Set 6: Knowledge Representation: The Propositional Calculus
Outline

• Representing knowledge using logic
  – Agent that reason logically
  – A knowledge based agent

• Representing and reasoning with logic
  – Propositional logic
    • Syntax
    • Semantic
    • validity and models
    • Rules of inference for propositional logic
    • Resolution
    • Complexity of propositional inference.

• Reading: Russel and Norvig, Chapter 7
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
Knowledge Representation

Defined by: syntax, semantics

Reasoning: in the syntactic level
Example: \( x > y, y > z \models x > z \)
The party example

• If Alex goes, then Beki goes: A → B
• If Chris goes, then Alex goes: C → A
• Beki does not go: not B
• Chris goes: C
• Query: Is it possible to satisfy all these conditions?

• Should I go to the party?
Example of languages

- **Programming languages:**
  - Formal languages, not ambiguous, but cannot express partial information. Not expressive enough.

- **Natural languages:**
  - Very expressive but ambiguous: ex: small dogs and cats.

- **Good representation language:**
  - Both formal and can express partial information, can accommodate inference

- Main approach used in AI: logic-based
Wumpus World test-bed

- **Performance measure**
  - gold +1000, death -1000
  - -1 per step, -10 for using the arrow

- **Environment**
  - Squares adjacent to wumpus are smelly
  - Squares adjacent to pit are breezy
  - Glitter iff gold is in the same square
  - Shooting kills wumpus if you are facing it
  - Shooting uses up the only arrow
  - Grabbing picks up gold if in same square
  - Releasing drops the gold in same square

- **Sensors:** Stench, Breeze, Glitter, Bump, Scream

- **Actuators:** Left turn, Right turn, Forward, Grab, Release, Shoot
Wumpus world characterization

- **Fully Observable** No – only *local* perception
- **Deterministic** Yes – outcomes exactly specified
- **Episodic** No – sequential at the level of actions
- **Static** Yes – Wumpus and Pits do not move
- **Discrete** Yes
- **Single-agent?** Yes – Wumpus is essentially a natural feature
Exploring a wumpus world
Exploring a wumpus world
Exploring a wumpus world
Exploring a wumpus world
Exploring a wumpus world
Exploring a wumpus world
Exploring a wumpus world
Exploring a wumpus world
Logic in general

- **Logics** are formal languages for representing information such that conclusions can be drawn.
- **Syntax** defines the sentences in the language.
- **Semantics** define the "meaning" of sentences;
  - i.e., define **truth** of a sentence in a world.
- **E.g.,** the language of arithmetic:
  - $x + 2 \geq y$ is a sentence; $x^2 + y > \{}$ is not a sentence.
  - $x + 2 \geq y$ is true iff the number $x + 2$ is no less than the number $y$.
  - $x + 2 \geq y$ is true in a world where $x = 7$, $y = 1$.
  - $x + 2 \geq y$ is false in a world where $x = 0$, $y = 6$. 
Entailment

- Entailment means that one thing follows from another:

\[ KB \models \alpha \]

- Knowledge base \( KB \) entails sentence \( \alpha \) if and only if \( \alpha \) is true in all worlds where \( KB \) is true

  - E.g., the KB containing “the Giants won” and “the Reds won” entails “Either the Giants won or the Reds won”

  - E.g., \( x+y = 4 \) entails \( 4 = x+y \)

- Entailment is a relationship between sentences (i.e., syntax) that is based on semantics
Models

• Logicians typically think in terms of models, which are formally structured worlds with respect to which truth can be evaluated.

• We say \( m \) is a model of a sentence \( \alpha \) if \( \alpha \) is true in \( m \).

• \( M(\alpha) \) is the set of all models of \( \alpha \).

• Then \( KB \models \alpha \) iff \( M(KB) \subseteq M(\alpha) \).

