Propositional logic, I

Readings :

- CHAPTER 13,14 from Nilsson
- CHAPTER 7 from Russell + Norvig
- Gödel, Escher, Bach: An eternal Golden Braid
Douglas R. Hofstadter

Introdution

- **Representations of the world**
  - » **Iconic (analogical):** simulation of relevant aspects of the environment and of the effects of the agent's actions on the environment.
  - » **Featured-based:** Vector of feature (attribute) values + declarative descriptions of the environment.
    - – **General laws**: "Round objects have no edges".
    - – **Negative information**: "We are not in Tokyo"
    - – **Uncertain information**: "We are either in Antwerp or in Dublin".
    - – **Constraints** on values of the variables: "Grades in this course are between 0 and 10".
- **Reasoning:** Information of the values of certain features can be inferred from the values of other features + knowledge about the problem.

<u>Example</u>: Block-lifting robot in block's world

Features x_1(BATTERY_OK), x_2 (LIFTABLE) x_3 (MOVES)

Robot can sense BATTERY_OK, MOVES, but not LIFTABLE

Knowledge: BATTERY_OK $\wedge$ LIFTABLE $\Rightarrow$ MOVES.

Robot reads {BATTERY_OK (1), MOVES(0)}

and infers LIFTABLE(0)

The elements of logic

- **A formal language**: Symbols + syntactic rules to combine symbols into grammatical sentences.
- **Inference rules**: Typographical rules (they do not use meaning) to construct new sentences from a given group of sentences.
- **Semantics**: Interpretation rules that associate sentences in the language with statements in the domain of discourse.

Forget meaning. Think of logic as a game.

BEWARE: It is easy (and often misleading) to slip into meaning-based (informal) deductions when you are doing a logic (formal) deduction.

Language, I

- **Atoms**: `T ,F`
  `P,Q,P1,P2,P3,...,Pn,...`

- **Parentheses:** (,)

- **Propositional connectives**:
  ¬, ∧, ∨, ⇒, ⇔ (in order of precedence)

- **Well-formed formulas** (wffs or **sentences)**
  - » **An atom is a wff.**
  - » **If** w_1 **and** w_2 are wffs, then so are the expressions

    $\neg w_1$ (negation of w_1).

    $(w_1 \wedge w_2)$ (conjunction of the conjuncts w_1 and w_2).

    $(w_1 \vee w_2)$ (disjunction of the disjuncts w_1 and w_2).

    $(w_1 \Rightarrow w_2)$ (Implication, conditional, rule or if-then statement:
    w_1 is the premise or antecedent,
    w_2 is the conclusion, or consequent).

    $(w_1 \Leftrightarrow w_2)$ (biconditional, or if-and-only-if or iff statement)

Language, II

- **Grammar (in Backus-Naur Form)**
  - » Sentence → AtomicSentence | ComplexSentence
  - » AtomicSentence → `T|F|` Symbol
  - » Symbol → `P|Q|P1|P2|P3|...|Pn,...`
  - » ComplexSentence → ¬Sentence
    (Sentence∧Sentence)
    (Sentence∨Sentence)
    (Sentence⇒Sentence)
    (Sentence⇔Sentence)

- **Note:** This grammar is very strict about parentheses. In the notes that follow, some parentheses will be omitted and the rules of precedence of connectives will be used. Parentheses (,) will be used only to alter precedence rules or to improve the readability of the formulas. They will be considered as extra-linguistic separators that group wffs into sub-wffs.

- **Note:** In US ⊃ is used instead of ⇒

Examples: ((P∧Q)⇒¬P), or P∧Q⇒¬P
(P⇒¬P) , or P⇒¬P
((P∨P)⇒¬P), or P∨P⇒¬P

Inference rules

- **Rules of inference:** Typographical rules that specify how to generate new wffs from a given set of wffs

- A possible set of **rules of inference R is**:
  Let w_1, w_2 be wffs
  (1) *Modus ponens* :
  w_2 can be inferred from w_1 and $w_1 \Rightarrow w_2$
  (2) $\wedge$ introduction :
  $w_1 \wedge w_2$ can be inferred from w_1 and w_2
  (3) Commutativity of $\wedge$:
  $w_2 \wedge w_1$ can be inferred from $w_1 \wedge w_2$
  (4) $\wedge$ elimination :
  w_1 can be inferred from $w_1 \wedge w_2$
  (5) $\vee$ introduction:
  $w_1 \vee w_2$ can be inferred from either w_1 or w_2
  (6) $\neg$ elimination :
  w_1 can be inferred from $\neg\neg w_1$

  This set of rules of inference **R** is **sound** but **not complete**
  ("sound" , "complete" are defined later)

Proofs and theoremhood

- **Proof:** The sequence of wffs
  $$\{w_1, w_2, \ldots, w_{n-1}, w_n\}$$
  is a proof (or deduction) of w_n from a set of wffs Δ with the set of inference rules *R* if each $w_i, i=1,2,\ldots,n-1$ is either in Δ or can be deduced from Δ with *R.*

