CS255 Giga Cheat Sheet

12 October 2020 11:17

Welcome,

This is the Giga Cheat Sheet: CS255 Artificial Intelligence Edition (v1)

Made with love by 丐山 ㄖ GG毛 丅

Textbook 1 - Foundations of Computation Agents, Poole and Mackworth
Textbook 2 - Artificial Intelligence: A Modern Approach

Revision Index

Other Reading

The following books give a good discussion of Bayesian AI and of knowledge representation:

- D. Barber, Bayesian Reasoning and Machine Learning, Cambridge University Press, 2012
- R. Brachman and H. Levesque, Knowledge Representation and reasoning, Morgan Kaufmann, 2004
- K. Korb, A. Nicholson, Bayesian Artificial Intelligence, Chapman and Hall, 2004
- Jackson P, Introduction to Expert Systems, Addison Wesley, 1999
- Giarratano J, Riley G, Expert Systems; Principles and Programming (4th ed), 2005.

Other Reading

These books are good background for CS255:

- Callan, Artificial Intelligence, Palgrave
- Ginsberg, Essentials of Artificial Intelligence, Morgan Kaufman
- Nilsson, Artificial Intelligence: A New Synthesis, Morgan Kaufman
- The library has many other AI books containing useful background material (especially for coursework).

Learning Outcomes for CS255

At the end of the module you will:

- develop an appreciation for knowledge based systems, intelligent agents and their architectures,
- understand a wide variety of knowledge representation and artificial intelligence approaches to planning,
- understand various methods for search (uninformed and informed), planning and reinforcement learning, and
- understand various methods for representing and reasoning under uncertainty.

Revision Index

30 April 2021 12:24

Artificial Intelligence Exam Content

Please note that although the notes themselves are exhaustive (in terms of the module content), this list is only a subset of the content that we could be examined on. However, I believe that understanding of the following list of topics should suffice for the exam, based on the revision lectures and past papers.

If you have any suggestions regarding adding or removing from this list, message me on Discord.

Introduction to AI ↗

Intro to Rationality,

AIFCA Sections 2.1–2.2
AIMA Sections 2.1–2.2

Architecture, Hierarchy

AIFCA Sections 2.3–2.5
AIMA Sections 2.3–2.5

Dimensions of Complexity
System architecture - body, controllers
Hierarchical Control - layers forming a hierarchy of controllers
Agent Types - simple reflex, reflex with state, goal based, utility based agents, learning agents

Uninformed Search ↗

AIFCA Sections 3.1-3.5
AIMA Sections 3.1–3.4

Lowest-Cost-First Search
Greedy best first graph search

Informed Search ↗

AIFCA Sections 3.6-3.8
AIMA Sections 3.5–3.6

A* Search
Depth-First Brand-And-Bound search
Direction of Search

Constraint Satisfaction ↗

AIFCA Sections 4.1-4.6
AIMA Chapter 6

Types of Constraints - MRV, degree, least constraining value
Backtracking Algorithms to solve constraint satisfaction problems
Conditioning - Cutset conditioning
Improving Search Efficiency

Arc Consistency
Variable Elimination

Local Search ↗

AIFCA Section 4.7
AIMA Sections 4.1-4.2

Hill-Climbing
Greedy Descent
Random Walk
Simulated Annealing
Genetic Algorithms

Adversarial Search ↗

AIFCA Sections 11.1–11.4
AIMA Sections 5.1–5.5

Minimax
Alpha-Beta Pruning

Planning ↗

AIFCA Sections 5.1-5.6, 6.1–6.3, 6.5
AIMA Sections 7.1–7.5, 7.7, 8.1–8.2, 10.1, 10.2, 10.4, 11.1–11.3

Search Vs. Planning
Situation Calculus
What actually is a plan?
Partial-Order Planning
Problems with POP
Conditional Planning
Clobbering - What is it and how do we avoid it?
Action Monitoring / Plan Monitoring

Knowledge Representation ↗

AIFCA Section 5
AIMA Chapters 7 and 12

Defining Knowledge and Reasoning
Expert Systems
Rules as Knowledge
Production Rules
Rule Based Systems
Forward & Backward Chaining
Conflict Resolution - recency, refractoriness, specificity

Bayesian AI ↗

AIFCA Section 8.1-8.4 & 9.1-9.3
AIMA Chapters 13 & 14

Bayesian Probability
Inference - Joint Probability Distribution
Bayes' Rule

Normalisation
Introduction to Bayesian Belief Networks
Probabilistic Inference in Belief Networks
Decision Making
Inference by Enumeration
Influence Diagrams
Variable Elimination

Reinforcement Learning ↗

AIFCA Sections 12.1-12.7, 9.5
AIMA Chapter 21

Utility, rewards, and values
Policies
Markov Decision Processes
State-Based RL :Q-Learning and SARSA
The Explore-Exploit dilemma and solutions
On-Policy & Off-Policy Learning

What is AI?

13 October 2020 13:02

1. "The automation of activities that we associate with human thinking, activities such as decision-making problem solving, learning.."
2. "The study of mental faculties through the use of computational models"
3. "The study of how to make computers do things which at the moment people are better"
4. "The branch of computer science concerned with the automation of intelligent behaviour"

- Thinking humanly
- Acting humanly
- Thinking rationally
- Acting rationally

Success measured in terms of human behaviour or against some ideal of rationality

If you can think intelligently, then you will act intelligently

The Turing Test

Acting humanly suggests the main components of AI
Knowledge, reasoning, language, learning

So, is this a good test of intelligence?

Not really – not reproducible and can't be mathematically analysed. Not an intelligence test.
Things that are considered intelligent like object recognition are not tested

Searle's Chinese Room

Thinking Rationally

- Derivation of conclusion from particular premises
- Various forms of logic, notation, and rules of inference
- Task or problem expressed in login -> AI program deduces solution

Problem: Difficult to express tasks in logic
Problem: How to cope with uncertainty
Problem computational expense

Acting Rationally

Acting so as to maximise goal achievement, given the available information and resources
Thought or inference not necessarily involved, e.g. reflex actions
The right thing: acting so as to maximise goal achievement, given the information and resources available

Building Rational Agents

An agent is an entity that perceives and acts
An agent can be viewed as a function from percept histories to actions
F: P -> A
Agents typically required to exhibit autonomy. For any given class of environments and tasks we seek the agents with the best performance
We want to design the best agent we can with our limited resources

Artificial Intelligence

Is the synthesis and analysis of computational agents that act intelligently

An agent acts intelligently if:

- Its actions are appropriate for its goals and circumstances
- It is flexible to changing environments and goals
- It learns from experience
- It makes appropriate choices given perceptual and computational limitations

Rational Agents

16 October 2020 13:46

Inputs

in the context of a self-driving car

- Abilities – e.g. steering, braking
- Goals – safety, timeliness
- Prior knowledge – what signs mean
- Stimuli – vision, lasers
- Past experience – effects of steering, friction of surfaces, how people move

Rational Action

- Goals can be defined as a performance measure, defining a numerical value for a given environment history
- Rational action is whichever action maximises the expected value of the performance measure given the percept sequence to date - I.e. doing the right thing
- Previous perceptions are typically important

Rational Agents are not omniscient – they don't know the actual outcome of their actions. They just do the best they can, given the current percepts

Actions that were expected to give a good return but failed to, can still be considered rational

Dimensions of Complexity

We can view the design space for AI as being defined by a set of dimensions of complexity. These dimensions define a **design space** for AI; different points in this space are obtained by varying the values on each dimension.

1. Modularity

Flat – one level of abstraction – adequate for simple systems. Continuous or discrete.
Modular – interacting modules that can be understood separately
Hierarchical – agent has modules that are recursively decomposed into modules – complex computers, biological systems etc

2. Planning horizon

Statis – world does not change
Finite – agent reasons about a fixed number of steps into the future
Indefinite – the agent thinks about a finite number of steps but we do not predetermine the number of steps
Infinite – the agent has to keep planning forever – process oriented

3. Representation

- Modern AI is about finding compact representations and exploiting the compactness for computational gain
- Explicitly – e.g. a chess board, one way the world could be
- Features of propositions – states can be described using features 30 binary features can represent 2^30 possible 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 man individuals
- When describing a complex world, the features can depend on **relations** and **individuals**. What

we call an *individual* could also be called a **thing**, an **object** or an **entity**. A relation on a single individual is a **property**. There is a feature for each possible relationship among the individuals.

E.g. With a light switch s_2
Instead of the feature, it could use the relation $position(s2, up)$. This relation enables the agent to reason about all switches or for an agent to have general knowledge about switches that can be used when the agent encounters a switch.

By reasoning in terms of relations and individuals, an agent can reason about whole classes of individuals without ever enumerating the features or propositions, let alone the states. An agent may have to reason about infinite sets of individuals, such as the set of all numbers or the set of all sentences. To reason about an unbounded or infinite number of individuals, an agent cannot reason in terms of states or features; it must reason at the relational level.

4. Computational Limits

Perfect rationality: the agent can determine the best course of action, without taking into account its limited computational resources
Bounded rationality – we have to make good decisions based on limited resources e.g. memory

To take into account bounded rationality, an agent must decide whether it should act or reason for longer. This is challenging because an agent typically does not know how much better off it would be if it only spent a little bit more time reasoning. Moreover, the time spent thinking about whether it should reason may detract from actually reasoning about the domain.

5. Learning from experience

The model is specified a priori

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

Usually some mix of prior knowledge is used – nature vs nurture

6. Uncertainty

Sensing and effect

In some cases, an agent can observe the state of the world directly. For example, in some board games or on a factory floor, an agent may know exactly the state of the world. In many other cases, it may only have some noisy perception of the state and the best it can do is to have a probability distribution over the set of possible states based on what it perceives. For example, given a patient's symptoms, a medical doctor may not actually know which disease a patient has and may have only a probability distribution over the diseases the patient may have.

The **sensing uncertainty dimension** concerns whether the agent can determine the state from the stimuli

In each dimension an agent can have

- No uncertainty – e.g. you will run out of power
- Disjunctive uncertainty – there is a set of states that are possible e.g. charge for 30 minutes, or you will run out of power
- Probabilistic uncertainty - Agents need to act even if they are uncertain. We need to predict what might happen in order to decide what to do - e.g. probability you will run out of power is 0.01 if you charge for 30 minutes and 0.8 otherwise

Probabilities can be learned from data and prior knowledge
Acting is gambling – if you don't use probabilities you will lose to agents that do

Sensing Uncertainty

- Fully observable – the agent can observe the state of the world
- Partially-observable – there can be a number of states possible given what the agent can perceive

Effect uncertainty

If an agent knew the initial state and its action, could it predict the resulting state?
Deterministic: the resulting state is determined from the action and the state
Stochastic – there is only a probability distribution over the resulting states.

7. Preference

What is the agent trying to achieve?

- Achieve goal is a goal – this can be a complex logical formula
- Complex preference – maybe involve trade-offs between desiderate, perhaps at different times
- Ordinal – the order is the only thing that matters
- Cardinal – counts matter e.g. we want exactly 0 crashes

8. Number of agents

Are there multiple agents?
Single agent – any other agents are part of the environment
Multiple agent reasoning – an agent reasons strategically about the reasoning of other agents

9. Interaction

When does the agent reason to determine what to do?

- Online – while interacting
- Offline – before acting

Example: State-space Search

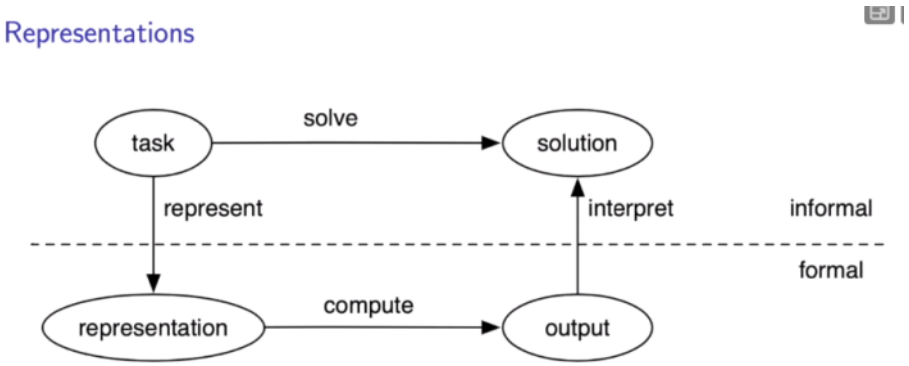
Dimension	Values
Modularity	*flat*, modular, hierarchical
Planning horizon	non-planning, finite stage, *indefinite stage*, infinite stage
Representation	*states*, features, relations
Computational limits	*perfect rationality*, bounded rationality
Learning	*knowledge is given*, knowledge is learned
Sensing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

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
Three values of dimensions promise to make reasoning simpler for the agent

- Hierarchical reasoning
- Individuals and relations
- Bounded rationality

Desirable Properties for a Representation



- determine what constitutes a solution
- represent the task in a way a computer can reason about
- use the computer to compute an output, which is answers presented to a user or actions to be carried out in the environment, and
- interpret the output as a solution to the task.

We have a task, that is represented in the AI system in some way.
The computation is done on this representation of the task, and then the output is relayed back out into the world.

A **representation** of some piece of knowledge is the particular data structures used to encode the knowledge so it can be reasoned with. A **knowledge base** is the representation of all of the knowledge that is stored by an agent.

We want our representation to have a few characteristics:

- Rich in data to express the knowledge needed to solve the problem
- As close to the problem as possible: compact, natural and maintainable
- Amenable to efficient computation – expresses features of the problem that can be exploited for computational gain - Able to trade off accuracy and computation time/space
- Able to be acquired from people, data, and past experiences.

Defining a Solution

- Given an informal description of a solution, what is a solution?
- Typically, much is left unspecified, but the unspecified parts can't be filled in arbitrarily
- Much work in AI is motivated by common-sense reasoning – the computer needs to make common-sense conclusions about unstated assumptions

Quality of Solutions

Does it matter if the answer is wrong or answers are missing?

- An optimal solution is a best solution according to some measure of solution quality
- A satisficing solution is one that is good enough according to some description of solutions that are adequate
- An approximately optimal solution is one whose measure of quality is close to the best theoretically possible. E.g. get within 10% of the optimal solution. This is sometimes still just as hard as getting the optimal solution, though.
- A probably solution is one that is likely to be a solution

Decision and Outcomes

- Good and bad decisions can have good and bad outcomes
- Information can be valuable because it leads to better decisions: value of information

- We can often trade off computation time and solution quality – An anytime algorithm can provide a solution in any time, but makes better decisions with more time

Agents are concerned not just about finding the right answer, but finding the information that will allow them to find the right answer

Choosing a representation
We need to represent a problem to solve it on a computer

Physical symbol system hypothesis
A symbol is a physical pattern that can be manipulated
A symbol system allows you to create/modify/delete symbols

The hypothesis states that a physical symbol system has the necessary and sufficient means for general intelligent action

Knowledge & Symbol Levels

Two levels of abstraction seem to be common among entities – biological and computational

- Knowledge level is about the external world – what the agent knows and what its goals are
- Symbol level is about what symbols the agent uses to implement the knowledge level – it is a level of description of an agent in terms of what reasoning it is doing

Mapping from Problem to Representation

- What level of abstraction of a problem to represent?
- What individuals and relations in the world to represent?
- How can an agent represent knowledge to ensure that the representation is natural, modular and maintainable?
- How can an agent acquire the information from data, sensing, experience, or other agents?

Choosing a Level of Abstraction

- High-level is easier to understand for a human
- a low-level description can be more accurate and more predictive, we lose details when we abstract away details in high level abstractions
- You may not know the information needed for a low-level description

A delivery robot can model the environment at a high level of abstraction in terms of rooms, corridors, doors, and obstacles, ignoring distances, its size, the steering angles needed, the slippage of the wheels, the weight of parcels, the details of obstacles, the political situation in Canada, and virtually everything else. The robot could model the environment at lower levels of abstraction by taking some of these details into account. Some of these details may be irrelevant for the successful implementation of the robot, but some may be crucial for the robot to succeed. For example, in some situations the size of the robot and the steering angles may be crucial for not getting stuck around a particular corner. In other situations, if the robot stays close to the centre of the corridor, it may not need to model its width or the steering angles.

Although no level of description is more important than any other, we conjecture that you do not have to emulate every level of a human to build an AI agent but rather you can emulate the higher levels and build them on the foundation of modern computers. This conjecture is part of what AI studies.

It is sometimes possible to use multiple levels of abstraction

Reasoning and Acting

Reasoning determines what action an agent should do

1. Design time reasoning – done by the designer of the agent
2. Offline computation – done by the agent before it has to act

3. Online computation – computation done by an agent receiving information and acting

Agent Architecture & Hierarchy

17 October 2020 16:05

Objectives:

- Learn what an agent is and agent functions
- Understand the different types of agent

Agents

By a hierarchic system, or hierarchy, we 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.

An **agent** is something that acts in an environment.

Agents interact with the environment with a body. An embodied agent has a physical body. A robot is an artificial purposive embodies agent.

Agents act in the world through their actuators, also called effectors.

Agent Systems

An agent system is made up of an agent and the environment in which it acts.
Agents in an agent system receive stimuli from their environment and carries out actions on the environment.

An agent is made up of a body and a controller. The controller receives percepts from the body and sends commands to the body.

The Agent Function

Agents are situated in time T that can be thought of as discrete, broken into sub-sections. T+1 is one of these intervals after T.

If time is dense, there is always another moment in time between any two other given moments – e.g. time is continuous.

Assume that T has a starting point, which we arbitrarily call 0.

Suppose P is the set of all possible percepts. A **percept trace**, or **percept stream**, is a function from T into P. It specifies what is observed at each time.

Suppose C is the set of all commands. A **command trace** is a function from T into C. It specifies the command for each time point.

A percept trace for an agent is thus the sequence of all past, present, and future precepts received by the controller. A command trace is the sequence of all past, present, and future commands issued by the controller. The commands can be a function of the history of percepts. This gives rise to the concept of a **transduction**, a function from percept traces into command traces.

A transduction is **causal** if, for all times t, the command at time t depends only on percepts up to and including time t. The causality restriction is needed because agents are situated in time; their command at any time cannot depend on future percepts.

A **controller** is an implementation of a causal transduction.

The **history** of an agent at time t is the percept trace of the agent for all times before or at time t and the command trace of the agent before time t.

Thus, a **causal transduction** maps the agent's history at time t into the command at time t. It can be seen as the most general specification of a controller.

Belief State

Although a causal transduction is a function of an agent's history, it cannot be directly implemented due to the fact that an agent does not have access to its entire history. It only has access to its current percepts and those that it has remembered.

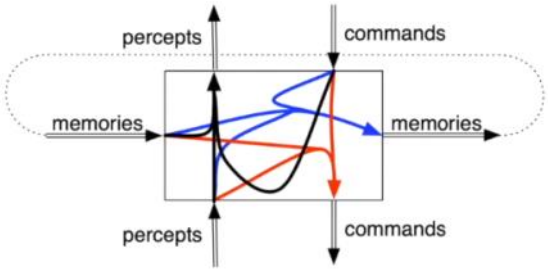
The memory or belief state of an agent at time t is all the information the agent has remembered from the previous times. An agent has access only to the history it has encoded in its belief state. Thus, the belief state encapsulates all the information about an agents history, that the agent can use for current and future commands. At any time, ag agent has access to its belief state and its current percepts.

The belief state can contain any information, subject to the agent's memory and processing limitations. This is a very general notion of belief.

The belief state function determines what we're going to remember - I.e. what will the next belief state look like
The command function uses the current belief state and the current percepts to decide on some action

If there are a finite number of possible belief states, the controller is called a **finite state controller** or a **finite state machine**. A **factored representation** is one in which the belief states, percepts, or commands are defined by features. If there are a finite number of features, and each feature can only have a finite number of possible values, the controller is a **factored finite state machine**. Richer controllers can be built using an unbounded number of values or an unbounded number of features. A controller that has an unbounded but countable number of states can compute anything that is computable by a Turing machine.

Functions Implemented in a Layer



- *memory function:* $remember(memory, percept, command)$
- *command function:* $do(memory, percept, command)$
- *percept function:* $higher\_percept(memory, percept, command)$

Hierarchical Control

There is much evidence that people have multiple qualitatively different levels. Kahneman [2011] presents evidence for two distinct levels: **System 1**, the lower level, is fast, automatic, parallel, intuitive, instinctive, emotional, and not open to introspection, and **System 2**, the higher level, is

slow, deliberate, serial, open to introspection, and based on reasoning.

In a hierarchical controller there can be multiple channels – each representing a feature – between layers and between layers at different times.

There are three types of inputs to each layer at each time:

1. the features that come from the belief state, which are referred to as the remembered or previous values of these features
2. the features representing the precepts from the layer below in the hierarchy
3. the features representing the commands from the layer above in the hierarchy.

There are three types of outputs from each layer at each time:

- the higher-level percepts for the layer above
- the lower-level commands for the layer below
- the next values for the belief-state features.

The low-level controllers can run much faster, and react quickly
They also deliver a simple view of the world to the high level controllers.

Types of Agent

We can view agents as being specific by the agent function that maps a percept sequence to a sequence of actions

Ideal rational agents do whatever action is expected to maximise performance measure on basis of percept sequence and built-in knowledge.
So in principle there is an ideal mapping of percept sequences to actions
The simple approach to this is a lookup table, however this is doomed to failure be

The lookup table suggests a notion of an ideal mapping
This would be the rational agent function
We want to implement the rational agent function by implementing a function that approximates the ideal mapping as closely as possible

1. Simple reflex agent – condition action rules
2. Reflex agents with state – retains knowledge about the world
3. Goal-based agents – have a representation of desirable states
4. Utility based agents – ability to discern some useful measure between possible means of achieving state
5. Learning agents – able to modify their behaviour based on their performance, or in light of new information

Goal Based Agents

If we want to solve more sophisticated problems, we might need to look at a GBA
It maintains some representation of the state, which gets updated. It has some knowledge about how the world works, and a set of goals

To achieve goals, you need a sequence of actions. We need to keep track of how close we are to achieving our goal. GBA's are much more flexible.

Utility Based Agents

Similar to GBA, but rather than purely thinking about what action we should do in terms of the goal, we think about the benefit a particular action has.

Sometimes there are many ways of achieving a goal, and many actions you can take to get there.

Utility based agents have a utility function that allows us between which goals we should achieve, and alternative ways of achieving a goal

Learning Agents

Can be built upon any of the previous types – all 4 can fit into the performance element

The performance element takes some input, and chooses some action on the environment.
The learning agent has a critic, a problem generator and a learning element

The learning element uses information about how the agent operates and the particular algorithms, and changes how the performance element operates. It does this by getting information on how it's doing from the critic, and new things to try out from the problem generator.

- Useful when we don't know much about the environment and the agent has to be able to learn.
- Learning provides an agent with a degree of autonomy
- Learning results from interaction between the agent and the world.
- The critic uses a performance standard to tell the agent how it is doing. Performance standard should be a fixed measure, external to the agent.
- Problem generator – responsible for suggesting actions in pursuit of new and informative experiences

Architecture 2

17 October 2020 18:05

- Agent Functions
- Types of Agent

We can view agents as being specific by the agent function that maps a percept sequence to a sequence of actions

Ideal rational agents do whatever action is expected to maximise performance measure on basis of percept sequence and built-in knowledge.
So in principle there is an ideal mapping of percept sequences to actions
The simple approach to this is a lookup table, however this is doomed to failure be

The lookup table suggests a notion of an ideal mapping
This would be the rational agent function
We want to implement the rational agent function by implementing a function that approximates the ideal mapping as closely as possible

Agent Types

1. Simple reflex agent – condition action rules
2. Reflex agents with state – retains knowledge about the world
3. Goal-based agents – have a representation of desirable states
4. Utility based agents – ability to discern some useful measure between possible means of achieving state
5. Learning agents – able to modify their behaviour based on their performance, or in light of new information

Goal Based Agents

If we want to solve more sophisticated problems, we might need to look at a GBA
It maintains some representation of the state, which gets updated. It has some knowledge about how the world works, and a set of goals

To achieve goals, you need a sequence of actions. We need to keep track of how close we are to achieving our goal. GBA's are much more flexible.

Utility Based Agents

Similar to GBA, but rather than purely thinking about what action we should do in terms of the goal, we think about the benefit a particular action has.

Sometimes there are many ways of achieving a goal, and many actions you can take to get there.

Utility based agents have a utility function that allows us between which goals we should achieve, and alternative ways of achieving a goal

Learning Agents

Can be built upon any of the previous types – all 4 can fit into the performance element

The performance element takes some input, and chooses some action on the environment.
The learning agent has a critic, a problem generator and a learning element

The learning element uses information about how the agent operates and the particular algorithms, and changes how the performance element operates. It does this by getting information on how it's

doing from the critic, and new things to try out from the problem generator.

Useful when we don't know much about the environment and the agent has to be able to learn.

Learning provides an agent with a degree of autonomy

Learning results from interaction between the agent and the world.

The critic uses a performance standard to tell the agent how it is doing. Performance standard should be a fixed measure, external to the agent.

Problem generator – responsible for suggesting actions in pursuit of new and informative experiences

Uninformed Search

18 October 2020 21:18

Objectives:

- Problem solving agents
- Problem types
- Problem formulation
- State space graphs
- Basic tree search and graph search algorithms

An agent could be programmed to act in the world to achieve a fixed goal or set of goals, but then it would not adapt to changing goals, and so would not be intelligent. *An intelligent agent needs to reason about its abilities and its goals* to determine what to do.

A problem-solving agent is a goal-based agent that will determine a sequence of actions that will achieve some goal state

There are 4 steps a problem solving agent must take

- Goal formulation
- Problem formulation
- Search – find the sequence of actions
- Execution

In the simplest scenario, an agent has no uncertainty, and has a state based model of the world, with a goal to achieve. This is either a flat representation or a single level of a hierarchy. The agent is able to determine how to achieve this goal by searching in its representation of the world state space for a way to get from its current state to a state that satisfies its goal.

Given a complete model, it tries to find a sequence of actions that will achieve its goal before it has to act in the world.

This problem is like finding a path from the start node to an end node in a directed graph.

Problem Types

- Deterministic, fully observable – single state problem
- Deterministic, partially observable – multi-state problem
- Stochastic, partially observable – contingency probably
- Unknown state space

Searching

"searching" in this chapter means searching in an internal representation for a path to a goal – from start node to end node.

Search underlies much of artificial intelligence. When an agent is given a problem, it is usually given a description that allows it to recognise a solution, but not an actual algorithm for a solution. The agent has to search for a solution itself.

The difficulty of search and the fact that humans are able to solve some search problems efficiently suggests that computer agents should exploit knowledge about special cases to guide them to a solution. This extra knowledge beyond the search space is called **heuristic knowledge**. This chapter considers one kind of heuristic knowledge in the form of an estimate of the cost from a node to a goal.

State Space

One general formulation of intelligent action is in terms of a state space. A state contains all of the information necessary to compute the effects of an action and to determine whether a states satisfies a goal.

When we do a state space search, we assume:

1. The agent has perfect knowledge of the state space and is planning for the case where it observes what state it is in: there is full observability
2. The agent has a set of actions that have known deterministic effects
3. The agent can determine whether a state satisfies a goal

A solution is a sequence of actions that will get the agent from the current state to the state that satisfies the goal.

A state space problem consists of:

- A set of states
- A distinguished state called the start state
- For each state, a set of actions available to the agent in that state
- An agent function that, given a state and an action, returns a new state
- A goal specified as a Boolean function that is true when state satisfies the goal, in which case we can say that s is a goal state
- A criterion that specifies the quality of an acceptable solution. For example, any sequence of actions that gets the agent to the goal state may be acceptable, or there may be costs associated with actions and the agent may be required to find a sequence that has minimal total cost. A solution that is best according to some criterion is called an optimal solution. We do not always need an optimal solution, for example, we may be satisfied with any solution that is within 10% of optimal.

Graph Searching

In this chapter, the problem of finding a sequence of actions to achieve a goal is abstracted as searching for paths in directed graphs.

To solve a problem, we first define the underlying search space, then apply a search algorithm to that search space. Many problem solving tasks are transformable into the problem of finding a path in a graph.

A directed graph consists of a set of nodes and a set of directed arcs between nodes. The idea is to find a path along these arcs from the start node to a goal node.

In representing a state-space problem, the states are represented as nodes, and the actions as arcs.

The abstraction is necessary because there may be more than one way to represent a problem as a graph. The examples in this chapter are in terms of state-space searching, where nodes represent states and arcs represent actions.

Formal Graph Searching

A directed graph has a set N of nodes and a set A of arcs, where an arc is an ordered pair of nodes.

Nodes are neighbours if there is an arc between them. Note that the relationship is not symmetrical – it doesn't necessarily go both ways.

A path from node s to node g is a sequence of nodes (n0, n1, … nk) such that s = n0 and g = nk, and there is an arc between each ni and ni+1.

A goal is a Boolean function on nodes. If goal(n) is true, we say that node n satisfies the goal, and n is a goal node.

To encode problems as graphs, one node is identifies as a start node. A solution is a path from the start node to a node that satisfies the goal.

Sometimes there is a cost, a non-negative number associated with the arcs. We write the cost of an arc as cost(ni, nj). The costs of arcs induce a cost of paths. Given a path p, cost of path p is the sum of the costs of all the arcs in the path.

An optimal solution is the solution that has the lowest cost.