  – E.g. \( KB = \) Giants won and Reds won \( \alpha = \) Giants won.
Entailment in the wumpus world

Situation after detecting nothing in [1,1], moving right, breeze in [2,1]

Consider possible models for $KB$ assuming only pits

3 Boolean choices $\Rightarrow$ 8 possible models
Wumpus models
Wumpus models

- \( KB = \) wumpus-world rules + observations
Wumpus models

- $KB = \text{wumpus-world rules} + \text{observations}$
- $\alpha_1 = \text{"[1,2] is safe"}$, $KB \models \alpha_1$, proved by model checking
Wumpus models

- $KB = \text{wumpus-world rules} + \text{observations}$
Wumpus models

- $KB = \text{wumpus-world rules} + \text{observations}$
- $\alpha_2 = \"[2,2] is safe\", KB \not\models \alpha_2$
Inference

$KB \vdash_i \alpha$ = sentence $\alpha$ can be derived from $KB$ by procedure $i$

Consequences of $KB$ are a haystack; $\alpha$ is a needle.
Entailment = needle in haystack; inference = finding it

**Soundness:** $i$ is sound if
whenever $KB \vdash_i \alpha$, it is also true that $KB \models \alpha$

**Completeness:** $i$ is complete if
whenever $KB \models \alpha$, it is also true that $KB \vdash_i \alpha$

Preview: we will define a logic (first-order 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 $KB$. 
Propositional logic: Syntax

- Propositional logic is the simplest logic – illustrates basic ideas

- The proposition symbols $P_1$, $P_2$ etc are sentences
  - If $S$ is a sentence, $\neg S$ is a sentence (negation)
  - If $S_1$ and $S_2$ are sentences, $S_1 \land S_2$ is a sentence (conjunction)
  - If $S_1$ and $S_2$ are sentences, $S_1 \lor S_2$ is a sentence (disjunction)
  - If $S_1$ and $S_2$ are sentences, $S_1 \Rightarrow S_2$ is a sentence (implication)
  - If $S_1$ and $S_2$ are sentences, $S_1 \Leftrightarrow S_2$ is a sentence (biconditional)
Propositional logic: Semantics

Each model specifies true/false for each proposition symbol

E.g. \( P_{1,2} \) false \( P_{2,2} \) true \( P_{3,1} \) false

With these symbols, 8 possible models, can be enumerated automatically.

Rules for evaluating truth with respect to a model \( m \):

- \( \neg S \) is true iff \( S \) is false
- \( S_1 \land S_2 \) is true iff \( S_1 \) is true and \( S_2 \) is true
- \( S_1 \lor S_2 \) is true iff \( S_1 \) is true or \( S_2 \) is true
- \( S_1 \Rightarrow S_2 \) is true iff \( S_1 \) is false or \( S_2 \) is true
  \( \text{i.e.,} \) is false iff \( S_1 \) is true and \( S_2 \) is false
- \( S_1 \Leftrightarrow S_2 \) is true iff \( S_1 \Rightarrow S_2 \) is true and \( S_2 \Rightarrow S_1 \) is true

Simple recursive process evaluates an arbitrary sentence, e.g.,

\[ \neg P_{1,2} \land (P_{2,2} \lor P_{3,1}) = true \land (true \lor false) = true \land true = true \]
# Truth tables for connectives

<table>
<thead>
<tr>
<th>$P$</th>
<th>$Q$</th>
<th>$\neg P$</th>
<th>$P \land Q$</th>
<th>$P \lor Q$</th>
<th>$P \implies Q$</th>
<th>$P \iff Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>false</td>
<td>false</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
</tr>
</tbody>
</table>
Wumpus world sentences

Let $P_{i,j}$ be true if there is a pit in $[i, j]$.
Let $B_{i,j}$ be true if there is a breeze in $[i, j]$.

- $\neg P_{1,1}$
- $\neg B_{1,1}$
- $B_{2,1}$

- "Pits cause breezes in adjacent squares"

$$B_{1,1} \iff (P_{1,2} \lor P_{2,1})$$
$$B_{2,1} \iff (P_{1,1} \lor P_{2,2} \lor P_{3,1})$$
Wumpus world sentences

Let $P_{i,j}$ be true if there is a pit in $[i, j]$.
Let $B_{i,j}$ be true if there is a breeze in $[i, j]$.

