Multi-agent Dynamical Systems with Reinforcement Learning

D. J. Albers
Joint work with James P. Crutchfield, Yuzuru Sato, and J. R. Albers
Dynamics Days
January 3, 2004
The Problem:

In the World: Ducks, Food, and 10,000 Lakes
- Robots and Blocks

In the Abstract: Dynamical Reconstruction = Learning
- Types of Information Captured in Particular Learning Constructions
- Long Term Versus Short Term Reconstruction
- Practical Training - Is the Most General Always the Most Representative or Easiest to Train
The World: Basic Components of a Reinforcement Learning Scheme

- **Agent**: the learner and action/decision maker (duck)

- **Environment**: everything that the agent interacts with, including all other agents in the population (other ducks + ponds)

- **Action**: an action taken by an agent, chosen from the set of all possible actions posed by the environment (selection of a pond)

- **Rewards**: a set of values, including positive, negative, and zero values, that an agent receives upon each action taken (a full or empty duck belly)

- **Population**: the set of all agents in the collective environment (all the ducks)
Multi-Agent Systems with Replicator Equations

“Memory versus Sensory Input”

\[
\frac{dx_i^n}{dx_i^n} = \beta_n [R_i^n - \sum_{j=1}^{M} x_j^n R_j^n] + \alpha_n I(x_i^n) \tag{1}
\]

Variables

- \( n = 1, \ldots, N \) indexes the agents in the population
- \( i = 1, \ldots, M \) indexes the choices (actions) possible by each agent
- \( R_i^n \) is the reward (punishment) agent \( n \) receives for choosing action \( i \)
- \( \alpha_n \) is the memory constant for agent \( n \) (controls memory decay)
- \( \beta_n \) is the learning constant for agent \( n \)
- \( x_i^n \) is the probability of agent \( n \) choosing action \( i \)
- \( I(x_i^n) = \sum_{j=1}^{M} x_j^n \log\left(\frac{x_j^n}{x_i^n}\right) \)
Multiple Agent Servicing Multiple Tasks

Three Schemes:

• Lone Ranger: \( R_i^n(t) = (x_i^n - \gamma) \) (\( \gamma = \frac{1}{M} \))
  
  – Ignore All Other Agents
  
  – Ignore Sites “Unknown” Sites

• Fashion Agent: \( R_i^n(t) = (x_i^n - \frac{1}{N-1} \sum_{k=1, k\neq i}^{M} x_k^n) \)
  
  – Follow (or Act Contrary) to the Crowd
  
  – Ignore Sites “Unknown” Sites

• Fashion Agent with a Conscience: \( R_i^n(t) = [\sigma_i - v_i^n(t)][x_i^n(t) - \sigma_i] \) (\( \sigma_i = \) desired service rate for site \( i \))
  
  – Avoid Overpopulated Locations
  
  – “All Seeing”
Two Preliminary Results

Basins of Attraction

Basin of Attraction Figures with 3 Sites Lone Ranger (left) and Fashion (right)

Memory versus Sensory Input

Bifurcation diagrams for $\beta$ ($\alpha = 0.4$, $a = 1$, $\gamma = 1/3$) and $\alpha$ ($\beta = 2$, $\gamma = 1/3$), 3 agents, 4 sites
Final Remarks

Why Do You Care?
Framework for Understanding Group Dynamics
Geometrical Understanding of Learning Schemes
A New Framework for Interpretation of Learning Dynamics

Future Directions
Time Dependent Service Rates ($\sigma$)
Introduction of Spatial Dependence
Other Reward Schemes
Bifurcation Theory (Center Manifold Analysis)

Hopes and Dreams
Introduction of More Complicated Agents
- Understand Differences and Similarities Between Representations
- Allow for a Better Understanding of Information Storage