Self-Interested Agents and Utility Theory

CPSC 532A Lecture 2

September 14, 2006

What is Multiagent Systems?

There are different levels of agency.

- A single-agent
  - Logic
  - Uncertainty
- A distributed single agent
- Multiple agents
  - What distinguishes these 'agents' from the setting above?
  - autonomy
  - asymmetric information
  - choose how to share it

Modeling Language

There are two fundamental approaches that are used in modeling multiagent systems (or AI systems generally):

- qualitative
  - usually uses some form of logic
- quantitative
  - usually Bayesian; probability theory and/or utility theory

Subject Matter

What subject matter does the theory describe?

- informational vs. motivational
  - knowledge and beliefs of agents
  - goals, preferences, utility functions
- individual-based or team-based
- strategic vs. non-strategic
  - non-strategic explicitly models agents' motivations; don't consider how or why they reach these motivations
  - strategic explicitly models agents' reasoning about their motivations

This course will focus on quantitative, motivational, individual-based, strategic theories. However, we'll briefly touch on some others.

Cooperative vs. Competitive MAS

Cooperative MAS:

- same desires: the strategic/non-strategic distinction is not very significant
- example: multirobot control, uncertain environment
- issues:
  - coordination
  - bandwidth, computational limits
- optimality well-defined

Competitive MAS:

- potentially different utility function (but may be the same)
- example: P2P file-sharing system on the internet

Resource Allocation in MAS

- easy in cooperative settings
  - optimality is well-defined
  - everyone wants the same thing
- difficult in competitive settings, because people can lie
  - mechanism design
  - maximizing payoff
  - design of agents
  - auctions: why important

Self-interested agents

- What does it mean to say that an agent is self-interested?
  - not that they want to harm other agents
  - not that they only care about things that benefit them
  - that the agent has its own description of states of the world that it likes, and that its actions are motivated by this description
- Utility theory:
  - quantifies degree of preference across alternatives
  - understand the impact of uncertainty on these preferences
  - utility function: a mapping from states of the world to real numbers, indicating the agent's level of happiness with that state of the world
  - Decision-theoretic rationality: take actions to maximize expected utility.

Example: friends and enemies

- Alice has three options: club (c), movie (m), watching a video at home (h)
- On her own, her utility for these three outcomes is 100 for c, 50 for m and 50 for h
- However, Alice also cares about Bob (who she hates) and Carol (who she likes)
  - Bob is at the club 60% of the time, and at the movies otherwise
  - Carol is at the movies 75% of the time, and at the club otherwise
- If Alice runs into Bob at the movies, she suffers disutility of 40; if she sees him at the club she suffers disutility of 90.
- If Alice sees Carol, she enjoys whatever activity she's doing 1.5 times as much as she would have enjoyed it otherwise (taking into account the possible disutility caused by Bob)
- What should Alice do (show of hands)?

What activity should Alice choose?

	$B = c$	$B = m$
$C = c$	15	150
$C = m$	10	100

$A = c$

	$B = c$	$B = m$
$C = c$	50	10
$C = m$	75	15

$A = m$

- Alice's expected utility for c:

$$0.25(0.6 \cdot 15 + 0.4 \cdot 150) + 0.75(0.6 \cdot 10 + 0.4 \cdot 100) = 51.75.$$

- Alice's expected utility for m:

$$0.25(0.6 \cdot 50 + 0.4 \cdot 10) + 0.75(0.6(75) + 0.4(15)) = 46.75.$$

- Alice's expected utility for h: 50.

Alice prefers to go to the club (though Bob is often there and Carol rarely is), and prefers staying home to going to the movies (though Bob is usually not at the movies and Carol almost always is).

Why utility?

- Why would anyone argue with the idea that an agent's preferences could be described using a utility function as we just did?
  - why should a single-dimensional function be enough to explain preferences over an arbitrarily complicated set of alternatives?
  - Why should an agent's response to uncertainty be captured purely by the *expected value* of his utility function?
- It turns out that the claim that an agent has a utility function is substantive.

Preferences and utility functions

Theorem (von Neumann and Morgenstern, 1944)

If an agent's preference relation satisfies the axioms Completeness, Transitivity, Decomposability, Monotonicity and Continuity then there exists a function $u : O \to [0, 1]$ with the properties that:

1. $u(o_1) \geq u(o_2)$ *iff the agent prefers* o_1 *to* o_2*; and*
2. *when faced about uncertainty about which outcomes he will receive, the agent prefers outcomes that maximize the expected value of* u*.*

- Is it possible to have utility functions on ranges other than $[0, 1]$?
  - Yes, and in fact any positive affine transformation of a utility function $au + b$, $a > 0$, yields another valid utility function.
- Why don't we use money instead of utility to measure happiness?