CS 798: Multiagent Systems

Preferences and Utility

Kate Larson

Cheriton School of Computer Science
University of Waterloo

January 6, 2010

Outline

Self-Interested Agents

We are interested in **self-interested** agents.

It does not mean that

- they want to harm other agents
- they only care about things that benefit them

It means that

- the agent has its *own* description of states of the world that it likes, and that its actions are motivated by this description

Self-Interested Agents

We are interested in **self-interested** agents.

It does not mean that

- they want to harm other agents
- they only care about things that benefit them

It means that

- the agent has its *own* description of states of the world that it likes, and that its actions are motivated by this description

Self-Interested Agents

We are interested in **self-interested** agents.

It does not mean that

- they want to harm other agents
- they only care about things that benefit them

It means that

- the agent has its *own* description of states of the world that it likes, and that its actions are motivated by this description

Example

Alice has three options

- Go to the club (C)
- Go to the movie (M)
- Stay at home (H)

Alice's decision depends on two other people, Bob and Chris.
Bob is Alice's enemy and Alice prefers it when he is not around.
Chris is Alice's friend, and makes events more fun.
Both Bob and Chris go to the club and the movies. In particular, Bob is at the club 60% of the time and at the movies 40% of the time. Chris is at the club 25% of the time and at the movies 75% of the time.
What should Alice do?

Preferences

Let o_1 and o_2 be two outcomes

- $o_1 \succeq o_2$ means o_1 is at least as desirable as o_2
  - the agent weakly prefers o_1 to o_2
- $o_1 \sim o_2$ means $o_1 \succeq o_2$ and $o_2 \succeq o_1$
  - the agent is indifferent between o_1 and o_2
- $o_1 \succ o_2$ means $o_1 \succeq o_2$ and $o_2 \not\succeq o_1$
  - the agent strictly prefers o_1 to o_2

Lotteries

Agents can also have preferences over *lotteries*

Definition

A *lottery* is a probability distribution over outcomes. It is written as

$$L = [p_1 : o_1, \ldots, p_n : o_n]$$

where o_i are outcomes and $p_i > 0$ such that

$$\sum_i p_i = 1.$$

Preference Axioms: Orderability

Definition (Orderability)

A preference relationship can be defined between every pair of outcomes:

$$\forall o_1, o_2 \; o_1 \succeq o_2 \text{ or } o_2 \succeq o_1.$$

Preference Axioms: Transitivity

Definition (Transitivity)

Preferences must be transitive:

$$\text{if } o_1 \succeq o_2 \text{ and } o_2 \succeq o_3 \text{ then } o_1 \succeq o_3.$$

Preference Axioms: Continuity

Definition (Continuity)

Assume that $o_1 \succ o_2$ and $o_2 \succ o_3$. Then there exists a $p \in [0, 1]$ such that

$$o_2 \sim [p : o_1, 1 - p : o_3].$$

Preference Axioms: Substitutability

Definition (Substitutability)

If $o_1 \sim o_2$ then for any $p \in [0, 1]$ and any outcome o_3

$$[p : o_1, 1 - p : o_3] \sim [p : o_2, 1 - p : o_3].$$

Preference Axioms: Monotonicity

Definition (Monotonicity)

If $o_1 \succ o_2$ and $p > q$ then

$$[p : o_1, 1 - p : o_2] \succ [q : o_1, 1 - q : o_2].$$

Preference Axioms: Decomposability

This is also known as the “no fun in gambling” rule since it says that two consecutive lotteries can be compressed into a single equivalent lottery.

Definition (Decomposability)

$$[p : o_1, 1-p[q : o_2, 1-q : o_3]]$$
$$\sim$$
$$[p : o_1, (1-p)q : o_2, (1-p)(1-q) : o_3].$$

Preferences and Utility Functions

Theorem (von Neumann and Morgenstern 1944)

If the axioms of preference are followed then there exists a function $u : O \mapsto \mathbb{R}$ such that $\forall o \in O$

- $o_1 \succeq o_2$ *if and only if* $u(o_1) \geq u(o_2)$
- $u([p_1 : o_1, \ldots, p_n : o_n]) = \sum_i p_i u(o_i)$

Proof idea:

- Define the utility of the best outcome to be $u(\overline{o}) = 1$ and the worst outcome to be $u(\underline{o}) = 0$
- Define the utility of each other outcome o as the p such that $o \sim [p : \overline{o}, 1 - p : \underline{o}]$.

Utility Theory

- The "units" do not matter
- Affine transformations do not really change anything;

$$U'(o) = aU(o) + b$$

will result in the same decision.

Note: Risk attitudes are important.

Utility Theory

- The “units” do not matter
- Affine transformations do not really change anything;

$$U'(o) = aU(o) + b$$

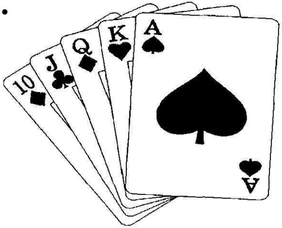
will result in the same decision.

Note: Risk attitudes are important.

What is game theory?

The study of games!

- Bluffing in poker
- What move to make in chess
- How to play Rock-Scissors-Paper

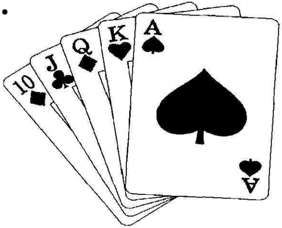


Also study of auction design, strategic deterrence, election laws, coaching decisions, routing protocols,...

