

Cours Rationalité

" Usage de la simulation agent

essayer de reproduire le comportement humain (exemple : John Duffy et la spéculation + principaux algorithmes d'apprentissage)

Comprendre la sélection d'une rationalité dans le temps

Modéliser l'apprentissage qui reproduise les comportements réels

Apprentissages

- " Apprentissage individuel : chaque individu apprend à se comporter “mieux” pour atteindre son objectif
 - " dilemme exploration - exploitation
 - " inclusion de rationalité complexe (other-regarding preferences)
- " Apprentissage social : tout le monde partage le même objectif et la même notion de “mieux” - la sélection (positive) des plus performants se fait au niveau de la population
 - " influence du mode de sélection
 - " processus de diffusion - apprentissage ??

En ce qui concerne les apprentissages individuels, Bourguine [1993] distingue plusieurs niveaux de rationalité des agents selon leur relation à leur environnement et leur capacité à modéliser le réel.

- Les agents *réactifs* réagissent de manière *fixe* à l'information provenant de leur environnement, sur le mode stimulus-réponse (réponse sensori-motrice ou " pavlovienne " héritée génétiquement) : il y a absence d'apprentissage.
- Les agents *hédoniques* apprennent (par auto-renforcement) à modifier leur comportement afin d'augmenter leur " plaisir ". Ils sont capables d'anticipations " hédoniques " et *d'adaptation lente à partir de leur expérience historique*, ce qui suppose un niveau de conscience plus élevé que l'agent réactif (consciousness).
- Les *agents éductifs* sont dotés d'une capacité de modélisation de leur environnement, ce qui suppose la capacité de former des représentations symboliques, de simuler les conséquences d'une action sur leur environnement, et donc un niveau de conscience plus élevé (awareness).

Selon une perspective plus proche des catégories de l'économiste, Walliser [1997] propose une typologie des processus qui permettent de converger vers un équilibre en théorie des jeux. Il en distingue quatre, soit par ordre décroissant des capacités cognitives attribuées aux agents :

- Dans un apprentissage ÉVOLUTIONNAIRE, chaque joueur joue une stratégie fixe qui se reproduit proportionnellement au gain obtenu lors de confrontations aléatoires (agent réactif).
- Dans un apprentissage COMORTEMENTAL, chaque joueur modifie sa stratégie compte tenu des résultats observés de ses propres actions dans le passé (agent hédonique).
- Dans un apprentissage EPISTEMIQUE, chaque joueur révisé ses croyances relatives aux stratégies des autres adversaires à partir des informations qu'il a pu observer (Fudenberg, Levine [1998]).
- Dans un processus EDUCTIF, chaque joueur dispose d'assez d'information pour simuler parfaitement le comportement des autres joueurs, ce qui conduit immédiatement à l'équilibre : il n'y a pas d'apprentissage.

Méthodes d'évolution individuelle des agents

- " Apprentissage par renforcement
 - " Fonction d'utilité > évolue chaque résultat et classe les actions qui y ont mené.
 - " Il existe plusieurs formes d'apprentissage, de taille de mémoire, et de règles de changement
 - " Cas particulier : "satisficing"
- " "Belief learning"
 - " Classifier system, fictitious play
 - " Croyances sur les propriétés de l'environnement (contexte)
- " Comparaison et copie des méthodes des autres
 - " Diffusion d'information ou de comportements
 - " Choisir les bons agents à copier (confiance, réseau)
 - " Mécanismes d'adoption de comportements
- " Algorithmes génétique ("social learning" au niveau de la population)

Routine-based (Melioration)

- choice between 2 activities a and a'
- $p(a,t)$: probability of action a at time t
- $dp(a,t)/dt = v(u_{moyt}(a) - u_{moyt}(a'))$
- $U_{moy}(a)$ is the average gain until t when performing action a
- v is monotonous with $v(0) = 0$
 $dp(a,t)/dt = p(a,t) \cdot (1 - p(a,t)) \cdot v(u_{moy}(a) - u_{moy}(a'))$

Possible to extend to more choices.

Possible to put weight when calculating the average (higher weight for recent gains).

BUT: memory costly.

