On Human-like Performance Artificial Intelligence: A Demonstration Using an Atari Game

Seng-Beng Ho, Xiwen Yang, Therese Quieta, Gangeshwar Krishnamurthy, Fiona Liausvia
AI Programme, A*STAR, Singapore
hosengbeng@gmail.com, Yang_Xiwen@scei.a-star.edu.sg, therese_quieta@ihpc.a-star.edu.sg, Gangeshwar_krishnamurthy@ihpc.a-star.edu.sg, liausviaf@ihpc.a-star.edu.sg

Abstract

Despite the progress made in AI, especially in the successful deployment of deep learning for many useful tasks, the systems involved typically require a huge number of training instances, and hence a long time for training. As a result, these systems are not able to rapidly adapt to changing rules and constraints in the environment. This is unlike humans, who are usually able to learn with only a handful of experiences. This hampers the deployment of, say, an adaptive robot that can learn and act rapidly in the ever-changing environment of a home, office, factory, or disaster area. Thus, it is necessary for an AI or robotic system to achieve human performance not only in terms of the "level" or "score" (e.g., success rate in classification, score in Atari game playing, etc.) but also in terms of the speed with which the level or score can be achieved. In contrast with earlier DeepMind's effort on Atari games, we describe a system that is able to learn causal rules rapidly in an Atari game environment and achieve human-like performance in terms of both score *and* time.

1. Introduction

Artificial intelligence (AI) has taken great strides in many domains of applications. However, there has been realization that even though many AI systems can perform certain tasks very well that normally require human intelligence, and sometimes even superseding human abilities in those tasks, their performance is not "human-like" in some aspects. For example, when deep learning is applied to pattern classification and recognition, the accuracy is very high and sometimes outstrips human performance. However, humans usually require only a few instances of training examples to learn to classify and recognize the objects involved with high accuracy, whereas deep learning systems typically require many orders of magnitude of the number of training examples needed by humans. Thus, we can distinguish two aspects of judging the capability of an intelligent system, human or artificial. There is the *level* of performance, which is often a percentage score on the success on some tasks, such as classification, and the other is the *time* taken to learn. Human-like performance means the system must perform well on both measures and this is critical in many real-world applications.

One notable example recently is the DeepMind's seemingly successful attempt in using deep reinforcement learning to play Atari games [1]. Their measure of success in playing these games focuses on the "score" measure – i.e., is the system able to score well, at human score levels. By that measure, they succeeded reasonably well - in more than 50% of the games involved, the system was able to score higher than that of humans. However, by the measure of *time*, DeepMind's system plays at a speed many orders of magnitudes slower than that of human players - Tsividis [2] pointed out this large discrepancy.

Instead of reinforcement learning, we believe human players use the learning and understanding of causality to learn how to play Atari games. Ho and Zhu's groups have developed a framework and method to learn causality from visual input [3-12]. In this paper, an Atari game, Space Invaders, is used to demonstrate that a framework based on learning of causal rules from the visual environment together with an AI problem solving process can achieve human-like performance *both* in terms of level (score) and *time taken* [12].

2. A Causal Learning and Problem Solving Framework

2.1. Basic Idea Behind Causal/Temporal Learning

Statistics takes a conservative stance with regards to correlations without intervention, which is that it may not imply causality [13]. Yang and Ho [14] take the stance that both causality and temporal correlations are important for AI's purposes. If one can establish the correlation between an intervention/action and a subsequent effect, thus establishing the causality between them, one can use it for (i) prediction – if the action is taken, the effect is expected; and (ii) problem solving – to achieve the effect, one can take the action. On the other hand, if a temporal correlation is observed between two events, the first of which is not an intervention/action taken by the system/human, then the temporal correlation is useful only for prediction – if the first event is observed, the second event is likely to follow.

2.2. Causal/Temporal Rule Learning

Fig. 1 illustrates the basic causal/temporal learning method used by Ho's group [6, 7, 14].

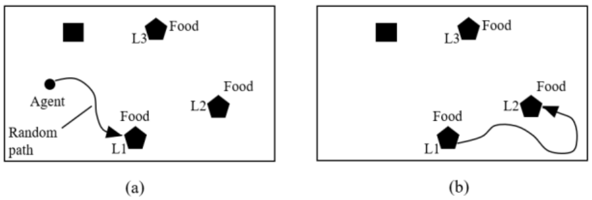

Figure 1: See text for explanation.

In Fig. 1(a) it is shown that an Agent explores around in an environment and accidentally *Touches* a piece of Food at time *T1* and at location *L1* and it finds itself experiencing an increase in energy. A causal rule can be learned as such: *At(Agent, L1, T1) & Touch(Agent, Food, L1, T1)* ➔ *Energy_Increase(Agent, L1, T1+Δ)*. This is a *specific* causal rule: it says as currently understood, the *Energy_Increase* can take place if the Agent *Touches* the Food at location *L1* and time *T1*.