That is, an optimal solution is a path p from the start node to a goal node such that there is no path p' from the start node to a goal node where the cost is lesser.

A cycle is a non-empty path where the end node is the same as the start node. A directed graph without any cycles is called a directed acyclic graph.

A tree is a DAG where there is one node with no incoming arcs and every other node has exactly one incoming arc. The node with no incoming arcs is the root of the tree. A node with no outgoing arcs is a leaf. In a tree, neighbours are called children.

In many problems, the search graph is not given explicitly, but is constructed as needed. For the search algorithms, all that is required is a way to generate the neighbours of a node, and to determine if the node is a goal node.

The forward branching factor of a node is the number of outgoing arcs. The backward branching factor is the number of incoming arcs to the node. These factors provide measures for the complexity of graph algorithms. We assume the branching factors are bounded, when we discuss the space and time complexity.

Tree Search & Graph Search

18 October 2020 23:28

Objectives:

- Graph search algorithms
- Basic tree search

Tree Search

In the vacuum example, there were 2 squares, and a binary dirt option – dirt or no dirt

It was simple to compute the state space diagram for this scenario.

In real life, e.g. a self-driving car on a street, there are going to be thousands of variables and thus an intractable number of potential states.

How do we deal with this?

Algorithm

- Offline, simulated exploration of the state space
- Starting with a start state, expand one of the explored states by generating its successors, e.g. find neighbours by considering possible actions to build a search tree.

```
function TREE-SEARCH(problem,strategy)
    returns solution, or failure
  initialise search tree using initial state of problem
  loop do
    if no candidates for expansion then return failure
    choose leaf node for expansion according to strategy
    if node contains a goal state
        then return corresponding solution
    else expand node and add resulting nodes to
        search tree
  end
```

Evaluating Search Strategies

- Completeness - does the algorithm always find a solution?
- Optimality – does the algorithm always find a least-cost solution?
- Time complexity – what is the maximum branching factor, depth of least cost solution, maximum depth of state space?
- Space complexity – what is the maximum number of nodes in memory?

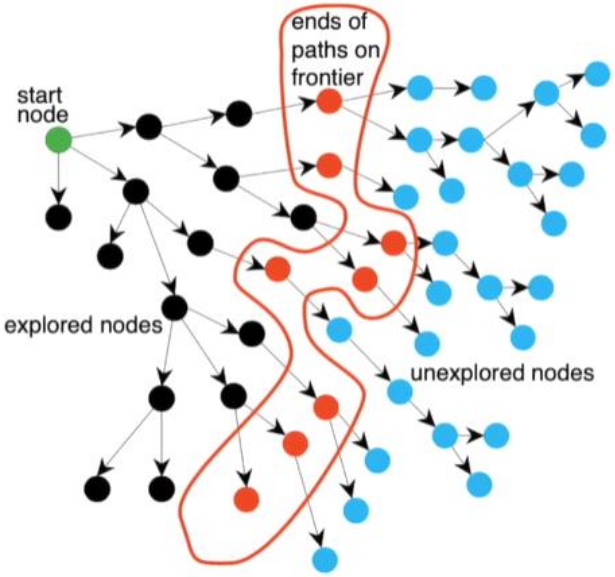
Graph Search

With a tree search, state spaces with loops give rise to repeated states that cause inefficiencies and infinite loops

Graph search is a practical way of exploring the state space that can account for such repetitions

1. Given a graph, incrementally explore paths from the start nodes

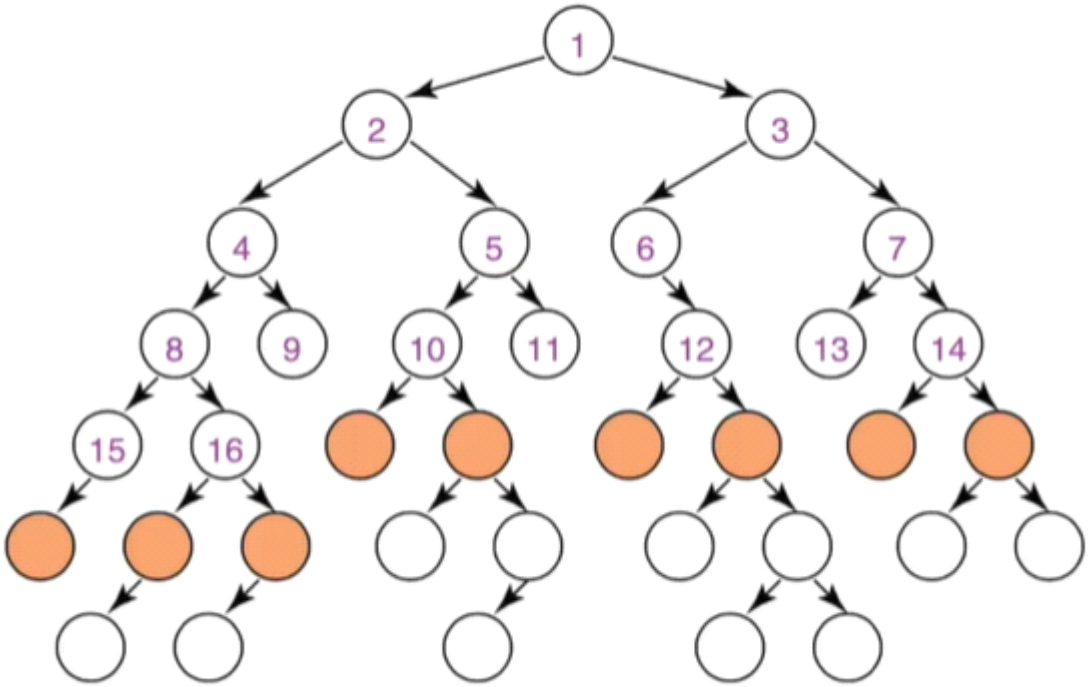
2. Maintain a frontier of paths
3. As search proceeds, the frontier expands into the unexplored nodes until a goal node is encountered
4. The way in which the frontier is expanded is defined by the search strategy



Starts with a set of nodes
We check if we're at the goal
We start with a frontier – at the start, we just have the start node
We check the frontier, and look for a goal. If none found, add next nodes to frontier and repeat.

Breadth-First Graph Search

Treat frontier as a queue
Select earliest elements added to frontier



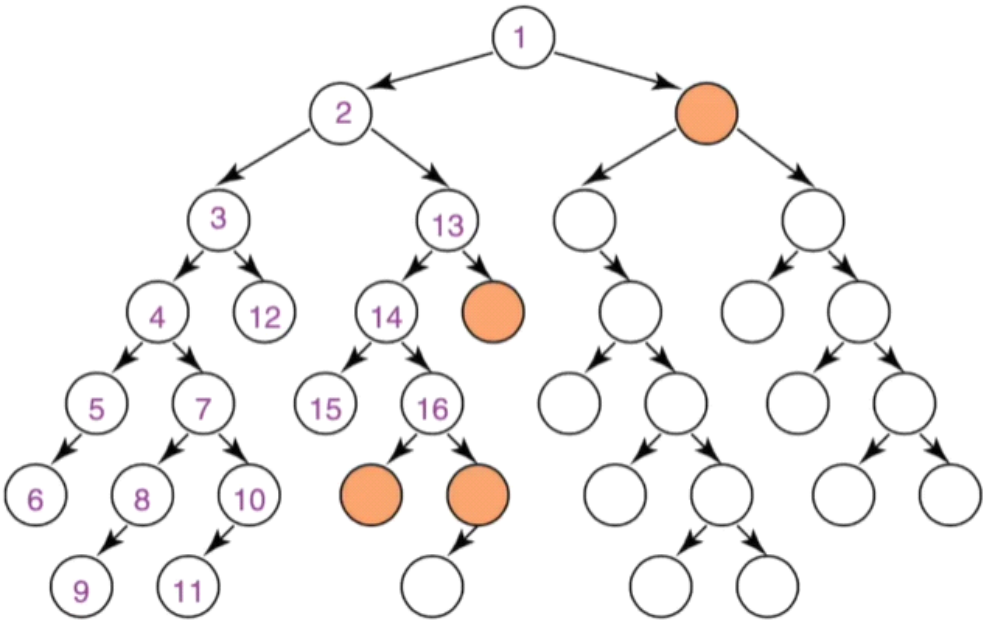
- If branching factor for all nodes is finite, bf graph search will find a solution
- Time complexity is exponential in path length - branching factor the power of n where n is the path length
- The space complexity is exponential in path length b^n
- The search is unconstrained by the goal

Depth-First Graph Search

Treat the frontier as a stack
Select the last element added to the frontier

If the list of paths on the frontier is p1, p2, p3...

- P1 is selected and the paths that extend p1 are added to the front of the stack in front of p2
- P2 is only selected when all of the paths from p1 have been explored



Complexity of a Depth-First Graph Search

- Not guaranteed to halt if we have cycles of infinite graphs
- The space complexity is linear in the number of arcs from the start of the current node
- If the graph is a finite tree, with a forward branching factor <= b and all paths from the start having at most k arcs, worst case time complexity is O(b^k)
- The search is unconstrained by the goal

Lowest-Cost-First Search

- Sometimes there are costs associated with arcs
- The cost of a path is the sum of the costs of its arcs
- An optimal solution is one with the minimum cost
- At each stage, the lowest-cost-first search selects a path on the frontier with the lowest cost
- The frontier is a priority queue ordered by path cost
- The first path to a goal is a least-cost path to a goal node

- When arcs costs are equal then breadth-first search

Summary of Uninformed Search Strategies

Strategy	Frontier Selection	Complete	Halts	Space
Breadth-first	First node added	Yes	No	Exp
Depth-first	Last node added	No	No	Linear
Lowest-cost-first	Minimal $cost(p)$	Yes	No	Exp

Complete — guaranteed to find a solution if there is one (for graphs with finite number of neighbours, even on infinite graphs)
Halts — on finite graph (perhaps with cycles)
Space — as a function of the length of current path.

Informed Search

23 October 2020 21:44

Objectives:

- Best first search
- A* search
- Pruning search
- Depth bounded search
- Brand and bound search
- Heuristics

AIFCA 3.6-3.8

Uninformed search is generally very inefficient; if we have extra information about the problem we should use it
Improve search using problem specific knowledge
This is still offline problem solving since we have complete knowledge of the problem and solution

Heuristic

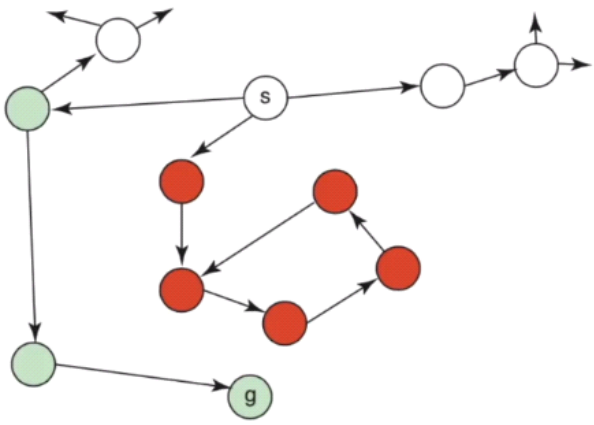
- idea: don't ignore the goal when selecting paths
- Often there is extra knowledge that can be used to guide the search: heuristics
- H(n) is an estimate of the cost of the shortest path from node n to a goal not
- H can be extended to paths
- H(n) is an underestimate if there is no path from n to a goal with cost strictly less than h(n)
- An admissible heuristic is a nonnegative heuristic function that is an underestimate of the actual cost of the path to the goal

Example Heuristic Function

- If the nodes are points on a Euclidean plane and the cost is the distance, h(n) can be the straight-line distance from n to the closest goal
- If the nodes are locations and cost is time, we can use the distance to a goal divided by max speed

Best-First Search Informed Search

- We can use the heuristic function to determine the order of the stack representing the frontier
- Idea: Select the path or node that is closet to a goal according to the heuristic function
- Heuristic depth-first search selects a neighbour so that the neighbour is selected first
- Greedy best fs selects a path on the frontier with the lowest heuristic value
- Best first search treats the frontier as priority
- queue ordered by h



The above is bad for a simple best-first search: heuristic depth first search will select the node below s and never terminate. Greedy best-first search will cycle between the nodes below s never finding an alternative route.

Complexity of Greedy best-first search

- The space complexity is exponential in path length b^n
- Time complexity is exponential in the path length
- Not guaranteed to find a solution, even if one exists
- It does not always find the shortest path

A* Search

It can be seen as an extension of Edsger Dijkstra's 1959 algorithm. A* achieves better performance by using heuristics to guide its search.

- Uses both path cost like in lowest-cost-first, and heuristic values, like in greedy best-first search
- Cost(p) is the cost of path p
- H(p) estimates the cost from the end of p to a goal
- Let(p) = cost(p) + h(p)
- F(p) estimates the total path cost of going from a start node to a goal via p

- A* is a mix of lowest-cost-first and best-first-search
- It treats the frontier as a priority queue ordered by f(p)
- It always selects the node on the frontier with the lowest estimated distance from the start to a goal node constrained to go via that node.

A search algorithm is *admissible* if, whenever a solution exists, it returns an optimal solution

A* is admissible if

1. The arc costs are greater than 0
2. The branching factor is finite
3. The heuristic h is a non-negative and an underestimate

Why is it Admissible?

- If a path p to a goal is selected from a frontier, can there be a shorter path to a goal?
- Suppose path p' is on the frontier. Because p was chosen before p', and $h(p) = 0$:

$$cost(p) \leq cost(p') + h(p')$$

- Because h is an underestimate:

$$cost(p') + h(p') \leq cost(p'')$$

  for any path p'' to a goal that extends p'
- So $cost(p) \leq cost(p'')$ for any other path p'' to a goal.

- The frontier always contains the initial part of a path to a goal, before that goal is selected
- A* halts, as the costs of the paths on the frontier keeps increasing, and will eventually exceed any finite number
- Note admissibility does not guarantee that every node selected from the frontier is on an

optimal path
- Although it does ensure that the first solution found will be optimal, even in graphs with cycles

Iterative deepening A*

IDA* performs repeated depth-bounded depth-first searches. Instead of the bound being on the number of arcs in the path, it is a bound on the value of f(n). The threshold is initially the vlaue of f(s), where s is the start node. IDA* then carries out a depth-first depth-bounded search but never expands a path with a high f-value than the current bound.

How can a better heuristic function help?

- A* expands all paths from the start in the set
- A* also expands some paths from the set
- Increasing h while keeping it admissible reduces the size of the first of these sets
- If the second set is large there can be significant variability in the space and time of A*

Complexity of A*

Exponential time complexity
Exponential space complexity – you keep all nodes in memory

Strategy	Selection from Frontier	Path found	Space
Breadth-first	First node added	Fewest arcs	Exponential
Depth-first	Last node added	No	Linear
Iterative deepening	—	Fewest arcs	Linear
Greedy best-first	Minimal $h(p)$	No	Exponential
Lowest-cost-first	Minimal cost (p)	Least cost	Exponential
A^*	Minimal cost $(p) + h(p)$	Least cost	Exponential
IDA*	—	Least cost	Linear

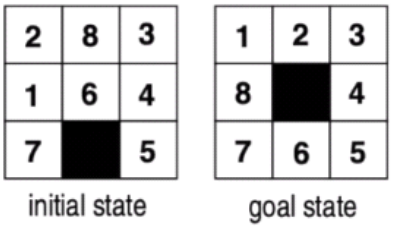
Heuristics

04 November 2020 20:32

A **heuristic function** $h(n)$, takes a node n and returns a non-negative real number that is an estimate of the cost of the least-cost path from node nn to a goal node. The function $h(n)$ is an **admissible heuristic** if $h(n)$ is always less than or equal to the actual cost of a lowest-cost path from node n to a goal.

How do we design our heuristics to make the best search algorithm possible?

Admissible heuristics: Example



- 8-puzzle is just hard enough to be interesting
- Branching factor 3(ish), typical solution around 20 steps
- Exhaustive search: 3^{20} states $(= 3.5 \times 10^7)$
- Eliminating repeating states: $9! = 362880$ states
- Need a decent heuristic.

Possible heuristics

H(n) = number of misplaced tiles
H(n) = Manhattan distance

Characterising Heuristics

- If A* tree-search expands N nodes and solution is depth d, then the effective branching factor * is the branching factor a uniform tree of depth d would have to contain N nodes.
- The closer the effective branching factor is to 1, then the larger the problem that can be solved.
- Can estimate b* experimentally (usually fairly consistent over problem instances)
- Q) is h2 always better than h1? If h(n) is bigger than another h(n) for all n then that heuristic is said to dominate the other heuristic

Deriving Heuristics

Can derive admissible heuristics of the exact solution cost of a relaxed version of the problem
A problem with less restrictions on operators is a relaxed problem
E.g. 8 puzzle, we can relax it and say that a tile can move from A to B if B is blank

If one dominates, use that. Else, calculate the h value for each state and use maximum value. If all options are admissible, then the chosen one will be admissible

Subproblems

Derive admissible heuristics from solution cost of a subproblem of given problem
Cost of subproblem = lower bound on cost of complete problem

We can store exact solution costs in a database which we can use to lookup values

We can combine pattern databases

Disjoint pattern databases

If we can divide up the problem so moves only affect a single subproblem

Statistical approach

Run search over training problems and gather statistics

Pruning

04 November 2020 20:40

The preceding algorithms can be improved by taking into account multiple paths to a node. We consider two pruning strategies. The simplest strategy is to prune cycles; if the goal is to find a least cost path, there is no use considering paths with cycles. The other strategy is only ever to consider one path to a node, and to prune other paths to that node.

Cycle Pruning

- The search can prune a path that ends in a node already on the path, without removing an optimal solution
- In depth-first methods, checking for cycles can be done in a constant time in path length
- For other methods checking for cycles can be done in linear time in path length

A simple method of pruning the search while guaranteeing that a solution will be found in a finite graph, is to ensure that the algorithm does not consider neighbours that are already on the path from the start.

Cycle pruning checks whether the last node on the path already appears earlier on the path from the start node to that node. Paths where the end node is already in the path are not added to the frontier, or are discarded when removed from the frontier.

The complexity of cycle pruning depends on which search method is used. For DFS, overhead can be constant if a hash function is used that sets a bit when the node is in the path. Alternatively, for methods that have exponential space, cycle pruning takes time linear in the length of the path being searched. These algorithms cannot do better than simply searching up the initial path being considered, checking to ensure that they do not add a node that already appears in the path.

Multiple Path Pruning & A*

- prune a path to a node if we've already found a path to that node
- This is done using a closed list of nodes at the end of expanded nodes
- When a path is selected, if the end node is in the closed list then the path is discarded, otherwise the node is added to the closed list and the algorithm proceeds as before..

Multiple-path pruning is implemented by maintaining an explored set (**closed list**) of nodes that are at the end of paths that have been expanded. The explored set is initially empty. When a path is selected, if the node at the end is already in the closed list then we can discard the path. Otherwise, we add the node at the end to the closed list and the algorithm proceeds.

This does not guarantee that we don't discard the least cost path. Something more sophisticated would need to be done in order to ensure the optimal solution is found. To ensure that the search algorithm can still find a lowest-cost path to a goal, we can do one of the following:

1. Make sure that the first path found to any node is a lowest-code path to that node, then prune all subsequent paths found.
2. If the algorithm finds a lower-cost path, remove all paths that used the higher-cost path to the node
3. Whenever the search finds a lower-cost path to a node than a path to that node already found, it could incorporate a new initial section on the paths that have extended the initial path.

- Suppose path p' to n' was selected, but there is a lower-cost path to n'. Suppose this lower-cost path is via path p on the frontier
- Suppose path p ends at node n
- p' was selected before p (i.e. $f(p') \leq (p)$), so: $cost(p') + h(p') \leq cost(p) + h(p)$
- Suppose $cost(n, n')$ is the actual cost of a path from n to n'. The path to n' via p is lower cost than via p' so: $cost(p) + cost(n, n') < cost(p')$
- From these equations: $cost(n, n') < cost(p') - cost(p) \leq h(p) - h(p') = h(n) - h(n')$
- We can ensure this doesn't occur if $|h(n) - h(n')| \leq cost(n, n')$.

Problem: What if a subsequent path to n is shorter than the first path to n?

1. Ensure this doesn't happen by making sure that the shortest path to a node found first
2. Remove all paths from the frontier that use the longer path

A* does not guarantee that when a path to a node is selected for the first time it is the lowest cost path to that node. Note that the admissibility theorem guarantees this for every path to a goal node but not for every path to any node. Whether it holds for all nodes depends on the properties of the heuristic function.

Consistent Heuristics

A non-negative function h(n) on node n that satisfies the constraint

$h(n) \leq cost(n, n') + h(n')$ for any two nodes n' and n, where $cost(n, n')$ is the cost of the least-cost path from n to n'.

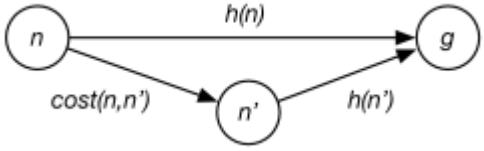
We can guarantee that a heuristic is consistent if it satisfies the monotone restriction

$h(n) \leq cost(n, n') + h(n')$ for any arc $\langle n,n' \rangle$

It is easier to check the monotone restriction as it only depends on the arcs, whereas consistency depends on all pairs of nodes.

Monotone Restriction

- Heuristic function h satisfies the monotone restriction if h(m) - h(n) <= cost(m,n) for every arc <m,n>
- If h satisfies the monotone restriction it is consistent meaning h(n) <= cost(n,n') + h(n') for any two nodes n and n'
- A* with a consistent heuristic and multiple path pruning always finds the shortest path to a goal
- This is a strengthening of the admissibility criterion



Consistency and the monotone restriction can be understood in terms of the triangle inequality which specifies that the length of any side of a triangle cannot be greater than the sum of lengths of the other two sides.

With a consistent heuristic, multiple-path pruning can never prevent A search from finding an optimal solution.*

This can be proved using proof by contradiction (see end of 3.7.1)

A* **search** in practice includes multiple-path pruning; if A* is used without multiple-path pruning, the lack of pruning should be made explicit. It is up to the designer of a heuristic function to ensure that the heuristic is consistent, and so an optimal path will be found.

Multiple-path pruning is **preferred over cycle pruning for breadth-first methods** where virtually all of the nodes considered have to be stored anyway.

Depth-first search does not have to store all of the nodes at the end of paths already expanded; storing them in order to implement multiple-path pruning makes depth-first search exponential in space. For this reason, **cycle pruning is preferred over multiple-path pruning for depth-first search**.

More Sophisticated Search

04 November 2020 21:19

Objectives:

- Understand that there are more sophisticated ways of searching
- Understand what Depth-First-Branch-And-Bound is
- Understand Iterative Deepening
- Become familiar with bi-directional search and island driven search

We can refine the aforementioned strategies using some more intelligent techniques

Depth-First Brand-And-Bound search

- Comines DFS with heuristic information
- Finds optimal solution most useful when there are multiple solutions and we want an optimal one
- Use the space of a DFS

- Suppose bound is the cost of the lowest cost path found to a goal so far
- What if the search encounters a path p such that the cost of p plus the heuristic applied to p is greater than the bound? - This means we can prune path p
- If we find a non-pruned path to the goal then we can set the bound equal to the cost of that path and then remember this as the best solution
- We should use DFS for linear space use
- We can guarantee an optimal solution with this algorithm

How do we initialise the bound? We can start it at infinity, or if we have an estimate we can use that

Notes on DFS BNB

- Cycle pruning works well with DFS BNB
- Multiple path pruning is not appropriate as storing explored set defeats space saving of dfs
- Can be combined with iterative deepening to increase the bound until either a solution is found or to show there is no solution

Context: Bounded Depth First Search

A bounded depth-first search takes a bound and does not expand paths that exceed the bound
Explores part of the search graph
Uses space linear in the depth of the search

The bound has to be the same or greater than the cost of getting to an optimal goal.
If we don't know that then we can do an iterative deepening search

Context: Iterative Deepening Search

1. Starts with a bound b=0
2. Do a bounded DFS with bound b
3. If a solution is found return that solution
4. Otherwise increment b and repeat

- This will find the same first solution as BFS
- Since using a depth-first search iterative deepening uses linear space
- Iterative Deepening has an asymptotic overhead of (b/(b-1)) times the cost of expanding the nodes at depth k using a BFS
- When b = 2 there is an overhead factor of 2, when b = 3 there is an overhead of 1.5 and as b

get higher the overhead factor reduces

Direction of Search

Bidirectional Search

- Search backward from the goal and forward from the start simultaneously
- This is effective since 2b^(k/2) < b^k and so can result in an exponential time saving
- The main problem is ensuring that the frontiers meet
- This is often used with one breadth-first method that builds a set of locations that can lead to the goal and in the other direction another method can be used to find a path to these interesting locations

Island Driven Search

- Find a set of islands between s and g
- There are m smaller problems rather than 1 big problem
- This can be effective since mb^(k/m) < b^k
- The problem is to identify the islands that the path must pass through it is difficult to guarantee optimality

Dynamic Programming

- For statically stored graphs build a table of dist(n), the actual distance of the shortest path from node n to a goal
- This can be built backwards from the goal
- This can be used locally to determine what to do

There are two main problems

- It requires enough space to store the graph
- The dist function needs to be recomputed for each goal

Constraint Satisfaction Problems

31 October 2020 15:06

Instead of reasoning explicitly in terms of states, it is typically better to describe states in terms of **features** and to reason in terms of these features. Features are described using **variables**. Often features are not independent and there are **hard constraints** that specify legal combinations of assignments of values to variables

When the goal state is defined in terms of restraints and you're looking for a solution that satisfies the constraints

- Types of CSP's
- Backtracking Search
- Arc-consistency in a constraint graph
- Domain splitting to solve
- Using CSP Problem structure
- Variable elimination for CSP

What is a constraint satisfaction problem?

- A CSP is characterized by a set of variables
- Each variable has an associated domain of possible values
- There are hard constraints on various subsets of the variables which specify legal combinations of values for the variables
- A solution to the CSP is an assignment of a value to each variable that satisfies all the constraints

CSP's as optimisation problems

- For optimisation problems there is a function that gives a cost for each assignment of a value to each variable
- A solution is an assignment of values to the variables that minimizes the cost function

Types of Variables

1. Discrete Variable – domain is finite or countably infinite
2. Binary Variable – a discrete variable with two values in its domain e.g. Boolean variable
3. Continuous Variable – not discrete e.g. a variable that corresponds to a real line

Given these variables, an assignment on the set of variable is a function from the variables into the domains of the variables:

$\{X1, X2, \dots, Xk\}$ as $\{X1 = v1, X2 = v2, \dots, Xk = vk\}$, where vi is in $dom(X_i)$

A possible world is a complete assignment of variables. That is, a function from variables into values that assigns a value to every variable

We use variables because we can reason about many worlds with just a few variables

Types of Constraints

In many domains , not all possible assignments of values to variables are permissible
A constraint specifies legal combinations of assignments of values to some of the variables.

A scope is a set of variables

A relation on a scope is a function that returns true or false to an assignment on a scope
A constraint is a scope and a relation on S. It involves each of the variables in its scope

A model is a possible world that satisfies all of the constraint

Constraints are defined by their **intension** in terms of formulas or by their **extension**, listing all the assignments that are true. Constraints defined extensionally can be seen as relations of legal assignments as in relational databases.

- Unary constraints – involve a single variable e.g. some variable can't be green or > 10
- Binary constraints – involve pairs of variables that can't be equal
- Higher-order constraints – involve 3 or more variables
- Preference or soft constraints – e.g. 1005 is better than 0805

Constraint Satisfaction Problems

- A set of variables
- A domain for each variable
- A set of constraints

A finite CSP has a finite set of variables and a finite domain for each variable

Given a CSP, we can do some useful tasks:

- Determine or not whether there is a model
- Find a model
- Count the number of models
- Enumerate all the models
- Find the best model given a measure of how good models are
- Determine whether some statement holds in all models

Real World CSP's

- Assignment problems
- Timetabling
- Hardware configuration
- Spreadsheets
- Floor planning

Solving CSP's

- A CSP can be solved by graph-searching
- A node is an assignment of values to some of the variables
- The start node is the empty assignment
- A goal node is a total assignment that satisfies the constraints

Generate-and-Test Algorithms

A finite CSP could be solved by an exhaustive algorithm. The assignment space D is the set of total assignments. The generate and test algorithm can be run to just return the first variable found.

Generate the assignment space I.e. the set of total assignments
Test each assignment with constraints

$$\begin{aligned}\mathbf{D} &= \mathbf{D}_A \times \mathbf{D}_B \times \mathbf{D}_C \times \mathbf{D}_D \times \mathbf{D}_E \\ &= \{1,2,3,4\} \times \{1,2,3,4\} \times \{1,2,3,4\} \\ &\quad \times \{1,2,3,4\} \times \{1,2,3,4\} \\ &= \{\langle 1,1,1,1,1\rangle, \langle 1,1,1,1,2\rangle, \ldots, \langle 4,4,4,4,4\rangle\}.\end{aligned}$$

How many assignments need to be tested for n variables, each with a domain of d? D to the power of n.

