There is nothing so powerful as truth, and often nothing so strange.
- Daniel Webster (1782-1852)

Overview
- Introduction & Agents
- Search, Heuristics & CSPs
- Adversarial Search
- Logical Knowledge Representation
- Planning & MDPs
- Reinforcement Learning
- Uncertainty & Bayesian Networks
- Machine Learning
- NLP & Special Topics

KR Hypothesis
Any intelligent process will have ingredients that
1) We as external observers interpret as knowledge
2) This knowledge plays a formal, causal & essential role in guiding the behavior

- Brian Smith (paraphrased)

Some KR Languages
- Propositional Logic
- Predicate Calculus
- Frame Systems
- Rules with Certainty Factors
- Bayesian Belief Networks
- Influence Diagrams
- Semantic Networks
- Concept Description Languages
- Non-monotonic Logic

Knowledge Representation
- represent knowledge in a manner that facilitates inferencing (i.e. drawing conclusions) from knowledge.
- Typically based on
  - Logic
  - Probability
  - Logic and Probability

Basic Idea of Logic
- By starting with true assumptions, you can deduce true conclusions.
Knowledge bases

- Knowledge base = set of sentences in a formal language
- Declarative approach to building an agent (or other system):
  - Tell it what it needs to know
  - Then it can ask itself what to do - answers should follow from the KB
- Agents can be viewed at the knowledge level
  i.e., what they know, regardless of how implemented
- Or at the implementation level
  i.e., data structures in KB and algorithms that manipulate them

Deep Space One

- Autonomous diagnosis & repair “Remote Agent”
- Compiled schematic to 7,000 var SAT problem

Muddy Children Problem

- Mom to N children “Don’t get dirty”
- While playing, K≥1 get mud on forehead
- Father: “Some of you are dirty!”
  - Noone raises hand
- Father: “Raise your hand if you are dirty”
  - Noone raises hand
  - …
- Father: “Raise your hand if you are dirty”
  - All dirty children raise hand

Components of KR

- Syntax: defines the sentences in the language
- Semantics: defines the “meaning” of sentences
- Inference Procedure
  - Algorithm
  - Sound?
  - Complete?
  - Complexity
- Knowledge Base

Propositional Logic

- Syntax
  - Atomic sentences: P, Q, ...
  - Connectives: ∧, ∨, ¬, →
- Semantics
  - Truth Tables
- Inference
  - Modus Ponens
  - Resolution
  - DPLL
  - GSAT
- Complexity
Propositional Logic: Syntax

- **Atoms**
  - P, Q, R, ...
- **Literals**
  - P, ¬P
- **Sentences**
  - Any literal is a sentence
  - If S is a sentence
    - Then (S ∧ S) is a sentence
    - Then (S ∨ S) is a sentence
- **Conveniences**
  - P → Q same as ¬P ∨ Q

Propositional Logic: SEMANTICS

- “Interpretation” (or “possible world”)
  - Assignment to each variable either T or F
  - Assignment of T or F to each connective via defns

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>Q</th>
<th>P ∧ Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>Q</th>
<th>P ∨ Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

Satisfiability, Validity, & Entailment

- **S is satisfiable** if it is true in some world
- **S is unsatisfiable** if it is false all worlds
- **S is valid** if it is true in all worlds
- **S1 entails S2** if wherever S1 is true S2 is also true

Examples

- P → Q
- X → X
- S ∧ (W ∧ ¬S)
- T ∨ ¬T

Notation

- Implication (syntactic symbol)
- Proves: S1 |-i S2 if inference algo, i, says ’S2’ from S1
- Entails: S1 |= S2 if wherever S1 is true S2 is also true
Resolution
If the unicorn is mythical, then it is immortal, but if it is not mythical, it is a reptile. If the unicorn is either immortal or a reptile, then it is horned.

\[ \neg M \lor I \oplus (\neg I \lor H) \]

\[ M = \text{mythical} \]
\[ I = \text{immortal} \]
\[ R = \text{reptile} \]
\[ H = \text{horned} \]

Prop. Logic: Knowledge Engr
1) One of the women is a biology major
2) Lisa is not next to Dave in the ranking
3) Dave is immediately ahead of Jim
4) Jim is immediately ahead of a bio major
5) Mary or Lisa is ranked first

