

學習理論：制約學習， 計牌學習與EWA學習

Joseph Tao-yi Wang
5/22/2009

大綱：學習理論

1. 學習：看到先前的結果之後，你會如何反應...
2. 目前已經有研究的學習規則：
 1. 制約學習(Reinforcement)
 2. 信念學習 (Belief learning)
 3. 混合式EWA：前面兩者的混合
 4. 其他：演化學習(Evolutionary)、預期學習(anticipatory learning)、模仿學習(imitation)、學習方向理論(learning direction theory)、規則學習(rule learning)等
3. 未來研究方向：
 1. 新的學習理論：你能開發新的學習理論嗎？
 2. 應用：我們如何應用這些工具來解釋現象？

看到上次結果之後，你會...

- 假設你正在做右邊這個獵鹿賽局
- 上次的結果是(B, L)
- 你這次會選擇什麼？

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

- 換另一個策略？
- 還是堅持原來策略？

看到上次結果之後，你會...

- 一個事先輸入程式的機器人不會改變策略
- 例如：演化學習

- 但是人類會考慮改變
- 你會如何改變策略？

- 制約學習：
 - 過去的報酬會「制約」你的決定
 - 「不太高明」的思考...

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

制約學習(reinforcement)

- 給定過去歷史，你會改變每個策略對你的「吸引力(attractions)」
- 制約：
 - 過去的報酬會「制約」你的決定
 - 也可能順便(spillovers)制約「附近」的策略 ϵ
- 例如：累積式制約
- $A^B(t) = \varphi A^B(t-1) + (1 - \epsilon) * 1$
- $A^T(t) = \varphi A^T(t-1) + \epsilon * 1$

制約學習

- (更一般化的)累積式制約：
 - $A^B(t) = \varphi A^B(t-1) + (1 - \epsilon) * 1 * [1 - \rho(t-1)]$
 - $A^T(t) = \varphi A^T(t-1) + \epsilon * 1 * [1 - \rho(t-1)]$
- 另外還有
- 加權平均式制約：
 - $A^B(t) = \varphi A^B(t-1) + \frac{(1 - \varphi)}{1 - \epsilon} (1 - \epsilon) * 1$
 - $A^T(t) = \varphi A^T(t-1) + \frac{(1 - \varphi)}{\epsilon} \epsilon * 1$

看到上次結果之後，你也可能會...

- 你會否改變你對「別人會做什麼」的預期？
 - 信念學習模型
- 計牌學習(Fictitious play)
 - 紀錄對手出手頻率
 - 例如：剪刀、石頭、布
- Cournot 最適反應
 - 假定對方這次的策略跟上次完全一樣

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

加權計牌學習

- 其他加權方式？一般的「加權計牌」學習
 - 計牌學習：所有歷史均有同樣加權 ($\rho=1$)
 - Cournot：只在乎最後一期 ($\rho=0$)
- 預測對手策略的事前機率(prior)：
 - $P_{t-1}(L) = 3/5, P_{t-1}(R) = 2/5$
- 事後機率(posterior)：
 - $P_{t-1}(L) = (3\rho + 1) / (5\rho + 1)$
 - $P_{t-1}(R) = (2\rho + 0) / (5\rho + 1), \rho = \text{遞減參數}$

加權計牌學習

- 事後機率(posterior)：
 - $P_{t-1}(L) = (3\rho + 1) / (5\rho + 1)$
 - $P_{t-1}(R) = (2\rho + 0) / (5\rho + 1)$
- 用此機率(信念)來計算期望報酬，即為該策略的「吸引力(attractions)」：
- $A^T(t) = [3(3\rho + 1) + 0(2\rho + 0)] / (5\rho + 1)$
- $A^B(t) = [1(3\rho + 1) + 1(2\rho + 0)] / (5\rho + 1)$
- 註：此計算和過去的實際報酬完全無關！

有沒有人兩種都做的？

- 制約學習模型下，人們不會改變信念
 - 但真人會改變信念！
- 計牌學習模型下，人們不會對實際報酬做出反應
 - 但真人會對誘因反應！
- EWA：上述兩者混合
 - Camerer and Ho (Econometrica, 1999)

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

「經驗加權吸引力」學習模型 Experience-Weighted Attraction

- 引入 δ ：人們心目中對於放棄之報酬(因為沒有選擇該策略)的加權
- Law of effect vs. Law of simulated effect
- $A^B(t) = [\varphi N(t-1) A^B(t-1) + 1] / N(t)$
- $A^T(t) = [\varphi N(t-1) A^T(t-1) + 3\delta] / N(t)$
而 $N(t) = \varphi(1 - \kappa) N(t-1) + 1$
- $N(t)$: 經驗加權(遞增的)

EWA的特例：制約學習

- $A^B(t) = [\varphi N(t-1) A^B(t-1) + \pi(B,L)] / N(t)$
- $A^T(t) = [\varphi N(t-1) A^T(t-1) + \pi(T,L) \delta] / N(t)$
而 $N(t) = \varphi(1 - \kappa) N(t-1) + 1$
- $\delta = 0, N(0) = 1$: 制約學習！
- $\kappa = 1$: (簡單的)累積式制約
- $\kappa = 0$: (加權)平均式制約
- 加權權重是 $\varphi / (\varphi + 1)$ 和 $1 / (\varphi + 1)$

EWA的特例：加權計牌學習

- $A^B(t) = [\varphi N(t-1) A^B(t-1) + \pi(B,L)] / N(t)$
- $A^T(t) = [\varphi N(t-1) A^T(t-1) + \pi(T,L) \delta] / N(t)$
而 $N(t) = \varphi (1 - \kappa) N(t-1) + 1$
- $\delta = 1, \kappa = 0$: 加權計牌學習!