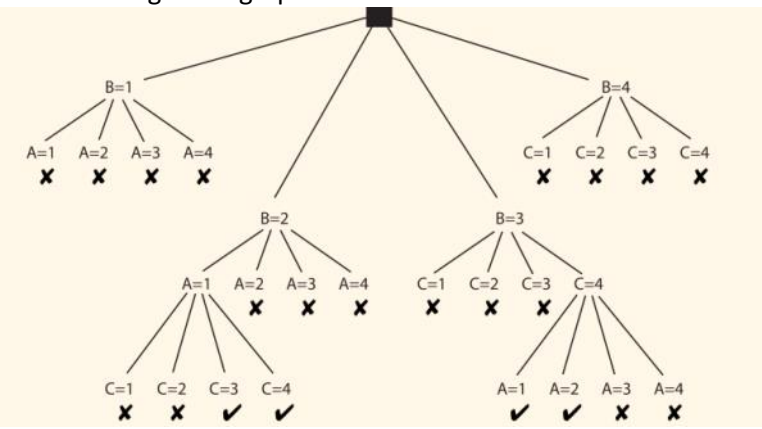
Algorithms for CSP's are trying to cut down this assignment space as it is so large.

Backtracking Algorithms

Generate-and-Test algorithms assign values to all variables before checking constraints. Because individual constraints only involve a subset of the variables, some constraints can be tested before all of the variables have been assigned values. If a partial assignment is inconsistent with the constraint then any total assignment that extends the partial assignment will also be inconsistent.

We can use backtracking instead, where we construct a search space that can be explored by the search algorithms discussed previously.

The following is the graph when A < B and B < C



1. Systematically explore D by instantiating the variables one at a time
2. Evaluate each constraint predicate as soon as all its variables are bound
3. Any partial assignment that does not satisfy the constraint can be pruned.

- Every solution appears at depth n so we can use a DFS
- Path is irrelevant so can use complete state formulation
- Branching factor b = (n-l)d at depth l hence n! x d^n leaves – top level branching factor is nd since any of d values can be assigned to any of n variables' next level branching factor si s(n-1) d and so on
- Variable assignments are commutative e.g. x=5, y=10 === y=10, x=5
- Only need to consider assignments to a single variable at each node
- Backtracking search si the basic uninformed algorithm for CSP's
- Can solve n-queens for n = 25

This is much more efficient that a generate-and-test algorithm because there we only test the constraints at the leaf nodes, whereas in a backtracking algorithm we prune all nodes immediately that violate a constraint, leaving us only with valid avenues to a possible model.

Backtracking Search

```
function BACKTRACKING-SEARCH(csp) returns solution/failure
  return RECURSIVE-BACKTRACKING({}, csp)

function RECURSIVE-BACKTRACKING(assignment,csp) returns solution/failure
  if assignment is complete then return assignment
  var ← SELECT-UNASSIGNED-VARIABLE(VARIABLES[csp], assignment, csp)
  for each value in ORDER-DOMAIN-VALUES(var, assignment, csp) do
    if value is consistent with assignment given CONSTRAINTS[csp] then
      add {var = value} to assignment
      result ← RECURSIVE-BACKTRACKING(assignment,csp)
      if result ≠ failure then return result
      remove {var = value} from assignment
  return failure
```

Improving Search Efficiency

1. Which variable should be assigned next? MRV & degree heuristic
2. What order should we try its values? LCV
3. Can we detect inevitable failure early? Consistency Algorithms
4. Can we take advantage of the problem structure? Cutset conditioning & variable elimination

Minimum remaining values (MRV)

Choose the variable with the fewest legal values

- Also called fail-first heuristic: will pick variable most likely to cause failure – if exists a variable with 0 possible assignments will pick and fail immediately

Degree Heuristic

Tie breaker among MRV variables

- Choose the variable with the most constraints on remaining variables
- Attempts to reduce branching factor of future choices

Least Constraining Value LCV

- Given a variable, choose the least constraining value: the one that rules out the fewest values in the remaining variables

Combining these we can solve n queens for n = 1000

Consistency Algorithms

Although DFS over the search space of assignments is usually a substantial improvement over generate and test, it still has various inefficiencies that can be overcome.

Idea: prune the domains as much as possible before selecting values from them

- A variable is domain consistent if no value of the domain of the node is ruled impossible by any of the constraints
- Example: if domain is Domain(B) = {1,2,3,4} and we have a rule that B can't equal 3, then D is not domain consistent

Constraint Network

1. There is a circular node for each variable

2. There is a rectangular node for each constraint
3. There is a domain of values associated with each variable node
4. There is an arc from variable X to each constraint that involves X

Arc Consistency

If there is a constraint that acts on some variables X, Y, Z, then the arc <X, c> can be called arc consistent if for each value of x in the domain(X) there are some values y and z in the domain(Y) and domain(Z) such that if we set X = x, Y = y, Z = z, the constraint is satisfied.

A network is arc-consistent if every arc is arc-consistent

What if an arc is not consistent?

If the arc <X, c> is not consistent then there are some values of X for which there are no values for Y, Z for which the constraint holds. In this case, all values of X in the domain(X) for which there are no corresponding values for the other variables can be deleted from the domain(X) to make the arc consistent.

When a value is removed from the domain of a variable it is possible that it will make some other arcs that were previously consistent, no longer consistent.

Arc Consistency Algorithm

See figure 4.3 in textbook

The arcs can be considered in turn making each arc consistent
When an arc has been made arc consistent, does it ever need to be checked again? YES – an arc needs to be revisited if the domain of one of following variables is reduced

Three possible outcomes when all arcs are consistent

- One domain is empty – no solution
- Each domain has a single value – unique solution
- Some domains have more than one value – there may or may not be a solution

Domain Splitting

Another method for simplifying the network is domain splitting, or case analysis
The idea is to split the problem into a number of disjoint cases and solve each case separately. The set of all solutions to the initial problem is the union of the solutions to each case.

In the simplest case, imagine we have a variable X with a domain of {t,f}. All solutions either have X = t, or X = f. One way to find the solutions is to set X = t, find all of the solutions with this assignment, then assign X = f, and find all those solutions.

If we make the problem into a smaller set of subproblems, we can get massive improvements in efficiency:

- Suppose each subproblem has c of n total variables
- There are n/c subproblems each of which takes at most d^c to solve. Worse case solution is therefore n/c x d^c, I.e linear in n

If the variable has > 2 elements, we can split it in a number of ways.

- Split it into a case for each value
- Always split the domain down the middle into 2 disjoint sets

We can be more efficient by interleaving arc consistency with the search
We would solve a CSP by using arc consistency to simplify the network before each step of domain splitting

After domain splitting, we do not need to start arc consistency from scratch. We can simply check the arcs that are possibly no longer arc consistent as a result of the split.

Complexity of Generalized Arc Consistency Algorithm

- If there are c binary constraints, and the domain of each variable is of size d. There are 2c arcs.
- Checking an arc <X, r(X,Y)> involves in the worst case iterating through each value in the domain of Y for each value in the domain of X, which takes d^2 time.
- This arc may need to be checked once for every element in the domain of Y, thus GAC for binary variables can be done in time O(cd^3) which is linear in C – the number of constraints
- The space used is O(nd) where n is the number of variables and d is the domain size.

Hard and Soft Constraints

Given a set of variables, assign a value to each variable that either:

- Satisfies some set of constraints – satisfiability problems – hard constraints
- Minimizes some cost function where each assignment of values to variables has some cost – optimisation problems – soft constraint
- Many problems are a mix of hard and soft constraints

Tree-Structured CSP's

- Theorem: If the constraint graph has no loops, the CSP can be solved in O(nd^2)
- Compare this to general CSP's, where worst case time is O(d^n)
- This property also applies to logical and probabilistic reasoning: an important example of the relation between syntactic restrictions and the complexity of reasoning.

Algorithm

- Choose variable as root, order variables from root to leaves such that ever nodes parent precedes it in the ordering
- For j from n down to 2, remove inconsistent domain elements for the parent.
- For j from 1 to n, assign X consistently with Parent

Nearly Tree-Structured CSP's

- Conditioning – instantiate a variable, prune its neighbours domains, I.e. assign variable so remained is a tree.
- Cutset conditioning: instantiate a set of variables such that the remaining constraint graph is a tree
- Cutset size c -> runtime O(d^c x (n-c)d^2) very fast for small c.

Variable Elimination

Arc consistency simplifies the network by removing values of variables. A complementary method is variable elimination which simplifies the network by removing variables.

The idea is to remove the variables one by one. When removing a variable X, VE constructs a new constraint on some of the remaining variables, reflecting the effects of X on all the other variables.

This new constraint replaces all of the constraints that involve X, forming a reduced network that does not involve X.

When we eliminate X, the influence of X on the remaining variables is through the constraint relations that involve X. First, the algorithm collects all of the constraints that involve X.

Variable Elimination Algorithm

If there is only one variable return the intersection of the unary constraints that contain it

- Select a variable x
- Join the constraints in which X appears, forming constraint R1
- Project R1 onto its variables other than X, forming R2
- Replace all of the constraints in which X appears by R2
- Recursively solve the simplified problem, forming R3
- Return R1 joined with R3

- When there is a single variable remaining, if it has no values, the network was inconsistent
- The variables are eliminated according to some elimination ordering
- Different elimination orderings result in different size intermediate constraints

Figure 4.6 gives a recursive algorithm for variable elimination to find all the solutions for a CSP.

The number of variables n the largest relation returned for a particular variable ordering is called the **treewidth of the graph for that variable ordering**. The **treewidth of a graph is the minimum treewidth for an ordering**.

The complexity of VE is exponential in treewidth and linear in the number of variables.

There are some heuristics that exist that can help us get the minimum treewidth:

1. Min-factor – at each stage, select the variable that results in the smallest relation
2. Minimum deficiency or minimum fill – at each stage, select the variable that adds the fewest arcs to the remaining constraint network. The intuition is that it is okay to remove a variable that results in a large relation as long as it does not make the network more complicated.

Example: eliminating C

$r_1 : C \neq E$

C	E
3	2
3	4
4	2
4	3

$r_2 : D < C$

C	D
3	2
4	2
4	3

$r_3 : r_1 \bowtie r2$

C	D	E
3	2	2
3	2	4
4	2	2
4	2	3
4	3	2
4	3	3

$r_4 : \pi_{\{D,E\}} r_3$

D	E
2	2
2	3
2	4
3	2
3	3

↪ new constraint

NB: $r_1 \bowtie r_2$ = join of r_1 and r_2, $\pi_S(r)$ = projection of r onto S.

Local Search

05 November 2020 13:11

Objectives:

1. Hill climbing
2. Greedy descent
3. Randomized algorithms
4. Simulated annealing
5. Genetic algorithms

Iterative Improvement Algorithms

Cases where we don't care about the path – we just want a goal state

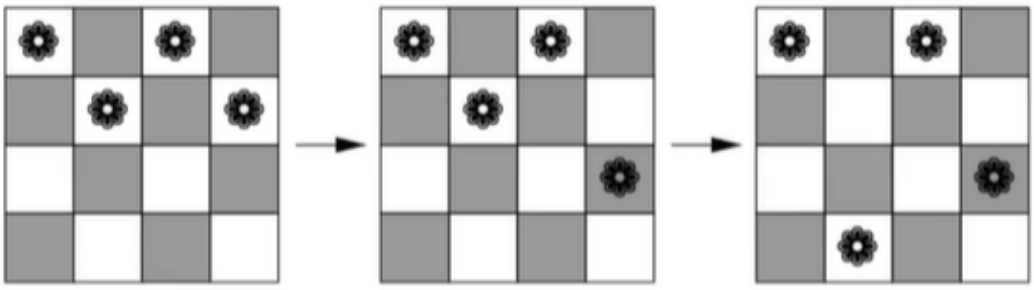
Systematically searching large spaces is not practical in many cases. We can use these iterative improvement algorithms to deal with these large search spaces.

The methods find solutions quickly on average, but do not guarantee that a solution will be found, even if one exists – they are therefore not able to prove a solution exists. They are useful when we already know a solution exists or is very likely to exist.

Local search methods start with a total assignment of a value to each variable and try to improve this assignment iteratively by taking improving steps, by taking random steps, or by restarting with another total assignment.

- Uses a single current state (not multiple paths) and typically moves to neighbours of state
- Not systematic, but low memory usage and can find solutions in continuous spaces
- Useful for optimisation problems including CSP's

The state space is the set of all configurations. The goal is a particular configuration
e.g. the travelling salesman or n-queens

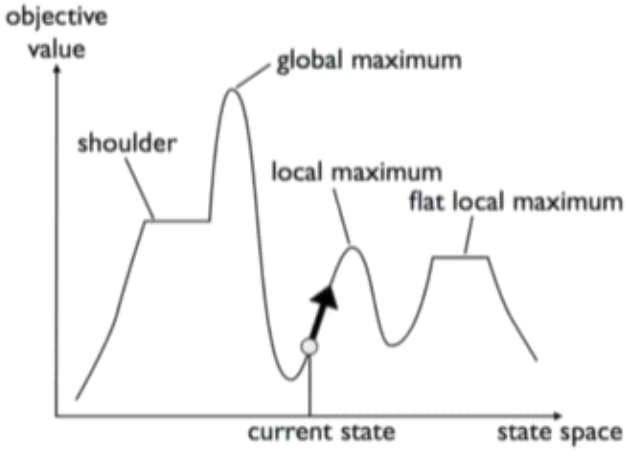


we can use iterative improvement algorithms on these problems, by keeping the current state and trying to improve it.

These algorithms often have constant space

Hill-Climbing

- Greedy local search – always tries to improve the current state, or reduce the cost if evaluation function is cost
- Each iteration moves in the direction of increasing value, or decreasing if evaluation is cost
- No search tree – just keep current state and its cost
- If > 1 alternative with equal cost, choose at random



Problems include local maxima where we have a local peak that is lower than the highest peak, but the algorithm will halt with a suboptimal solution

Also ridges, plateaux – flat area will conduct a random walk

Greedy Descent

Maintain an assignment of values for each variable
Repeatedly select a variable to change and a value for that variable

- Aim is to find an assignment with zero unsatisfied constraints
- Given an assignment of a value to each variable, a conflict is a violated constraint
- The goal is an assignment with zero conflicts
- Heuristic function to be minimized the number of conflicts

To choose a variable to change and add a new value:

- Find a variable-value pair that minimizes the number of conflicts
- Select a variable that participates in the most conflicts
- Select a value that minimizes the number of conflicts
- Select a variable that appeaars in any conflict
- Select a value that minimizes the number of conflicts
- Select a variable at random
- Select a value that minimizes the number of conflicts
- Select a variable and value at random, accept this change if it does not increase the number of conflicts

Complex Domains

- When the domains are small or unordered, the neighbours of an assignment can correspond to choosing another value for one of the variables
- When the domains are large and ordered, the neighbours of an assignment are the adjacent values for one of the variables
- If the domains are continuous, gradient descent changes each variable proportionally to the gradient of the heuristic function in that direction

The value of variable X_i goes from v_i to $v_i - \eta \frac{\partial h}{\partial X_i}$ where η is the step size.

Randomized Greedy Descent

As well as downward steps, we can allow for:

- Random steps: move to a random neighbour
- Random restart: reassign random values to all variables

Stochastic Local Search

Combination of:
Greedy descent – move to a lowest neighbour
Random walk
Random restart

Random Walk

Randomly sometimes choose a random variable value pair
When selecting a variable then a value:

- Sometimes choose any variable that participates in the most conflicts
- Sometimes choose any variable that participates in any conflict
- Sometimes choose any variable

Sometimes choose the best value and sometimes choose a random value

Simulated Annealing

1. Pick a variable at random and new value at random
2. If it is an improvement, adopt it
3. If it is not an improvement, adopt it probabilistically depending on a temperature parameter, T which is reduced over time

e.g. if T = 10 then there is a big chance we'll accept the proposed random change, but if T = 0.1 it is very small

- To prevent cycling we can maintain a tabu list of the last k assignments
- Do not allow an assignment that is already on the tabu list
- If k = 1, we do not allow an assignment of the same value to the variable chosen
- Can we very expensive if k is large

Simulated annealing requires an annealing schedule which specifies how T is reduced as the search progresss. Geometric cooling one of the most widely used schedules.

Parallel Search

- A total assignment is called an individual
- Idea: maintain a population of individuals instead of one
- At every stage, update each individual in the population
- Whenever an individual is a solution, it can be reported
- Like k restarts but uses k times the minimum number of steps

Beam Search

- Like parallel search, with k individuals but choose the k best out of all the neighbours
- When k = 1, it is a greedy descent
- When k = infinity it is a BFS
- The value of k lets us limit space and parallelism

This can be made into a stochastic beam search is we probabilistically choose the k individuals at the next generation

The probability that a neighbour is chosen is proportional to its heuristic value
This maintains diversity among individuals

The heuristic value reflects the fitness of the individual

Like **asexual reproduction**: each individual mutates and the fittest ones survive

Genetic Algorithms

- Related to stochastic beam search
- Successor states obtained from two parents
- Start with a population of k individuals
- Each individual represented as a string over a finite alphabet, e.g. string of 0's and 1's. N-queens would be [1 5 6 3 4 2 3 8] for one potential layout, for example.

Fitness Function

- Each individual in the population is evaluated by a fitness function
- Fitness function should return higher values for better states
- Fitness function determines probability of being chosen for reproduction
- Pairs of individuals chosen according to these probabilities – those below a threshold can be culled

Crossover

For each chosen pair, a random crossover point is chosen from the string representation
Offspring are generated by crossing the parents strings at a chosen point
First child gets first part of string from 1 and second from 2, and second gets the opposite

We then subject the new individuals to mutation

Adversarial Search

05 November 2020 15:22

Objectives:

- Adversarial Search
- Minimax
- Alpha-Beta Pruning
- Imperfect Decisions
- Games with Chances

A competitive multi-agent environment where goals are in conflict – gives rise to games

Other agents – opponents – which introduce uncertainty

- An adversarial search agent must deal with contingency
- High complexity and time sensitive – typically have to make a best guess based on experience and time available
- Chess has a branching factor of 35 and a game is 100 moves – thus we have 35^100 nodes – we have to make the best move given the situation

Uncertainty

- From the opponent trying to make the best move for themselves
- Randomness – e.g. throwing a dice
- Insufficient time to determine consequences

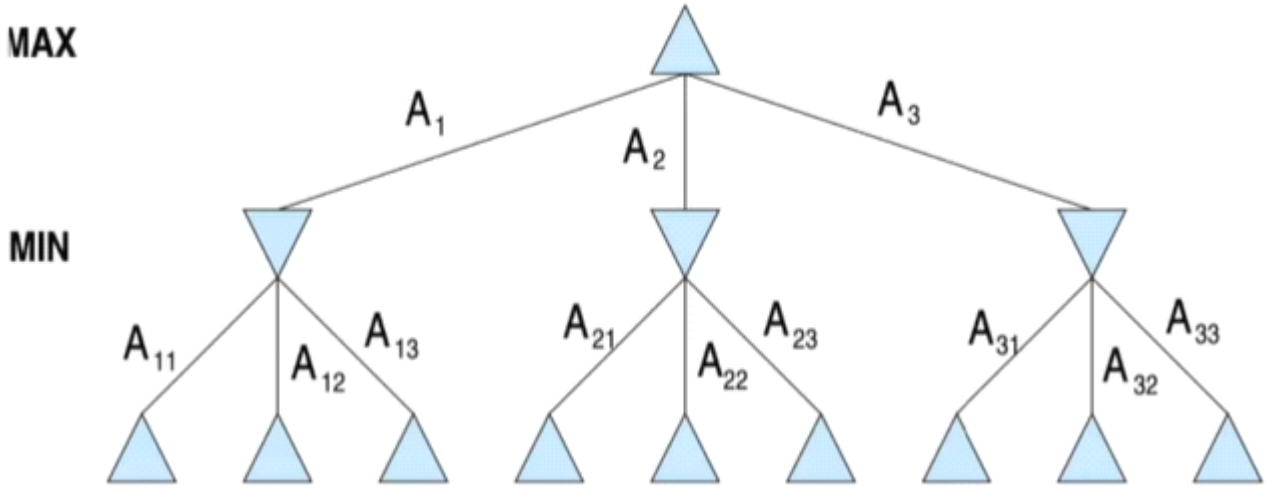
Formal View of a Game

- Initial state – the board position and player to move
- Set of operators defining legal moves and resulting states
- A terminal test to determine a terminal state
- A utility function or payoff function – gives a numeric value for terminal states e.g. +1, -1, 0 for chess win lose draw

We can build a game tree based on initial state and operators for each player

Ply

A single move in a 2 player game takes 2 half-moves, or is "2 ply"



Minimax

In the case where two agents are competing so that a positive reward for one is a negative reward for the other, we have a two-agent zero-sum game. The value of such a game can be characterized by a single number that one agent is trying to maximise and the other agent is trying to minimize. Having a single value fora two-agent zero sum game leads to a minimax strategy. Each node is either a max node, if it is controlled by the agent trying to maximise, or is a min node if it is controlled by the agent trying to minimize.

- Gives and optimal strategy for max
- Choose move with the highest minimax value

The minimax value of a state is the utility (for Max) of being in that state assuming both players play optimally from that state until the end of game

Obtains best achievable payoff against best play

Algorithm

- Generate a complete game tree
- Use utility function to rate terminal states
- Use utility of terminal states to give utility of nodes one level up
- Continue backing up tree until reaching the root
- Max should choose move that leads to highest utility

This is the minimax decision:

Maximises utility under the assumption that the opponent will play to minimise it

- Complete if tree is finite
- Optimal against an optimal opponent
- Space - O(bd) because DFS
- Time – O(b^d) - a killer for real games
- Minimax requires a complete search tree which is not practical
- Forms basis for realistic algorithms

Multi-player Minimax

Can extend minimax to multiple players by using vectors of utilities

Alpha-Beta Pruning

Complete search tree is impractical – alternative is to prune branches that will not influence decision

- Consider a node n that might be chosen
- If a better choice exists, then n will never be reached so prune it
- As soon as we discover a better choice than n, prune it

Minimax is a DFS, and ABP gets its name from the parameters backed up the path

- Alpha = value of best choice along the path for Max (highest utility)
- Beta = value of best choice for min (lowest utility for Max)

ABP updates alpha and beta as it searches, pruning as soon as value of current node is known to be worse than the current alpha or beta for max or min respectively
Pruning is done by terminating the recursive call

Order of Examining Successors

- The effectiveness of ABP is dependent on order of examining successors
- So, we should try to examine the best successors first
- If we could do this, the ABP looks at O(b^(d/2)) instead of O(b^d) for minimax – roughly twice the lookahead
- For random order successors, APB looks at O(b^(3d/4) nodes
-

In practice, a simple ordering function can give significant advantage
e.g. for chess we might look at capture moves first then threats, then forward moves, then backward moves

Imperfect Decisions

APB prunes much of the search tree, while minimax needs the complete tree
But APB still needs to search to the terminal states for some of the tree which is still impractical

Alternative: cut off the tree earlier, using:

- A heuristic evaluation function to get a value for states
- A cut-off test to determine when to stop going down the tree

Evaluation function gives an estimate for expected utility for a given position

Cutting off tree turns nonterminal nodes into terminal leaves.

Eval function should:

- Order terminal states as per utility function
- Approximate actual utility state

Uncertainty is unavoidable since we are not considering a complete tree

- Most evaluation functions calculate features of a state e.g. number of each piece
- These give equivalence classes for states which will lead to a win, draw or less with some probability
- Combine features with a weighted linear function
- Assumes features are independent

Cutting Off Search

Simplest approach is to set a fixed depth – cut-off test succeeds at depth d
More robust approach is to use an iterative deepening, continue until out of time, then return the best move found so far
Both of these approaches are unreliable

Solution: only apply eval function to quiescent positions – those whose value is unlikely to change significantly in the near future

- Non-quiescent positions expanded until quiescent positions reached
- This extra search is called quiescent search
- Quiescent search restricted to certain types of more to quickly resolve uncertainties in position
- Horizon problem: faced with unavoidable damaging move from opponent, a fixed depth search is fooled into viewing stalling moves as avoidance

Singular extension search as a means of avoiding horizon problem
Singular extension is a move that is clearly better than all others
In chess, can search to see whether opponent can advance pawn to 8th row, turning it into a queen

Forward pruning: immediately prune some moves from a node with no further considerations

- Only safe in special cases like when the two moves are symmetric or equivalent and only

consider on of them, or nodes are very deep in search tree

Games with Chance

- Many games contain chance
- Legal moves are dependent on roll of dice, so cannot construct a complete game tree
- Have to include chance nodes to solve this
- These nodes can be labelled with the possibility of the expected value taken over the chance nodes

Knowledge Bases

21 December 2020 14:26

Objectives:

- Understand Knowledge Bases
- Learn how a Knowledge-Based Agent functions
- Understand inference and inference engines

Knowledge Base – a database for the system that contains all the facts and beliefs that the system knows. It is a representation of the systems idea of the world
Inference engine – domain-independent algorithms – a mechanism for reasoning about those beliefs

The knowledge is *domain specific* but the inference engine is *domain independent*

Knowledge Bases = **sentences in a formal knowledge representation language** - but implementation could be anything – linked lists, arrays, databases etc

1. A declarative approach to building an agent – we tell is what it needs to know – it can then ask itself what to do. Answers should follow from the KB through inference
2. **TELL** and **ASK** are standard names for adding sentences and querying KB
3. Result of ASK must follow from previous TELLS as determined by inference mechanism

Each time the agent program is called, it firstly TELLS the knowledge base what it perceives, and then ASKS the knowledge base what action it should perform

We have to build agents that can take a knowledge base and use some inference mechanism to perform actions – **planning is about building that inference engine**

Knowledge Based Agents

- Can reason using inference and their knowledge.
- Can accept new tasks in the form of goals
- Can adapt to environmental change by updating knowledge
- Are able to infer unseen properties of the world from perceptions
- Can often find better solutions than simple search

Moreover, significantly more flexible with respect to **adopting new goals**, **partially observable environments, dynamic environments**

Characterising KBA's

1. **Knowledge Level** – what is known? Allows us to work at an abstract level of ASK and TELL (AKA epistemological level) e.g. a taxi might know that the Golden Gate Bridge links San Francisco to Marin County
2. **Logical Level** – knowledge encoded in formal sentences e.g. links(GGB, SF, M)
3. **Implementation Level** – data structures in KB and algorithms that manipulate them e.g. this could be a 1 in a 2D array of places, where 1 means Links X to Y

Simple KBA

The agent must be able to:

- Represent states and actions
- Incorporate new percept's
- Update internal representations of the world

- Deduce hidden properties of the world
- Deduce appropriate actions

Algorithm:

```
function KB-AGENT(percept) returns an action
  static : KB, a knowledge base
    t, a time counter, initially 0
  TELL(KB,MAKE-PERCEPT-SENTENCE(percept, t))
  action ← ASK(KB, MAKE-ACTION-QUERY(t))
  TELL(KB,MAKE-ACTION-SENTENCE(action, t))
  t ← t + 1
  return action
```

- Knowledge base may contain initial background knowledge
- Each iteration, TELL KB of perceptions, ASK What actions to perform
- Note: TELL and ASK refer to KB – they are internal
- Representation details hidden by MAKE-PERCEPT-SENTENCE and MAKE-ACTION-QUERY allow us to work at knowledge level
- Inference details hidden in TELL and ASK

Wumpus World

- Managed to get the goal because we can make inferences about knowledge gained from perceptions
- Combining knowledge obtained at different times and in different places allows us to infer more about the world
- Using lack of a particular perception rather than just the existence of a perception allows us to extract more knowledge from the world
- We rely on persistence of knowledge – the world is not fully observable

First-Order Logic

22 December 2020 14:29

A logic that is sufficient for building Knowledge Based Agents

Before, we've used propositional logic as our representation language because it is one of the simplest languages that demonstrates all the important points. Unfortunately, *propositional logic has a very limited ontology*, making only the commitment that the world consists of facts. This makes it difficult to represent even something simple

FOL or *First-Order Predicate Calculus makes a stronger set of ontological commitments*. The main one is that the world consists of objects, that is things with individual identities and properties that distinguish them from other objects.

Among these objects, various relations hold. Some of these relations are functions - relations for which there is only one value for a given input. It is easy to start listing examples of objects, properties, relations and functions

- Objects - people, houses, numbers, theories, colours, baseball games, wars
- Relations - brother of, bigger than, inside of, has colour, occurred after, owns
- Properties - red, round, prime
- Functions - father of, best friend, third inning of, one more than

FOL makes *no commitments to time, categories and events*. A logic that tried to, would only have limited appeal as there are so many different ways of interpreting them. Thus, FOL remains neutral and gives us the freedom to describe these things in a way that is appropriate for the domain. Freedom of choice is a general characteristic of FOL

Syntax and Semantics

In propositional logic, every expression is a sentence, which represents a fact. First-Order Logic has sentences, but it also has terms, which represent objects.
Terms are built from constant symbols, variables, and function symbols.