- **Theoremhood**:
  If there is a proof of w_n from Δ with inference rules *R* , then w_n is a **theorem** of Δ, with inference rules *R*
  $$\Delta \vdash_R w_n$$

Semantics: Truth tables

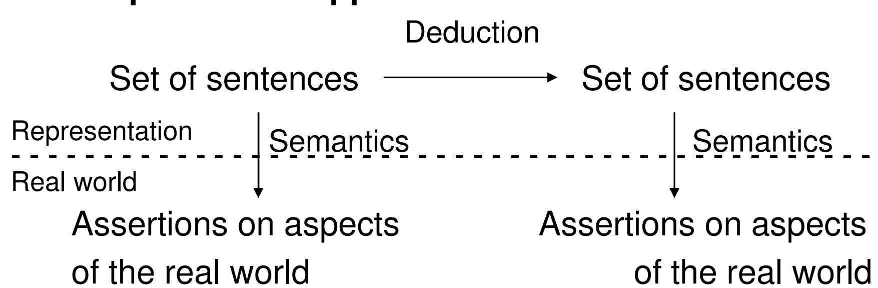
- **Truth tables:** Rules that specify the **truth value of a wff from the truth values of the atoms** that form the wff.
  - » $\neg w_1$ has value *True* if w_1 has value *False.* It has value *False* if w_1 has value *True.*
  - » $w_1 \wedge w_2$ has value *True* only if both w_1 and w_2 have value *True*; otherwise its value is *False.*
  - » $w_1 \vee w_2$ has value True if either w_1 or w_2 have value True; otherwise its value is *False.*
  - » $w_1 \Rightarrow w_2$ has the same truth value as $(\neg w_1 \vee w_2)$
  - » $w_1 \Leftrightarrow w_2$ has the same truth value as $(w_1 \Rightarrow w_2) \wedge (w_2 \Rightarrow w_1)$

w_1	w_2		$\neg w_1$	$w_1 \wedge w_2$	$w_1 \vee w_2$	$w_1 \Rightarrow w_2$	$w_1 \Leftrightarrow w_2$
True	*True*		*False*	*True*	*True*	*True*	*True*
True	*False*		*False*	*False*	*True*	*False*	*False*
False	*True*		*True*	*False*	*True*	*True*	*False*
False	*False*		*True*	*False*	*False*	*True*	*True*

Semantics: Interpretation

- **Interpretation:**
  - » An interpretation is a set of **associations** of **<u>atoms</u> to <u>propositions in the world</u>**.
    - – In an interpretation, the proposition associated to an atom is called the **denotation of that atom**.
  - » **Under** a given **interpretation atoms have truth values** (*True* or *False*) that are **determined by the truth or falsity of the corresponding proposition** in the world.
  - » The special **atom T** has always the value ***True***.
  - » The special **atom F** has always the value ***False***.

- **Semantics (meaning of a wff) ≡ interpretation + application of truth tables.**



Satisfiability

- **Satisfiability:**
  - » An **interpretation satisfies** a **set of wffs** Δ if **all wffs** in the set have value ***True*** under that interpretation. The **interpretation is** said to be **a model of** Δ.
  - » A **set of wffs** Δ is **satisfiable** if there exists at least one interpretation under which **all wffs** in the set Δ **have value *True***.
  - » A **wff w** is **satisfiable** if the **set {w} is satisfiable**.
- **Inconsistency**:
  - » A or **set of wffs** is said to be **inconsistent** or **unsatisfiable** if there is no interpretation under which all wffs in the set are satisfied.
  - » A **wff w is** said to be **unsatisfiable** if the **set {w}** is **unsatisfiable**.

  E.g. $F, P\wedge\neg P, \{P\vee Q, P\vee\neg Q, \neg P\vee Q, \neg P\vee\neg Q\}$

- **Validity:** A wff is said to be valid if it has the value *True* under all possible interpretations.

  Ex. $T, T\vee P, \neg P\vee P, P\Rightarrow P, P\Rightarrow(Q\Rightarrow P), ((P\Rightarrow Q)\Rightarrow P)\Rightarrow P$
  - » A valid wff is a **tautology** (it is devoid of meaning about the world).
- **Metatheorem 1:** if $\neg w$ is unsatisfiable, then the wff w is valid, and viceversa.

Equivalence

Two wffs are said to be logically **equivalent (≡)** if and only if they have the **same truth values under all possible interpretations**.

Using the truth tables the following **properties** obtain

- » **Absorption:** $(w_1\wedge T) \equiv w_1;\ (w_1\vee F) \equiv w_1$
- » **Contradiction (excluded middle):**

  $(w_1\wedge\neg w_1) \equiv F;\quad (w_1\vee\neg w_1) \equiv T$

  $(w_1\wedge F) \equiv F;\quad (w_1\vee T) \equiv T$
- » **Idempotency:** $(w_1\wedge w_1) \equiv w_1;\quad (w_1\vee w_1) \equiv w_1$
- » **Double negation elimination:** $\neg\neg w_1\equiv w_1$
- » **De Morgan's laws:**