- 作業...

- 提示: $N(t) = 1 + \varphi + \dots + \varphi^{t-1}$; 事後機率是

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

EWA的特例：加權計牌學習

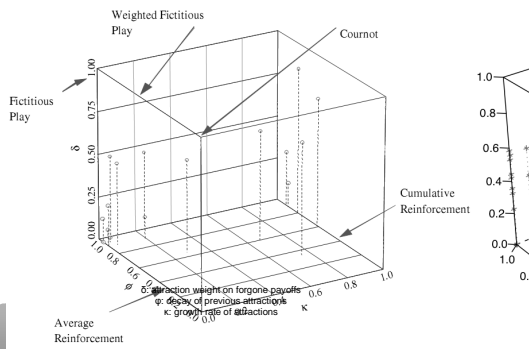
- $A^B(t) = [\varphi N(t-1) A^B(t-1) + \pi(B,L)] / N(t)$
- $A^T(t) = [\varphi N(t-1) A^T(t-1) + \pi(T,L) \delta] / N(t)$
而 $N(t) = \varphi (1 - \kappa) N(t-1) + 1$
- $\delta = 1, \kappa = 0$: 加權計牌學習!

- $\varphi = 1$: 計牌學習

- $\varphi = 0$: Cournot 最適反應

EWA參數方塊

Camerer, Wang, Ho (Economic Journal, 2008)

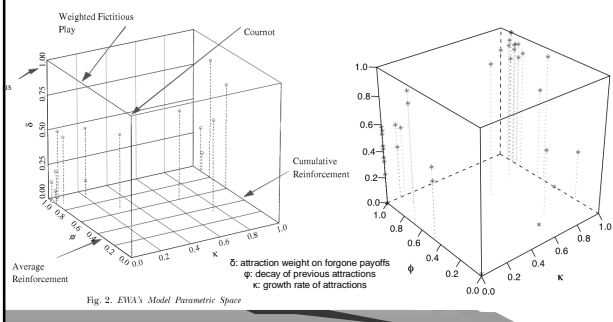


Interpretation of EWA Parameters

- δ : 過去吸引力的效果遞減
- κ : 吸引力成長率
- $N(t)$: 最初吸引力強度 (以「經驗」為單位)
- φ : $N(t)$ 的權重

EWA參數方塊

Camerer, Wang, Ho (EJ08') vs. Knoepfle, Wang, Camerer (JEEA09')



EWA模型的預測力

- 在35個賽局中，EWA模型大多都能更準確預測實驗結果，唯一的例外是有混合策略均衡的賽局
- See Camerer and Ho (book chapter, 1999)
- Econometrica 那篇的「加長版」
- BGT第六章有兩個例子：
- 產業發展分水嶺實驗
- 選美結果預測實驗

EWA模型的預測力

- 過適(Overfitting)：參數太多？
- 可以檢定：
 - 限制 vs. 不限制模型的預測力(fit)
- 預測新資料的能力(out-of-sample prediction)：
 - 用部分資料來估計參數，用以預測剩下的新資料
 - 不會有overfitting的問題(因為是預測新資料)
- 只有一個參數的「自我調整(self-tuned EWA)」
 - 此一簡化版的EWA預測能力和制約學習、計牌學習依樣好，預測新資料亦同

其他學習模型

- 其他的學習模型有：
- 預期學習(複雜學習)：
 - 複雜的人們會預期別人也會學習-對Cournot作出最適反應等(level-k)
 - 複雜EWA: Camerer, Ho, Chong (JET 2002)
- 模仿學習：模仿最厲害的人(或最一般的人)
- 學習方向理論：往最適反應的方向移動
- 規則學習：學習用哪一種規則來學習
 - Stahl (GEB 2000)

未來研究方向

- 我們目前會的就到這裡了
- 新的可能研究方向有：
 - 「資訊的吸收」如何幫助我們研究人們如何學習？
 - 學習方向理論和模仿學習都還有待發展
- 聖盃：人們到底是怎麼學習的？

未來研究方向

- 我們如何來應用這些工具？
- 學習理論的一些計量性質：
 - Salmon (Econometrica 2001): 用特定的學習規則來模擬資料，再用上述方法估計模擬資料
 - Identification (無法正確地估計出真實參數) 最糟糕是在有混合策略均衡或策略數字很少的實驗
 - EWA 對 δ 的估計尚可，其他參數則要1000回合才能準確估計(30回合不行)
- 此辦法可用來測試新的實驗設計

結論

- 學習：人們如何對歷史作出反應
- 制約學習
- 信念學習
 - 計牌學習(fictitious play)，Cournot最適反應等
- EWA：混合模型
 - 即使在預測新資料上(out-of-sample)也較佳
- 測試設計：模擬資料並估計之