1. Choose Vocabulary
   Universe: Lisa, Dave, Jim, Mary
   LD = “Lisa is immediately ahead of Dave”
   D = “Dave is a Bio Major”

2. Choose initial sentences (wffs)

Reasoning Tasks
- Model finding
  - KB = background knowledge
  - S = description of problem
  - Show (KB \land S) is satisfiable
  - A kind of constraint satisfaction
- Deduction
  - S = question
  - Prove that KB \models S
  - Two approaches:
    1. Rules to derive new formulas from old (inference)
    2. Show (KB \land \neg S) is unsatisfiable

Special Syntactic Forms
- General Form:
  \[ ((q \land \neg r) \to s) \land \neg (s \land t) \]
- Conjunction Normal Form (CNF)
  \[ \neg q \land \neg r \land \neg s \land \neg t \]
  - Set notation: \{ (\neg q, r, s), (\neg s, \neg t) \}
  - empty clause () = false
- Binary clauses: 1 or 2 literals per clause
  \[ \neg q \land r \land \neg s \land t \]
- Horn clauses: 0 or 1 positive literal per clause
  \[ \neg q \land \neg r \land s \land t \]
  \[ (q \land r) \to s \land (s \land t) \to false \]

Propositional Logic: Inference

A mechanical process for computing new sentences
1. Backward & Forward Chaining
2. Resolution (Proof by Contradiction)
3. GSAT
4. Davis Putnam

Inference 1: Forward Chaining

Forward Chaining
Based on rule of modus ponens
If know P₁, ... Pₙ & know (P₁ \land ... \land Pₙ) \to Q
Then can conclude Q

Backward Chaining: search
start from the query and go backwards
Analysis

- Sound?
- Complete?

Can you prove

\[
\{ \} \models \neg Q \lor Q
\]

- If KB has only Horn clauses & query is a single literal
  - Forward Chaining is complete
  - Runs linear in the size of the KB

Propositional Logic: Inference

A mechanical process for computing new sentences

1. Backward & Forward Chaining
2. Resolution (Proof by Contradiction)
3. GSAT
4. Davis Putnam

Conversion to CNF

\[ B_{1,4} = (P_2 \lor P_3) \]

1. Eliminate \( \Rightarrow \), replacing \( \alpha \Rightarrow \beta \) with \((\alpha \Rightarrow \beta) \land (\beta \Rightarrow \alpha)\).

\[ (B_{1,4} = ((P_2 \lor P_3) \land ((P_2 \lor P_3) = B_{1,4}) \]

2. Eliminate \( \Leftarrow \), replacing \( \alpha \Leftarrow \beta \) with \( \neg \alpha \lor \neg \beta \).

\[ (\neg (P_2 \lor P_3) \lor (\neg (P_2 \lor P_3) \lor B_{1,4})) \]

3. Move \( \land \) inwards using de Morgan’s rules and double-negation:

\[ (\neg B_{1,4} \lor P_2 \lor P_3) \land (\neg P_2 \lor \neg P_3 \lor B_{1,4}) \]

4. Apply distributivity law (\( \lor \) over \( \land \)) and flatten:

\[ (\neg B_{1,4} \lor P_2 \lor P_3) \land (\neg P_2 \lor B_{1,3}) \land (\neg P_3 \lor B_{1,3}) \]

Inference 2: Resolution

[Robinson 1965]

\[
\{(p \lor \alpha), (\neg p \lor \beta \lor \gamma)\} \models_R (\alpha \lor \beta \lor \gamma)
\]

Correctness

If \( S_1 \models_R S_2 \) then \( S_1 \models S_2 \)

Refutation Completeness:

If \( S \) is unsatisfiable then \( S \models_R () \)

Resolution

If the unicorn is mythical, then it is immortal; but if it is not mythical, it is a reptile. If the unicorn is either immortal or a reptile, then it is horned.

Prove: the unicorn is horned.

Prove: the unicorn is horned.

\[
\begin{align*}
M &= \text{mythical} \\
I &= \text{immortal} \\
R &= \text{reptile} \\
H &= \text{horned}
\end{align*}
\]

Resolution as Search

- States?
- Operators