Learning: Reinforcement, Fictitious Play, and EWA

Joseph Tao-yi Wang

4/24/2008

Outline: Learning

1. Learning: What you do after you see "results"…
2. What we know now: (various learning rules)
   1. Reinforcement
   2. Belief learning
   3. EWA: a hybrid of reinforcement and belief learning
   4. Others: Evolutionary, anticipatory learning, imitation, learning direction theory, rule learning, etc.
3. Further research:
   1. Beyond: New direction for research in learning
   2. Application: How can we use these tools?

What you do after you see…

- Suppose you are playing "stag hunt"
- (B, L) happened last time
- What would you do now?

- Change strategy?
- Stick to it?

	L	R
T	3, 3	0, 1
B	1, 0	1, 1

What you do after you see…

- A robot (pre-programmed) would stick to it
  – Evolutionary approach
- But humans think twice
- How would you switch?
- Reinforcement:
  – Choices "reinforced" by previous payoffs
  – "Very bad" reasoning

	L	R
T	3, 3	0, 1
B	1, 0	1, 1

Reinforcement Learning

- Update attractions (tendency to play a certain strategy) after given history
- Reinforcement:
  – Choices "reinforced" by previous payoffs
  – Allow spillovers to "neighboring strategies" ε
- Example: (cumulative)
- $A^B(t) = \varphi A^B(t-1) + (1 - \varepsilon) [1 - \rho(t-1)]$
- $A^T(t) = \varphi A^T(t-1) + \varepsilon$

Reinforcement Learning

- Cumulative:
- $A^B(t) = \varphi A^B(t-1) + (1 - \varepsilon) [1 - \rho (t-1)]$
- $A^T(t) = \varphi A^T(t-1) + \varepsilon$

- Weighted averages:
- $A^B(t) = \varphi A^B(t-1) + (1 - \varphi) (1 - \varepsilon)$
- $A^T(t) = \varphi A^T(t-1) + (1 - \varphi) \varepsilon$

What "else" could you do…

- Would you "update your beliefs about what others do"?
  - Belief learning models
- Fictitious play
  - Keep track of frequency
  - Ex: rock-paper-scissors
- Cournot best-response
  - What you did last time is what you'll do now

	L	R
T	3, 3	0, 1
B	1, 0	1, 1

Weighted Fictitious Play

- Other weights? Weighted fictitious play
  - Fictitious play: weigh all history equally ($\rho=1$)
  - Cournot: focus only on the last period ($\rho=0$)
- Prior:
  - $P_{t-1}(L) = 3/5$, $P_{t-1}(R) = 2/5$
- Posterior:
  - $P_{t-1}(L) = (3\rho + 1) / (5\rho + 1)$
  - $P_{t-1}(R) = (2\rho + 1) / (5\rho + 1)$, ρ = decay factor

Weighted Fictitious Play

- Posterior:
  - $P_{t-1}(L) = (3\rho + 1) / (5\rho + 1)$
  - $P_{t-1}(R) = (2\rho + 0) / (5\rho + 1)$
- Use this belief to compute payoffs:
- $A^T(t) = [\ 3\ (3\rho + 1) + 0\ (2\rho + 0)\ ] / (5\rho + 1)$
- $A^B(t) = [\ 1\ (3\rho + 1) + 1\ (2\rho + 0)\ ] / (5\rho + 1)$
- Note: Actually payoff received play no role

Could you being doing both?

- Reinforcement does not update beliefs
  - But people DO update!
- Fictitious play doesn't react to actual payoffs
  - But people DO respond
- EWA: a hybrid of two
  - Camerer and Ho (Econometrica, 1999)

	L	R
T	3, 3	0, 1
B	1, 0	1, 1

Experience-Weighted Attraction

- Add δ : the weight players give to forgone payoffs from unchosen strategies
  - Law of effect vs. Law of simulated effect
- $A^B(t) = [\ \varphi\ N(t-1)\ A^B(t-1) + 1] / N(t)$
- $A^T(t) = [\ \varphi\ N(t-1)\ A^T(t-1) + 3\delta] / N(t)$
  where $N(t) = \varphi(1 - \kappa)\ N(t-1) + 1$
- $N(t)$: Experience weight (weakly increasing)

EWA Special Case: Reinforcement

- $A^B(t) = [\ \varphi\ N(t-1)\ A^B(t-1) + \pi(B,L)\ ] / N(t)$
- $A^T(t) = [\ \varphi\ N(t-1)\ A^T(t-1) + \pi(T,L)\ \delta] / N(t)$
  where $N(t) = \varphi(1 - \kappa)\ N(t-1) + 1$
- $\delta = 0$, $N(0) = 1$: Reinforcement!
- $\kappa = 1$: (Simple) cumulative reinforcement
- $\kappa = 0$: (Weighted) average reinforcement
  - Weights are $\varphi/ (\varphi + 1)$ and $1/(\varphi + 1)$