- $\neg P_{1,1}$
- $\neg B_{1,1}$
- $B_{2,1}$

- "Pits cause breezes in adjacent squares"

<table>
<thead>
<tr>
<th>$B_{1,1}$</th>
<th>$B_{2,1}$</th>
<th>$P_{1,1}$</th>
<th>$P_{1,2}$</th>
<th>$P_{2,1}$</th>
<th>$P_{2,2}$</th>
<th>$P_{3,1}$</th>
<th>$KB$</th>
<th>$\alpha_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>false</td>
</tr>
</tbody>
</table>

Alpha1 = no pit in (1,2)
alpha2: no pit in (2,2)
Inference by enumeration

- Depth-first enumeration of all models is sound and complete

```plaintext
function TT-ENTAILS?(KB, α) returns true or false
  symbols ← a list of the proposition symbols in KB and α
  return TT-CHECK-ALL(KB, α, symbols, [])

function TT-CHECK-ALL(KB, α, symbols, model) returns true or false
  if EMPTY?(symbols) then
    if PL-TRUE?(KB, model) then return PL-TRUE?(α, model)
  else return true
  else do
    P ← FIRST(symbols); rest ← REST(symbols)
    return TT-CHECK-ALL(KB, α, rest, EXTEND(P, true, model) and TT-CHECK-ALL(KB, α, rest, EXTEND(P, false, model))
```

- For $n$ symbols, time complexity is $O(2^n)$, space complexity is $O(n)$
Logical equivalence

Two sentences are **logically equivalent** iff true in same models: \( \alpha \equiv \beta \iff \alpha \vDash \beta \) and \( \beta \vDash \alpha \)

- \((\alpha \land \beta) \equiv (\beta \land \alpha)\) commutativity of \(\land\)
- \((\alpha \lor \beta) \equiv (\beta \lor \alpha)\) commutativity of \(\lor\)
- \(((\alpha \land \beta) \land \gamma) \equiv (\alpha \land (\beta \land \gamma))\) associativity of \(\land\)
- \(((\alpha \lor \beta) \lor \gamma) \equiv (\alpha \lor (\beta \lor \gamma))\) associativity of \(\lor\)
- \(\neg(\neg\alpha) \equiv \alpha\) double-negation elimination
- \((\alpha \Rightarrow \beta) \equiv (\neg\beta \Rightarrow \neg\alpha)\) contraposition
- \((\alpha \Rightarrow \beta) \equiv (\neg\alpha \lor \beta)\) implication elimination
- \((\alpha \iff \beta) \equiv ((\alpha \Rightarrow \beta) \land (\beta \Rightarrow \alpha))\) biconditional elimination
- \(\neg(\alpha \land \beta) \equiv (\neg\alpha \lor \neg\beta)\) de Morgan
- \(\neg(\alpha \lor \beta) \equiv (\neg\alpha \land \neg\beta)\) de Morgan
- \((\alpha \land (\beta \lor \gamma)) \equiv ((\alpha \land \beta) \lor (\alpha \land \gamma))\) distributivity of \(\land\) over \(\lor\)
- \((\alpha \lor (\beta \land \gamma)) \equiv ((\alpha \lor \beta) \land (\alpha \lor \gamma))\) distributivity of \(\lor\) over \(\land\)
Validity and satisfiability

A sentence is valid if it is true in all models,
e.g., True,  A ∨ ¬A,  A ⇒ A,  (A ∧ (A ⇒ B)) ⇒ B

Validity is connected to inference via the Deduction Theorem:
KB ⊢ α if and only if (KB ⇒ α) is valid

A sentence is satisfiable if it is true in some model
e.g., A ∨ B,  C

A sentence is unsatisfiable if it is true in no models
e.g., A ∧ ¬A

Satisfiability is connected to inference via the following:
KB ⊢ α if and only if (KB ∧ ¬α) is unsatisfiable
Validity