FOL BNF

1. Sentence -> AtomicSentence | Sentence Connective Sentence | Quantifier Variable, ... Sentence | !Sentence | (Sentence)
2. AtomicSentence => Predicate(Term,...) | Term = Term
3. Term -> Function(term, ...) | Constant | Variable
4. Connective -> => | /\ | \/ | <=>
5. Quantifier -> For All | For Some
6. Constant -> A | John
7. Predicate -> Before | HasColor | Raining
8. Function -> Mother | LeftLegOf

Constant

Which object in the world is referred to by each constant symbol? Each constant symbol names exactly one object, nbut not all objects need to have names, and some can have several names. Thus, the symbol john, in one particular interpretation might refer to a specific king, but the symbol king could refer

Predicate Symbols

An interpretation specifies that a predicate symbols refers to a particular relation in the model

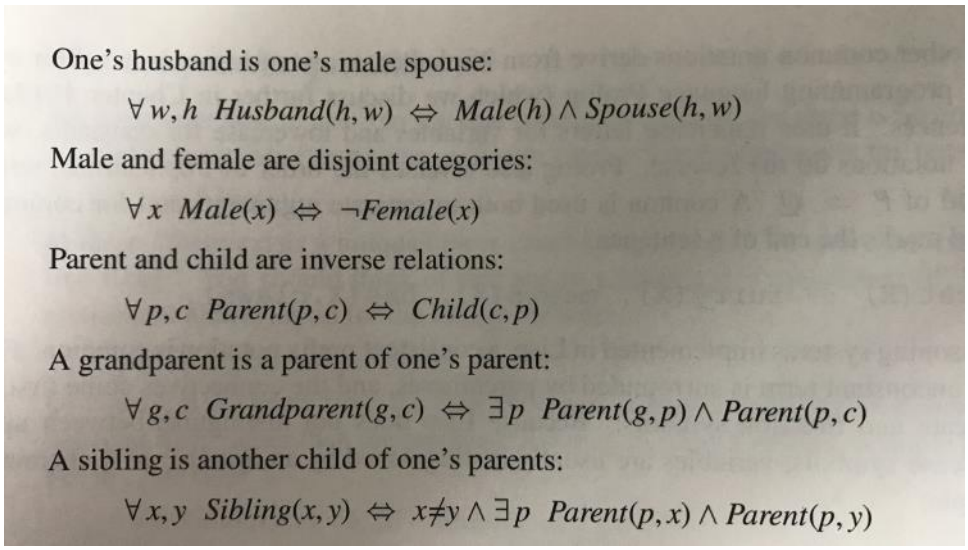
Brother might refer to the relation of brotherhood
A relation is defined by the set of tuples of objects that satisfy it
Thus, a predicate symbol is a symbol that points to this list of satisfying tuples

Function Symbols

- Some relations are functional - that is, any given object is related to exactly one other object by the relation.
- For example, any angle has only one number that is its cosine, and person has only oner person that is his or her father.
- In such cases, it is often more convenient to define a function symbol e.g. cosine that refers to the appropriate relation between angles and numbers.
- In the model, the mapping is just a set of n+1 tuples with a special property, namely that the last element of each tuple is the value of the function for the first n elements.
- A table of cosines is an example of this set of tuples.
- Unlike predicate symbols which hare used to state that relations hold among certain objects, function symbols are used to refer to particular objects without using their names.

Using First-Order Logic

In knowledge representation, a domain is a section of the world about which we wish to express some knowledge


One's husband is one's male spouse:

$$\forall w,h \;\; Husband(h,w) \Leftrightarrow Male(h) \wedge Spouse(h,w)$$

Male and female are disjoint categories:

$$\forall x \;\; Male(x) \Leftrightarrow \neg Female(x)$$

Parent and child are inverse relations:

$$\forall p,c \;\; Parent(p,c) \Leftrightarrow Child(c,p)$$

A grandparent is a parent of one's parent:

$$\forall g,c \;\; Grandparent(g,c) \Leftrightarrow \exists p \;\; Parent(g,p) \wedge Parent(p,c)$$

A sibling is another child of one's parents:

$$\forall x,y \;\; Sibling(x,y) \Leftrightarrow x \neq y \wedge \exists p \;\; Parent(p,x) \wedge Parent(p,y)$$

Representation, Reasoning & Logic

22 December 2020 12:08

The object of knowledge representation is to *express knowledge in computer-tractable form*, such that it can help the agent to perform well.

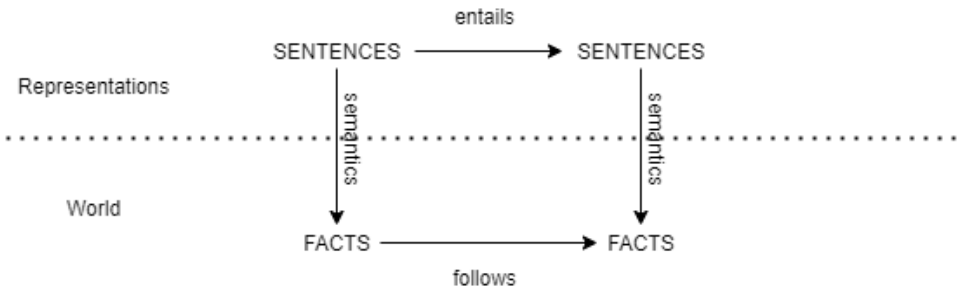
There are two aspects that are key in knowledge representation - **syntax** and **semantics**

- The syntax of a language describes the possible configurations that can constitute sentences. E.g. although X, Y and > are valid symbols in a language, X Y > might not be syntactically correct, whereas X > Y might be.
- Semantics determines the facts in the world to which the sentences refer. Without semantics a sentence is just an arrangement of electrons or a collection of marks on a page. With semantics, each statement makes a claim about the world. With semantics, we can say that when a particular configuration exists within an agent, the agent believes the corresponding sentence.

Provided we have a precisely syntax and semantics, we can call the language a logic.

From the syntax and semantics, we can derive an inference mechanism, that uses the logic.

- Facts are parts of the world
- Representations are encoded in some way that can be physically stored in an agent
- Because sentences are physical configurations of parts of the agent, reasoning must be a process of constructing new physical configurations from old ones. Proper reasoning should ensure that the new configurations represent facts that actually follow from the facts that the old configurations represent



An inference procedure can do two things:

1) Given a KB, generate new sentences that are entailed by KB
2) Given a KB and a new sentence alpha, decide whether or not KB entails alpha.

Entailment

We want to generate new sentences that are necessarily true, given that the old sentences are true. This relationship between sentences is called **entailment**, and mirrors the relations of facts following from each other.

KB |= alpha

Entailment is important since it provides a strong way of showing that if certain propositions are true, then some other proposition must be true

- KB entails sentence alpha if and only if alpha is true in all worlds where KB is true

- e.g. if the KB has "James is male" and "James is 34" then the KB entails "James is male or 34"
- Semantics give mapping of sentences to facts
- Logical inference generates sentences that are entailed by existing sentences and should ensure relationship mirrored in real world
- By considering the semantics of a language we can extract the proof theory of the language – what reasoning steps are sound

Inference procedures that generate only entailed sentences are sound or truth preserving

Inference

The term "inference" generally covers any processes by which conclusions can be reached. We are mainly concerned with sound reasoning, which is called "logical inference", or "deduction". Logical inference is a process that implements the entailment relation between sentences.

if i can derive alpha from KB, then we would write:

$$\mathrm{KB} \vdash_i \alpha$$

means that sentence "alpha can be derived from KB by inference procedure i", or "i derives alpha from KB"

- **Soundness**: i is sound if whenever KB |-i alpha, it is also true that KB |= alpha
- **Completeness**: i is complete if whenever KB |= alpha, it is also true that:
  - $\mathrm{KB} \vdash_i \alpha$

We need a logic which is expressive enough to say almost anything of interest and for which there exists a sound and complete inference procedure. That is, the procedure will answer any question whose answer follows from what is known by the KB

Logics

A formal system for describing states of affairs, consisting of

1) the syntax of the language, which describes how to make sentences
2) the semantics of the language, which states the systematic constraints on how sentences relate to states of affairs
3) The proof theory, a set of rules for deducing the entailments of a set of sentences.

there are two main types of logic - propositional and first-order logic

Propositional logic

Symbols represent whole propositions, or facts. We combine propositions with Boolean connectives to generate sentences with more complex meanings.

- Negation - if S is a sentence, then not S is a sentence
- Conjunction - if S and B are sentences then S and B is a sentence
- Disjunction - if S and B are sentences then S or B is a sentence
- Implication - if S and B are sentences then S ==> B is a sentence
- Equivalence is S and B are sentences then S <=> B is a sentence

This creates a lot of problems - we need a lot of rules and logic to create a semi competent agent.

E.g. in Wumpus World, rule "Don't go forward if a wumpus is in front of you" requires 64 rules - 16 squares with 4 orientations

First-Order Logic

Represents the world in terms of objects and predicates on objects, e.g. properties of objects or relations between objects, as well as using connectives and quantifiers, which allow sentences to be written about everything in the universe at once.

Summary

We have introduced the idea of a KBA, and showed how we can define a logic with which the agent can reason about the world and be guaranteed to draw correct conclusions, given correct premises. We have also showed how an agent can turn this knowledge into action

- Intelligent agents need knowledge about the world in order to reach good decisions
- Knowledge is contained in agents in the form of **sentences** in a **knowledge representation language**, stored in a knowledge base
- A knowledge based agent is composed of a knowledge base and an **inference mechanism**. It operates by storing sentences about the world in its knowledge base, using the inference mechanism to infer new sentences, and using them to decide what action to take.
- A representation language is defined by its syntax and semantics, which specify the structure of sentences and how they relate to facts in the world
- The interpretation of a sentence is the fact to which it refers/ *If it refers to a fact that is part of the actual world then it is true.*
- *Inference is the process of deriving new sentences from old ones*/ We try to design sound inference processes that derive true conclusions given true premises. An inference process is complete if it can derive all true conclusions from a set of premises.
- A sentence that is *true in all worlds under all interpretations is called valid*. If an implication sentence can be shown to be valid, then we can derive its consequent if we know its premise. The ability to show validity independent of meaning is essential
- Different logics make different commitments about what the world is made of and what kinds of beliefs we can have regarding facts
- Logics are useful for the commitments they do not make, because the lack of commitment gives the knowledge base writer more freedom
- Propositional logic commits only to the existence of facts that may or may not be the case in the world being represented. It has a simple syntax and semantics, but suffices to illustrate the process of inference.
- Propositional logic can accomodate certain inferences needed by a logical agent, but quickly becomes impractical for even very small worlds.

Questions

What if we expand our knowledge base from our initial knowledge using entailment, but then the premises change. Does it invalidate all the entailed knowledge?

E.g. we instantiate our agent with some knowledge about the world in its KB, like "the earth is spherical", "pi = 3.141592", "g = 9.81", and from this knowledge the agent can deduce other things. What happens if we then discover that the earth is flat, or g = 10? This would invalidate all the new knowledge that the agent has.

What does it do? Does it keep a track of where it gained its knowledge from, and delete all knowledge that stems from the incorrect fact? This could get confusing if you deduce facts from your existing knowledge, then use those facts to deduce more information, and so on and so on, until you have a sprawling KB based on a faulty premise.

Planning

18 December 2020 17:12

1. Knowledge bases
2. Reasoning using knowledge and inference
3. Search vs. Planning
4. Partial-order Planning
5. Conditional Planning
6. Monitoring and Replanning

Search Vs. Planning

In most situations, the branch factor of the search problem is so large that search becomes virtually useless.

General search problems..

1. In search, we must specify an initial state, operators, and optionally a heuristic function
2. Branching factor may be huge depending on how we specify these operators
3. Path length may be very long, and thus there are too many states to consider
4. The agent is forced to construct a full sequence of actions and must decide what to do in initial state first.

Difficulties with heuristics...

1. Heuristics can only choose which state is closer to a goal, but cannot eliminate actions from consideration
2. Evaluation function ranks these guesses, but must still consider them all
3. We need to work on the appropriate part of the sequence.

The main problem with basic problem solving agents is that they consider actions in sequence, starting from the initial state. Until the agent has worked out HOW to obtain the items we're looking for, it cannot really decide where to go. The agent therefore needs a more flexible way of structuring its deliberations so that it can in a non-linear fashion.

3 Key Ideas Behind Planning

The **first key idea** in planning is that we 'open up' the representation of states, goals and actions.

1. Planning algorithms use descriptions in some formal language - usually first-order logic
2. States and goals are represented by sets of sentences
3. Actions are represented by logical descriptions of preconditions and effects

This allows the planner to make direct connections between states and actions.
e.g. if the agent knows that the goal is a conjunction that includes have(milk), and it knows that buy(x) achieves have(x) then the agent knows that it is worthwhile to consider a plan that includes buy(milk). It need not consider other irrelevant actions such as buy(orange)

The **second key idea** is that the planner is free to add actions to the plan wherever they are needed, rather than in an incremental sequence starting at the initial state. For example, the agent may decide it needs to buy(milk) even before it has decided how to do such a thing.

There is no necessary *connection between the order of planning and the order of execution*.

The **third key idea** is that most parts of the world are independent of most other parts, and thus it makes it feasible to take a conjunctive goal and solve it using a divide-and-conquer strategy. A subplan involving going to the supermarket can be used to achieve the first two conjuncts and another subplan involving going to the hardware store can be used to achieve the third. The

supermarket subplan can be further divided into a milk subplan and a bananas subplan. We can then put all the subplans together to solve the whole problem.

Planning Systems

Open up action, state and goal representations to allow selection - represent in first-order logic

- States and goals = sets of sentences
- actions = description of preconditions and effects

Planning systems allow a planner to make direct connections between states and actions

- We can divide-and-conquer by subgoaling
- Planner can consider several smaller easier problems, and then combine solutions
- Works because little interaction between subplans, otherwise cost of combining solution outweighs the gain e.g. no help for 8 puzzle to consider each tile separately
- Relax requirement for sequential construction of solutions
- Allows planner to add actions where needed, so can make obvious or important decisions first to reduce branching factor

There is no connection between the order of planning and execution. We can do this because of logic - At(Supermarket) represents a class of states, but search requires a complete state description, and so we could not do this.

In the real world, planning tends to do better than search

SPA Algorithm

- Update KB
- if not already executing a plan, generate a goal and construct a plan to achieve it
- Agent must be able to cope if the goal is infeasible or achieved (set action to NoOp)
- Once agent has a plan, it will execute to completion
- Minimal interaction with environment: perceive to determine initial state but then just execute plan - no relevance checks

Simple Planning Agent

```
function SIMPLE-PLANNING-AGENT(percept)
    returns an action
  static : KB, a knowledge base
    p, a plan, initially NoPlan
    t, a time counter, initially 0
  local G, a goal
    current, a current state description
  TELL(KB,MAKE-PERCEPT-SENTENCE(percept,t))
  current ← STATE-DESCRIPTION(KB,t)
  if p = NoPlan then
    G ← ASK(KB,MAKE-GOAL-QUERY(t))
    p ← IDEAL-PLANNER(current,G,KB)
  if p = NoPlan or p empty then action ← NoOp else
      action ← FIRST(p)
      p ← REST(p)
  TELL(KB,MAKE-ACTION-SENTENCE(action,t))
  t ← t + 1
  return action
```

Situation Calculus

- A way of describing change in first-order logic
- World viewed as a sequence of situations, snapshots of the state of the world
- Situations generated from previous situations by actions
- **Fluent**: Functions and predicates that change with time given a situation argument
- Those that do not change are called **eternal** or **atemporal**
- Change represented by function Result(action, situation) which denotes the result of performing action in situation

- *Possibility axioms* — Describes when it is possible to execute an action (Precondition $\implies$ Poss(a,s)) e.g. $At(Agent, x, s) \wedge Adjacent(x, y) \implies Poss(Go(x, y), s)$
- *Effect axioms* — changes due to action (Poss(a,s) $\implies$ changes), e.g. $Poss(Go(x, y), s) \implies At(Agent, y, Result(Go(x, y), s))$

Planning In Situation Calculus

- Planning can be seen as a logical inference problem using situation calculus
- Logical sentences to describe initial state, goal and operators
- Initial state - sentence about the situation - At home, no milk, no bananas etc
- Goal state - logical query for suitable situations - At home and have milk and have bananas etc

- *Given*
  - Initial State: $At(Agent, [1, 1], S_0) \wedge At(G, [1, 2], S_0) \wedge Gold(G)$
  - Possibility Axioms
    - $At(Agent, x, s) \wedge Adjacent(x, y) \implies Poss(Go(x, y), s)$
    - $Gold(g) \wedge At(agent, x, s) \wedge At(g, x, s) \implies Poss(Grab(g), s)$
  - Effect Axioms
    - $Poss(Go(x, y), s) \implies At(Agent, y, Result(Go(x, y), s))$
    - $Poss(Grab(g), s) \implies Holding(g, Result(Grab(g), s))$
- *Goal State*: $\exists seq,\ At(G, [1, 1], Result(seq, S_0))$
- From first Possibility Axiom, $Poss(Go(x, y), s)$
- From first Effect Axiom, $At(Agent, y, Result(Go(x, y), s))$
- So can Agent grab the gold?

Nothing in the KB base says that the location of the gold remains unchanged

Frame Axioms

These tell us how the non-changes due to action.
If some object is at some state, and is not the agent and is not being held by the agent, then the object is still in situation S when the agent moves.

If we have f fluents (things that can change) and a actions, it requires O(AF) frame axioms

The Frame Problem

- representational - proliferation of frame axioms (original frame problem)
- representation problem now largely solved
- inferential - having to carry properties through inference steps, even if remain unchanged
- inferential problem avoided by planning; we do not address it for inference systems.

We solve the representational frame problem with **successor state axioms**

- Each axiom is about a predicate (not an action per se)
- General form: p true afterwards = (an action made P true OR P true already and no action made p false)
- We need a successor state axiom for each predicate that can change over time
- Axiom must list all ways the predicate can become true or false

Practicality of Planning

With first order logic, predicates and situational axioms and calculus, we theoretically have all that is required - but this is unpractical (time, space, semi-decidability) etc.

Thus, we need a restricted language. This reduces the number of possible solutions to search through.

Actions represented in a restricted language, allows creation of efficient planning algorithms
So we need a language and a planning algorithm for that language.

STRIPS is a restricted language that lends itself to efficient planning algorithms, while retaining much of the expressiveness of situation calculus representations

Basic Representations for Planning

Representing States and Goals

States are represented by conjunctions of function-free ground literals - or predicates applied to constant symbols, possible negated.

These state descriptions do not need to be complete. An incomplete description, corresponds to a set of possible complete states for which the agent would like to obtain a successful plan.

Many systems adopt the negation as failure convention - if a state description does not mention a given positive literal
Remember - a goal given to a planner asks for a sequences of actions that makes the goal true if executed, while a query given to a theorem proves asks whether the query is true given the KB

State example: At(Home AND !Have(milk) AND !Have(bananas)
Goal example - we want to be at a shop that sells milk: At(x) AND Sells(x, Milk)

Representing Actions

STRIPS operators have 3 components

- **action description** - what an agent actually returns to the environment in order to do something
- **precondition** - conjunction of atoms (positive literals) that says what must be true before the operator can be applied
- **effect** of an operator is a conjunction of literals (positive or negative) that describes how the situation changes when the operator is applied

Op(ACTION:Go(There), PRECONDITION:At(here) AND Path(here, there), EFFECT:At(there) AND ! At(here))

An operator with variables is known as an operator schema, because it does not correspond to a single executable action but rather to a family of actions, one for each different instantiation of the variables.

We say that an operator is applicable in a state s if there is some way to instantiate the variables in o so that every one of the preconditions of o is true in s.

For example, if the initial state includes the literals At(home), path(home, supermarket) then the action Go(supermarket) is applicable and the resulting situation contains the literals !at(home), At(Supermarket), Path(Home, Supermarket)

Representation for Plans

If we're going to search through plan space, we need to be able to represent them. We can settle on a good representation for plans by considering partial plans for a simple problem: putting on a pair of shoes.

least-commitment - only make choices about things you care about currently, leaving the other choices to be worked out later. This is good for search since you are likely to make the wrong choice and have to backtrack later. A least commitment planner could leave the ordering of the two steps unspecified. This helps us avoid bad plans, by delaying the decision making process until we have more information, allowing the agent to make better choices.

A planner that can represent plans in which some steps are ordered with respect to each other and other steps are unordered is called **partial order planner.**
The alternative planner that plans with a simple list of steps, is a **total-order planner.** A totally ordered plan that is derived from a plan P by adding ordering constraints is called a linearization.

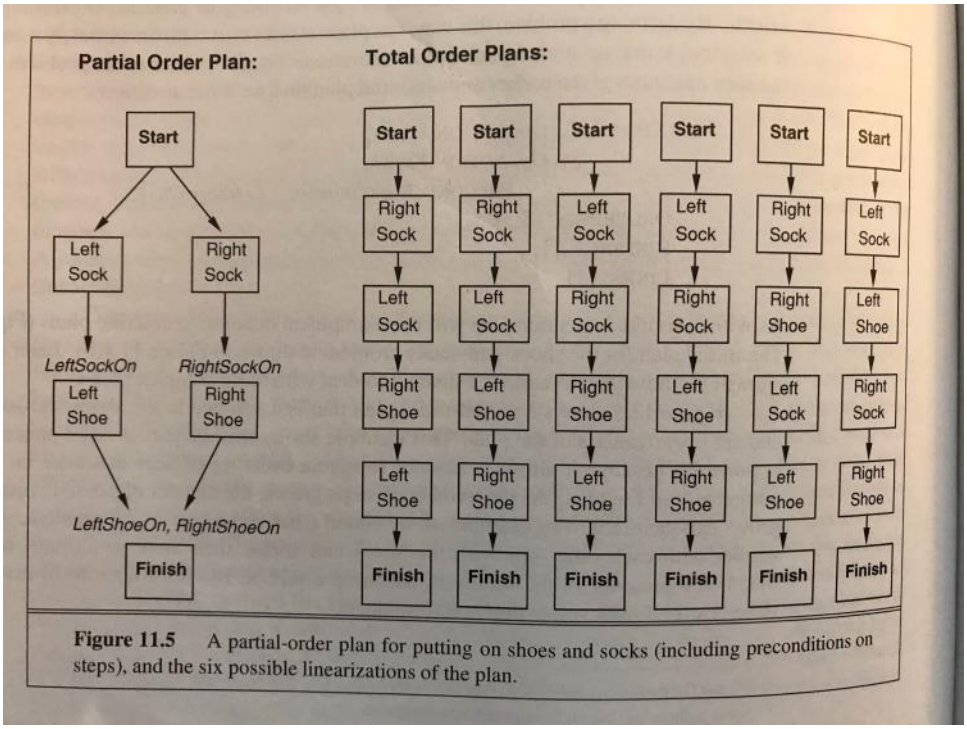


Figure 11.5 A partial-order plan for putting on shoes and socks (including preconditions on steps), and the six possible linearizations of the plan.

Situation Space vs Plan Space

There are a lot of situations that can occur in a world. A path through this space constitutes a plan for a problem. If we wanted, we could take a problem described in the STRIPS languages and solve it by starting at the initial state and applying operators one at a time until we reached a state that includes all the literals in the goal. We could use standard search methods for this

An algorithm that did this would be a problem solver, but it could also be considered a planner. The algorithm would operate in **situation space**. It would be a **progression planner** because it would search forward from the initial situation to the goal situation. Obviously, the branching factor for this method makes it problematic.

One way to cut the branching factor is to search backwards, from a goal state to the initial state. This is a **regression planner**. This approach is possible because the operators contain enough information to regress from a partial description of a result state to a partial description of the state before an operator is applied. We cannot get complete descriptions of states this way, but luckily we don't need to.

This regression approach is desirable because usually, the initial state has many applicable operators that could potentially be used, whereas to move back from the goal state there are typically only a few conjuncts, each of which will only have a few operators.

Searching backwards is hard because we have to achieve a *conjunction of goals*, rather than just one.

Alternatively, we can search through the **space of plans**, rather than the space of plans rather than the space of situations. That is, *we start with a simple, incomplete plan*, which we call a partial plan. Then we *consider ways of expanding the partial plan* until we come up with a complete plan that solves the problem . The operators in this search are operators on plans:

- Adding a step
- Imposing an ordering that puts one step before another
- Instantiating a previously unbound variable.

The solution is the final plan, and the path taken to achieve it is irrelevant.

Refinement operators on plans – take a partial plan and add a constraint. These eliminate plans from the set, but never add new ones
Modification operators – anything that's not a refinement operator is a modification operator. Some planners work by creating an incorrect plan then debugging it with modifications operators.

Plan

- A plan is a data structure consisting of:
- A set of plan steps - each step is one of the operators for the problem
- A set of step ordering constraints e.g. s must occur before x
- A set of variable binding constraints - in the form v = x, where v is a var in some step, and x is either a constant or another variable
- A set of causal links - s_i --c--> s_j, read as "si achieves c for sj". Causal links serve to record the purpose of steps in the plan: here a purpose of s_i is to achieve the precondition of s_j

The initial plan before any refinements take place, simple describes the unsolved problem. It has two steps - Start and Finish and the constraint that start is before finish. There are not links or bindings.

Plan Solution

A solution is a plan that an agent can execute, that guarantees achievement of the goal.
To check a plan is valid, it makes sense to insist on a fully instantiated, totally ordered plan.
However, this is unsatisfactory:

1. Agents can perform tasks in parallel so it makes sense to allow solutions with parallel actions
2. There are many linearisation's of a plan – it is more natural to just return the PO plan than arbitrarily pick a plan
3. If we're creating plans that will be combined with larger plans, it makes sense to maintain flexibility

Therefore, we accept plans that are partially ordered, and

- Complete – every precondition of very step is achieved by some other step – A step achieves a condition if the condition is one of the effects of the step, and if no other step can possible cancel out the condition.
- Consistent – a consistent plan is one in which there are no constrictions in the ordering or binding constraints. A contradiction occurs when both Si must be before Sj and Sj must be before Si – remember plans are transitive.

Partial-Order Planning

31 December 2020 18:53

Quick Recap

You should understand:

1. The idea of an agent that can reason
2. The idea of a KB and an agent performing inference
3. How we adapt that into something that makes plans
4. How plans can do things differently to search
5. Why we might prefer planning over search
6. What the issues are with planning
7. Situation calculus and what it's used for
8. STRIPS and what it can do

Closed-world assumption - most planners assume that if state descriptions do not mention a positive literal, we can assume it to be false - this can be dangerous
Goals - conjunctions of literals, may contain variables
Planner - a system you can ask for a sequence of actions that make the goal true if executed

Operators comprise three components

- **action** e.g. buy(x) - buy an x
- **precondition** e.g. At(p), Sells(p, x) - we're at p and p sells x
- **effect** e.g. Have(x) - thus we now have x

The goal is to gradually move from incomplete and vague plan to complete and correct plans

Key Terminology

- **Partial plan** – an incomplete plan that we consider ways of expanding, until we come up with a plan that solves the problem
- **Operator schema** – an operator that takes variables – basically a function.
- **Least commitment** – one should only make choices about things that you currently care about, leaving other choices to be worked out later
- **Partial order** – A planner that can represent plans in which some steps are ordered with respect to each other, and some are unordered, is called a partial order planner.
- **Total order** – a plan that has all the steps in some order – there is no ambiguity about when the steps will take place.
- **Linearisation** – a totally ordered plan that is derived from a partially ordered plan is called a linearisation of the plan.
- **Fully instantiated plans** – a plan in which every variable is bound to a constant
- **Causal links** – a causal link is written as "Si achieves c for Sj". They serve to record the purpose of steps in the plan – here the purpose of Si is to achieve the precondition c of Sj

Plan

A plan is formally a data structure consisting of;

1. A set of plan steps – each step is an operator to the problem
2. A set of step ordering constraints – each ordering constraint is in the form Step I before Step j, meaning that step I has to take place before step j, although not necessarily directly before.
3. A set of variable binding constraints – each variable constraint is of the form v = x where v is a variable in some step, and x is either a constant or another variable
4. A set of causal links – as described above

Outline of POP

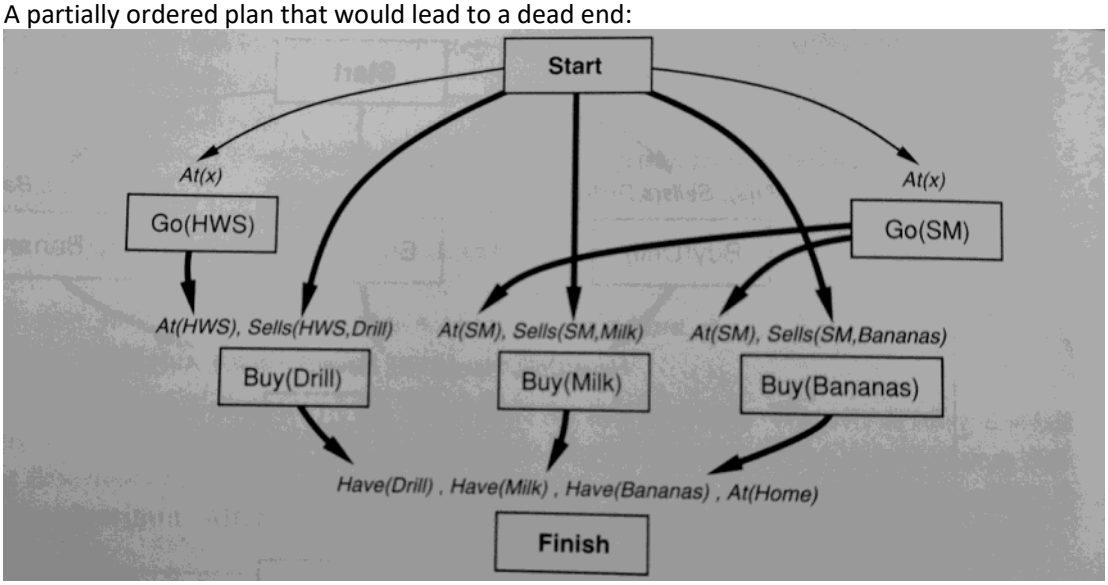
We sketch an outline for a partial-order regression planner that searches through plan space.