Erev-Roth

- N actions or strategies
- $q_{ij}(t)$ agent i has a tendency to play j at time t
- At start $q_{ij}(1) = q_{ik}(1) = S(1)$ for all j, k and i
 - $P_{ij}(t) = q_{ij}(t) / \sum q_{ik}(t)$
 - $P_{ij}(t) = \exp(q_{ij}(t)) / \sum \exp(q_{ik}(t))$
- If i plays k and gets Π , et $R(\Pi) = \Pi - \Pi_{\min}$
 - $q_{ij}(t+1) = (1 - \Phi)q_{ij}(t) + E_{k(j,R(\Pi))}$
 - with $E_{k(j,R(\Pi))} = (1-\varepsilon) R(\Pi)$ if $j=k$
 $E_{k(j,R(\Pi))} = (\varepsilon / (N-1)) \cdot R(\Pi)$ otherwise

Parameters : S (the higher S is, the slower learning gets), Φ is reduction of experience, ε tendency to experiment, size of the set of possible experiments

Bush-Mosteller

- N actions or strategies
- $p(a,t)$: probability of action a at time t (usually initialisation: $1/N$ for all).
- If $\Pi > 0$
 - $p(a,t+1) = p(a,t) - v(\Pi) \cdot p(a,t)$ if a not chosen
 - $p(a,t+1) = p(a,t) + v(\Pi) \cdot (1-p(a,t))$ if a was chosen
- Si $\Pi < 0$
 - $p(a,t+1) = p(a,t) + v(-\Pi) \cdot (p(a,t) \cdot p(a(t),t)) / (1 - p(a(t),t))$
 - $p(a,t+1) = p(a,t) - v(-\Pi) \cdot (p(a,t))$ if a was used

Parameters : v , size of the set of possible actions.

Satisficing (Simon, 57)

- Not looking for the best but for “good enough”, defined with aspiration levels

$$z(t+1) = \lambda.z(t) + (1-\lambda).\Pi(t)$$

(Evolves in time with gains)

- Can be associated to imitation

$$z_i(t+1) = \lambda_1.z_i(t) + \lambda_2.(\Pi_i(t)-z_i(t)) + \lambda_3.\Pi_{soc}(t)$$

- $\lambda_1 + \lambda_2 + \lambda_3 = 1$

- Then abandoned choice is replaced by another one, following a routine mechanism or random or whatever

Parameters: λ , λ_i , method to replace the abandoned action

Imitation

- Imitation :
 - exp average of gains for all actions of all agents (sum of weight is 1)
 - choice of action that brings higher gain among other agents

Parameters: perception network, initial values

BUT: the knowledge of others' gains is necessary (application?)

Fictitious play

- Memory of past actions of others - assumption that the frequency will be the same.
- Deduces an “action profile” for all others with expected probability
 - $E(p(a_{i-}, t)) = 1/t \sum \delta(a_{i-}(\tau) = a_{i-})$
 - with $\delta(a_{i-}(\tau) = a_{i-})$ is 1 si $a_{i-}(\tau) = a_{i-}$ and 0 otherwise
- Since all gains are known, it is possible to calculate expected gains for each action profile.
- Choice of the action that gives the best expected gain
 - $E(\Pi_i(a_i, t)) = \sum \Pi_i(a_i, a_{i-}) \cdot E(p(a_{i-}, t))$
- (possible to use exponential average)
BUT: memory costly

Classifier system

- Set of rules = [condition-action]
- Rule R: $(c_1, c_2, \dots, c_N) \rightarrow (a_1, a_2, \dots, a_q) / \text{strength}$
- Facing situation (c_1, c_N) , rules are selected considering
 - strength
 - specificity (more general \rightarrow less chosen).

$$\text{Bid}(R) = g_1 \cdot (g_2 + g_3 \cdot \text{specificity}(R)) \cdot \text{Strength}(R(t)) + \varepsilon$$

- And strength evolves

$$\text{Strength}(R(t+1)) = \text{Strength}(R(t)) + \Pi(t) - B(R)$$

Pb for economics: only works for a set of pre-established rules. In “real” classifier system, there is a genetic algorithm to diversify rules while selecting them.

Often: addition of noise to select bids.

Memetics

Inspiration of genetics in building of representation

the noosphere (Morin), « memes » (Dawkins), epidemiology of representations (Sperber).

Hales (following Bura): « Memes » are on animats using an environment and subject to selection.

Each meme : propensity to mutate and to reproduce; fights with neighbors and strengthening.

3 stages : calculation of satisfaction, mutation, replication.

If satisfied → increased aggressiveness and decreased mutation

If not satisfied → the reverse is true

Existence of a meta-meme: open-mindedness that suppresses that phenomena

Results

Scenario : Just enough food, too much food, predators.

Stabilisation of size of animat population able to occupy an area (carrying capacity).

killing memes population can grow but is a sign of instability in the system

open-mindedness meme help global survival of the population