<table>
<thead>
<tr>
<th>$P$</th>
<th>$H$</th>
<th>$P \lor H$</th>
<th>$(P \lor H) \land \neg H$</th>
<th>$((P \lor H) \land \neg H) \Rightarrow P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>True</td>
</tr>
<tr>
<td>False</td>
<td>True</td>
<td>True</td>
<td>False</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>True</td>
<td>True</td>
<td>False</td>
<td>True</td>
</tr>
</tbody>
</table>

**Figure 6.10** Truth table showing validity of a complex sentence.
Truth Tables

• Truth tables can be used to compute the truth value of any wff (well formed formula)
• Can be used to find the truth of \( ((P \rightarrow R) \rightarrow Q) \vee \neg S \)
• Given n features there are \( 2^n \) different worlds, different interpretations.
• Interpretation: any assignment of true and false to atoms
• An interpretation satisfies a wff (sentence) if the sentence is assigned true under the interpretation
• A model: An interpretation is a model of a sentence if the sentence is satisfied in that interpretation.
• Satisfiability of a sentence can be determined by the truth-table
  – Bat_on and turns-key_on \( \rightarrow \) Engine-starts
• A sentence is unsatisfiable or inconsistent it has no models
  – \( P \land (\neg P) \)
  – \( (P \lor Q) \land (P \lor \neg Q) \land (\neg P \lor Q) \land (\neg P \lor \neg Q) \)
Rules of inference

◊ Modus Ponens or Implication-Elimination: (From an implication and the premise of the implication, you can infer the conclusion.)
\[
\begin{align*}
\alpha & \Rightarrow \beta, \quad \alpha \\
\hline 
\alpha & \Rightarrow \beta
\end{align*}
\]

◊ And-Elimination: (From a conjunction, you can infer any of the conjuncts.)
\[
\begin{align*}
\alpha_1 \land \alpha_2 \land \ldots \land \alpha_n \\
\hline 
\alpha_i
\end{align*}
\]

◊ And-Introduction: (From a list of sentences, you can infer their conjunction.)
\[
\begin{align*}
\alpha_1, \alpha_2, \ldots, \alpha_n \\
\hline 
\alpha_1 \land \alpha_2 \land \ldots \land \alpha_n
\end{align*}
\]

◊ Or-Introduction: (From a sentence, you can infer its disjunction with anything else at all.)
\[
\begin{align*}
\alpha_i \\
\hline 
\alpha_i \lor \alpha_2 \lor \ldots \lor \alpha_n
\end{align*}
\]

◊ Double-Negation Elimination: (From a doubly negated sentence, you can infer a positive sentence.)
\[
\begin{align*}
\neg \neg \alpha \\
\hline 
\alpha
\end{align*}
\]

◊ Unit Resolution: (From a disjunction, if one of the disjuncts is false, then you can infer the other one is true.)
\[
\begin{align*}
\alpha \lor \beta, \quad \neg \beta \\
\hline 
\alpha
\end{align*}
\]

◊ Resolution: (This is the most difficult. Because \( \beta \) cannot be both true and false, one of the other disjuncts must be true in one of the premises. Or equivalently, implication is transitive.)
\[
\begin{align*}
\alpha \lor \beta, \quad \neg \beta \lor \gamma \\
\hline 
\alpha \lor \gamma
\end{align*}
\] or equivalently
\[
\begin{align*}
\neg \alpha \Rightarrow \beta, \quad \beta \Rightarrow \gamma \\
\hline 
\neg \alpha \Rightarrow \gamma
\end{align*}
\]

Figure 6.13  Seven inference rules for propositional logic. The unit resolution rule is a special case of the resolution rule, which in turn is a special case of the full resolution rule for first-order logic discussed in Chapter 9.
Proof methods

• Proof methods divide into (roughly) two kinds:

  – Application of inference rules
    • Legitimate (sound) generation of new sentences from old
    • Proof = a sequence of inference rule applications
      Can use inference rules as operators in a standard search algorithm
    • Typically require transformation of sentences into a normal form

  – Model checking
    • truth table enumeration (always exponential in $n$)
    • improved backtracking, e.g., Davis--Putnam-Logemann-Loveland (DPLL),
      Backtracking with constraint propagation, backjumping.
    • heuristic search in model space (sound but incomplete)
      e.g., min-conflicts-like hill-climbing algorithms
Resolution in Propositional Calculus

• **Using clauses as wffs**
  – Literal, clauses, conjunction of clauses (cnfs) \((P \lor Q \lor \neg R)\)
• **Resolution rule:**
• **Resolving** \((P \lor Q)\) and \((P \lor \neg Q)\) \(\models \neg P\)
  – Generalize modus ponens, chaining .
  – Resolving a literal with its negation yields empty clause.
• **Resolution is sound**
• **Resolution is NOT complete:**
  – \(P\) and \(R\) entails \(P \lor R\) but you cannot infer \(P \lor R\) From \((P \land R)\) by resolution
• **Resolution is complete for refutation:** adding \((\neg P)\) and \((\neg R)\) to \((P \land R)\) we can infer the empty clause.
• **Decidability of propositional calculus by resolution refutation:** if a sentence \(w\) is not entailed by \(KB\) then resolution refutation will terminate without generating the empty clause.
Resolution

Conjunctive Normal Form (CNF—universal)

\[ \text{conjunction of disjunctions of literals} \]

\[ \text{clauses} \]

E.g., \((A \lor \neg B) \land (B \lor \neg C \lor \neg D)\)

Resolution inference rule (for CNF): complete for propositional logic

\[
\frac{\ell_1 \lor \cdots \lor \ell_k, \quad m_1 \lor \cdots \lor m_n}{\ell_1 \lor \cdots \lor \ell_{i-1} \lor \ell_{i+1} \lor \cdots \lor \ell_k \lor m_1 \lor \cdots \lor m_{j-1} \lor m_{j+1} \lor \cdots \lor m_n}
\]

where \(\ell_i\) and \(m_j\) are complementary literals. E.g.,

\[
\frac{P_{1,3} \lor P_{2,2}, \quad \neg P_{2,2}}{P_{1,3}}
\]

Resolution is sound and complete for propositional logic
Conversion to CNF

\[ B_{1,1} \iff (P_{1,2} \lor P_{2,1}) \]

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

\[ (B_{1,1} \Rightarrow (P_{1,2} \lor P_{2,1})) \land ((P_{1,2} \lor P_{2,1}) \Rightarrow B_{1,1}) \]

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

\[ (\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land ((\neg P_{1,2} \lor P_{2,1}) \lor B_{1,1}) \]

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

\[ (\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land ((\neg P_{1,2} \land \neg P_{2,1}) \lor B_{1,1}) \]

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

\[ (\neg B_{1,1} \lor P_{1,2} \lor P_{2,1}) \land ((\neg P_{1,2} \lor \neg B_{1,1}) \lor (\neg P_{2,1} \lor B_{1,1})) \]
Resolution algorithm

- Proof by contradiction, i.e., show $KB \land \neg \alpha$ unsatisfiable

```plaintext
function PL-RESOLUTION(KB, \alpha) returns true or false
    clauses ← the set of clauses in the CNF representation of $KB \land \neg \alpha$
    new ← { }
    loop do
        for each $C_i, C_j$ in clauses do
            resolvents ← PL-RESOLVE($C_i, C_j$)
            if resolvents contains the empty clause then return true
            new ← new \cup resolvents
        if new \subseteq clauses then return false
    clauses ← clauses \cup new
```
Resolution example

• $KB = (B_{1,1} \iff (P_{1,2} \lor P_{2,1})) \land \neg B_{1,1}, \quad \alpha = \neg P_{1,2}$
Soundness of resolution

