Reinforcement Learning In Multiagent Systems

Anthony Andrade

Overview

- Multiagent Learning
- Reinforcement Learning
- Examples

Multiagent Learning

- Centralized Vs Decentralized
  - Definitive Characteristics

Centralized Vs Decentralized

- Centralized
  - executed in all parts by a single agent
  - requires no interaction with other agents
- Decentralized
  - several agents engaged in the same learning process

Characteristics Describing Strictly Decentralized learning

- Degree of decentralization
- Interaction-specific features
- Involvement-specific features
- Goal-specific features

Degree of decentralization

- Distributedness
- Parallelism

Interaction-specific features

- level of interaction
  - Observation to complex dialogues
- persistence of interaction
  - Short-term to long-term
- frequency of interaction
  - Low to high
- pattern of interaction
  - Unstructured to hierarchical
- variability of interaction
  - Fixed to changeable

Involvement-specific features

- relevance of involvement
- role played
  - generalist (centralized learning)
  - specialist (decentralized learning)

Goal-specific features

- type of improvement
  - individual improvement
  - group improvement
- compatibility of the learning goals
  - conflicting goals
  - complementary goals

Credit Assignment Problem

- The problem of assigning credit for an overall performance change
- A fundamental learning problem
  - Inter-agent CAP
  - Intra-agent CAP

Inter-agent CAP

- assignment for an overall performance change to the external actions of the agents
- the degree to which an agent's action changes overall performance
- particularly difficult in multiagent systems
- Who did it?

Intra-agent CAP

- assignment for a particular external action of agent to its underlying internal inferences and decision
- The knowledge, inferences, and decisions that led to an action
- How did the agent do it?

Reinforcement Learning

- An agent's goal is to maximize the utility of its actions
- An agent predicts the best action to execute in the current situation and executes it.
- The agent then adjusts its estimates of the executed action's utility based on environmental feedback
- The agent may also adjust the rates of the actions that led up to the current action

Reinforcement Learning (cont.)

- can include a model of the environment.
- Represented by a 4-tuple (S, A, P, r)
  - S set of states
  - A set of actions
  - P probability of moving from one state to another given a particular action
  - r reward function

Reinforcement Learning (cont.)

- policy maps current state to desirable action(s)
- π Policy that maps the current state to desirable actions

Q-Learning

- Essentially finds a policy for agent without the use of an explicit model
- Instead of a model, it stores an estimate for each state-pair

Learning Classifier Systems

- adjusts rule strengths from environmental feedback
- discovers new rules through a genetic algorithm

Bucket Brigade Algorithm

- rule strength for classifier firing is increased by environmental feedback
- rule strength is slightly decreased when fired, the amount is reassigned to the rule fired before that rule

Isolated, Concurrent Reinforcement Learners

- Agent seeks to maximize environmental feedback
- Other agents are not explicitly modeled
- RL is well suited to situations where information about the domain and the capabilities of other agents is limited.

Why not communicate

- Doesn’t guarantee coordination
- Can distract an agent
- Agents can become overly reliant on communication

Features that determine good CIRL domains

- Agent coupling
  - Tightly coupled
  - ***Loosely coupled***
- Agent Relationships
  - Cooperative
  - ***Indifferent***
  - ***Adversarial***

Features that determine good CIRL domains (cont.)

- Feedback Timing
  - ***Immediate***
  - Delayed
- Optimal behavior combinations
  - Single
  - ***Multiple***

CIRL Conclusions

- As long as favorable features exist, agents can acquire coordination knowledge for friends and foes
- Cooperative situations
  - Complimentary policies
    - Role specialization
- Coordination knowledge transfers
  - When used in a similar situation

Interactive Reinforcement Learning of Coordination

- Explicit Communication to decide on both group and individual actions
- Uses a modification of the Bucket Brigade Algorithm for learning and a contract net for coordination
  - Action Estimation Algorithm (ACE)
  - Action Group Estimation Algorithm (AGE)

Cellular Channel Allocation

- Cells
  - Particular geographical area over which communication will occur
- Channels
  - Different frequencies used to transfer calls
- Minimum Separation Distance
  - The minimum number of cells that must separate two cells using the same channel