The planner starts with an initial plan representing the start and finish steps, and on each iteration, adds one more step. If this leads to an inconsistent plan, it backtracks and tries another branch of the search space

To keep the search focused, the planner only considers adding steps that serve to achieve a precondition that has not yet been achieved.

If you have two conditions that contradict each other – e.g. go(HardwareStore) and go(superMarket) which both require at(Home), then you reach a dead end. There is no way to go(superMarket) if we're at(hardwareStore) if the precondition for go(superMarket) is at(home). Thus, we have a flawed plan.

A partially ordered plan that would lead to a dead end:



Interestingly, the planner could notice this partial plan is flawed without wasting a lot of time. The key is that the causal links in a partial plan are protected links. A causal link is protected by ensuring that threats (steps that might delete or clobber the protected condition) are ordered to come before or after the new step.

Clobbering

If we place the clobbering step that threatens a condition prior to the step, then we call it a demotion
If we place the step that threatens the condition c after the causal link that protects that condition, we call it a promotion

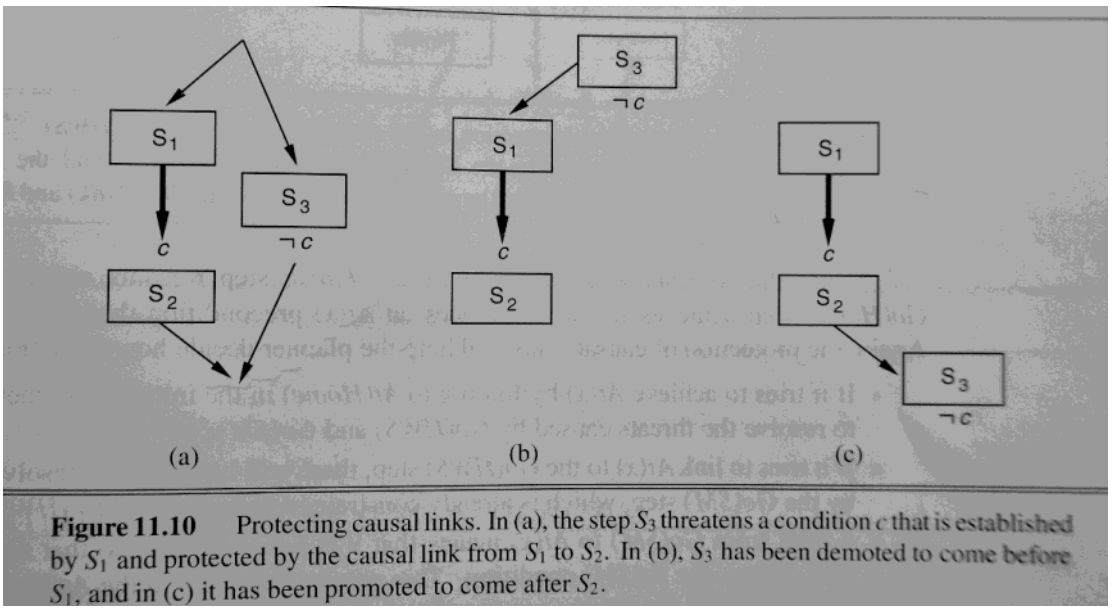


Figure 11.10 Protecting causal links. In (a), the step S_3 threatens a condition c that is established by S_1 and protected by the causal link from S_1 to S_2. In (b), S_3 has been demoted to come before S_1, and in (c) it has been promoted to come after S_2.

Here we have an example of causal link protection in a shopping plan:

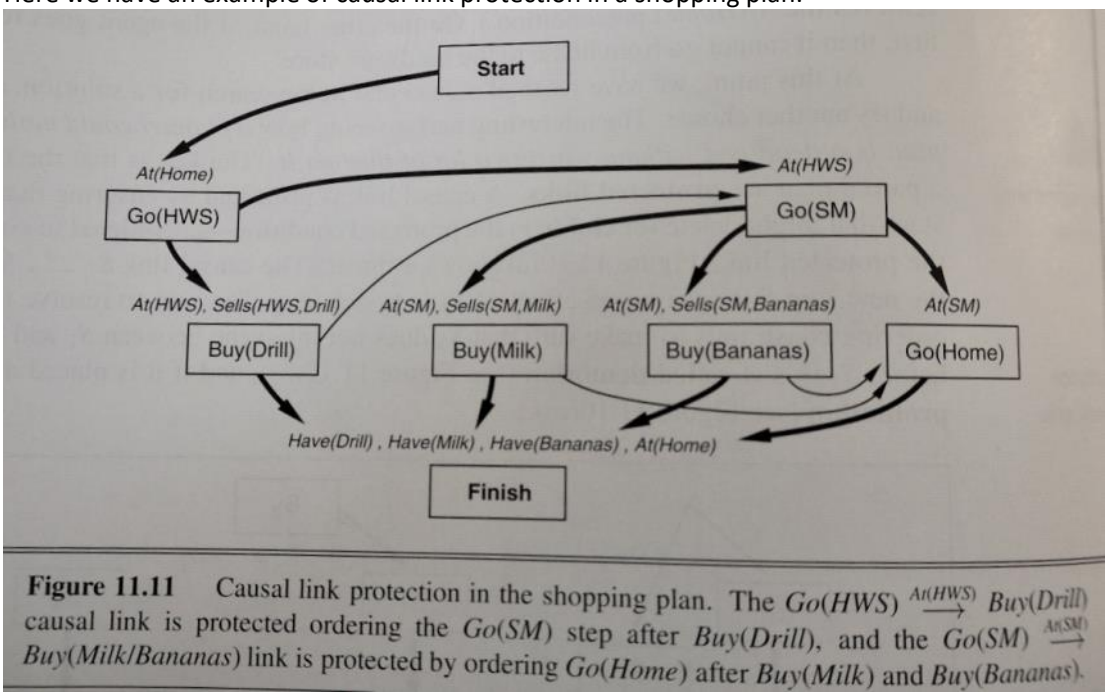
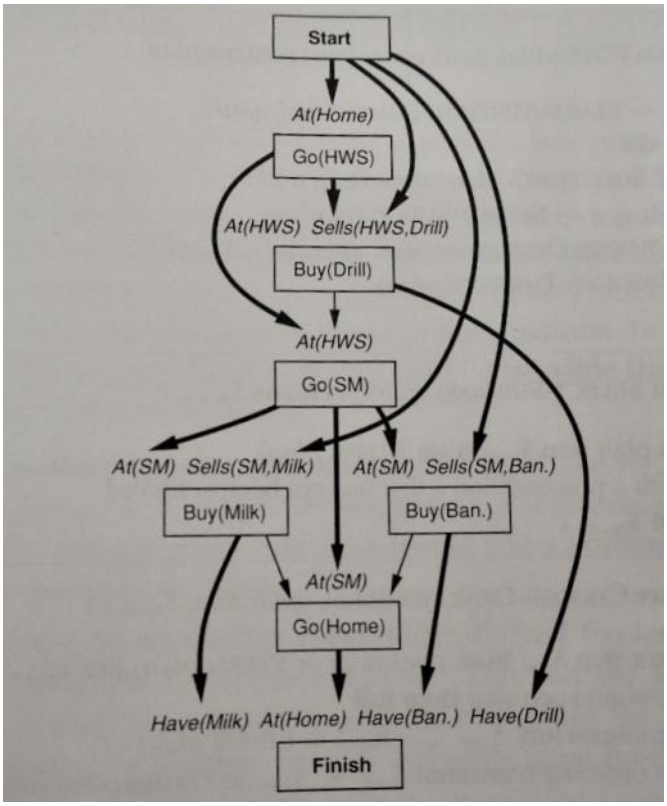


Figure 11.11 Causal link protection in the shopping plan. The $Go(HWS) \xrightarrow{At(HWS)} Buy(Drill)$ causal link is protected ordering the $Go(SM)$ step after $Buy(Drill)$, and the $Go(SM) \xrightarrow{At(SM)} Buy(Milk/Bananas)$ link is protected by ordering $Go(Home)$ after $Buy(Milk)$ and $Buy(Bananas)$.

This is what a finished solution to the shopping problem might look like. Note that the result is almost a totally ordered plan, the only ambiguity being that Buy(Milk) and Buy(bananas) can come in any order.



Our partial order planner can now take a problem that would require thousands of search states for a problem-solving approach, and solve it with only a few search states. Moreover, the least commitment nature of the planner means it only needs to search at all in the places where subplans interact with each other. Finally, the causal links allow the planner to recognise when to abandon a doomed plan without wasting a lot of time expanding irrelevant parts of the plan.

POP Algorithm

```
function POP(initial,goal,operators)
    returns a plan
  plan ← MAKE-MINIMAL-PLAN(initial,goal)
  loop do
    if SOLUTION?(plan) then return plan
    S_need, c ← SELECT-SUBGOAL(plan)
    CHOOSE-OPERATOR(plan,operators,S_need,c )
    RESOLVE-THREATS(plan)
  end
```

```
function SELECT-SUBGOAL(plan)
    returns plan step and precondition (S_need, c)
  pick step S_need from STEPS(plan)
    with precondition c that has not been achieved
  return S_need, c
```

```
function CHOOSE-OPERATOR(plan,operators, S_need,c)
  pick step S_add from operators or STEPS(plan)
      that has effect c
  if no such step then fail
  add causal link S_add --c--> S_need to LINKS(plan)
  add ordering constraint S_add ≺ S_need to ORDERINGS(plan)
  if S_add newly added step from operators then
    add S_add to STEPS(plan)
    add Start ≺ S_add ≺ Finish to ORDERINGS(plan)
```

```
procedure RESOLVE-THREATS(plan)
  for each S_threat that threatens a link
      S_i --c--> S_j in LINKS(plan) do
    choose either
      Demotion: Add S_threat ≺ S_i to ORDERINGS(plan)
      Promotion: Add S_j ≺ S_threat to ORDERINGS(plan)
  if not CONSISTENT(plan) then fail
```

POP starts with a minimal partial plan, and at each step extends the plan by achieving a precondition c of a step S_need.

It does this by *choosing some operator*, either from the existing steps of the plan or form the pool of operators, that achieves the precondition. *It records the causal link for the newly achieved precondition, and then resolves any threats to causal links.*

The new step may *threaten an existing causal link* or an existing step may threaten the new causal link. If at any point the algorithm fails to find a relevant operator or resolve a threat, it backtracks to a previous choice point. An important subtlety is that the selection of a step and precondition in select-subgoal is not a candidate for backtracking.

The reason is that every precondition needs to be considered eventually, and the handling of preconditions is commutative: Handling c1 and then c2 leads to exactly the same set of possible plans as handling c2 and then c1. So we can just pick a precondition and move ahead without worrying about backtracking. The pick we make effects only the speed, and not the possibility of finding a solution.

Notice that we start at the goal and move backwards until we have a solution. Thus, POP is a sound and complete regression planner. Every plan it returns is a solution.

Problems with POP and STRIPS

STRIPS cannot express:

Hierarchical plans

- we often want to specify plans at different levels of detail.
- We want to have several levels before reaching executable actions.
- This makes computation manageable and resulting plan understandable – analogous to high level and low level machine code.
- This allows human specification of abstract partial plans to guide the planner because we often want to give the planner guidance.

Complex conditions

- Operations have variables, but we have no quantification
- STRIPS use of variables is limited – we cannot express that pick(bag) also causes all objects in the bag to be lifted
- Operators are unconditional - we cannot express actions having different effects according to conditions

Solution – conditional effects: avoids premature commitment – have effect when condition

Time

- In situation calculus, time is discrete, and actions occur instantaneously
- We need to represent that actions take time, may only be applicable at certain times, and that the goal may have a deadline

Resources

- Real problems have limited resources – financial time, quantity of materials, machinery available
- Actions have a cost that we need a way to represent
- Actions descriptions need to represent the requirements of performing the action
- Planner needs to take cost into consideration

Solution: extend STRIPS and modify POP accordingly

1. **Hierarchical decomposition** can be added to STRIPS in the form of non-primitive operators. We can then modify the planner to replace non-primitives with decomposition. We do this by adding abstract operators
2. **Abstract operators** that can be decomposed into steps
3. Decompositions are predetermined and stored in a library of plans – works best when there are several possible decompositions

Broadening Operator Descriptions

We want to make operator descriptions more expressive to widen the applicability of POP

1. Conditional effects – avoids premature commitment - e.g. have effect WHEN
2. Allow negated goals – ability to call CHOOSE-OPERATOR with !p instead of p
3. Disjunctive preconditions – allow SELECT-SUBGOAL to make a nondeterministic choice between disjuncts - use principle of least commitment
4. Disjunctive effects – introduces nondeterminism – may be able to address with coercion
5. Universally quantified preconditions – instead of clear(b) we can do "for all blocks, remove blocks from b
6. Resource constraints – add numeric-value measures such as money(200) or fuel(20) - a measure fluent
7. Temporal constraints – we can treat time as a resource – remember we can do things concurrently so time is the max of the jobs not the sum

With resource constraints, we want to plan for scarce resources first, delaying choice of causal links where possible. We can check for failure then without a finished plan, so we don't have to try all operators.

Conditional Planning

So far we have assumed that the world is fully observable, static and deterministic – we also assumed that action descriptions are correct and complete. Unfortunately, the real world is not like this – typically have to deal with both incomplete and incorrect information.

In the real world we have incomplete information – we don't know the preconditions of an operator, and we don't know the effects of an operator e.g. inflate(tire) might cause inflated(tire), slowhiss(tire), burst(tire), breakPump(tire) etc

We can also have incorrect information. We might think we have a spare tire, then find we don't.

We can never finish listing all the required preconditions and possible conditional outcomes of actions in the real world

Conditional Planning

- Plan to obtain information
- Subplan for each contingency that could happen
- Expensive because it plans for many unlikely cases

Replanning

Action Monitoring:

- Preconditions of next action note met

Plan monitoring:

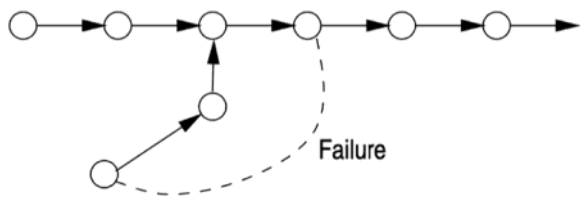
- Effects of an action might not be as predicted
- Failure = preconditions of remaining plan are not met
- Preconditions = causal links at current time
- Check current preconditions with perceived state

In each case, we fail and need to replan.

There are 2 types of plan failure –

1. **bounded indeterminacy**: when unexpected effects of actions can be enumerated – we can use conditional planning to deal with this
2. **Unbounded indeterminacy**: set of possible outcomes is too large to enumerate – we can plan for (at most) a limited number of contingencies and must replace when things go wrong – e.g. driving

Replanning



A naive approach would be to replan from scratch, but this is inefficient and we might get stuck in a cycle of a bad plan

A better approach would be to plan to get back on track by connecting the old plan as show above
Ideally we introduce learning so the agent doesn't loop forever

Alternative Solutions to Replanning

Coercion – force the world into a particular state to reduce uncertainty

Abstraction – ignore details of a problem that may be unknown
Aggregation – Treat a large number of objects which have individual uncertainty as a single, aggregate practicable object

Reactive Planning

Abandon domain-independent planning and use domain specific knowledge - the agent has procedural knowledge – library of partial plans representing a collection of behaviours

Knowledge Representation

06 January 2021 10:41

Objectives:

- Defining knowledge and reasoning
- Expert systems
- Knowledge as rules
- Rule based systems
- Forward and backward chaining
- Conflict resolution
- Assumption-based reasoning

Defining Knowledge & Reasoning

What is knowledge?

A *relation defined by the propositional attitude between a knower and a proposition* (a simple declarative sentence)
e.g. John knows that Abraham Lincoln was president, or John hopes to see a shooting star tonight

What matters is whether the proposition is true or false - this is what defines the state of the world according to an agent

Not all knowledge is propositional - e.g. "Jon knows Mary well"

Reasoning

- Explicitly representing all propositions believed to be true is difficult
- Reasoning bridges the gap between what is represented (known true propositions) and what is believed by an agent
- Let KB be a set of propositions believed to be true and alpha be a proposition not in KB
- alpha is said to be logically entailed by KB, KB | = alpha is we believe alpha to be implicitly true given the propositions in the KB
- Knowledge representation language need to have a well-defined notion of entailment
    - What does it mean for a proposition to be true or false?
    - Consider what else we can decide is true or false based on that knowledge

The expressiveness of a representation language has a direct impact on the computational complexity of the reasoning process

Can we compute all the propositions that are entailed by the KB? (logically complete)
Can we guarantee that any proposition believed to be true as a result of reasoning is indeed true? (logically sound)

Knowledge Representation Hypothesis

Any mechanically embodied intelligent process will be comprised of structural ingredients that

- we as external observers naturally take to represent a propositional account of knowledge that the overall process exhibits
- independent of such external semantic attribution, play a formal but causal and essential role in engendering the behaviour that manifests that knowledge

In other words, a process contains a collection of propositions which it believes to be true, and reasons with the propositions during its operation

Propositions

Typically it is more natural to represent knowledge as a logical formulae rather than a table of information

- with formulae, it is easier to check correctness and debug formulae
- can incrementally add to a formulae easily
- can extend with infinitely many variables and domains

For efficient reasoning, we can exploit the Boolean nature of such formulae

A proposition is a statement that is true or false, which naturally can be represented as a logical formulae

Semantics

When creating a KB, we must choose the variables that will be used to build propositions
These should have meaning to the KB designer

We then give the system knowledge about the domain and can then make enquiries

- knowledge will take the form of a clause, h <- b which consists of two parts
- the body, b is a logical formulae that can evaluate to true or false
- The head, h is a variable that can be derived to be true as a result of b being true

Notably, the system doesn't understand the meaning of the symbols

View of Semantics

Users view

- the user must decide the task domain - intended interpretation
- A variable must be associated with each proposition you want to represent
- tell the system clauses that are true in the intended interpretation - this is known as axiomatizing the domain
- if KB |= alpha then alpha must be true in the intended interpretation
- Users interpret the systems responses using the intended interpretation of the symbols

Computers View

- the system does not have access to the intended interpretation
- It is only aware of the KB
- The system can determine if a particular formula is a logical consequence of KB, but does not understand what that formula really means

Summary of Knowledge Representation

- We talked about defining knowledge and reasoning
- We now know what it means for a system to reason about something
- There is a disconnect between the semantic meaning we give to symbols, and how the computer interprets the symbols

Expert Systems

A computer system with a KB that takes some knowledge and reasons about it, and provides advice, responses and courses of action

- Simulates human reasoning

- Performs reasoning over a representation of human knowledge
- Solves problems with heuristic or approximate methods

Who is an Expert?

- Process knowledge focussed on a specific domain
- Capable of solving problems
- Capable of explaining how they solve problems

Example - The MYCIN System

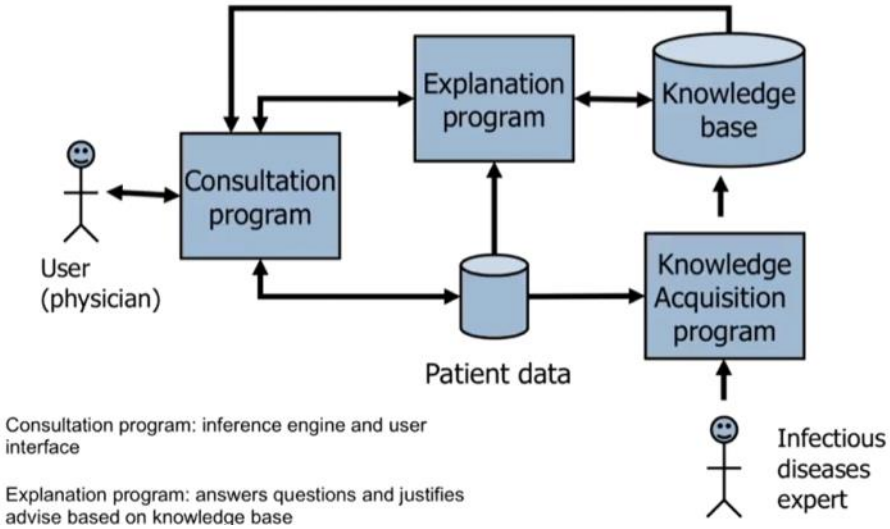
Developed by Stanford

Provides advice to a physician on selection of antibiotics for treating infections

- generates hypotheses and weights them based on evidence provided
- Makes therapy recommendations

The Problem

Blood infections therapy
Therapy process - identify organism involved
choose the most appropriate drug or combination of drugs



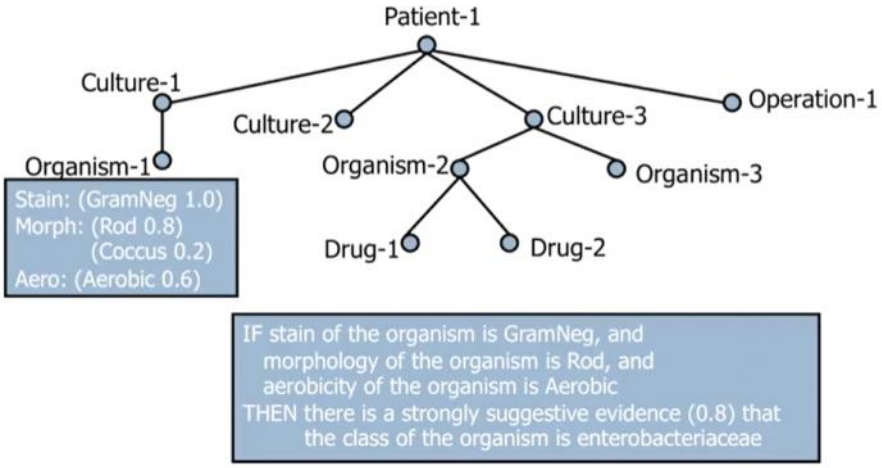
MYCIN Knowledge Representation

- KB contains rules in the form of *if condition and condition 2 ...*
- The rule tally states how certain the conclusion is, given that the conditions are satisfied
- The certainty associated with a conclusion is a function of the combined certainties of the conditions are the rule tally

Also stored in the KB

- Lists such as a list of all known organisms
- Knowledge tables containing clinical data
- A classification of clinical parameters according to the context in which they apply - e.g. are they attributes of patient or organism?

MYCIN Patient Data



MYCIN Control Structure

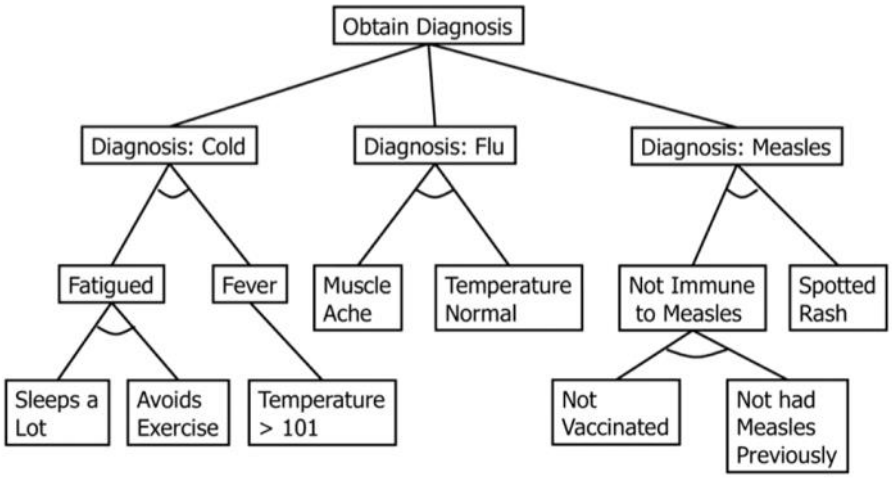
Consultation is a search through a tree of goals

- child nodes in the tree are sub goals that must be achieved to achieve the goal represented by the parent node

MYCIN top level goal specified as the following rule

- If there is an organism which requires therapy, and consideration has been given to any other organism requiring therapy
- THEN compile a list of possible therapies and determine the best one in the list
- The action part of this rule is the root node of the tree of goals

Leaves of the 3 are facts such as lab data - satisfy this goal by asking the user



Rule Based Systems

Rules as Knowledge

Information can be stored as **propositions**

People tend to associate intelligent behaviour with regularities in behaviour - rational people act consistently

Production rules are a formalism that has been used in automata theory, formal grammars and the design of programming languages

In expert systems, this is referred to as a **condition action rule** or situation action rule

Canonical Systems

- Formal system based on:
  - An alphabet, A, for making strings
  - Some strings that are taken as axioms
  - A set of productions of the form:
    - $\alpha_1\$_1 \ldots \alpha_m\$_m \rightarrow \beta_1\$'_1 \ldots \beta_n\$'_n$
    - Grammar rules for manipulating strings of symbols
    - Known as rewrite rules
    - Should look slightly familiar
- Example:
  - A = a,b,c
  - Axioms: a,b,c,aa,bb,cc
  - Productions:
    - $\$ \rightarrow a\$a$
    - $\$ \rightarrow b\$b$
    - $\$ \rightarrow c\$c$
  - Generates all and only the palindromes based on the alphabet A, through application of the productions

Knowledge Representation

- Alphabet replaced by a Vocabulary that consists of:
  - A set O of names of objects in the domain
  - A set A of attributes of the objects
  - A set V of values that these attributes can take
- Grammar for generating symbol structures
  - object-attribute-value triples
  - (o,a,v), o∈O, a∈A and v∈V
  - Example
    - (ORGANISM-1,morphology,rod)
    - (ORGANISM-1 (morphology rod) (aerobicity aerobic))

Working Memory

- A store of facts (assertions/propositions)
  - Define the initial state of the KB/model
  - Rules define operators allowing transitions from one state to another
- Each fact is referred to as a working memory element
- Described using the vocabulary and grammar of the system
- Can be interpreted as an existential sentence in FOL
  - (student (name john) (department computerScience))
  - $\exists x[student(x) \wedge (name(x) = john) \wedge (department(x) = computerScience)]$

Key Points

- Facts are working memory elements - we describe them with the grammar of the system
- We can describe the fact with an existential sentence in First Order Logic

A Production Rule

if P_1 *and* P_2 ... *and* P_m *are TRUE*
then perform actions Q_1 *and* Q_2 ... *and* Q_n

- Two-part structure
  - An antecedent set of conditions (the *if* part of the rule)
    - A *condition* is represented by by an object-attribute-specification vector
    - *type* $attribute_1$:$specification_1$ ... $attribute_k$:$specification_k$
    - The set of conditions is interpreted *conjunctively*
    - A condition (that is not negated) must match a WME
    - Matching implies the type is identical
    - Each attribute-specification pair in the condition has a corresponding attribute-value pair in the WME, where the value matches the specification
    - If there is a WME for each condition, the consequent action will be performed
  - A consequent set of actions (the *then* part of the rule)
    - Actions operate (add/delete/modify facts) on working memory

CLIPS Syntax

- Assume the Working Memory contains the following facts:
  - (patient (name Jones) (organism organism-1))
  - (organism (name organism-1) (morphology rod) (aerobicity aerobic))
- If KB contains the rule:
  - (defrule diagnosis
    (patient (name ?) (organism ?org))
    (organism (name ?org) (morphology rod) (aerobicity aerobic))
    =>
    (assert
    (organism (name ?org) (identity enterobacteriaceae) (confidence 0.8)))
- Then:
  - Match WME of type patient with first condition
  - Bind variable org, if not already bound, to organism-1
  - Match WME of type organism
- Rule fires to assert the new fact (consequent) in Working Memory
- organism (name organism-1) (identity enterobacteriaceae) (confidence 0.8)

The Rule Interpreter

Recognise-act cycle:

- Match the antecedent condition of rules against elements in the working memory
- if more than one rule antecedent matches ("can fire") choose one of the rules based on some conflict resolution strategy
- Apply the rule (add and delete elements of working memory)
- Back to Match

Cycle halts when no rules become active or if action of rule fired is to halt

Controlling Inference Behaviour

- Global control - domain independent, hard coded within the inference engine
- Local Control - domain dependent, coded in the form of meta rules - reason about which rule to fire rather than about objects in the domain

Summary of Rule-Based Systems

We have set up our rule based systems with a working memory
we can build rules that allow us to reason
we know the process that we use to work out what conclusions can be drawn from KB (antecedence)

Forward & Backward Chaining

- When facts match a rules condition, the rule fires
- Producing a solution from a KB is akin to a logical proof
- We are demonstrating that something logically follows from the initial state of the system
- There is no way for information to get added that isn't a consequence of the existing data
- We call the series of rules that we first the inference chain
- There are two main ways to build a chain - forward chaining and backward chaining