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
<th>$\alpha \lor \beta$</th>
<th>$\neg \beta \lor \gamma$</th>
<th>$\alpha \lor \gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td>False</td>
<td>False</td>
<td>True</td>
<td>False</td>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td>False</td>
<td>True</td>
<td>False</td>
<td>True</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>False</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>True</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
</tbody>
</table>

**Figure 6.14** A truth table demonstrating the soundness of the resolution inference rule. We have underlined the rows where both premises are true.
The party example

• If Alex goes, then Beki goes: $A \rightarrow B$
• If Chris goes, then Alex goes: $C \rightarrow A$
• Beki does not go: $\neg B$
• Chris goes: $C$
• Query: Is it possible to satisfy all these conditions?

• Should I go to the party?
Example of proof by Refutation

• Assume the claim is false and prove inconsistency:
  – Example: can we prove that Chris will not come to the party?

• Prove by generating the desired goal.

• Prove by refutation: add the negation of the goal and prove no model

• Proof:
Proof by refutation

• Given a database in clausal normal form KB
  – Find a sequence of resolution steps from KB to the empty clauses
• Use the search space paradigm:
  – States: current cnf KB + new clauses
  – Operators: resolution
  – Initial state: KB + negated goal
  – Goal State: a database containing the empty clause
  – Search using any search method
Proof by refutation (contd.)

• Or:
  – Prove that KB has no model - PSAT
    • A cnf theory is a constraint satisfaction problem:
      – variables: the propositions
      – domains: true, false
      – constraints: clauses (or their truth tables)
      – Find a solution to the csp. If no solution no model.
      – This is the satisfiability question
      – Methods: Backtracking arc-consistency ≈ unit resolution, local search
Resolution refutation search strategies

- **Ordering strategies**
  - Breadth-first, depth-first
  - I-level resolvents are generated from level-(I-1) or less resolvents
  - Unit-preference: prefer resolutions with a literal

- **Set of support**:
  - Allows resolutions in which one of the resolvents is in the set of support
  - The set of support: those clauses coming from negation of the theorem or their descendents.
  - The set of support strategy is refutation complete

- **Linear input**
  - Restricted to resolutions when one member is in the input clauses
  - Linear input is not refutation complete
  - Example: (PVQ) (P V not Q) (not P V Q) (not P V not Q) have no model
Complexity of propositional inference

- Checking truth tables is exponential
- Satisfiability is NP-complete
- However, frequently generating proofs is easy.
- Propositional logic is monotonic
  - If you can entail alpha from knowledge base KB and if you add sentences to KB, you can infer alpha from the extended knowledge-base as well.
- Inference is local
  - Tractable Classes: Horn, 2-SAT
- Horn theories:
  - Q ← P₁, P₂, ..., Pₙ,
  - Pi is an atom in the language, Q can be false.
  - : only implications. Clauses have a positive literal
- Solved by modus ponens or “unit resolution”.
Forward and backward chaining

Horn Form (restricted)
\[ \text{KB} = \textit{conjunction of Horn clauses} \]
Horn clause =
\[ \diamond \text{ proposition symbol; or} \]
\[ \diamond \text{ (conjunction of symbols) } \Rightarrow \text{ symbol} \]
E.g., \( C \land (B \Rightarrow A) \land (C \land D \Rightarrow B) \)

Modus Ponens (for Horn Form): complete for Horn KBs
\[
\alpha_1, \ldots, \alpha_n, \quad \alpha_1 \land \cdots \land \alpha_n \Rightarrow \beta
\]
\[
\beta
\]
Can be used with forward chaining or backward chaining.
These algorithms are very natural and run in linear time
Forward chaining algorithm

function PL-FC-ENTAILS?(KB, q) returns true or false

local variables: count, a table, indexed by clause, initially the number of premises
inferred, a table, indexed by symbol, each entry initially false
agenda, a list of symbols, initially the symbols known to be true

while agenda is not empty do
    p ← POP(agenda)
    unless inferred[p] do
        inferred[p] ← true
        for each Horn clause c in whose premise p appears do
            decrement count[c]
            if count[c] = 0 then do
                if HEAD[c] = q then return true
                PUSH(HEAD[c], agenda)
    return false

