USA, 2012.
12. Hinrichs, T. R. and Forbus, K. D., "Goes First : Teaching Simple Games through Multimodal Interaction," *Advances in Cognitive Systems,* vol. 3, pp. 31-46, 2014.
13. Stearns, B. and Laird, J. E., " Modeling Instruction Fetch in Procedural Learning," in *16th International Conference on Cognitive Modelling (ICCM)*, Madison, WI, 2018.
14. Oh, J. H., Suppe, A., Duvallet, F., Boularias, A., Navarro-Serment, J. V. L. E., Romero, R. Dean, O., Lebiere, C., Hebert, M. and Stentz, A., "Toward Mobile Robots Reasoning Like Humans," in *Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*, January, 2015.
15. Osborne, M. J. and Rubinstein, A., A Course in Game Theory, Cambridge, MA: MIT Press, 1994.
16. Lee, K. and Hwang, J.-H., "Human-Robot Interaction as a Cooperative Game," *O. Castillo (ed.) et al., Trends in intelligent systems and computer engineering (IMECS 2007),* 2008.
17. Wagner, A. R., "Using Games to Learn Games: Game-Theory Representations as a source for Guided Social Learning," in *Seventh International Conference on Social Robotics (ICSR 2016)*, Kansas City, MO, 2016.
18. "Connect Four Demo," Rethink Robotics, [Online]. Available: http://sdk.rethinkrobotics.com/wiki/Connect_Four_Demo.
19. Clark, H. H. and Brennan, S. E., "Grounding in communication," *Perspectives on socially shared cognition,* vol. 13, pp. 127-149, 1991.