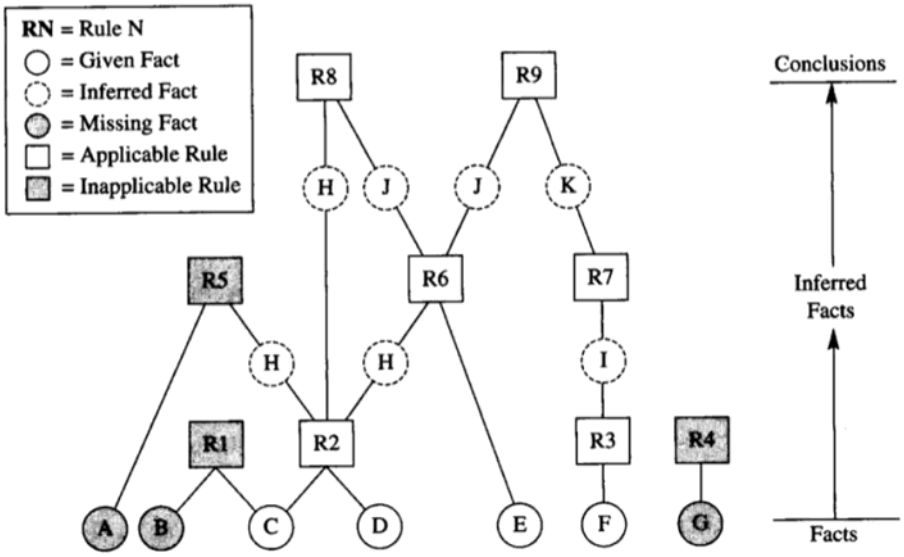
Forward Chaining

- Data driven - start with known data (facts)
- facts that can be inferred will be inferred even if they are not related to the goal

bottom-up ground proof procedure

Is a simple procedure of matching production rules that can be fired

- Select rules that produce new WME for the KB
- This is repeated until no rules can fire, then we can check to see if the desired solution is now in the KB
- *Is both logically sound and complete*
- Can spend a lot of time processing rules that do not contribute to the goal



Backward Chaining

Goal driven - system has a goal and the inference engine attempts to find the evidence to prove it
Only use data which is needed

top-down definite clause proof procedure

- Start from the query and work backwards to determine if it is a logical consequence of the KB
- Our query will be a clause made up of a number of WME, that we want the KB to contain

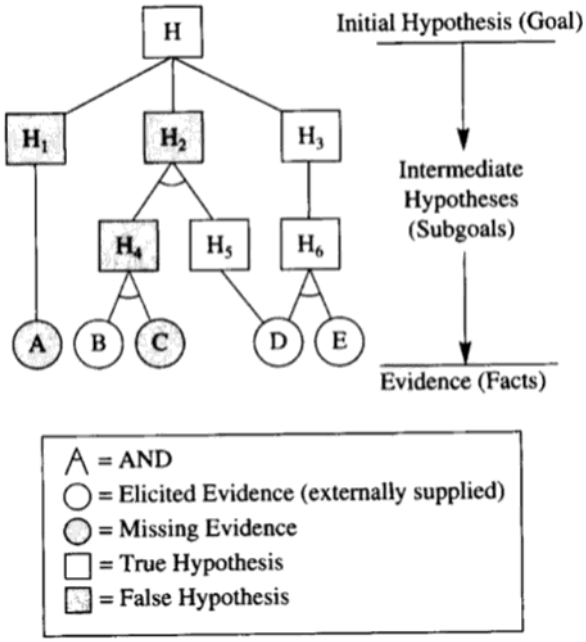
- We select an element of the clause and find a production rule that results in that element being added to the KB
- We replace the element with the condition of the production rule
- We repeat this until the query clause is entirely made up of elements that already exist in the KB

Backwards chaining is non-deterministic, based on the choice of production rules

Can stop early if one of the elements of the query cannot be derived

Because our queries are made up of conjunctives, if one element cannot be derived from the KB, *then the whole query cannot be derived*

However, when choosing the production rule that derives the element, we may also hit a dead end
- this does not mean we have to stop, merely we must select a new production rule
- we can only definitively say an element cannot be derived if we have explored all the clauses



Askable-clauses
- The only way to receive new information is from the user
- It can be tedious to have the user input all the information they know
- Instead ,we can define some clauses as askable, meaning it is information that KB can ask the user about

Conflict Resolution
non-syntactic errors in rule based systems:
- ❖ an incorrect answer is produce - a clause that should be false has been interpreted to be true
- ❖ an answer was not produced
- ❖ a program gets stuck in an infinite loop
- ❖ the system asks irrelevant questions - this requires investigation and assessment of the KB

Debugging Incorrect Answers
- suppose some clause / variable was proved false in the intended interpretation
- there must be some rule in the KB that was used to prove that clause

- either one of the variables in the rule is false in the intended interpretation - we can debug this by asking the user
- all of the variables are true in the intended interpretation - the rule itself is wrong and should be reassessed

Missing Answers

If a variable is true in the intended interpretation but could not be proved, then either

- there is no appropriate rule for the variable
- there is a rule and it did not fire when it should have
- this means one of the variables of the condition should be true but isn't
- this means we can solve this recursively, finding all the variables that should be true but do not have a rule

Infinite Loops

- A KB can get stuck in an infinite loop if there are rules that are cyclical - if you converted the KB into a directed graph, we will see a cycle
- A cyclical nature is symptomatic of another bug
- Rules should be reassessed to help ensure an acyclic KB
- Forward chaining cannot get stuck in infinite loops it only fires if it will add new elements to the working energy

Conflict Resolution

The firing of a rule may affect the activation of other ules as it affects the facts stored in the database

The method for choosing which rule to fire when more than one rule can be fired in a given inference cycle is called conflict resolution

Basic Conflict Resolution

- Fire rules in order of appearance in the KB - rules order will have a strong impact on the outcome - implicit knowledge
- Rule priority - make explicit the order in which the rules may fire

Specificity

- Fire the most specific rule
- More conditions means that it's more relevant to the current situation

Recency

- Fire the rule that use the data most recently entered into the working memory

Refractoriness

- A rule is only allowed to fire once on the same data
- Prevents loops in inference

Meta Knowledge

- Knowledge about knowledge
- Concerns the use and control of domain knowledge in an expert system
- Represented in the form of metarules
- Determine the strategy for the use of task specific rules in the expert system
- May be domain independent but more likely to be domain dependent

Examples

- Domain independent meta rule

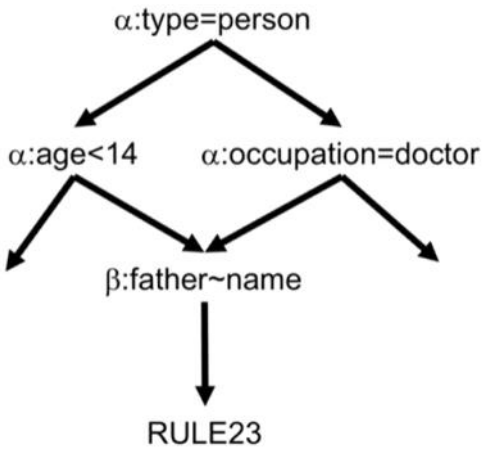
  IF (1) there are rules which do not mention the current goal in the premise
  and (2) there are rules which mention the current goal in their premise
  THEN it is definite (1.0) that the former should be done before the later

- Domain specific

  IF (1) the infection is a pelvic-abscess
  and (2) there are rules which mention in their premise enterobacteriaceae
  and (3) there are rules which mention in their premise gram-positive rods
  THEN there is suggestive evidence (0.4) that the former should be done before the later

Efficient Rule Matching

- Observations
  - The working memory is only modified very slightly in each recognize-act cycle
  - Many rules share conditions
- RETE algorithm
  - Create a network from rule antecedents (offline)
    - Two types of nodes
    - α node: represents simple, self-contained tests
    - β node: variables create constraints between different conditions
  - Example:
    - Rule23:
    - If (person name:x age:(¡14) father:y) (person name:y occupation:doctor) THEN ...



RETE

During operation:

- Tokens representing new or changed WME's are passed through the network
- Tokens that make it through the network satisfy the rule
- If a token cannot move through the network it is reassessed when the corresponding WME is modified

Rule-Based Systems: Pros and Cons

- Easy mapping between expert humans and Boolean rules
- Each rule represents an independent piece of knowledge
- Separation of knowledge from the processing/control structure
- Ability to represent and reason with uncertain knowledge
- Exhaustive search through all rules during each inference cycle
- The Knowledge Acquisition Bottleneck
  - No independent learning
- Brittle

Assumption-Based Reasoning

Often we want agents to make assumptions rather than doing deduction from their knowledge
Abduction - an agent makes assumptions to explain observations

- for examples, it hypothesizes what could go wrong with a system to produce the observed symptoms
- Default reasoning - an agent makes assumptions of normality to make predictions

- for example, a delivery robot may want to assume Mary is in her office even if it is not true

Design & Recognition

Two different tasks use assumption-based reasoning

- **Design** - aim to design an artefact or plan
- **Recognition** - aim to find out what is true

Assumption-Based Framework

- A set of closed formula, F, called the facts
- A set H, called the possible hypothesis or assumable

Bayesian AI

07 January 2021 10:42

Objectives:

- Bayesian Probability
- Inference using joint probability distributions
- introduction to Bayesian Belief Networks
    - entailed independence relations
- Inference in Bayesian Belief Networks
    - exact inference
- Making decisions with outcome probabilities
    - decision trees
    - influence diagrams

Reading: AIFCA 8.1-8.4 & 9.1-9.3, AIAMA: Ch. 13 & 14

Probability Recap

A problem with FOL is that agents never have access to the whole truth about their environment. In almost every case, even in simple worlds, there will be important questions to which the agent cannot find a categorical answer. The agent has to act under *uncertainty*.

- At best, our agent can only provide a **degree of belief**
- The tool we use for dealing with degrees of belief is **probability theory**.
- This provides a way of summarising the uncertainty that comes from our laziness and ignorance.
- Degree of truth, as opposed to degree of belief, is the subject of fuzzy logic, which is covered later.

Just as entailment status can change when more sentences are added to the KB, probabilities can change when more evidence is acquired.

All probability statements must therefore indicate the evidence with respect to which the probability is being assessed. As the agent receives new percepts its probability assessments are updated to reflect the new evidence.

- Before the evidence is obtained, we talk about **prior** or **unconditional probability**
- After evidence is obtained, we talk about **posterior** or **conditional probability**

Uncertainty and Rationality

To make choices between options, an agent must first have preferences between possible outcomes of various plans.

We will be using **utility theory** to represent and reason with preferences. The term utility is used in the sense of "*the quality of being useful*".
Preferences as expressed by utilities, are combined with probabilities in the general theory of rational decisions called decision theory.

- **Utility theory -** every state has a degree of usefulness to an agent, and the agent will prefer states with higher utility.

- **Decision theory = probability theory + utility theory**

An agent is rational if and only if it choose the action that yields the highest expected utility, average over all the possible outcomes of the action - This is called the principle of **Maximum Expected Utility** (MEU).

Probabilities and utilities are therefore combined in the evaluation of an action by weighting the utility of a possible outcome by the probability that it occurs.

Prior Probability

We use the notation **P(A)** for the *unconditional* or *prior probability* that the condition A is true.

For example if Cavity denotes the probability that a patient has a cavity, then P(Cavity) = 0.1 means that in the absence of any other information, the agent will assign a probability of 0.1 to the event that the patient has a cavity.

Remember, P(A) can only be used when there is no other information.

As soon as some more information, **B**, is known, we have to reason with the conditional probability of **A given B**, instead of **P(A)**.

The proposition that is the subject of a probability statement can be represented by a proposition symbol, as in the **P(A)** example. Propositions can also include equalities involving random variables.

For example, if we are concerned about the random variable Weather,
- P(Weather = Sunny) = 0.7
- P(Weather = Rain) = 0.2
- P(Weather = Cloud) = 0.08
- P(Weather = Snow) = 0.02

Each random variable X has a domain of possible values, **<x1, x2, ... , xn>** that it can take on.

We can view proposition symbols as random variables as well, if we assume that they have a domain **<true, false>.**

Thus, **P(Cavity)** can be viewed as **P(Cavity == True)** and **P(!Cavity) = P(Cavity == False)**

If we want to talk about the probabilities for a random variable, we can do so with P(Weather) which denotes a vector of values for the probabilities of each state of weather. Given the preceding values, for example, we would write:
- **P(Weather) = <0.7, 0.2, 0.08, 0.02>**

This statement defines a probability distribution for the random variable Weather

You can look at the probability of many random variables at once with **P(Weather, Cavity)**, which creates a 4x2 table of probabilities, containing all the combinations of the random variables.

Conditional Probability

Once the agent has obtained some evidence concerning the previously unknown propositions making up the domain, prior probabilities are no longer applicable. Instead, we use conditional or posterior probabilities, with the notation **P(A|B)**. This is read as "*The probability of A given that all we know is B*"

e.g. **P(Cavity|Toothache) = 0.8** reads "*The probability of a cavity given that we know the patient has a toothache = 0.8*"

As soon as we know C, we can't compute **P(A|B)**, we must instead compute **P(A|B /\ C)**

We can think of the prior probability as just a conditional probability that looks like **P(A |)**, where the probability is conditioned on no evidence.

We can also use the **P** notation with conditional probabilities. **P(X|Y)** is a 2D table giving the values of **P(X = xi | Y = yj)** for each possible **i,j**. Conditional probabilities can be defined in terms of unconditional probabilities

P(A|B) = P(A/\B) / P(B) holds whenever **P(B) > 0**, and can be written as **P(A/\B) = P(B|A)P(B)**

P(A/\B) = **P(A|B)P(B)** = **P(B|A)P(A)**

We can also extend our P notation to handle equations like these, providing conciseness.

e.g. **P(X, Y) = P(X|Y)P(Y)** which denotes a set of equations relating to the corresponding individual entries in the tables (not a matrix multiplication of the tables)

Thus, one of the equations might be

P(X = x1 /\ Y = y2) = P(X = x1 | Y = y2) P(Y = y2)

In general, if we are interested in the probability of a proposition **A**, and we have accumulated evidence **B**, then the quantity we must calculate is **P(A|B)**. Sometimes we will not have this conditional probability, available directly in the KB, and we must resort to probabilistic inference.

- Let α and β be two propositions such that $P(\beta) \neq 0$. Then the conditional probability of α given β, denoted by $p(\alpha|\beta)$, is defined as:
- $p(\alpha|\beta) = p(\alpha \wedge \beta)/p(\beta)$
- This is also known as the **posterior probability** of α being true when β (described as the *evidence*) is true.
  - $P(\alpha)$ is also known as the **prior probability** and is equivalent to $P(\alpha|\ true)$
  - Reflects our background knowledge about the chance of α being true
- Exercise: Given that the card drawn from the top of the pack is from a black suit, what is the probability of the card being a King or Queen?
- Note that conditional probability is **not** a measure of causality

- Not all random variables or events affect the probability of each other.
- If we have two random variables, X and Y then:
  - If $p(X|Y) = p(X)$ we say that X and Y are independent of each other, as Y occurring has not affected the chance of X occurring.
  - $p(X|Y) = p(X \wedge Y)/p(Y) = p(X) \implies p(X \wedge Y) = p(X) * p(Y)$
  - This is also known as **unconditional independence**
- X and Y are **conditionally independent** given random variable Z if:
  - $p(X \wedge Y|Z) = p(X|Z) * p(Y|Z)$
  - $p(X|Y \wedge Z) = p(X|Z)$
  - $p(Y|X \wedge Z) = p(Y|Z)$
- Assuming independence is a useful tool, as it means we do not need a list of exhaustive conditional probabilities, which can often be infeasible to compute.

Probability Theorems

- Total Probability
  - Given a set of disjoint events A_i that partition the Sample Space: $p(\Omega) = \sum_i p(A_i)$
  - Also: $p(B) = \sum_i p(B \wedge A_i) = \sum_i p(B|A_i)p(A_i)$
- Product Rule
  - $p(A \wedge B) = p(B|A)p(A)$
- Chain Rule (Generalization of the product rule)
  - If we rearrange the definition of conditional probability, we find that $P(\alpha \wedge \beta) = P(\alpha|\beta) * P(\beta)$
  - This means any conjunction of propositions and can be expressed as a product of conditional probabilities
    - $P(a_1 \wedge a_2 \wedge ... \wedge a_i) = P(a_1) * P(a_1|a_2) * P(a_3|a_1 \wedge a_2) * ... * P(a_i|a_1 \wedge ... \wedge a_{i-1})$

Joint Probability Distribution

A joint completely specifies an agent's probability assignments to all propositions in the domain (both simple and complex)

A probabilistic model of a domain consists of a set of random variables that can take on particular values with certain probabilities. Let **X1 ... Xn** be the variables. An atomic event is an assignment of particular values to all the variables, in other words a complete specification of the state of the domain.
The joint probability distribution **P(X1 ... X2)** assigns probabilities to all possible atomic events. Recall that **P(Xi)** is a one dimensional vector of probabilities for the possible values of the variable **Xi.**

Then the joint is an n-dimension table with a value in every cell, giving the probability of that particular state occurring.

	toothache	!toothache
cavity	0.04	0.06
!cavity	0.01	0.89

- Any conjunction of atomic events is necessarily **false**.
- Since they are collectively exhaustive, their disjunction is necessarily **true**.

Adding across a row or columns gives the unconditional probability of a variable, e.g.

P(Cavity) = 0.06 + 0.04 = 0.1

Thus the probability that the patient has a cavity given the evidence that they have a toothache

P(Cavity|Toothache) = P(Cavity /\ Toothache) / P(Toothache) = 0.04/(0.04 + 0.01) = 0.8

Computing the joint is useful, but expensive. Bayes' rule gets us around this cost.

Bayes' Rule

Recall the two forms of the product rule:

- **P(A /\ B) = P(A|B)P(B)**
- **P(A /\ B) = P(B|A)P(A)**

Equating the RHS we get

$$\mathbf{P(B|A)} = \frac{P(\boldsymbol{A}|\boldsymbol{B})P(\boldsymbol{B})}{\mathbf{P(A)}}$$

Bayes' Rule needs 3 terms - a conditional probability and 2 unconditional probabilities - just to compute one conditional probability

Example

- A doctor knows that meningitis causes a stiff neck, 50% of the time
- The doctor knows some unconditional facts, the prior probability of a patient having meningitis is 1/50,000
- The prior probability of a patient having a stiff neck is 1/20

Let S = stiff neck
Let M = meningitis

P(S|M) = 0.5
P(M) = 1/50000
P(S) = 1/20

P(M|S) = P(S|M) * P(M) / P(S) = (0.5 * 1/50000) / (1/20) = 0.0002

This is very useful to know, since if there is suddenly an outbreak of meningitis, we can simply run the equation with new numbers and work out the new relationship between the two probabilities.

Normalisation

The goal of a normalising constant is to reduce any probability function to a probability density function with a total probability of one.

For example:

P(M|S) = P(S|M)P(M) / P(S)
P(W|S) = P(S|W)P(W) / P(S)

if we divide the top by the bottom we get P(S|M)P(M) / P(S|W)P(W) - the relative likelihood of whiplash W compared to meningitis M.

In some cases, relative likelihood is sufficient for decision making, but when the two possibilities yield radically different utilities for various treatment actions, one needs exact values in ordre to make rational decisions.

Combining Evidence

Given many variables, we might need an exponential number of probability values to complete a task. At this point we may as well go back to the joint.

in many domains, we can simplify the application of Bayes' rule so that it requires fewer probabilities in order to compute a result.

The first step is to take a slightly different view of the process of incorporating multiple pieces of evidence. The process of Bayesian updating incorporates evidence on piece at a time, modifying the previously held belief in the unknown variable.

In Bayesian updating, when each new piece of evidence is observed the belief in the unknown variable is multiplied by a factor that depends on the new evidence.

Working out this multiplication factor depends not just on the new evidence, but also on the evidence already obtained.

The key observation here is that of conditional independence

P(Catch|Cavity /\ Toothache) == P(Catch|Cavity)

The probability of the probe catching does not depend on the presence of a toothache. Thus we can simplify our expression.

Conditional Independence

A useful way to limit the amount of information required is to assume that each variable only directly depends on a dew other variables. This uses assumptions of conditional independence. Not only does it reduce how many numbers are required to specify a model, but also the independence structure may be exploited for efficient reasoning.

Random variable X is conditionally independent of random variable Y given a set of random variables **Zs** if
P(X|Y), Zs = P(X|Zs)

whenever the probabilities are well defined. This means that for all x in the domain of X, for all y in the domain of Y and for all z in the domain of z, if **P(Y = y /\ | Zs = s) > 0**, then

P(X = x | Y = y /\ Zs = z)
is equal to
P(X = x | Zs = z)

In other words, given a value of each variable in Zs, knowing Y's value does not affect the belief in the value of X

Example

consider the probabilistic model of students and exams. it is reasonable to assume that the random variable Intelligence is independent of ***Works_hard,*** *given no other observations. If you find a student that works hard, it does not tell you anything about their level of intelligence.*

the answers to the exams, ***Answers****, would depend on whether the student is intelligent and works hard. Thus, given* ***Answers****, intelligent would be dependent on* ***Works_hard;*** *if you found someone had insightful answers, and did not work hard, your belief that they are intelligent would go up.*

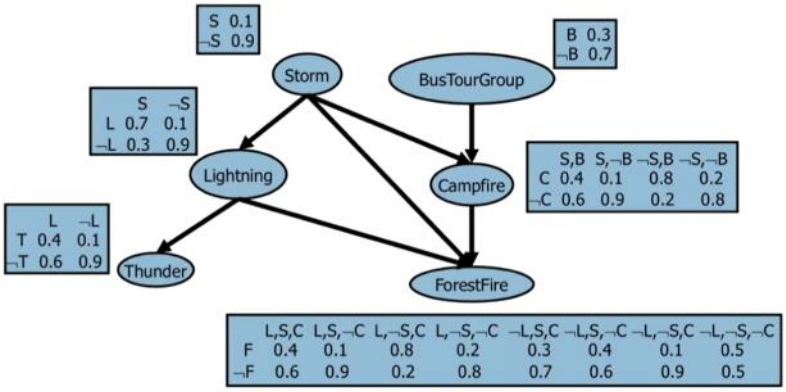
The grade on the exam, ***Grade****, should depend on the student's answers, not on the intelligence or whether they work hard. Thus* ***Grade*** *would be independent of* ***Intelligence*** *given* ***Answers****. However, if the answers were not observed,* ***Intelligence*** *will affect* ***Grade*** *because highly intelligent students would be expected to have different answers than less intelligent students. Thus* ***Grade*** *is dependent on I****ntelligence*** *given no observations.*

Conditional independence is a useful assumption that is often natural to assess and can be exploited in inference. It is very rare that we should have a table of probabilities of worlds and assess independence numerically.

Summary

- Probabilities represent an inability to reach a definite decision regarding truth
- Basic probability statements include prior probabilities and conditional probabilities.
- The axioms of probability specify constraints on reasonable assignments of probabilities to axioms. An agent violating the axioms will behave irrationally and can be manipulated.
- The joint probability distribution specifies the probability of each complete assignment of values to random variables. It is usually far too large to create or use.
- Bayes' rule allows unknown probabilities to be computed from known, stable ones.
- In the general case, combining many pieces of evidence may require assessing a large number of conditional probabilities, as in the joint probability distribution
- Conditional independence brought about by direct causal relationships in the domain allow Bayesian updating to work effectively even with multiple pieces of evidence.

- The computation, from observed evidence, of posterior probabilities for query propositions
  - The full joint probability distribution is the Knowledge Base
- The General Inference Procedure is as follows:
  - Let X be the query variable
  - Let E be the evidence variables and e be the observed values for them
  - Let Y be the unobserved variables
  - We need to calculate $p(X|e)$
  - $p(X|e) = \alpha p(X|e) = \alpha \sum_y p(X, e, y)$
  - α is the normalization constant, $1/p(e)$
  - $p(e) = \sum_{x,y} p(x, e, y)$


S 0.1
¬S 0.9
Storm
BusTourGroup
B 0.3
¬B 0.7
S ¬S
L 0.7 0.1
¬L 0.3 0.9
Lightning
Campfire
S,B S,¬B ¬S,B ¬S,¬B
C 0.4 0.1 0.8 0.2
¬C 0.6 0.9 0.2 0.8
L ¬L
T 0.4 0.1
¬T 0.6 0.9
Thunder
ForestFire
L,S,C L,S,¬C L,¬S,C L,¬S,¬C ¬L,S,C ¬L,S,¬C ¬L,¬S,C ¬L,¬S,¬C
F 0.4 0.1 0.8 0.2 0.3 0.4 0.1 0.5
¬F 0.6 0.9 0.2 0.8 0.7 0.6 0.9 0.5

Bayesian Belief

08 January 2021 15:08

Introduction

A belief network is a representation of a particular independence among variables. Belief networks should be viewed as a modelling language.

Many domains are concisely and naturally represented by exploiting the independencies that belief networks compactly represent.

Once the network structure and the domains of the variables for a belief network are defined, which numbers are required (the conditional probabilities) are prescribed. The user cannot simply add arbitrary conditional probabilities but must follow the network's structure. If the numbers required of a belief network are provided and are locally consistent, the whole network will be consistent.

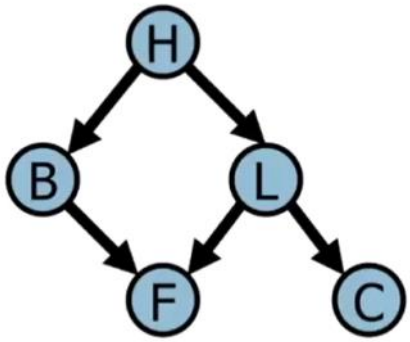
In contrast, the maximum entropy or random worlds approaches infer the most random worlds that are consistent with a probabilistic knowledge base. They form a probabilistic knowledge representation of the second type. For the random worlds approach, any numbers that happen to be available are added and used.

However, if you allow someone to add arbitrary probabilities, it is easy for the knowledge to be inconsistent with the axioms of probability. Moreover, it is difficult to justify an answer as correct if the assumptions are not made explicit.

Markov Condition

Before we define a Bayesian Belief Network, we need to understand the Markov Condition.

- Variables are often related through an inference chain
  - A history of smoking effects the probability of lung cancer, which in turn effects the existence of fatigue.
- Suppose we have a joint probability distribution P of the random variables in some set V and a directed acyclic graph $G =< V, E >$
  - (G, P) is said to satisfy the Markov Condition if for each variable $X \in V$, X is conditionally independent of the set of all its non-descendants given the set of all its parents, $I_p(\{X\}, ND_x | PA_x)$
  - ND_x is the set of all non-descendants of X
  - PA_x is the set of all parents of X



- If (G, P) satisfies the Markov condition, then P is equal to the product of the conditional distributions of all nodes given values of their parents, whenever these conditional distributions exist
  - Allows the number of parameters to be determined to be much smaller
  - Only the conditional probabilities $p(X|PA_x)$ need to be determined
  - If each node is binary and has at most one parent, less than $2n - 1$ parameters need to be determined as opposed to $2^n - 1$
- But, if we need to know P in the first instance to know that (G, P) satisfies the Markov condition, how have we reduced the numbed of parameters to determine?
- Given a DAG, G, in which each node is a random variable, and a conditional probability distribution of each node given values of its parents in G
  - the product of the conditional distributions yields a joint probability distribution P of the variables and (G, P) satisfies the Markov condition.

Belief Networks

The notion of conditional independence is used to give a concise representation of many domains. The idea is that, given a random variable X, there may be a few variables that DIRECTLY affects the X's value, in the sense that X is conditionally independent of other variables given these variables. The set of locally affecting variables is called the Markov Blanket. This locality is exploited in a belief network.

A belief network is a directed model of conditional dependence among a set of random variables. The conditional independence in a belief network takes in an ordering of the variables and results in a directed graph.

Defining a Belief Network

To define a BN on a set of random variables **{X1, X2, X3}** first select a total ordering of the variables, say **X1, X2, X3**. The chain rule shows how we can then decompose a conjunction into conditional probabilities.

$$P(X_1 = v_1 \wedge X_2 = v_2 \wedge \cdots \wedge X_n = v_n)$$
$$= \prod_{i=1}^{n} P(X_i = v_i \mid X_1 = v_1 \wedge \cdots \wedge X_{i-1} = v_{i-1}).$$

In terms of random variables and probability distributions..

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1}).$$

Define the parents of random variable Xi (written parents(Xi) to be a minimal set of predecessors of Xi in the total ordering such that the other predecessors of Xi are conditionally independent of Xi given parents(Xi). Thus Xi probabilistically depends on each of its parents but is independent of its other predecessors.

$$P(X_i \mid X_1, \ldots, X_{i-1}) = P(X_i \mid parents(X_i)).$$

Putting the chain rule and the parents definition together, we arrive at

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid parents(X_i)).$$

The probability over all of the variables P(X1, X2, ... Xn) is called the joint probability distribution. A BN defines a factorization of the joint probability distribution into a product of conditional probabilities.

A belief network, also called a Bayesian Network, is an acyclic direct graph where the nodes are random variabels. There is an arc from each element of parents(Xi) into Xi. Associated with the belief network is a set of conditional probability distributions that specify the conditional probability of each variable given its parents which includes the prior probability of those variables with no parents.

Thus a belief network consists of:

1. A DAG - where each node is labelled by a random variable
2. A domain for each random variables
3. A set of conditional probability distributions giving P(X | parents(X)) for each variable X

A belief network is acyclic by construction.

Remember that different orderings of variables can result in different belief networks. In particular, which variables are parents is dependent on ordering since only predecessor nodes can be parents of a variable. Some of the orderings may result in networks with fewer arcs than others, which generally speaking is a good thing since it simplifies the network.