What is game theory?

The study of games!

- Bluffing in poker
- What move to make in chess
- How to play Rock-Scissors-Paper

- 

Also study of auction design, strategic deterrence, election laws, coaching decisions, routing protocols,...

What is game theory?

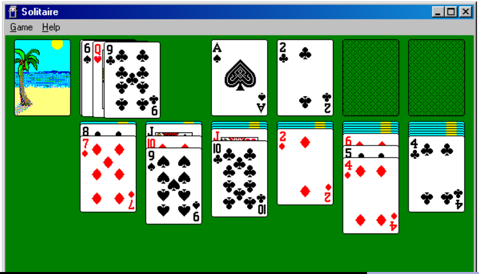
Game theory is a formal way to analyze **interactions** among a **group** of **rational** agents who behave **strategically**.

What is game theory?

Game theory is a formal way to analyze **interactions** among a **group** of **rational** agents who behave **strategically**.

Group: Must have more than one decision maker

- Otherwise you have a decision problem, not a game



Solitaire is not a game.

What is game theory?

Game theory is a formal way to analyze **interactions** among a **group** of **rational** agents who behave **strategically**.

Interaction: What one agent does directly affects at least one other agent

Strategic: Agents take into account that their actions influence the game

Rational: An agent chooses its best action (maximizes its expected utility)

Example

Pretend that the entire class is going to go for lunch:

1. Everyone pays their own bill
2. Before ordering, everyone agrees to split the bill equally

Which situation is a game?

Impact

Influential in a variety of fields, including

- economics
- political science
- linguistics
- psychology
- biology
- computer science
- ...

2 branches

- Non-cooperative: basic unit is the individual
- Cooperative: basic unit is the group

Impact

Influential in a variety of fields, including

- economics
- political science
- linguistics
- psychology
- biology
- computer science
- ...

2 branches

- Non-cooperative: basic unit is the individual
- Cooperative: basic unit is the group

Normal Form

A normal form game is defined by

- Finite set of agents (or players) N, $|N| = n$
- Each agent i has an action space A_i
  - A_i is non-empty and finite
- Outcomes are defined by action profiles, $a = (a_1, \ldots, a_n)$, where a_i is the action taken by agent i
- Each agent has a utility function $u_i : A_1 \times \ldots \times A_n \mapsto \mathbb{R}$

Normal Form

A normal form game is defined by

- Finite set of agents (or players) N, $|N| = n$
- Each agent i has an action space A_i
  - A_i is non-empty and finite
- Outcomes are defined by action profiles, $a = (a_1, \ldots, a_n)$, where a_i is the action taken by agent i
- Each agent has a utility function $u_i : A_1 \times \ldots \times A_n \mapsto \mathbb{R}$

Normal Form

A normal form game is defined by

- Finite set of agents (or players) N, $|N| = n$
- Each agent i has an action space A_i
  - A_i is non-empty and finite
- Outcomes are defined by action profiles, $a = (a_1, \ldots, a_n)$, where a_i is the action taken by agent i
- Each agent has a utility function $u_i : A_1 \times \ldots \times A_n \mapsto \mathbb{R}$

Normal Form

A normal form game is defined by

- Finite set of agents (or players) N, $|N| = n$
- Each agent i has an action space A_i
  - A_i is non-empty and finite
- Outcomes are defined by action profiles, $a = (a_1, \ldots, a_n)$, where a_i is the action taken by agent i
- Each agent has a utility function $u_i : A_1 \times \ldots \times A_n \mapsto \mathbb{R}$

Normal Form

A normal form game is defined by

- Finite set of agents (or players) N, $|N| = n$
- Each agent i has an action space A_i
  - A_i is non-empty and finite
- Outcomes are defined by action profiles, $a = (a_1, \ldots, a_n)$, where a_i is the action taken by agent i
- Each agent has a utility function $u_i : A_1 \times \ldots \times A_n \mapsto \mathbb{R}$

Normal Form

A normal form game is defined by

- Finite set of agents (or players) N, $|N| = n$
- Each agent i has an action space A_i
  - A_i is non-empty and finite
- Outcomes are defined by action profiles, $a = (a_1, \ldots, a_n)$, where a_i is the action taken by agent i
- Each agent has a utility function $u_i : A_1 \times \ldots \times A_n \mapsto \mathbb{R}$

Examples

Prisoners' Dilemma

	C	D
C	a,a	b,c
D	c,b	d,d

$c > a > d > b$

Pure coordination game

$\forall$ action profiles $a \in A_1 \times \ldots \times A_n$ and $\forall i, j$, $u_i(a) = u_j(a)$.

	L	R
L	1,1	0,0
R	0,0	1,1

Agents do not have conflicting interests. There sole challenge is to coordinate on an action which is good for all.

Zero-sum games

$\forall a \in A_1 \times A_2, u_1(a) + u_2(a) = 0$. That is, one player gains at the other player's expense.

Matching Pennies

	H	T
H	1,-1	-1, 1
T	-1,1	1,-1

	H	T
H	1	-1
T	-1	1

Given the utility of one agent, the other's utility is known.

More Examples

Most games have elements of both cooperation and competition.

BoS

	H	S
H	2,1	0,0
S	0,0	1,2

Hawk-Dove

	D	H
D	3,3	1,4
H	4,1	0,0