Reinforcement learning: a new approach for the Cultural Geography model

LtCol (HEA) Sotiris Papadopoulos
July 15, 2010
Outline

- Background
- Research Questions
- Methodology
- The utility-based agent
- Components
- The Experiment
- Remaining work
- Questions – Remarks – Suggestions
The Cultural Geography (CG) model is a government owned, open source prototype agent-based model of civilian populations currently implemented in Java and using Simkit as the simulation engine.

The model aims, through the implementation of social and behavioral science, to track individual, group-level and population-wide changes on positions related to various issues.

At its current stage the model examines the issues of security, elections and infrastructure.
Cultural Geography Model

Other Actors
- Theory of Planned Behavior
- Insurgents
- HNSF
- CF

Conflict Ecosystem
- Events
  - Events cause updates to issue stance

Infrastructure
- Structural
- Commodities
- Markets
- Services

Civilian Populace
- Social Network
  - Tribal/Political
  - Homophily
- Influence
- Education
- Trust
- Age

Influencing Groups

Influencing Groups

Entity Stereotype
- Age
- Tribe
- Politics
- Education
- Demography

Human Cognition
- Issue Stance
  - Beliefs
  - Interest
- Values
- Narrative Identity

1a Population Stereotypes

1b Human Cognition

2

3

4

1c Education

Events cause updates to issue stance

Courtesy of TRAC Monterey
The purpose of this study is to enhance the functionality of the Cultural Geography model by:

- Introducing the concept of reinforcement learning inside the model.
- Using utility theory for the decision making process.
- Constructing Decision Networks, as a result of the above mentioned process.
- Giving the user total control of the agents’ functionality.
Research Questions

- Are utility based agents appropriate for representing decision making in social networks as compared to the non utility based agents previously used within the CG model?
- How sensitive are the decisions chosen to differences in utility functions?
- What is the effect of exploration vs exploitation in the decision making process?
Methodology

- Literature review of reinforcement learning and utility theory
- Development of an Agent Template for implementation
- Improvement and finalization of the template
- Development of an experimental design
- Comparative analysis with different utility functions and roles (as time allows)
The Utility–based Agent

- An agent that decides about his actions based on a clearly defined utility value.
- All candidate actions are assigned a utility value that is discounted accordingly based on each action’s usage in the past.
- The mean utility for each action is the action’s activation level.
- The agent chooses the best course of action based on the action’s activation level and using the Boltzmann distribution for enhancing diversity in action choosing.
The Utility–based Agent

- New elements that were introduced:
  - Five new classes
  - A new Data Logger
  - An updated setup Excel spreadsheet that contains the new components
  - Changes were made in various existing components of the CG model to support the new agent

- The new code has already been incorporated in the core CG model and is currently in use.
  - Supporting PAKAF Multi–Level Assessment for ISAF J–2, MG Flynn.
Components

Simple Utility Agent Umpire

- Action Energy
- Agent Percept
- Agent Action
- Firing Time
- Point Utility
An experiment was designed by TRAC Monterey, to test the functionality of the new version of the CG model. Among the factors that were chosen for the design, the following two factors are of particular interest for this study:

- Discount rate (impacts utility calculation)
- Temperature (impacts balance between explore & exploit during agent decision making)
The Experiment

Preliminary results show a considerable decrease in the time it takes an agent to decide and act, as opposed to the previous incarnation of the CG model.

Also, there is strong evidence suggesting that the agents “learn” over time. This is supported by taking a simple look of their decisions as the simulation progresses.
As expected, the “learning curve” of the agents changes as we vary the values of temperature, meaning as we move from exploration to exploitation.

Analysis of the results is still ongoing.
Conclusions & Future work

- Use a method other than the Boltzmann distribution for the decision making process.

- Allow an agent to use multiple (weighted) percepts for each utility calculation. All percepts and their respective weights will be defined by the user.

- Create a function for the decrease of temperature over time (to move gradually from exploration to exploitation).
Questions – Suggestions