EWA S.C.: Weighted Fictitious Play

- $A^B(t) = [\ \varphi\ N(t\text{-}1)\ A^B(t\text{-}1) + \pi(B,L)\ ]\ /\ N(t)$
- $A^T(t) = [\ \varphi\ N(t\text{-}1)\ A^T(t\text{-}1) + \pi(T,L)\ \delta]\ /\ N(t)$
  where $N(t) = \varphi(1 - \kappa)\ N(t\text{-}1) + 1$
- $\delta = 1, \kappa = 0$: Weighted Fictitious Play!
  - Good Homework exercise…
  - Hint: $N(t)=1 + \varphi + \ldots + \varphi^{t-1}$; Posterior is

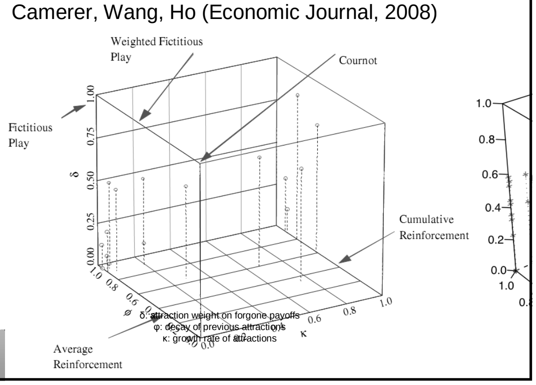
$$P_t(L) = \frac{I(L,h(t)) + (\varphi + ... + \varphi^{t-1}) \cdot P_{t-1}(L)}{1 + \varphi + ... + \varphi^{t-1}}$$

EWA S.C.: Weighted Fictitious Play

- $A^B(t) = [\ \varphi\ N(t\text{-}1)\ A^B(t\text{-}1) + \pi(B,L)\ ]\ /\ N(t)$
- $A^T(t) = [\ \varphi\ N(t\text{-}1)\ A^T(t\text{-}1) + \pi(T,L)\ \delta]\ /\ N(t)$
  where $N(t) = \varphi(1 - \kappa)\ N(t\text{-}1) + 1$
- $\delta = 1, \kappa = 0$: Weighted Fictitious Play!
  - φ = 1: Fictitious Play
  - φ = 0: Cournot best-response

The EWA Cube

Camerer, Wang, Ho (Economic Journal, 2008)



Interpretation of EWA Parameters

- δ: Decay of previous attractions
- κ: The rate attractions grow
- $N(t)$: The strength of initial attractions (in units of "experience-equivalence")
  - φ: Weight in $N(t)$

The EWA Cube

Camerer, Wang, Ho (EJ08') vs. Wang, Knoepfle, Camerer 08'

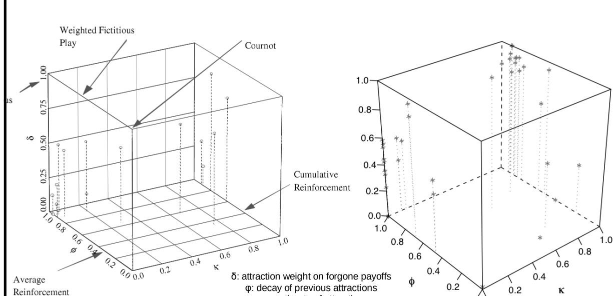


Fig. 2. *EWA's Model Parametric Space*

Prediction Power of EWA

- EWA generally improves accuracy in about 35 games (except for mixed ones)
  - See Camerer and Ho (book chapter, 1999)
  - "Long version" of the Econometrica paper?
- BGT, Ch. 6 provides two examples:
  - Continental Divide
  - p-Beauty Contest

Prediction Power of EWA

- Overfitting: Too many parameters?
- Can be tested:
  – Restricted fit vs. Unrestricted
- Better Out-of-sample Prediction Power:
  – Estimate parameters and predict “new data”
  – Not prone to overfitting (because of new data)
- 1-parameter “self-tuned EWA” works too:
  – EWA “Lite” Does as good as reinforcement or fictitious play, even on data with new games

Other Learning Rules

- Other rules include:
- Anticipatory learning (Sophistication):
  – Sophisticated players are aware that others are learning –BR to Cournot, etc. (level-k)
  – Soph. EWA: Camerer, Ho, Chong (JET 2002)
- Imitation: Imitate average or “best” player
- Learning direction theory: Move toward BR
- Rule learning: Learn which “rule” to use
  – Stahl (GEB 2000)

Further research

- Here is where we stand.
- Are there new direction for research in learning?
  - How does “information acquisition” help us study how people learn?
  - Learning direction theory and imitation are still “loose ends” to be filled

Holy Grail: How do people “actually” learn?

Further research

- How can we use these tools?
- Econometric Properties of learning rules:
  – Salmon (Econometrica 2001): Simulate data via certain learning rules and estimate them
  – Identification is bad for mixed strategy equilibrium and games with few strategies
  – EWA estimation does well on δ, others are okay if 1000 periods (not 30 periods)
- Can use this to “test designs”

Conclusion

- Learning: How people react to past history
- Reinforcement
- Belief Learning
  – Fictitious play, Cournot, etc.
- EWA: a Hybrid model
  – Performs better even “out-of-sample”
- Design tests: simulate and estimate