Learning objectives

At the end of the class you should be able to:

- characterize simplifying assumptions made in building AI systems
- determine what simplifying assumptions particular AI systems are making
- suggest what assumptions to lift to build a more intelligent system than an existing one

Dimensions

- Research proceeds by making simplifying assumptions, and gradually reducing them.
- Each simplifying assumption gives a dimension of complexity
  - multiple values in a dimension: from simple to complex
  - simplifying assumptions can be relaxed in various combinations

Dimensions of Complexity

Dimension	Values
Modularity	flat, modular, hierarchical

Dimensions of Complexity

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage

Dimensions of Complexity

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations

Dimensions of Complexity

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality

Dimensions of Complexity

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned

Dimensions of Complexity

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable

Dimensions of Complexity

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic

Dimensions of Complexity

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences

Dimensions of Complexity

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, adversary, multiple agents

Dimensions of Complexity

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, adversary, multiple agents
Interaction	offline, online

Modularity

- Model at one level of abstraction: flat
- Model with interacting modules that can be understood separately: modular

Modularity

- Model at one level of abstraction: flat
- Model with interacting modules that can be understood separately: modular
- Model with modules that are (recursively) decomposed into modules: hierarchical

Modularity

- Model at one level of abstraction: flat
- Model with interacting modules that can be understood separately: modular
- Model with modules that are (recursively) decomposed into modules: hierarchical
- Example: Planning a trip from here to a see the Mona Lisa in Paris.

Modularity

- Model at one level of abstraction: flat
- Model with interacting modules that can be understood separately: modular
- Model with modules that are (recursively) decomposed into modules: hierarchical
- Example: Planning a trip from here to a see the Mona Lisa in Paris.
- Flat representations are adequate for simple systems.
- Complex biological systems, computer systems, organizations are all hierarchical

Modularity

- Model at one level of abstraction: flat
- Model with interacting modules that can be understood separately: modular
- Model with modules that are (recursively) decomposed into modules: hierarchical
- Example: Planning a trip from here to a see the Mona Lisa in Paris.
- Flat representations are adequate for simple systems.
- Complex biological systems, computer systems, organizations are all hierarchical
- A flat description is either continuous or discrete. Hierarchical reasoning is often a hybrid of continuous and discrete.

By a hierarchic system, or hierarchy, I mean a system that is composed of interrelated subsystems, each of the latter being in turn hierarchic in structure until we reach some lowest level of elementary subsystem. In most systems of nature it is somewhat arbitrary as to where we leave off the partitioning and what subsystems we take as elementary. Physics makes much use of the concept of "elementary particle," although the particles have a disconcerting tendency not to remain elementary very long . . . Empirically a large proportion of the complex systems we observe in nature exhibit hierarchic structure. On theoretical grounds we would expect complex systems to be hierarchies in a world in which complexity had to evolve from simplicity.

– *Herbert A. Simon, The Sciences of the Artificial, 1996*

Planning horizon

...how far the agent looks into the future when deciding what to do.

- Static: world does not change

Planning horizon

...how far the agent looks into the future when deciding what to do.

- Static: world does not change
- Finite stage: agent reasons about a fixed finite number of time steps

Planning horizon

...how far the agent looks into the future when deciding what to do.

- Static: world does not change
- Finite stage: agent reasons about a fixed finite number of time steps
- Indefinite stage: agent reasons about a finite, but not predetermined, number of time steps

Planning horizon

...how far the agent looks into the future when deciding what to do.

- Static: world does not change
- Finite stage: agent reasons about a fixed finite number of time steps
- Indefinite stage: agent reasons about a finite, but not predetermined, number of time steps
- Infinite stage: the agent plans for going on forever (process oriented)

Clicker Question

The planning horizon dimension has values static, finite stage, indefinite stage, infinite stage.

The planning horizon of an agent is:

- A What functions the agent is able to carry out
- B The stage of life it is in
- C What it heads towards
- D How far into the future it considers the consequences of its actions
- E Where the agents's sun sets

Representation

Much of modern AI is about finding compact representations and exploiting the compactness for computational gains.
A agent can reason in terms of:

- Explicit states — a state is one way the world could be

Representation

Much of modern AI is about finding compact representations and exploiting the compactness for computational gains.
A agent can reason in terms of:

- Explicit states — a state is one way the world could be
- Features or propositions.
  - States can be described using features.
  - 30 binary features can represent $2^{30} = 1,073,741,824$ states.