• Forward chaining is sound and complete for Horn KB
Forward chaining

- Idea: fire any rule whose premises are satisfied in the $KB$,
  - add its conclusion to the $KB$, until query is found
Forward chaining example

\[ P \Rightarrow Q \]
\[ L \land M \Rightarrow P \]
\[ B \land L \Rightarrow M \]
\[ A \land P \Rightarrow L \]
\[ A \land B \Rightarrow L \]
\[ A \]
\[ B \]
Forward chaining example
Forward chaining example
Forward chaining example
Forward chaining example
Forward chaining example
Forward chaining example
Forward chaining example
Backward chaining (BC)

Idea: work backwards from the query $q$:

- to prove $q$ by BC,
  check if $q$ is known already, or
  prove by BC all premises of some rule concluding $q$

Avoid loops: check if new subgoal is already on the goal stack

Avoid repeated work: check if new subgoal

1. has already been proved true, or
2. has already failed
Backward chaining example

\[ P \Rightarrow Q \]
\[ L \wedge M \Rightarrow P \]
\[ B \wedge L \Rightarrow M \]
\[ A \wedge P \Rightarrow L \]
\[ A \wedge B \Rightarrow L \]
\[ A \]
\[ B \]
Backward chaining example
Backward chaining example
Backward chaining example
Backward chaining example
Backward chaining example
Backward chaining example
Backward chaining example
Backward chaining example
Backward chaining example
Forward vs. backward chaining

• FC is data-driven, automatic, unconscious processing,
  – e.g., object recognition, routine decisions

• May do lots of work that is irrelevant to the goal

• BC is goal-driven, appropriate for problem-solving,
  – e.g., Where are my keys? How do I get into a PhD program?

• Complexity of BC can be much less than linear in size of KB
Efficient propositional inference

Two families of efficient algorithms for propositional inference:

Complete backtracking search algorithms
• DPLL algorithm (Davis, Putnam, Logemann, Loveland)

• Incomplete local search algorithms
  – WalkSAT algorithm
The DPLL algorithm

Determine if an input propositional logic sentence (in CNF) is satisfiable.

Improvements over truth table enumeration:

1. Early termination
   A clause is true if any literal is true.
   A sentence is false if any clause is false.

2. Pure symbol heuristic
   Pure symbol: always appears with the same "sign" in all clauses.
   e.g., In the three clauses (A v ¬B), (¬B v ¬C), (C v A), A and B are pure, C is impure.
   Make a pure symbol literal true.

3. Unit clause heuristic
   Unit clause: only one literal in the clause
   The only literal in a unit clause must be true.
The DPLL algorithm

function DPLL-SATISFIABLE?(s) returns true or false
  inputs: s, a sentence in propositional logic
  clauses ← the set of clauses in the CNF representation of s
  symbols ← a list of the proposition symbols in s
  return DPLL(clauses, symbols, [])

function DPLL(clauses, symbols, model) returns true or false
  if every clause in clauses is true in model then return true
  if some clause in clauses is false in model then return false
  P, value ← FIND-PURE-SYMBOL(symbols, clauses, model)
  if P is non-null then return DPLL(clauses, symbols−P, [P = value | model])
  P, value ← FIND-UNIT-CLAUSE(clauses, model)
  if P is non-null then return DPLL(clauses, symbols−P, [P = value | model])
  P ← FIRST(symbols); rest ← REST(symbols)
  return DPLL(clauses, rest, [P = true | model]) or DPLL(clauses, rest, [P = false | model])
The WalkSAT algorithm

• Incomplete, local search algorithm

• Evaluation function: The min-conflict heuristic of minimizing the number of unsatisfied clauses