After another instance of the event shown in Fig. 1(b), a general rule *At(Agent, Any L, Any T) & Touch(Agent, Food, Same Any L, Same Any T)* ➔ *Energy_Increase(Agent, same Any L, Same Any T) is learned.*

2.3. The Atari Game Space Invaders

In this section, we describe the basic approach of the game playing system [12]. Fig. 2(a) shows a screen shot of the Space Invaders game. In Fig. 2(b), we show that the symbolic predicate description of the scene of the Space Invaders game is extracted through a "vision module." The data is organized in a temporal form: at each time frame, there is a predicate description of every entity and their associated parameters (Fig. 1, previous section). So, for example, the description of a few of the Space Invaders and the Player at time frame *Time(t1)* would be: *Time(t1) – At(Invader(ID=10), x1, y1), (At(Invader(ID=11), x2, y2)… At(Player(x10, y10))….* Learning then takes place on this level of environmental description.

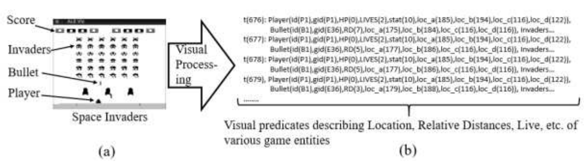

Figure 2: (a) The Space Invaders Game. (b) The symbolic predicates extracted from the scene.

2.4. A System for Causal/Temporal Learning, Reasoning, and Problem Solving

In traditional AI, a General Problem Solver (GPS) acts on the facts and knowledge involved in particular domains to derive solutions for particular situations [15]. Our approach is similar here, except that the knowledge, in the form of causal rules, is learned from causal/temporal learning, while in traditional AI, the knowledge involved was typically hand-coded. Fig. 3(a) shows this basic structure [12].

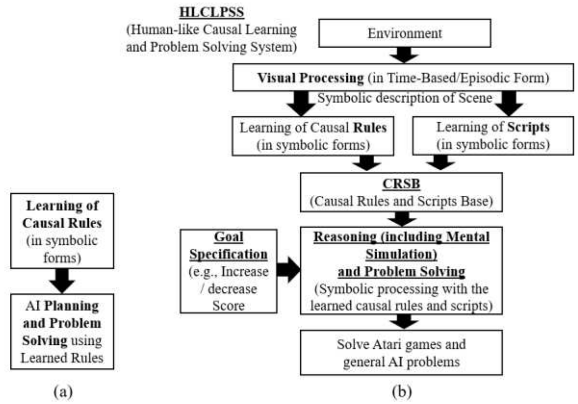

Figure 3: (a) The basic overall causal learning and problem solving framework. (b) The detailed processing modules of the Human-like Causal Learning and Problem Solving System (HLCLPSSS). [12]

Fig. 3(b) shows that the processing begins with the Environment, which for Space Invaders would be the Space Invaders' visual scene. A Visual Processing module converts that to a time-based, episodic form as described above (Fig. 2(b)).

Next, causal rules, much like those discussed in connection with Fig. 1 are learned and encoded. The system also learns and encodes Scripts - sequences of any 5 actions observed in the environment are stored as Scripts. This vastly cuts down the search space of the problem solving process. Following this, the system carries our reasoning and problem solving, including mental simulation.

2.5. Goal-Directed Problem Solving

There are two kinds of goals - Goal to achieve a desired state (Increase of Score) and Goal to avoid an undesired state (Destruction of Player). These are described as follows:

Goal to Achieve a Desired State and the Associated Learning Process

Fig. 4 illustrates a typical situation in Space Invaders in which there is a desired Goal to achieve.

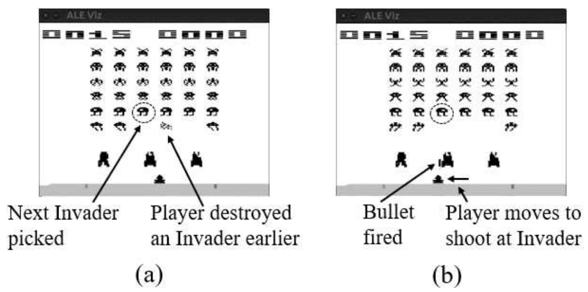

Figure 4: To achieve a desired goal in Space Invaders: See text for explanation

In Fig. 4(a), it is shown that the Player is not in a position to fire a bullet to destroy an Invader. The Player carries out a backward chained problem solving process and obtains a solution – move to a location at which the Invader is in the direct line of fire and fire a bullet to destroy the Invader.