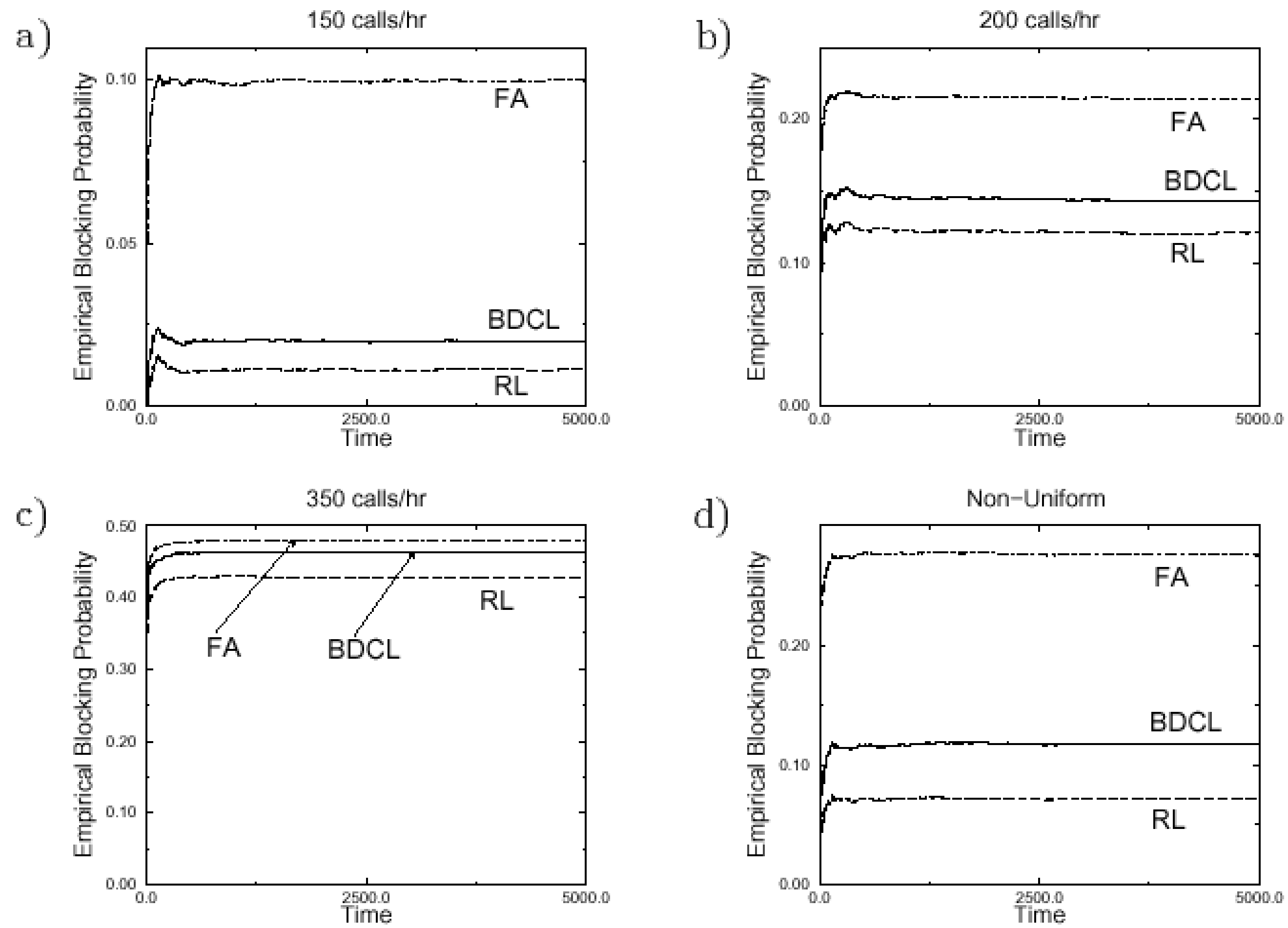
Cellular Channel Allocation

- The Problem
  - As new calls come in, keep the channel assignment optimal for that area, so as to drop as few calls as possible

Algorithms

- Fixed Assignment (FA)
  - In use in many cellular systems today
- Borrowing with Directional Channel Locking (BDCL)
  - Complicated and computationally expensive
  - Regarded as a powerful heuristic
- Reinforcement Learning
  - Based on Temporal Difference RL, TD(0)

Performance of FA, BDCL, & RL



Results

- RL out performed both Fixed Assignment and Borrowing with Directional Channel Locking

Demo

- Cellular Channel Allocation Java Demo