Representation

Much of modern AI is about finding compact representations and exploiting the compactness for computational gains.
A agent can reason in terms of:

- Explicit states — a state is one way the world could be
- Features or propositions.
  - States can be described using features.
  - 30 binary features can represent $2^{30} = 1,073,741,824$ states.
- Individuals and relations
  - There is a feature for each relationship on each tuple of individuals.
  - Often an agent can reason without knowing the individuals or when there are infinitely many individuals.

Computational Limits

- Perfect rationality: the agent can determine the best course of action, without taking into account its limited computational resources.

Computational Limits

- Perfect rationality: the agent can determine the best course of action, without taking into account its limited computational resources.
- Bounded rationality: the agent must make good decisions based on its perceptual, computational and memory limitations.

Learning from experience

Whether the model is fully specified a priori:

- Knowledge is given.

Learning from experience

Whether the model is fully specified a priori:

- Knowledge is given.
- Knowledge is learned from data or past experience.

Whether the model is fully specified a priori:

- Knowledge is given.
- Knowledge is learned from data or past experience.

. . . always some mix of prior (innate, programmed) knowledge and learning (nature vs nurture).

Learning from experience

Whether the model is fully specified a priori:

- Knowledge is given.
- Knowledge is learned from data or past experience.

. . . always some mix of prior (innate, programmed) knowledge and learning (nature vs nurture).

- Learning is impossible without prior knowledge (bias).

Uncertainty

There are two dimensions for uncertainty. In each dimension an agent can have

- No uncertainty: the agent knows what is true
- Disjunctive uncertainty: there is a set of states that are possible
- Probabilistic uncertainty: a probability distribution over the worlds.

Why probability?

- Agents need to act even if they are uncertain.
- Predictions are needed to decide what to do:
  - definitive predictions: you will be run over tomorrow
  - disjunctions: be careful or you will be run over
  - point probabilities: probability you will be run over tomorrow is 0.002 if you are careful and 0.05 if you are not careful

Why probability?

- Agents need to act even if they are uncertain.
- Predictions are needed to decide what to do:
  - definitive predictions: you will be run over tomorrow
  - disjunctions: be careful or you will be run over
  - point probabilities: probability you will be run over tomorrow is 0.002 if you are careful and 0.05 if you are not careful
- Acting is gambling: agents who don't use probabilities will lose to those who do.
- Probabilities can be learned from data and prior knowledge.

Sensing Uncertainty

Whether an agent can determine the state from its stimuli:

- Fully-observable: the agent can observe the state of the world.

Sensing Uncertainty

Whether an agent can determine the state from its stimuli:

- Fully-observable: the agent can observe the state of the world.
- Partially-observable: there can be a number states that are possible given the agent's stimuli.

Effect Uncertainty

If an agent knew the initial state and its action, could it predict the resulting state?

Effect Uncertainty

If an agent knew the initial state and its action, could it predict the resulting state?

The dynamics can be:

- Deterministic: the resulting state is determined from the action and the state

Effect Uncertainty

If an agent knew the initial state and its action, could it predict the resulting state?

The dynamics can be:

- Deterministic: the resulting state is determined from the action and the state
- Stochastic: there is uncertainty about the resulting state.

Clicker Question

A domain for transporting parcels betwen cities where the location of each truck and each parcel is known, but trucks can get into accidents is:

- A Stochastic and Partially Observable
- B Stochastic and Fully Observable
- C Deterministic and Fully Observable
- D Deterministic and Partially Observable
- E None of the above or more than one of the above

Clicker Question

Teaching students concepts to get them to understand is:

A Stochastic and Partially Observable

B Stochastic and Fully Observable

C Deterministic and Fully Observable

D Deterministic and Partially Observable

E None of the above or more than one of the above

Clicker Question

Poker (from each player's point of view) is:

- A Stochastic and Partially Observable
- B Stochastic and Fully Observable
- C Deterministic and Fully Observable
- D Deterministic and Partially Observable
- E None of the above or more than one of the above

Clicker Question

A deterministic agent is:

- A is determined to get to its goal
- B does not know exactly which state it is in
- C could predict the next state if it knew its current state and the action to be taken
- D no longer has any termites
- E knows what state it is in

Need for preferences

Alice . . . went on "Would you please tell me, please, which way I ought to go from here?"
"That depends a good deal on where you want to get to," said the Cat.
"I don't much care where —" said Alice.
"Then it doesn't matter which way you go," said the Cat.