The learning process proceeds as follows. In an initial "exploration phase" (much like the exploration phase of reinforcement learning), the Player fires at random, and occasionally a bullet would hit an Invader and destroy the Invader. After a few instances of similar experience, a causal rule such as this is learned: *At(Player_Bullet, Any Location, Any Time) & Contact(Player_Bullet, Invader(Any ID), Any Same Location, Any Same Time)* ➔ *Destroyed(Invader(Any Same ID), Any Same Location, Any Same Time + 1)*. (In our implemented system, the learned rules may not look as "clean", as there are other "noisy" conditions that "creep" into the rule, but they suffice for problem solving purposes and this clean rule is good for illustrating the basic idea here.). When an Invader is destroyed, the Score goes up, and that is a desired Goal.

After all these causal rules have been learned, the system is ready to carry out backward chained problem solving. At all times, the system is in the mode of looking for ways to achieve a desired Goal, and in this case, it would be *Increase_Score* or *Destroyed(Invader(Any ID, Any Location, Any Time).*

Goal to Avoid an Undesirable State and the Associated Learning Process

Fig. 5 illustrate a typical situation in Space Invaders in which there is an undesired Goal to avoid.

In Fig. 5(a), it is shown that an Invader fires a bullet at the Player. Using *mental simulation* based on the earlier learned, known rules of the bullet's behavior, the system knows that some time in the future the bullet will hit the Player (because it is in the bullet's path) and the Player will be destroyed. The system therefore concocts a plan to *prevent* this from happening. The solution is to move left a little bit as shown in Fig. 5(b). In Fig. 5(c) it is shown that the bullet and hence the destruction of the player is successfully avoided.

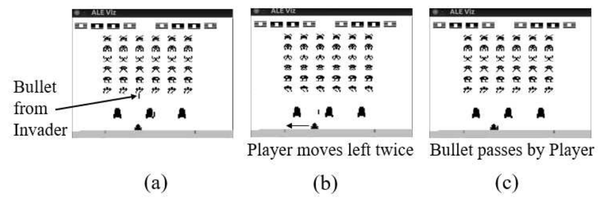

Figure 5: Player avoids an undesired goal. See test for explanation.

In the beginning of the Space Invaders game, the situation of experiencing the undesired goal is first learned in a few instances in which the Invaders fire bullets and they destroy the Player (this is not random – the game engine deliberately does that). The entire script learned is:

Appear(Bullet, Loc1) ***and*** *Move(Bullet, Loc2)* ***and*** *Move(Bullet, Loc3)* ***and…***
Contact(Bullet, Player, Loc10) ➔ *Destroyed(Player, Loc10)* (1)

Through contrapositive reasoning, this is converted into:

Not(Destroyed(Player, Loc10) ➔
Not(Appear(Bull Appear(Bullet, Loc1) ***or***
Not(Move(Bullet, Loc2) ***or***
Not(Move(Bullet, Loc3) ***or***
Not(Contact(Bullet, Player, Loc10)) (2)

which means that *any* of the actions taken to negate the original events in the sequence is sufficient to achieve a *negation* of *Destroyed(Player, Loc10),* which is the desired Goal of avoiding an undesired state.

The system then queries its Causal Rule and Script Base (CRSB) in Fig. 3 to see if there is any ready solution to effect at least one of the negations. If not, then it attempts random actions to see if that can achieve the Goal. It turns out that in this case by randomly emitting a sequence of left and right movements of the Player, a *Not(Contact(Bullet, Player, Loc10))* can be achieved – typically by moving left or right by a few pixels. This involves a search process with a small state space.

3. Results of Human-like Performance Space Invaders Game Playing System

The various causal learning, reasoning, and problem solving processes, including internal mental simulations processes (Fig. 3(b)), have been implemented and tested on the Space Invaders game [12]. Fig. 6 shows the results.

In Fig. 6, the results from 3 trials played by our Human-like Causal Learning and Problem Solving System

(HLCLPSS), are shown along with (i) human novice performance; and (ii) DeepMind's deep reinforcement learning results, as reported in their paper [1], time-scaled based on the total number of video frames needed before certain performance is achieved (video frame is 30 frames per second). We also executed DeepMind's publicly accessible code to obtain its performance in the early part of the game – up to 20 hours of play time. There is a level of score, about 200, when the Player shoots at random with no goal-directed behavior, and it is shown as a line labeled "Avg Random Play".