Example

Example 8.13. *Consider the four variables of Example 8.12, with the ordering:* $Intelligent, Works\_hard, Answers, Grade$. *Consider the variables in order.* $Intelligent$ *does not have any predecessors in the ordering, so it has no parents, thus* $parents(Intelligent) = \{\}$. $Works\_hard$ *is independent of* $Intelligent$, *and so it too has no parents.* $Answers$ *depends on both* $Intelligent$ *and* $Works\_hard$, *so*

$$parents(Answers) = \{Intelligent, Works\_hard\}.$$

$Grade$ *is independent of* $Intelligent$ *and* $Works\_hard$ *given* $Answers$ *and so*

$$parents(Grade) = \{Answers\}.$$

The corresponding belief network is given in Figure 8.2.

This graph defines the decomposition of the joint distribution:

$$\begin{aligned} P(Inte & lligent, Works\_hard, Answers, Grade) \\ = \; & P(Intelligent) * P(Works\_hard) * P(Answers \mid Intelligent, Works\_hard) \\ & * P(Grade \mid Answers) \end{aligned}$$

In the examples below, the domains of the variables are simple, for example the domain of $Answers$ *may be* $\{insightful, clear, superficial, vacuous\}$ *or it could be the actual text answers.*

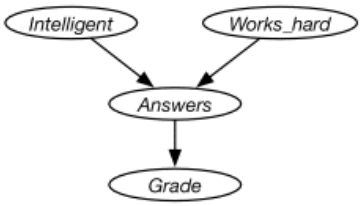


Figure 8.2: Belief network for exam answering of Example 8.13

Observations and Queries

A belief network specifies a joint probability distribution from which arbitrary conditional probabilities can be derived. The most common probabilistic inference task is to compute the posterior distribution of a query variable, or variables, given some evidence, where the evidence is a conjunction of assignments of values to some of the variables.

Example 8.14. *Before there are any observations, the distribution over intelligence is* $P(Intelligent)$*, which is provided as part of the network. To determine the distribution over grades,* $P(Grade)$*, requires inference.*

If a grade of A *is observed, the posterior distribution of* $Intelligent$ *is given by:*

$$P(Intelligent \mid Grade = A)\,.$$

If it was also observed that $Works\_hard$ *is false, the posterior distribution of* $Intelligent$ *is:*

$$P(Intelligent \mid Grade = A \wedge Works\_hard = false)\,.$$

Although $Intelligent$ *and* $Works\_hard$ *are independent given no observations, they are dependent given the grade. This might explain why some people claim they did not work hard to get a good grade; it increases the probability they are intelligent.*

Constructing Belief Networks

Book

To represent a domain in a belief network, the designer must consider the following:

1. What are the relevant variables?
    a. What the agent may *observe in the domain - each feature should be a variable*, because the agent must be able to condition on all of its observations
    b. What information the agent is interested in knowing the **posterior probability** of. Each of these needs to be a *variable that can be queried*
    c. Other hidden variables that will not be observed or queried but make the model simpler. These variables either account for dependencies, reduce the size of the specification of the conditional probabilities, or better model how the world is assumed to work
2. What values should these variables take?
    a. For each variable, the designer should *specify what it means to take each value in its domain*. What must be true in the world for a variable to have a particular value? This must satisfy the **clarity principle**: an omniscient agent should be able to know the value of a variable.
3. What is the r*elationship between he variables*? This should be expressed by adding arcs in the graph to define the parent relation
4. How does the *distribution of a variable depend on its parents*? This is express in terms of the conditional probability distributions.

See examples here.

- Given an ordering of nodes $\{X_1, X_2, ..., X_n\}$
- Process each node in order
  - Add it to the existing network
  - Add arcs from a minimal set of parents such that the parent set renders the current node independent of every other node preceding it
  - Define $PA_{X_i} \subseteq \{X_1, X_2, ..., X_{i-1}\}$
- Define the CPT for X_i
- Note
  - The resulting network, given any node ordering, can define the same joint probability distribution
  - Topology may be very different
  - Some networks will be more compact than others
  - Compact networks are desirable as they are more tractable
  - Dense networks fail to represent independencies or causal dependencies

Representing Conditional Probabilities and Factors

A conditional probability distribution is a function on variables - given an assignment to the values of the variables, it gives a number.

Even with a small number of nodes, this conditional probability table can grow very large.

The relationships between parents and child nodes usually fall into one of several categories that have canonical distributions - i.e. they fit some standard pattern.

The simplest example is provided by deterministic nodes who depend solely on their parent nodes.

Uncertain relationships can be characterized by noisy logical relationships.

Noisy-OR

A generalisation of the logical OR. The noisy-or adds some uncertainty to the standard logical-OR.

The model makes 3 assumptions:

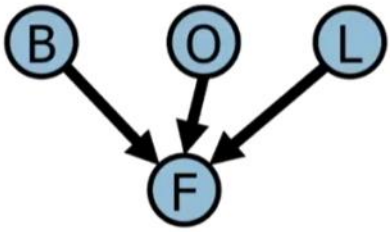
- each cause has an independent chance of causing the effect
- all the possible causes are listed (we can add a leak-node to catch all miscellaneous causes)
- it assumes whatever inhibits the cause from creating the effect is independent of what inhibits another cause from causing the effect. These inhibitors are not represented as nodes but as "noise parameters".

e.g. if P(Fever|Cold) = 0.4, then the noise parameter is 0.6.

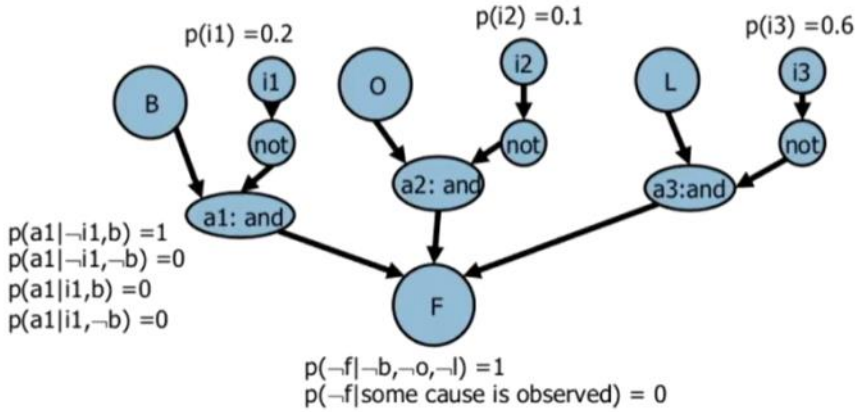
If exactly one parent is true, then the output is false with probability equal to the noise parameter for that node. In general, the probability that the output node is False is just the product of the noise parameters for all the input nodes that are true.

- Local probability distributions can get large as they are $O(2^n)$
- We can approximate these distributions by using canonical interaction models that require fewer parameters
- Noisy-OR:
  - Describes a set of n causes ($x_i's$) and their common effect (y)
  - Assumes each x_i is sufficient to cause the effect, y, in the absence all other causes.
  - The ability of x_i to cause y is *independent* of the presence of the other causes
- Only need k parameters:
  - $p_i = p(y|\neg x_1, ..., \neg x_{i-1}, x_i, \neg x_{i+1}, ..., x_k)$

- Consider the BBN representing the relationship between Fatigue, Lung Cancer, Bronchitis and Other Causes
- Causal Inhibition: Each cause has an inhibitor, that inhibits the expression of the cause
  - The effect is observed if and only if the inhibitor is disable
  - Bronchitis will result in Fatigue if and only if the mechanism that inhibits Bronchitis from causing Fatigue is not present
- The inhibiting mechanism of one cause is independent of the mechanism of other causes (Exception independence)
- The effect can happen only if at least one of its causes is present and *not* being inhibited (Accountability)
  - $p(\neg B, \neg O, \neg L, F) = 0$



- Nodes whose value is exactly specified by the parent nodes are called deterministic nodes, i.e. a_1, a_2 and a_3 in the BBN below
- An inhibitor has a probability of being "observed"



Inference

09 January 2021 12:51

Bayesian probability
BBN and how to build them

Now we will discuss how to use inference in Bayesian Belief Networks

Probabilistic Inference in Belief Networks

Exact Inference

Probabilities are computed exactly

- a simple version of this is enumeration
- we can also use variable elimination, which is a method that exploits conditional independence

Approximate Inference

Approximates the probabilities and are characterized by different guarantees they provide

- they produce guaranteed bounds on the probabilities i.e. the exact probability will fall between a given range
- They may produce probabilistic bounds on the error i.e. the error is within 0.1 of the error 95% of the time. Such algorithms also guarantee that, as time increases, the probability estimates will converge to the exact answer.

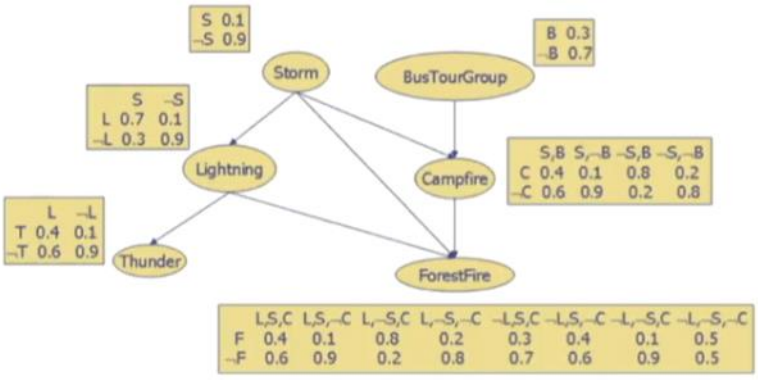
Type of reasoning

- Diagnostic reasoning - reasoning from symptom to cause
- Predictive Reasoning - reasoning from cause to symptom
- Intercausal Reasoning - reasoning about the mutual causes of a common effect
- Combined Reasoning - if the query variable is a parent of some observed variables and the descendent of other observed variables

Inference by Enumeration

- This involves enumerating through every world that is consistent with the evidence
- $p(X|e) = \alpha p(X|e) = \alpha \sum_y p(X, e, y)$
- $\alpha = 1/p(e) = 1/\sum_{x,y} p(X, e, y)$
- Remember that y is the set of hidden variables
  - These are variables in the network that are not included in the evidence or our query set, X.
  - If there are 2 hidden variables in the network, (A, B), we have to consider four different worlds.
  - $p(X|E \wedge A \wedge B)$
  - $p(X|E \wedge A \wedge \neg B)$
  - $p(X|E \wedge \neg A \wedge B)$
  - $p(X|E \wedge \neg A \wedge \neg B)$

Example



- Given the BBN pictured:
  - What is the probability of a forest fire, i.e. $X = F$, given that:
    - ★ A Storm (S) took place
    - ★ Lightning (L) was observed
    - ★ No Bus Tour Group (B) visited the forest
- $p(F|S, L, \neg B) = ?$

We have 3 values for our variables. We have 2 other variables that we don't know the value of.

If we sum the probabilities of all the possible resulting states, we can work out the probability of F.

The possible values of C and T are: C && T, C && ¬T, ¬C && T, ¬C && ¬T

- Calculate the posterior probability $p(F|S \wedge L \wedge \neg B)$
  - The event $E = S \wedge L \wedge \neg B \wedge F$
  - $\{S \wedge L \wedge \neg B \wedge C \wedge T \wedge F,$
    $S \wedge L \wedge \neg B \wedge C \wedge \neg T \wedge F,$
    $S \wedge L \wedge \neg B \wedge \neg C \wedge T \wedge F,$
    $S \wedge L \wedge \neg B \wedge \neg C \wedge \neg T \wedge F\}$
- $p(E) = \sum_{x \in E} p(x)$

So if we sum those states...

$$p(E) = \sum_{x \in E} p(x)$$

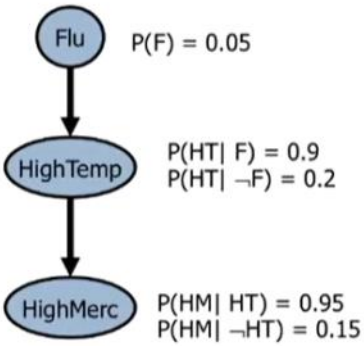
- $p(S \wedge \neg B \wedge L \wedge C \wedge T \wedge F)$
  $= p(S)p(\neg B)p(L|S)p(C|S \wedge \neg B)\ p(T|L)p(F|S \wedge L \wedge C)$
  $= 0.1 \times 0.7 \times 0.7 \times 0.1 \times 0.4 \times 0.4$
  $= 0.000784$
- $p(S \wedge \neg B \wedge L \wedge C \wedge \neg T \wedge F)$
  $= p(S)p(\neg B)p(L|S)p(C|S \wedge \neg B)\ p(\neg T|L)p(F|S \wedge L \wedge C)$
  $= 0.1 \times 0.7 \times 0.7 \times 0.1 \times 0.6 \times 0.4$
  $= 0.001176$
- $p(S \wedge \neg B \wedge L \wedge \neg C \wedge T \wedge F)$
  $= p(S)p(\neg B)p(L|S)p(\neg C|S \wedge \neg B)\ p(T|L)p(F|S \wedge L \wedge \neg C)$
  $= 0.1 \times 0.7 \times 0.7 \times 0.9 \times 0.4 \times 0.1$
  $= 0.001764$
- $p(S \wedge \neg B \wedge L \wedge \neg C \wedge \neg T \wedge F)$
  $= p(S)p(\neg B)p(L|S)p(\neg C|S \wedge \neg B)\ p(\neg T|L)p(F|S \wedge L \wedge \neg C)$
  $= 0.1 \times 0.7 \times 0.7 \times 0.9 \times 0.6 \times 0.1$
  $= 0.002646$

Note the evidence never changes, and F (our query variable) never changes.

Inference in a Chain of 3 Nodes

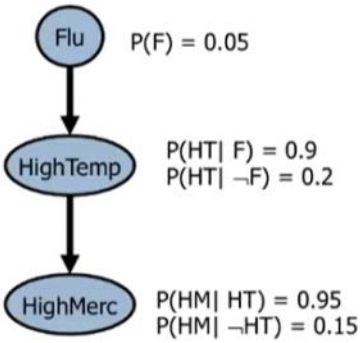
Example 1

- If Flu is Observed, we can use the chain rule: $p(HM|F) = \sum_{HighTemp} p(HM|HT)p(HT|F) = 0.9 \times 0.95 + 0.1 \times 0.15 = 0.87$
- Recalled that if A and C are conditionally independent: $p(C|A) = p(C|B)p(B|A) + p(C|\neg B)p(\neg B|A)$



Example 2

- If HighMecury readings are Observed, we can use Bayes rule and the Chain rule:
- $p(F|HM) = \alpha p(F)p(HM|F) = \alpha p(F) \sum_{HighTemp} p(HM|HT)p(HT|F) = \alpha \times 0.05 \times 0.87 = 0.0435\alpha$
- $p(\neg F|HM) = \alpha p(\neg F)p(HM|\neg F) = \alpha p(\neg F) \sum_{HighTemp} p(HM|HT)p(HT|\neg F) = \alpha \times 0.95 \times 0.31 = 0.2945\alpha$
- $\alpha = \frac{1}{0.0435+0.2945} \implies p(F|HM) = 0.1287$



Variable Elimination

Uses Bayesian Belief Networks. Another form of exact inference.

Adapted from a solution to finding solutions to CSP's.
Algorithm is based on the notion that a belief network specifies a factorisation of the joint probability distribution - this is more efficient than enumeration
Remember that a conditional probability is a function on a variable Y and some set of evidence variables
Also recall that the probability of the possible outcomes for Y, given a specific set of value assignments to the evidence, should sum to 1.
We may call a function on a set of variables a factor and we say that the scope of the factor is the set of variables it involves.
The conditional probability P(X|Y,Z) could be described as a factor with scope X, Y, Z

X	Y	$P(Z = t \mid X, Y)$
t	t	0.1
t	f	0.2
f	t	0.4
f	f	0.3

Expressing Factors

We can express out factors as an array
- if there is an ordering of the variables, e.g. alphabetical
- and the values in the domains are mapped into non-negative integers then
- there is a unique representation of each factor as a one-dimensional array

For example, we can represent the conditional probability table pictured as [0.1, 0.2, 0.4, 0.3]

We can perform a number of operations on factors - conditioning, summing and multiplying

Conditioning

If we have observed a variable, we can define a new factor with a new domain. The domain will be a subset of the domain prior to discovering the value of the variable.

As we discover more and more variables, the subset of possible values becomes smaller and smaller:

$r(X,Y,Z) =$

X	Y	Z	val
t	t	t	0.1
t	t	f	0.9
t	f	t	0.2
t	f	f	0.8
f	t	t	0.4
f	t	f	0.6
f	f	t	0.3
f	f	f	0.7

$r(X=t,Y,Z) =$

Y	Z	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

$r(X=t,Y,Z=f) =$

Y	val
t	0.9
f	0.8

$r(X=t,Y=f,Z=f) = 0.8$

Multiplying Factors

If two factors share a variable within their scope, we can multiply them together to make a new factors with a scope equivalent to the union of the original two factor's scope.

$f_1 =$

A	B	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

$f_2 =$

B	C	val
t	t	0.3
t	f	0.7
f	t	0.6
f	f	0.4

$f_1 * f_2 =$

A	B	C	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

Summing Factors

We eliminate a chosen variable, by adding together the outcomes for each possible value of the variable.
As with conditioning, it allows the removal of a variable in order to simplify our conditional table.

This is known as summing out a variable.

$f_3 =$

A	B	C	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

$\sum_B f_3 =$

A	C	val
t	t	0.57
t	f	0.43
f	t	0.54
f	f	0.46

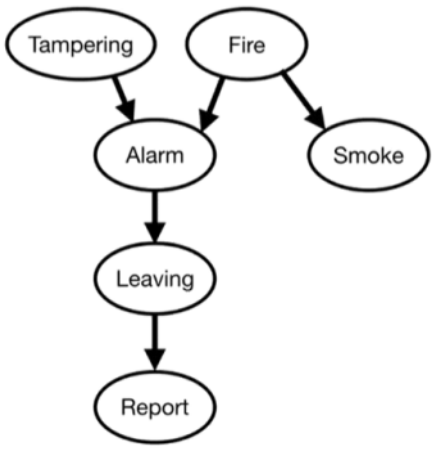
Back to Variable Elimination...

For a given query, we can represent the conditional probabilities of a BBN as a series of factors
The algorithm for solving a query is then as follow

- construct a factor for each conditional probability condition
- eliminate each non-query variable
    - if the variable is observed, its value is set to the observed value in each of the factors in which the variable appears
    - otherwise sum out the variable
- multiply the remaining factors and then normalise

Variable Elimination Example

- Consider the pictured network, it represents a number of conditional probabilities:
    - $p(Tampering), p(Fire), p(Alarm|Tampering, Fire)$
      $p(Smoke|Fire), p(Leaving|Alarm), p(Report|Leaving)$
- How would we solve the query
  $P(Tampering|Smoke = True, Report = True)$?
- We perform step 1: construct our factors:
    - $f_0(Tampering), f_1(Fire), f_2(Alarm, Tampering, Fire)$
      $f_3(Smoke, Fire), f_4(Leaving, Alarm),$
      $f_5(Report, Leaving)$
- First, we set our observed variables, Smoke and Report, to their value, and eliminate them from the factors:
    - $f_0(Tampering), f_1(Fire), f_2(Alarm, Tampering, Fire)$
      $f_3(Fire), f_4(Leaving, Alarm), f_5(Leaving)$



- Next, we eliminate the variable Fire.
- This involves multiplying together all factors that contain Fire, and then summing Fire out.
- $f_1(Fire) * f_2(Tampering, Alarm, Fire) * f_3(Fire) = f_{1,2,3}(Fire, Tampering, Alarm)$
- $\sum_{Fire} f_{1,2,3} = f_6(Tampering, Alarm)$

- We can repeat this same process with the factors containing Alarm, to eliminate that variable.
- We now have the following factors:
    - $f_0(Tampering), f_5(Leaving), f_7(Tampering, Leaving)$
- We're trying to find *Tampering*, so we must now eliminate *Leaving*
- Again, by multiplying and summing out *Leaving*, we have:
    - $f_0(Tampering), f_8(Tampering)$
- Finally, we multiply these factors together:
    - $f_9(Tampering) = f_0(Tampering) * f_8(Tampering)$
- Remember that although these factors have the same scope, their values will be different, as they are constructed using different evidence.

- Finally, we can find the posterior distribution over Tampering by:
- $f_9(Tampering)/\sum_{Tampering} f_9(Tampering)$
- The denominator here represents the prior probability of the evidence, in this case $Smoke = True, Report = True$.

Decision Making

- BBN's provide a mechanism to conduct Bayesian inference on large sets of random variables.
- These likelihood's of the outcomes must be used in some way to make a decision, a process not specifically support by a BBN. - for example, decide a treatment regime for a patient.
- We need to explicitly consider represent actions under consideration and utility of the resultant outcomes.
- We consider two possible tools for this decision making process
    - decision trees
    - influence diagrams, or decision networks

Expected Utility

- given a set of outcomes, of a particular action A, a utility function assigns a utility (value - measure of desirability) to each outcome
- Assuming that we have a probability distribution over the set of outcomes, the expected utility of taking the action A is defined as

- $EU(A) = \sum_i P(O_i|A) \times U(O_i|A)$

The assumption is that a Bayesian decision maker wants to maximise their expected utility through actions.

Combining utility Theory with Probability theory

- given evidence E, which action A is expected to deliver the most value
    - $EU(A|E) = \sum_i P(O_i|E, A) \times U(O_i|A)$

Utility can be a number of things, and is often represented as money

Outcomes

To inform our agent we must define relations between outcomes

- To inform our agent, we must define relations between outcomes.
- Consider two outcomes, σ_1 and σ_2
    - $\sigma_1 \succeq \sigma_2$ - σ_1 is **weakly preferred** to σ_2, meaning that σ_1 is at least as desirable as σ_2.
    - $\sigma_1 \sim \sigma_2$ - means that $\sigma_1 \succeq \sigma_2$ and $\sigma_2 \succeq \sigma_1$, and that the outcomes are equally preferred, meaning we are **indifferent** to which one we choose.
    - $\sigma_1 \succ \sigma_2$ - σ_1 is **strictly preferred** to σ_2 i.e. this means that we are not indifferent, and that we do not weakly prefer σ_2 to σ_1.
- These relations should be:
    - Complete - An agent has preferences between all pairs of outcomes
    - Transitive - If $\sigma_1 \succ \sigma_2$ and $\sigma_2 \succ \sigma_3$ then $\sigma_1 \succ \sigma_3$

Decision Trees

Decision Trees contain two types of nodes:

1. Chance nodes - represented by a circle
    a. representing a random variable

   b. edges emanating from change node represent the possible outcomes of the random variable and are labelled with the probability of the outcome
2. Decision node - represented by a square
   a. Representing a decision to be made
   b. edges emanating from such a node represent a set of mutually exclusive and exhaustive actions that the decision maker can make

The expected utility EU
- Of a change node is defined as the expected value of the utilities associated with its outcomes
- Of a decision alternative (action) is the expected utility of the chance node encountered if the action is taken
- Of a decision node is the maximum utility of all its alternatives

Normative theory - the assumption that the agent wants to maximise expected utility
Prospect theory - the final end state is not what people have preferences over - rather what matters is how much the choice differs from the current situation - 10k means more to someone with $500 than someone with $100,000.

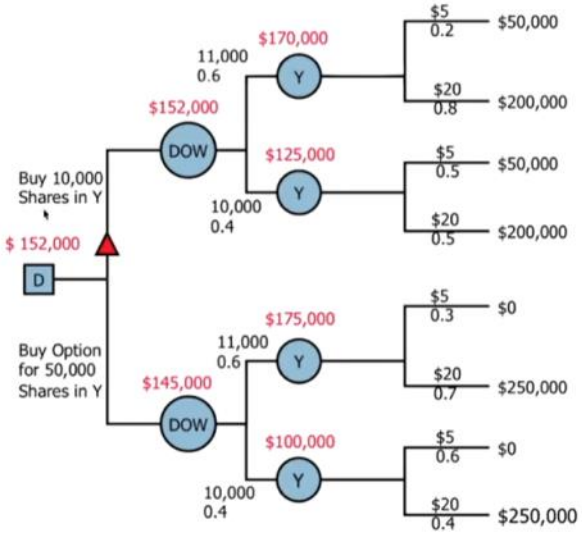
Representing Risk Aversion

Include it in the utility function

- One way to model an individual's attitude to risk is with a utility function that maps money to utility
  - For example using a exponential utility function, where R is the risk tolerance. The larger the value of R, the more risk-tolerant the user
  - $U_R(x) = 1 - e^{-x/R}$
  - For R = 500:
    - $EU(BuyX) = 0.25 * U_{500}(500) + 0.25 * U_{500}(1000) + 0.5 * U_{500}(2000) = 0.865$
    - $EU(Bank) = U_{500}(1050) = 0.877$
  - For R = 1000:
    - $EU(BuyX) = 0.25 * U_{1000}(500) + 0.25 * U_{1000}(1000) + 0.5 * U_{1000}(2000) = 0.688$
    - $EU(Bank) = U_{500}(1050) = 0.65$

A more complex example

- Sue is considering buying 10,000 shares in company Y @ $10/share
  - Her buying this number of shares will affect the price of Y
- She believes that the overall value of the DOW industrial average will also affect the price of Y
  - She believes that, in one month the DOW will either be at 10,000 or 11,000
  - And that Y will be at $5 or $20 per share
- Alternatively, she can buy an option of Y worth $100,000
  - Allows her to buy 50,000 shares in Y for $15/share in one month
- Her beliefs are:
  - $p(DOW = 11,000) = 0.6$
  - $p(Y = \$5 | Decision = buy, DOW = 11,000) = 0.2$
  - $p(Y = \$5 | Decision = buy, DOW = 10,000) = 0.5$
  - $p(Y = \$5 | Decision = option, DOW = 11,000) = 0.3$
  - $p(Y = \$5 | Decision = option, DOW = 10,000) = 0.6$



Influence Diagrams

- Decision trees are fine but grow exponentially
- Influence diagrams suffer from neither of these issues
- Contains 3 nodes: chance (circle), decision (square), and utility (diamond)

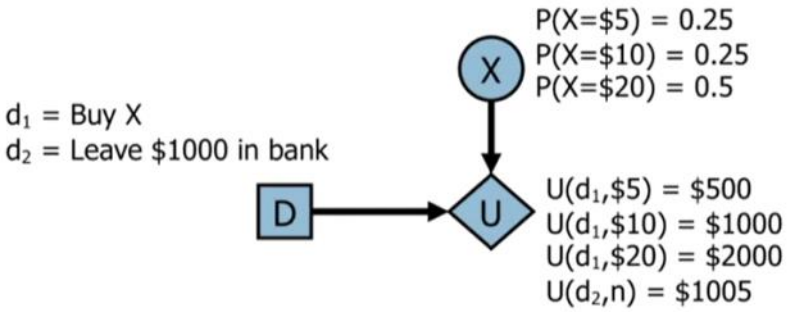
A [decision network](#) (aka influence diagram) is a graphical representation of a finite sequential decision problem. They extend belief networks to include decision variables and utility.
A decision network extends the single-stage decision network to allow for sequential decisions and allows both chance nodes to be parents of decision nodes.

Different edges have different meanings

- Edge to chance node - value of node is probabilistically dependent on the value of the parent
- Edge to decision node: value of the parent is known at the time the decision is made
    - If parent is a decision node, the edge represents a decision sequence
- Edge to a utility node - value of node is deterministically dependent on the value of the parent

The chance nodes satisfy the **Markov Condition**

These diagrams are *Bayesian Networks augmented with a utility node and ordered decision nodes.*



- Solving an influence diagram
    - Which decision choice has the maximum utility? i.e. $max(EU(d_1), EU(d_2))$
    - $EU(d_1) = E(U|d_1) = P(X = \$5) \times U(d_1, \$5) + P(X = \$10) \times U(d_1, \$10) + P(X = \$20) \times U(d_1, \$20) = \$1375$
    - $EU(d_2) = E(U|d_2) = \$1005$
- The decision is d_1.

Evaluating Influence Diagrams

with 1 d node:

1. add any evidence (set probability of value of random variables observed to 1)
2. for each action value in the decision node
   a. set decision node to that value
   b. calculate posterior probabilities of nodes that are parent of the utility node
   c. calculate the expected utility of the action
3. Return action with highest utility

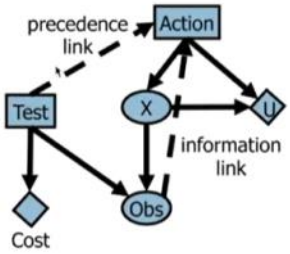
Information Links

Links from chance nodes to decision nodes
Indicate that the chance node must be known before a decision is made corresponding to that decision node
= can be used to explicitly calculate what decision should be made, given the different values for a chance node

- Evaluating Influence Diagram with Information Links
  - Add any evidence (set probability of value of random variables observed to 1)
  - For each combination of values of the parents of the decision node, for each action value in the decision node:
    - Set decision node to that value
    - Calculate posterior probabilities of nodes that are parents of the utility node
    - Calculate the expected utility of the action
  - Record the resulting expected utility for the action
  - Return the table of actions and associated expected utility values (decision table)

Sequential Decision Making

- Typical Example is a Test-Action Influence Diagram
  - Test Decision Node must be evaluated first
  - The cost associated with a test is included as a separate utility node
  - If the decision is to run a test, evidence will be obtained as a result of the test
    - The chance node representing the observation, Obs, has an information link to the Action decision
    - One value, unknown, of Obs will represent the decision to not test within CPT
    - $p(Obs = unknown | Test = no) = 1$
    - $p(Obs = unknown | Test = yes) = 0$
- Evaluating such an influence diagram:
  - Evaluate decision network with any available evidence
  - Enter test decision as evidence
  - If test decision is not 'yes', use value 'unknown'
  - Evaluate Action decision



Summary

- In this topic, we have covered:
  - Definitions of probability functions, spaces, distributions and propositions
  - Probabilistic inference
  - Theorems of probability, including Bayes Rule
  - Bayesian Belief Networks, their construction and use
  - Variable elimination
  - Utility functions
  - Influence networks, their definition and how to evaluate them

Reinforcement Learning

10 January 2021 10:10

Objectives:

- Utility, rewards, and values
- Decision-theoretic planning and Markov Decision Processes
- State-based reinforcement learning algorithms: Q-learning and SARSA
- The explore-exploit dilemma and solutions
- On-poly and off-policy reinforcement learning

AIFCA 12.1-12.7, 9.5, AIMA 21

See also https://www.davidsilver.uk/teaching/ for more detailed explanations of MDP's and Policies.