• Balance between greediness and randomness
The \texttt{WalkSAT} algorithm

\begin{verbatim}
function \texttt{WalkSAT}(\texttt{clauses}, \texttt{p}, \texttt{max-flips}) returns a satisfying model or \texttt{failure}
inputs: \texttt{clauses}, a set of clauses in propositional logic
         \texttt{p}, the probability of choosing to do a “random walk” move
         \texttt{max-flips}, number of flips allowed before giving up

\texttt{model} \leftarrow \text{a random assignment of true/false to the symbols in clauses}
\texttt{for} \texttt{i} = 1 \text{ to } \texttt{max-flips} \text{ do}
   \text{if model satisfies clauses then return model}
   \text{clause} \leftarrow \text{a randomly selected clause from clauses that is false in model}
   \texttt{with probability} \texttt{p} \text{ flip the value in model of a randomly selected symbol}
   \text{from clause}
   \text{else flip whichever symbol in clause maximizes the number of satisfied clauses}
\texttt{return failure}
\end{verbatim}
Hard satisfiability problems

• Consider random 3-CNF sentences. e.g.,

\[(\neg D \lor \neg B \lor C) \land (B \lor \neg A \lor \neg C) \land (\neg C \lor \neg B \lor E) \land (E \lor \neg D \lor B) \land (B \lor E \lor \neg C)\]

\[m = \text{number of clauses}\]
\[n = \text{number of symbols}\]

– Hard problems seem to cluster near \(m/n = 4.3\) (critical point)
Hard satisfiability problems
Hard satisfiability problems

- Median runtime for 100 satisfiable random 3-CNF sentences, $n = 50$
Inference-based agents in the wumpus world

A wumpus-world agent using propositional logic:

\[ \neg p_{1,1} \]
\[ \neg w_{1,1} \]
\[ b_{x,y} \iff (p_{x+1,y} \lor p_{x-1,y} \lor p_{x,y+1} \lor p_{x,y-1}) \]
\[ s_{x,y} \iff (w_{x+1,y} \lor w_{x-1,y} \lor w_{x,y+1} \lor w_{x,y-1}) \]
\[ w_{1,1} \lor w_{1,2} \lor \ldots \lor w_{4,4} \]
\[ \neg w_{1,1} \lor \neg w_{1,2} \]
\[ \neg w_{1,1} \lor \neg w_{1,3} \]
\[ \ldots \]

\[ \Rightarrow 64 \text{ distinct proposition symbols, 155 sentences} \]
function PL-WUMPUS-AGENT( percept) returns an action
inputs: percept, a list, [stench,breeze,glitter]
static: KB, initially containing the “physics” of the wumpus world
x, y, orientation, the agent’s position (init. [1,1]) and orient. (init. right)
visited, an array indicating which squares have been visited, initially false
action, the agent’s most recent action, initially null
plan, an action sequence, initially empty

update x,y,orientation, visited based on action
if stench then TELL(KB, S_{x,y}) else TELL(KB, \neg S_{x,y})
if breeze then TELL(KB, B_{x,y}) else TELL(KB, \neg B_{x,y})
if glitter then action ← grab
else if plan is nonempty then action ← POP(plan)
else if for some fringe square [i,j], ASK(KB, (\neg P_{i,j} \land \neg W_{i,j})) is true or
for some fringe square [i,j], ASK(KB, (P_{i,j} \lor W_{i,j})) is false then do
plan ← A*-GRAPH-SEARCH(Route-PB([x,y], orientation, [i,j],visited))
action ← POP(plan)
else action ← a randomly chosen move
return action
Summary

Logical agents apply inference to a knowledge base
to derive new information and make decisions

Basic concepts of logic:
- syntax: formal structure of sentences
- semantics: truth of sentences wrt models
- entailment: necessary truth of one sentence given another
- inference: deriving sentences from other sentences
- soundness: derivations produce only entailed sentences
- completeness: derivations can produce all entailed sentences

Wumpus world requires the ability to represent partial and negated information, reason by cases, etc.

Forward, backward chaining are linear-time, complete for Horn clauses
Resolution is complete for propositional logic

Propositional logic lacks expressive power