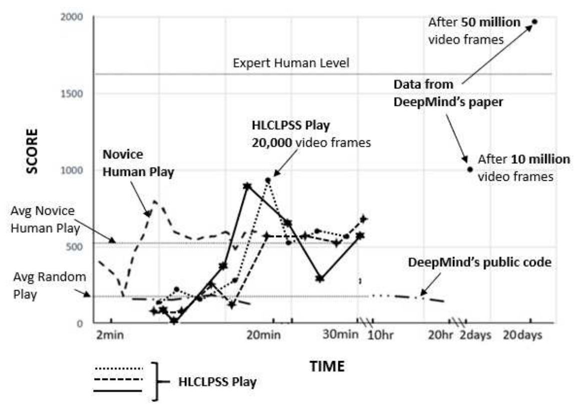

Figure 6: Results. See text for explanation [12].

The results show that HLCLPSS is able to achieve very close to novice human player's performance with respect to score and speed, the speed of which is many orders of magnitude faster than that of DeepMind's system.

4. Conclusion

In this paper, we first define what we mean by human-like performance AI, which is a system that is not only able to achieve human performance in terms of "level" or "score" (like the percentage accuracy in classification or the game score in a computer game), but it must also achieve the level or score in reasonably short, human-like time frame. Then we describe a causal learning and problem solving framework to demonstrate how, when applied to an Atari game Space Invaders, it is able to achieve human-like performance – achieving human-like game *score* in human-like *time* frame.

What have demonstrated in this paper is the ability of the system to reach human-like performance at the human *novice* level. We are currently continuing to enrich the basic framework of Fig. 3 to allow the system to reach human *expert* level performance, within human-like learning *time* frames.

Future research will apply the basic system to more Atari games to further explore some fundamental issues, as well as to apply the basic causal learning and problem solving framework to real world robotic situations.

References

[1] Mnih, V., et al.: Human-level control through deep reinforcement learning. Nature 518, 529-533 (2015).

[2] Tsividis, P. A., Pouncy, T., Xu, J. L., Tenenbaum, J. B., Gershman, S. J.: Human learning in Atari. In: AAAI Spring Symposium Technical Report, pp. 643-646. AAAI: Palo Alto, California (2017).

[3] Fire, A., Zhu, S.-C.: Learning perceptual causality from video. ACM Transactions on Intelligent Systems and Technology, 7(2), 23 (2012).

[4] Fire, A., Zhu, S.-C,: Inferring hidden statuses and actions in video by causal reasoning, In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 9-16. IEEE Press, Piscataway, New Jersey (2017).

[5] Ho, S.-B.: On effective causal learning. In: Proceedings of the 7th International Conference on Artificial General Intelligence, pp. 43-52. Springer-Verlag, Berlin. (2014).

[6] Ho, S.-B.: Principles of Noology: Toward a Theory and Science of Intelligence. Springer International Publishing, Switzerland (2016).

[7] Ho, S.-B.: Deep thinking and quick learning for viable AI. In: Proceedings of the Future Technologies Conference, pp. 156-164. IEEE Press, Piscataway, New Jersey (2016).

[8] Ho, S.-B.: The role of synchronic causal conditions in visual knowledge learning. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 9-16, IEEE Press. Piscataway, New Jersey (2017).

[9] Ho, S.-B., Liausvia, F.: A Ground level causal learning algorithm. In: Proceedings of the IEEE Symposium Series on Computational Intelligence for Human-like Intelligence, pp. 110-117. IEEE Press, Piscataway, New Jersey (2016).

[10] Ho, S.-B., Liausvia, F.: On inductive learning of causal knowledge for problem solving. In: Technical Reports of the Workshops of the 31st AAAI Conference on Artificial Intelligence, pp. 735-742. AAAI, Palo Alto, California (2017).

[11] Ho, S.-B., Edmonds, M., S.-C., Zhu: Actional-perceptual causality: Concepts and inductive learning for AI and robotics. Presented at the Workshop on Perspectives on Robot Learning: Imitation and Causality, Robotics: Science and Systems Conference, Carnegie Mellon University, Pittsburgh, Pennsylvania (2018).

[12] Ho, S.-B., Yang, X., Quieta, T., Krishnamurthy, G.. Liausvia, F.: On Human-like Performance Artificial Intelligence – A Demonstration Using an Atari Game. In: International Conference on Artificial Intelligence and Security, New York (2019).

[13] Agresti, A., Franklin, C.: Statistics: The Art and Science of Learning from Data (3rd ed.). Pearson, Boston (2007).

[14] Yang, X., Ho, S.-B.: Learning correlations and causality through an inductive bootstrapping process. In: IEEE Symposium on Computational Intelligence, IEEE Press, Piscataway, New Jersey (2018).

[15] Newell, A., Shaw, J.C., Simon, H.A.: Report on a general problem-solving program. In: Proceedings of the International Conference on Information Processing, pp. 256–264 (1959).