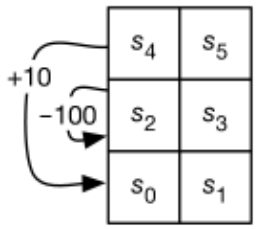
A RL agent acts in an environment, observing its state and receiving rewards. From its perceptual and reward information, it must determine what to do.

The learning agent is given the possible states and the set of actions it can carry out
At each time the agent observes the state of the environment and the reward received. We are assuming the environment is fully-observable.
At each time, after observing the state and reward, the agent carries out an action.
The goal of the agent is to maximize its discounted reward, for some discount factor.

RL can be formalized in terms of Markov Decision Processes, in which the agent initially only knows the set of possible states and the set of possible actions.

The dynamics, $P(s'|a,s)$, and the reward function, $R(s,a)$, are not given to the agent. As in an MDP, after each action, the agent observes the state it is in and receives a reward.

We can imagine this as a tiny RL environment:



The agent knows only that it has four actions, and what state it is in.
he agent does not know how the states are configured, what the actions do, or how rewards are earned.

RL is Hard

- The **credit assignment problem** or the blame attribution problem is the problem of determining which action was responsible for a reward or punishment. The action responsible may have occurred a long time before the reward was received. Moreover, not just a single action but rather a combination of actions carried out in the appropriate circumstances may be responsible for a reward.
- even if the dynamics of the world do not chance, the effect of an action of the agent depends on what the agent will do in future. What may initially seem like a bad thing for the agent to do may end up being an optimal action because of what the agent does in the future. This is common among planning problems, but it is complicated in the reinforcement learning context because the agent does not know a priori, the effects of its actions
- The **explore-exploit dilemma** - if an agent has worked out a good course of actions, should it continue to follow these actions? exploiting what it has determined? or should it explore more

to find better actions? An agent that never explores may act forever in a way that could have been much better if it had explored earlier.

Markov Decision Processes

The decision networks of the previous section were for finite-stage partially observable domains. Here, we consider indefinite horizon and infinite horizon problems

Often an agent must reason about an ongoing process or it does not know how many actions it will be required to do. These are called finite horizon problems when the process may go on forever or indefinite horizon problems when the agent will eventually stop but it does not know when.

For ongoing processes, it may not make sense to only the utility at the end, because the agent may never get to the end. Instead, an agent can receive a sequence of rewards, that incorporate the action costs in addition to any prizes or penalties that may be awarded. Indefinite horizon problems can be modelled using a stopping state. A stopping state is a state in which all actions have no effects; that is, when the agent is in that state, all actions immediately return to that state with a zero reward. Goal achievement can be modelled by having a reward for entering such a stopping state.

A Markov Decision Process can be seen as a Markov Chain augmented with actions and rewards or as a decision network extended in time. At each stage, the agent decides which action to perform; the reward and resulting state depend on both the previous state and the action performed.

We only consider stationary models where the state transitions and the rewards do not depend on the time.

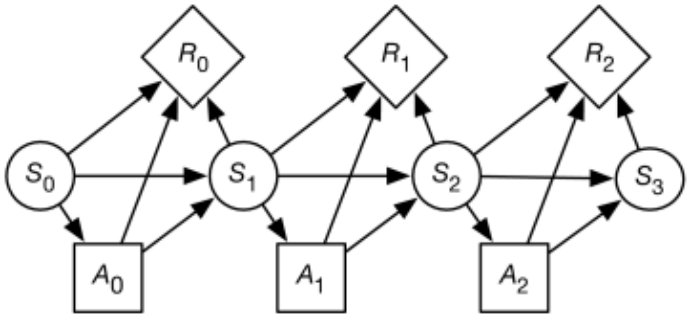
A **Markov decision process** or an **MDP** consists of

- S, a set of states of the world
- A, a set of actions
- $P : S \times S \times A \to [0, 1]$, which specifies the **dynamics**. This is written as $P(s' \mid s, a)$, the probability of the agent transitioning into state s' given that the agent is in state s and does action a. Thus,

$$\forall s \in S \; \forall a \in A \; \sum_{s' \in S} P(s' \mid s, a) = 1.$$

- $R : S \times A \times S \to \Re$, where $R(s, a, s')$, the **reward function**, gives the expected immediate reward from doing action a and transitioning to state s' from state s. Sometimes it is convenient to use $R(s, a)$, the expected value of doing a in state s, which is $R(s, a) = \sum_{s'} R(s, a, s') * P(s' \mid s, a)$.

A Markov Decision Process can be depicted using a decision network:



With decision networks, the designer has to consider what information is available to the agent

when it decides what to do. There are two common variations:

- **Fully Observable Markov Decision Processes** - the agent gets to observe the current state when deciding what to do
- **Partially Observable Markov Decision Processes** - a combination of an MDP and a hidden markov model. At each time, the agent gets to make some ambiguous and possibly noisy observations that depend on the state. The agent only has access to the history of rewards, observations and previous actions when making a decision. It cannot directly observe the current state.

Rewards

To decide what to do, the agent compares different sequences of rewards. The most common way to do this is to convert a sequence of rewards into a number called the value, the cumulative reward or the return. To do this, the agent combines an immediate reward with other rewards in the future. Suppose the agent receives the sequence of rewards r1, r2, r3, r4...

Total reward $V = \sum_{i=1}^{\infty} r_i$. In this case, the value is the sum of all of the rewards. This works when you can guarantee that the sum is finite; but if the sum is infinite, it does not give any opportunity to compare which sequence of rewards is preferable. For example, a sequence of $1 rewards has the same total as a sequence of $100 rewards (both are infinite). One case where the total reward is finite is when there are stopping states and the agent always has a non-zero probability of eventually entering a stopping state.

Average reward $V = \lim_{n \to \infty}(r_1 + \cdots + r_n)/n$. In this case, the agent's value is the average of its rewards, averaged over for each time period. As long as the rewards are finite, this value will also be finite. However, whenever the total reward is finite, the average reward is zero, and so the average reward will fail to allow the agent to choose among different actions that each have a zero average reward. Under this criterion, the only thing that matters is where the agent ends up. Any finite sequence of bad actions does not affect the limit. For example, receiving $1,000,000 followed by rewards of $1 has the same average reward as receiving $0 followed by rewards of $1 (they both have an average reward of $1).

Discounted reward $V = r_1 + \gamma r_2 + \gamma^2 r_3 + \cdots + \gamma^{i-1} r_i + \cdots$, where γ, the **discount factor**, is a number in the range $0 \leq \gamma < 1$. Under this criterion, future rewards are worth less than the current reward. If γ was 1, this would be the same as the total reward. When $\gamma = 0$, the agent ignores all future rewards. Having $0 \leq \gamma < 1$ guarantees that, whenever the rewards are finite, the total value will also be finite.

The discounted reward can be rewritten as

$$\begin{aligned} V &= \sum_{i=1}^{\infty} \gamma^{i-1} r_i \\ &= r_1 + \gamma r_2 + \gamma^2 r_3 + \cdots + \gamma^{i-1} r_i + \cdots \\ &= r_1 + \gamma \left(r_2 + \gamma \left(r_3 + \cdots \right) \right). \end{aligned}$$

Suppose V_k is the reward accumulated from time k:

$$\begin{aligned} V_k &= r_k + \gamma \left(r_{k+1} + \gamma \left(r_{k+2} + \cdots \right) \right) \\ &= r_k + \gamma V_{k+1}. \end{aligned}$$

To understand the properties of V_k, suppose $S = 1 + \gamma + \gamma^2 + \gamma^3 + \cdots$, then $S = 1 + \gamma S$. Solving for S gives $S = 1/(1-\gamma)$. Thus, with the discounted reward, the value of all of the future is at most $1/(1-\gamma)$ times as much as the maximum reward and at least $1/(1-\gamma)$ times as much as the minimum reward. Therefore, the eternity of time from now only has a finite value compared with the immediate reward, unlike the average reward, in which the immediate reward is dominated by the cumulative reward for the eternity of time.

In economics, γ is related to the interest rate: getting $\$1$ now is equivalent to getting $\$(1+i)$ in one year, where i is the interest rate. You could also see the discount rate as the probability that the agent survives; γ can be seen as the probability that the agent keeps going.

Policies

In a FOMDP, the agent gets to observe its currents state before deciding which action to carry out. For now, we assume that the MDP is FO. A policy specifies what the agent should do as a function of the state it is in. A stationary policy is a function:

$\pi : S \rightarrow A$

In a non-stationary policy the action is a function of the state and the time; we assume policies are stationary.

Given a reward criterion, a policy has an expected value for every state. Let:

$V^{\pi}(s)$

Be the expected value of following pi in state s. This specifies how much value the agent expects to receive from following the policy in that state.

Policy pi is an **optimal policy** if there is no policy pi' and no state s such that V^pi'(s) > V^pi(s). That is, it is a policy that has a greater or equal expected value at every state than any other policy.

if there are 100 states and 4 actions that can be performed in each state, there are 4^100 possible stationary policies. Each policy specifies an action for each state.

For infinite horizon problems, a stationary MDP always has an optimal stationary policy. However, for finite-state problems, a non-stationary policy might be better than all stationary policies.

For example, if the agent had to stop at time n, for the last decision in some state, the agent would act to get the largest immediate reward without considering the future actions, but for earlier decisions it may decide to get a lower reward immediately to get a larger reward in the future.

Value of a Policy

Consider how to compute the expected value using the discounted reward of a policy, given a discount factor of y. The value is defined in terms of two interrelated functions:

- $V^{\pi}(s)$ is the expected value of following policy π in state s.
- $Q^{\pi}(s,a)$, is the expected value, starting in state s of doing action a, then following policy π. This is called the **Q–value** of policy π.

You can define Q and V recursively in terms of each other.

If the agent is in state s', performs action a, and arrives in state s', it gets the immediate reward of R(s, a, s'), plus the discounted future reward:

$$\gamma V^{\pi}(s')$$

When the agent is planning it does not know the actual resulting state, so it uses the expected value, average over the possible resulting states:

$$\begin{aligned} Q^{\pi}(s,a) &= \sum_{s'} P(s' \mid s,a)\,(R(s,a,s') + \gamma V^{\pi}(s')) \\ &= R(s,a) + \gamma \sum_{s'} P(s' \mid s,a)\, V^{\pi}(s') \end{aligned}$$

where $R(s,a) = \sum_{s'} P(s' \mid s,a)\, R(s,a,s')$.

$V^{\pi}(s)$ is obtained by doing the action specified by π and then following π:

$$V^{\pi}(s) = Q^{\pi}(s,\pi(s)).$$

Value of an Optimal Policy

Let $Q^*(s,a)$, where s is a state and a is an action, be the expected value of doing a in state s and then following the optimal policy. Let $V^*(s)$, where s is a state, be the expected value of following an optimal policy from state s.

Q^* can be defined analogously to Q^π:

$$\begin{aligned} Q^*(s,a) &= \sum_{s'} P(s' \mid s,a)\,(R(s,a,s') + \gamma V^*(s')) \\ &= R(s,a) + \gamma \sum_{s'} P(s' \mid s,a)\,\gamma V^*(s')\,. \end{aligned}$$

$V^*(s)$ is obtained by performing the action that gives the best value in each state:

$$V^*(s) = \max_a Q^*(s,a)\,.$$

An optimal policy π^* is one of the policies that gives the best value for each state:

$$\pi^*(s) = \arg\max_a Q^*(s,a)$$

where $\arg\max_a Q^*(s,a)$ is a function of state s, and its value is an action a that results in the maximum value of $Q^*(s,a)$.

Value Iteration

Value iteration is a *method of computing an optimal policy for an MDP and its value*.

Value iteration starts at the end and works backwards, refining an estimate of either **Q*** or **V***.

There really is no end, so it uses an arbitrary end point. Let **Vk** be the value function assuming there are k stages to go

Let **Qk** be the Q-function assuming there are **k** stages to go. These can be defined recursively. Value iteration starts with an arbitrary function **V0**.

For subsequent stages, it uses the following equations to get the functions for **k+1** stages to go form the functions for k stages to go.

$$\begin{aligned} Q_{k+1}(s,a) &= R(s,a) + \gamma * \sum_{s'} P(s' \mid s,a) * V_k(s') \\ V_k(s) &= \max_a Q_k(s,a) \end{aligned}$$

It can either save the V[S] array or the Q[S,A] array. Saving the V array results in less storage, but it is more difficult to determine an optimal action, and one more iteration is needed to determine which waction results in the greatest value.

```
1: procedure Value_iteration(S, A, P, R)
2:     Inputs
3:         S is the set of all states
4:         A is the set of all actions
5:         P is state transition function specifying P(s' | s, a)
6:         R is a reward function R(s, a)
7:     Output
8:         π[S] approximately optimal policy
9:         V[S] value function
10:    Local
11:        real array V_k[S] is a sequence of value functions
12:        action array π[S]
13:    assign V_0[S] arbitrarily
14:    k := 0
15:    repeat
16:        k := k + 1
17:        for each state s do
18:            V_k[s] = max_a R(s, a) + γ * Σ_s' P(s' | s, a) * V_{k-1}[s']
19:    until termination
20:    for each state s do
21:        π[s] = arg max_a R(s, a) + γ * Σ_s' P(s' | s, a) * V_k[s']
22:    return π, V_k
```

Figure 9.16: Value iteration for MDPs, storing V

The above shows the value iteration algorithm when the V array is stored. This procedure converges no matter what the initial value function V0 is. An initial value function that approximates V* converges quicker than one that does not. The basis for many abstraction techniques for MDP's is to use some heuristic method to approximate V*, and to use this an initial seed for value iteration.

Consider this example:

Example 9.27. *Suppose Sam wanted to make an informed decision about whether to party or relax over the weekend. Sam prefers to party, but is worried about getting sick. Such a problem can be modeled as an MDP with two states, healthy and sick, and two actions, relax and party. Thus*

$$S = \{healthy, sick\}$$
$$A = \{relax, party\}$$

Based on experience, Sam estimate that the dynamics $P(s' \mid s, a)$ is given by

S	A	***Probability of*** $s' = healthy$
$healthy$	$relax$	*0.95*
$healthy$	$party$	*0.7*
$sick$	$relax$	*0.5*
$sick$	$party$	*0.1*

So, if Sam is healthy and parties, there is a 30% chance of becoming sick. If Sam is healthy and relaxes, Sam will more likely remain healthy. If Sam is sick and relaxes, there is a 50% chance of getting better. If Sam is sick and parties, there is only a 10% chance of becoming healthy.

Sam estimates the (immediate) rewards to be:

S	A	***Reward***
$healthy$	$relax$	*7*
$healthy$	$party$	*10*
$sick$	$relax$	*0*
$sick$	$party$	*2*

Thus, Sam always enjoys partying more than relaxing. However, Sam feels much better overall when healthy, and partying results in being sick more than relaxing does.

The problem is to determine what Sam should do each weekend.

Example 9.31. *Consider the two-state MDP of Example 9.27 with discount $\gamma = 0.8$. We write the value function as $[healthy\_value, sick\_value]$, and the Q-function as $[[healthy\_relax, healthy\_party], [sick\_relax, sick\_party]]$. Suppose initially the value function is $[0, 0]$. The next Q-value is $[[7, 10], [0, 2]]$, so the next value function is $[10, 2]$ (obtained by Sam partying). The next Q-value is then*

State	*Action*	***Value***
$healthy$	$relax$	*7+0.8*(0.95*10+0.05*2) = 14.68*
$healthy$	$party$	*10+0.8*(0.7*10+0.3*2) = 16.08*
$sick$	$relax$	*0+0.8*(0.5*10+0.5*2) = 4.8*
$sick$	$party$	*2+0.8*(0.1*10+0.9*2) = 4.24*

So the next value function is $[16.08, 4.8]$. After 1000 iterations, the value function is $[35.71, 23.81]$. So the Q function is $[[35.10, 35.71], [23.81, 22.0]]$. Therefore, the optimal policy is to party when healthy and relax when sick.

Asynchronous Value Iteration

A common refinement of this algorithm is **asynchronous value iteration**. Rather than sweeping through the states to create a new value function, AVI updates the states one at a time, in any order, and stores the values in a single array. AVI can store either the **Q[s,a]** array or the **V[s]** array.

it converges faster than the value iteration and is the basis of some RL algorithms. Termination can be difficult to determine if the agent must guarantee a particular error, unless it is careful about how the actions and states are selected. Often, this procedure is run indefinitely as an anytime algorithm where it is always prepared to give its vest estimate of the optimal action in a state when asked.

Asynchronous value iteration could also be implemented by storing just the $V[s]$ array. In that case, the algorithm selects a state s and carries out the update:

$$V[s] := \max_a R(s,a) + \gamma * \sum_{s'} P(s' \mid s,a) * V[s'].$$

Although this variant stores less information, it is more difficult to extract the policy. It requires one extra backup to determine which action a results in the maximum value. This can be done using

$$\pi[s] := \arg\max_a R(s,a) + \gamma * \sum_{s'} P(s' \mid s,a) * V[s'].$$

Example 9.33. *In Example 9.32, the state one step up and one step to the left of the +10 reward state only had its value updated after three value iterations, in which each iteration involved a sweep through all of the states.*

In asynchronous value iteration, the $+10$ reward state can be chosen first. Next, the node to its left can be chosen, and its value will be $0.7 * 0.9 * 10 = 6.3$. Next, the node above that node could be chosen, and its value would become $0.7 * 0.9 * 6.3 = 3.969$. Note that it has a value that reflects that it is close to a $+10$ reward after considering 3 states, not 300 states, as does value iteration.

Temporal Differences

To understand RL, we need to consider how to average values that arrive to an agent sequentially

Suppose we have some values v1, v2, v3 etc. The goal is to predict the next value, given the previous values. One way to do this is to have a running approximation of the expected value vi.

For example, given a sequence of students grades, and the aim to predict the next grade, we can sum the total and divide by the number of assessments to get an average.

If we get another grade, we need to add that to our data. We can do this by maintaining a running average.

Let A_k be an estimate of the expected value based on the first k data points $v_1, \ldots, v_k$. A reasonable estimate is the sample average:

$$A_k = \frac{v_1 + \cdots + v_k}{k} \ .$$

Thus,

$$\begin{aligned} k * A_k &= v_1 + \cdots + v_{k-1} + v_k \\ &= (k-1)\, A_{k-1} + v_k. \end{aligned}$$

Dividing by k gives

$$A_k = \left(1 - \frac{1}{k}\right) * A_{k-1} + \frac{v_k}{k} \ .$$

Let $\alpha_k = \frac{1}{k}$; then

$$\begin{aligned} A_k &= (1 - \alpha_k) * A_{k-1} + \alpha_k * v_k \\ &= A_{k-1} + \alpha_k * (v_k - A_{k-1}) \, . \end{aligned}$$

The difference, $v_k - A_{k-1}$, is called the **temporal difference error** or **TD error**; it specifies how different the new value, v_k, is from the old prediction, A_{k-1}. The old estimate, A_{k-1}, is updated by α_k times the TD error to get the new estimate, A_k. The qualitative interpretation of the temporal difference formula is that if the new value is higher than the old prediction, increase the predicted value; if the new value is less than the old prediction, decrease the predicted value. The change is proportional to the difference between the new value and the old prediction. Note that this equation is still valid for the first value, $k = 1$, in which case $A_1 = v_1$.

The above analysis assumes that the values have an equal weight, which isn't always true. For example, old values might be less useful than new values.

In RL, the values are estimates of the effect of actions; more recent values are more accurate than earlier values because the agent is learning, and so they should be weighted more.

We can do this with alpha as a constant **0 < alpha <= 1** that does not depend on **k**. Unfortunately, this does not converge to the average value when there is variability in the values in the sequence.

You could reduce alpha more slowly and potentially have the benefits of both approaches: weighting recent observations more and still converging to the average. You can guarantee convergence if:

$$\sum_{k=1}^{\infty} \alpha_k = \infty \text{ and } \sum_{k=1}^{\infty} \alpha_k^2 < \infty.$$

The first condition is to ensure that random fluctuations and initial conditions get averaged out, and the second guarantees convergence.

One way to give more weight to more recent experiences, but also converge to the average, is to set **αk=(r+1)/(r+k)** for some **r>0**. For the first experience **α1=1**, so it ignores the prior **A0**. If **r=9**, after 11 experiences, **α11=0.5** so it weights that experience as equal to all of its prior experiences. The

parameter r should be set to be appropriate for the domain.

Note that guaranteeing convergence to the average is not compatible with being able to adapt to make better predictions when the underlying process generating the values keeps changing.

On-Policy & Off-Policy Learning

11 January 2021 12:18

Q-Learning

An agent tries to learn the optimal policy from its history of interaction with the environment. A history of an agent is a sequence of state-action rewards:

$$\langle s_0, a_0, r_1, s_1, a_1, r_2, s_2, a_2, r_3, s_3, a_3, r_4, s_4 \ldots \rangle$$

which means that the agent was in state **s0** and did action **a0**, which resulted in it receiving reward **r1** and being in state **s1**, and then it did action **a1**, and so on.

We treat this history of interaction as a sequence of experiences, where an experience is a tuple

$$\langle s, a, r, s' \rangle$$

which means that the agent was in state **s**, took action a and then received reward **r**, ending up in state **s'**. These experiences will be the data from which the agent can learn what to do. As in decision-theoretic planning, the aim is for the agent to maximise its value which is usually the discounted reward.

Remember that **Q*(s, a)** is the expected value (cumulative discounted reward) of doing a in state **s**, and then following the optimal policy.

Q-learning uses temporal differences to estimate the value of **Q*(s, a)**

In Q-learning, the agent maintains a table of **Q[S, A]. Q[s,a]** represents its current estimate of **Q*(s,a)**

An experience provides on data point for the value of **Q(s,a)**. The data point is that the agent received the future value of

$r + \gamma V(s')$, where $V(s') = \max_{a'} Q(s', a')$;

This is the actual current reward plus the discounted estimated future value. This new data point is called a return. The agent can use the temporal difference equation to update its estimate for **Q(s,a)**

$$Q[s,a] := Q[s,a] + \alpha * \left(r + \gamma \max_{a'} Q[s', a'] - Q[s,a] \right)$$

or, equivalently,

$$Q[s,a] := (1 - \alpha) * Q[s,a] + \alpha * \left(r + \gamma \max_{a'} Q[s', a'] \right).$$

Consider this Q-learning controller. The do(a) line specifies the action the controller commands the body to do, and the reward and the resulting state are percepts the controller receives from the body.

```
1: controller Q-learning(S, A, γ, α)
2:     Inputs
3:         S is a set of states
4:         A is a set of actions
5:         γ the discount
6:         α is the step size
7:     Local
8:         real array Q[S, A]
9:         states s, s'
10:        action a
11:    initialize Q[S, A] arbitrarily
12:    observe current state s
13:    repeat
14:        select an action a
15:        do(a)
16:        observe reward r and state s'
17:        Q[s, a] := Q[s, a] + α * (r + γ * max_a' Q[s', a'] − Q[s, a])
18:        s := s'
19:    until termination
```

The Q-learning leans an approximation of the optimal Q-function as long as the agent explores enough, and there is no bound on the number of times it tries an action in any state (it does not always do the same subset of actions in a state)

Consider the example we looked at early with the {health, sick}, {relax, party} problem.

Example 12.3. *Consider the two-state MDP of Example 9.27. The agent knows there are two states* $\{healthy, sick\}$ *and two actions* $\{relax, party\}$*. It does not know the model and it learns from the* s, a, r, s' *experiences. With a discount,* $\gamma = 0.8$*,* $\alpha = 0.3$*, and* Q *initially 0, the following is a possible trace (to a few significant digits and with the states and actions abbreviated):*

s	a	r	s'	$Update = (1-\alpha) * Q[s,a] + \alpha(r + \gamma max_{a'} Q[s', a'])$
he	re	7	he	$Q[he, re] = 0.7 * 0 + 0.3 * (7 + 0.8 * 0) = 2.1$
he	re	7	he	$Q[he, re] = 0.7 * 2.1 + 0.3 * (7 + 0.8 * 2.1) = 4.07$
he	pa	10	he	$Q[he, pa] = 0.7 * 0 + 0.3 * (10 + 0.8 * 4.07) = 3.98$
he	pa	10	si	$Q[he, pa] = 0.7 * 3.98 + 0.3 * (10 + 0.8 * 0) = 5.79$
si	pa	2	si	$Q[si, pa] = 0.7 * 0 + 0.3 * (2 + 0.8 * 0) = 0.06$
si	re	0	si	$Q[si, re] = 0.7 * 0 + 0.3 * (0 + 0.8 * 0.06) = 0.014$
si	re	0	he	$Q[si, re] = 0.7 * 0.014 + 0.3 * (0 + 0.8 * 5.79) = 1.40$

With α *fixed, the* Q*-values will approximate, but not converge to, the values obtained with value iteration in Example 9.31. The smaller* α *is, the closer it will converge to the actual* Q*-values, but the slower it will converge.*

Exploration & Exploitation

The Q-learner controller does not specify what the agent should actually do. The agent learns a Q-function that can be used to determine an optimal action. There are two things that are useful for the agent to do:

- exploit the knowledge taht is has found for the current state s by doing one of the actions a that maximises Q[s,a]
- explore in order to build a better estimate of the optimal Q-function; it should sometimes select a different action from the one that it currently thinks is best

Epsilon-Greedy Strategy

Where **0 <= epsilon <= 1** is the explore probability, is to select the greedy action that maximises Q[s,a] all but **epsilon** of the time, and pick a random action epsilon of the time. It is possible to change **epsilon** over time which is intuitively correct since we'd want to act randomly at the start and gradually act more in line with our knowledge as we accumulate more information about the world.
However, this treats all of the actions, apart from the best action, equivalently.

Soft-Max

If there are few seemingly good actions it may be more sensible to select among the good actions to explore those further, rather than expending effort exploring actions that look less promising. One way to do this is to select an action a with a probability depending on the value of **Q[s,a]**. A common method is to use a Gibbs or Boltzmann distribution, where the probability of selecting action **a** in state **s** is proportional to

$$e^{Q[s,a]/\tau}$$

Thus in state **s** the agent selects action a with probability

$$\frac{e^{Q[s,a]/\tau}}{\sum_a e^{Q[s,a]/\tau}}$$

Where tau is the temperature specifying how randomly we should choose values. When t is high, the actions are chosen in almost equal amounts. As the temperature is reduce, we choose the actions with the highest **Q[s,a]** more often. When tau = 0, we choose the best action always.

Optimism in the Face of uncertainty

Initialise the Q-function to values that encourage exploration. If the Q-values are initialized to high values, the unexplored areas will look good, so that a greedy search will tend to explore. This does encourage exploration, but can cause the agent to hallucinate that some state-action pairs are good for a long time, with no real evidence. A state only begins to look bad when all its actions look bad.

In noisy environments, OITFOU can mean that a good action never gets explored more because by random chance, it gets a low Q-value from which it never recovers.

On-Policy & Off-Policy Learning

Q-learning is an off-policy learner.

- ✧ An off-policy learner learns the value of an optimal policy independently of the agent's actions, as long as it explores enough.
- ✧ An off-policy learner can learn the optimal policy even if it is acting randomly.

- ✧ A learning agent should, however, try to exploit what it has learned by choosing the best action, but it cannot just exploit because then it will not explore enough to find the best action.
- ✧ An off-policy learner does not learn the value of the policy it is following, because the policy it is following includes exploration steps.

However, there may be cases where ignoring what the agent actually does is dangerous: where there are large negative rewards. An alternative is to learn the value of the policy that the agent is actually carrying out, which includes exploration steps, so that it can iteratively be improved.

SARSA

An on-policy reinforcement learning algorithm that estimates the value of the policy being followed.

An experience in SARSA is of the form <s, a, r, s', a'> which means the agent was in state **s** and did action **a**, ending up with reward **r** in state **s**' from which it decided to do **a**'

This provides a new experience to update Q(s,a). The new value that this experience provides is:

$$r + \gamma Q(s', a').$$

The Q-values that SARSA computes depend on the current exploration policy which for example may be greedy with random steps. It can find a different policy that Q-learning in situations when exploring incur large penalties. For example, when a robot goes near the top of the stairs, even if this is an optimal policy, it may be dangerous for exploration steps.

SARSA will discover this and adopt a policy that keeps the robot away from the stairs. It will find a policy that is optimal, taking into account the exploration inherent in the policy.

SARSA is useful when deploying an agent that is exploring in the world. If you want to do offline learning, and then use that policy in an agent that does not explore, Q-learning may be more appropriate.