Will Reasoning Improve Learning?

By Nicolaas J. Vriend
*Economic Letters, 1997, vol. 55, no. 1, p. 9-18*

CSE 254 – Seminar on Learning Algorithms
Hector Jasso
May 2, 2001

Will Reasoning Improve Learning?

1. Ultimatum Game
2. Adding Reasoning to Learning
3. Effects of Reasoning
4. Conclusion
1. Ultimatum Game

Definition of Ultimatum Game

- Two players, A and B, and a pie.
- Player A proposes how to split the pie (example: A gets 80%; B gets 20%).
- Player B accepts/rejects the proposal. Accept = pie is split. Reject = pie is thrown away.

- Alternative Environments:
  1. A and B play exactly once.
  2. A and B play together repeatedly.
  3. A plays repeatedly with different partners.
What does Player A choose?

- **Game theoretic:**
  - Player A offers the *minimum*. Player B *accepts*.

- **Empirical evidence** (gathered in environment 3):
  - Player A offers somewhat less than half the pie to player B.

---

Game Theory

- Interdisciplinary approach to the study of *human behavior*.

- Disciplines involved: *mathematics, economics* and other *social and behavioral sciences*.

Game Theory and Economics

- The key link between neoclassical economics and game theory is rationality.

- Neoclassical economics assumes that people are rational in their choices.

- Game theory helps explore “abnormal” situations like restricted competition.

Are Humans Rational?

- Do humans choose strategies “rationally” when the outcome depends on the strategies of others or information is incomplete?

- Are people more cooperative/aggressive than would be “rational”? 
2. Adding Reasoning to Learning

Spectrum of Modes of Individual Behavior

- Players are **fully introspective** about themselves and others.

- Besides **reasoning**, players learn from **past experience**.

- Players do not reason, they only learn from **past experience** (reinforcement learning.).
Interest in Reinforcement Learning

• **Computer Science**: because of its success in performing difficult tasks.

• **Psychology**: for explaining empirical evidence of subjects in experiments.

• **Economics**: benchmark more attainable in reality than perfect rationality.

---

Actual Reinforcement Learning

• Player A behaves adaptively to her environment:

  – Player A will try any of \( k \) different actions, and repeat those that led to high payoffs in the past.
  – **Propensity** of trying option \( k \) is updated according to the payoff \( z \).
    \[
    q_k(t + 1) = q_k(t) + z
    \]
  – Probability of choosing option \( k \) is.
    \[
    p_k(t) = \frac{q_k(t)}{\sum_k q_k(t)}
    \]
Reinforcement Learning

- **Choose** an action according to probabilities.
- **Deduce** information about payoffs.
- **Update** propensities to choose actions.

Alternate types of step 2

A. **Actual reinforcement.**

B. **Vicarious reinforcement:** incorporate observation of other agents and advice from supervisor.

C. **Virtual reinforcement:** use imagination regarding un-chosen actions and foregone benefits.
Virtual Reinforcement

- In virtual reinforcement, player A can reason:
  - If B accepts offer x, B will also accept higher offers $x' > x$.
  - If B rejects offer x, B will also reject lower offers. $x' < x$.

- This introduces an asymmetry in the information obtained by player A!

3. Effects of Reasoning
Actual Reinforcement Learning

1. The pie has size $P$.
2. Possible offers: $x = 0, 1, 2, \ldots, P$.
3. Player B accepts every offer.
4. Player A tries every action equally often, say $n$ times.
5. Payoff for A: if player B accepts, payoff is $P$ – offer. If player B rejects, 0.
6. Propensity increases according to the reward.
7. Only actual reinforcement learning takes place.

---

Proposition: The most reinforced offer will be $x = 0$.

Proof: After trying each possible offer $n$ times,

$$r(x) = n \cdot (P - x)$$

which has a maximum at $x = 0$. 

---
1. The pie has size $P$.
2. Possible offers: $x = 0, 1, 2, \ldots, P$.
3. Player B accepts every offer.
4. Player A tries every action equally often, say $n$ times.
5. Payoff for A: if player B accepts, payoff is $P - \text{offer}$. If player B rejects, 0.
6. Propensity increases according to the reward.
7. Player A reasons: if B accepts $x$, then B would accept $x' > x$.

\section*{Proposition:} The most reinforced offer will be $x$, where $x > (P - 2)/2$ and $(x - 1) < (P - 2)/2$

\section*{Proof:} After trying each possible offer $n$ times,
\[ r(x) = n \cdot (x + 1) \cdot (P - x) \]
Taking the first difference gives:
\[ r(x + 1) - r(x) = n \cdot (P - 2x - 2) \]
Hence $r(x + 1) - r(x) < 0$ if $x > (P - 2)/2$
Reinforcement Learning

Fig. 1. Outcomes of two types of reinforcement process.

Relaxing the Assumptions

- Vary the size of $P$.
- Allow non-integer offers $P$.
- Player $B$ does not play perfect equilibrium game.
- Player $A$ does not try every strategy equally often.
- Non-linear environment.
- Average reinforcements.
4. Conclusion

When *virtual updating* is considered, information asymmetry introduces a *bias* away from the *perfect equilibrium strategy*.

On the path from *basic reinforcement learning* to *fully introspective reasoning*, virtual reinforcement leads to strategies farther away from the *game-theoretic rational strategy*. 
Conclusion

One has to be cautious with *ad hoc* models of learning and adaptive behavior, in particular with so-called “self-evident” improvements of learning.