Lewis Carroll, 1832–1898
Alice's Adventures in Wonderland, 1865
Chapter 6

Preference

What does the agent try to achieve?

- achievement goal is a goal to achieve. This can be a complex logical formula.

Preference

What does the agent try to achieve?

- achievement goal is a goal to achieve. This can be a complex logical formula.
- complex preferences may involve tradeoffs between various desiderata, perhaps at different times.

Preference

What does the agent try to achieve?

- achievement goal is a goal to achieve. This can be a complex logical formula.
- complex preferences may involve tradeoffs between various desiderata, perhaps at different times.
  - ordinal only the order matters

Preference

What does the agent try to achieve?

- achievement goal is a goal to achieve. This can be a complex logical formula.
- complex preferences may involve tradeoffs between various desiderata, perhaps at different times.
  - ordinal only the order matters
  - cardinal absolute values also matter

Preference

What does the agent try to achieve?

- achievement goal is a goal to achieve. This can be a complex logical formula.
- complex preferences may involve tradeoffs between various desiderata, perhaps at different times.
  - ordinal only the order matters
  - cardinal absolute values also matter

Examples: coffee delivery robot, medical doctor

Clicker Question

Sam prefers coffee to tea is:

A achievement goal

B ordinal preference

C cardinal preference

Clicker Question

Sam wants coffee is:

A achievement goal

B ordinal preference

C cardinal preference

Clicker Question

An agent that assigns numerical values to a set of features and acts to maximize the sum of the values has:

- A cardinal preferences
- B goals
- C ordinal preferences
- D a full-observable environment
- E a partially-observable environment

Number of agents

Are there multiple reasoning agents that need to be taken into account?

- Single agent reasoning: any other agents are part of the environment.

Number of agents

Are there multiple reasoning agents that need to be taken into account?

- Single agent reasoning: any other agents are part of the environment.
- Adversarial reasoning considers another agent, where when one agent wins, the other loses. two-player zero-sum game

Number of agents

Are there multiple reasoning agents that need to be taken into account?

- Single agent reasoning: any other agents are part of the environment.
- Adversarial reasoning considers another agent, where when one agent wins, the other loses. two-player zero-sum game
- Multiple agent reasoning: an agent reasons strategically about the reasoning of other agents, perhaps needing to coordinate or cooperate.

Agents can have their own goals: cooperative, competitive, or goals can be independent of each other

Interaction

When does the agent reason to determine what to do?

- reason offline: before acting
- reason online: while interacting with environment

Dimensions of Complexity

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

State-space Search

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

Deterministic Planning

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

Decision Networks

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

Markov Decision Processes (MDPs)

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

Decision-theoretic Planning

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

Reinforcement Learning

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

Classical Game Theory

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

Humans

Dimension	Values
Modularity	flat, modular, hierarchical
Planning horizon	non-planning, finite stage, indefinite stage, infinite stage
Representation	states, features, relations
Computational limits	perfect rationality, bounded rationality
Learning	knowledge is given, knowledge is learned
Sensing uncertainty	fully observable, partially observable
Effect uncertainty	deterministic, stochastic
Preference	goals, complex preferences
Number of agents	single agent, multiple agents
Interaction	offline, online

The dimensions interact in complex ways

- Partial observability makes multi-agent and indefinite horizon reasoning more complex

The dimensions interact in complex ways

- Partial observability makes multi-agent and indefinite horizon reasoning more complex
- Modularity interacts with uncertainty and succinctness: some levels may be fully observable, some may be partially observable

The dimensions interact in complex ways

- Partial observability makes multi-agent and indefinite horizon reasoning more complex
- Modularity interacts with uncertainty and succinctness: some levels may be fully observable, some may be partially observable
- Three values of dimensions promise to make reasoning simpler for the agent:

The dimensions interact in complex ways

- Partial observability makes multi-agent and indefinite horizon reasoning more complex
- Modularity interacts with uncertainty and succinctness: some levels may be fully observable, some may be partially observable
- Three values of dimensions promise to make reasoning simpler for the agent:
  - Hierarchical reasoning
  - Individuals and relations
  - Bounded rationality