  $\neg(w_1\vee w_2) \equiv \neg w_1\wedge\neg w_2;\quad \neg(w_1\wedge w_2) \equiv \neg w_1\vee\neg w_2$
- » **Law of the contrapositive:** $w_1\Rightarrow w_2 \equiv \neg w_2\Rightarrow \neg w_1$
- » **Commutativity:** $w_1\vee w_2 \equiv w_2\vee w_1;\ w_1\wedge w_2 \equiv w_2\wedge w_1$
- » **Associative laws:**

  $(w_1\wedge w_2)\wedge w_3 \equiv w_1\wedge(w_2\wedge w_3)\equiv w_1\wedge w_2\wedge w_3$ [conjunction]

  $(w_1\vee w_2)\vee w_3 \equiv w_1\vee(w_2\vee w_3)\equiv w_1\vee w_2\vee w_3$ [disjunction]
- » **Distributive laws:**

  $w_1\wedge(w_2\vee w_3) \equiv (w_1\wedge w_2)\vee(w_1\wedge w_3)$

  $w_1\vee(w_2\wedge w_3) \equiv (w_1\vee w_2)\wedge(w_1\vee w_3)$.

- If w_1 and w_2 are equivalent, then $(w_1\Leftrightarrow w_2)$ is valid.

Tricks

- The following substitutions transform the evaluation of truth values of a wff in propositional logic into an arithmetic problem

Logic	Artithmetic
False	0
True	1
$\wedge$	$\cdot$
$\vee$	$+$

- In some deductions, the notation $\overline{P}$ is sometimes used as short-hand for ¬P

Normal forms

- **A literal** is either an atom (positive literal), or an atom preceded by the negation sign (¬) (negative literal).
- **Clause**: A set of literals.
  A clause can also be represented as a disjunction of distinct literals. E.g. {P, Q, ¬R } ≡ P∨Q∨¬R;
- **Unit clause**: A clause (i.e. a set) containing a single literal.
- **Empty clause**: A clause containing no literals. { } ≡ F
- **Conjunctive normal form (CNF):** A wff written as a conjunction of clauses (disjunctions of literals) is said to be in conjunctive normal form.
  - » <u>**Metatheorem 2:**</u> There is an algorithm which transforms any wff in propositional calculus into an **equivalent** CNF
- **Disjunctive normal form (DNF):** A wff written as a disjunction of conjunctions of literals is said to be in disjunctive normal form.
  - » <u>**Metatheorem 2 (dual):**</u> There is an algorithm which transforms any wff in propositional calculus into an **equivalent** DNF.

Conversion to CNF

Algorithm which transforms any wffs in propositional calculus into an **equivalent** CNF

1. **Eliminate implication** connectives using

   $w_1 \Rightarrow w_2 \equiv \neg w_1 \vee w_2$

   $w_1 \Leftrightarrow w_2 \equiv (w_1 \Rightarrow w_2) \wedge (w_2 \Rightarrow w_1)$

2. **Reduce the scope of ¬**
   - » Using De Morgan's laws
   - » Eliminate repeated ¬ symbols.
     Apply this rule repeatedly until all ¬ symbols appear immediately preceding an atom

3. Convert to CNF using **associative/distributive laws.**

4. Simplify expression using **equivalence rules**.

Example: Conversion to CNF, I

- **Consider the proposition**

  $((P \Leftrightarrow Q) \Rightarrow (R \Rightarrow S)) \wedge (Q \Rightarrow \neg(P \wedge R))$

- Convert the proposition to **equivalent CNF**

  1. Eliminate implication symbols

     $(\neg((\neg P \vee Q) \wedge (\neg Q \vee P)) \vee (\neg R \vee S)) \wedge (\neg Q \vee \neg(P \wedge R))$

  2. Reduce the scope of ¬

     $((\neg(\neg P \vee Q) \vee \neg(\neg Q \vee P)) \vee (\neg R \vee S)) \wedge (\neg Q \vee \neg P \vee \neg R)$

     $(((\neg\neg P \wedge \neg Q) \vee (\neg\neg Q \wedge \neg P))) \vee (\neg R \vee S)) \wedge (\neg Q \vee \neg P \vee \neg R)$

     $(((P \wedge \neg Q) \vee (Q \wedge \neg P) \vee (\neg R \vee S))) \wedge (\neg Q \vee \neg P \vee \neg R)$

  3. Apply associative / distributive laws

     $((P \wedge \neg Q) \vee (Q \wedge \neg P) \vee (\neg R \vee S)) \wedge (\neg Q \vee \neg P \vee \neg R)$

     For $((P \wedge \neg Q) \vee (Q \wedge \neg P) \vee (\neg R \vee S))$:
     [matrix with implicit disjunctions between rows,
     implicit conjunctions between elements in the same row]

P	¬Q	(generate from this matrix 4 conjunctions of
Q	¬P	disjunctions, where each disjunction has
(¬R∨S)		one element from each row)

     $(P \vee Q \vee \neg R \vee S) \wedge (P \vee \neg P \vee \neg R \vee S) \wedge (\neg Q \vee Q \vee \neg R \vee S) \wedge$
     $(\neg Q \vee \neg P \vee \neg R \vee S) \wedge (\neg Q \vee \neg P \vee \neg R)$

  4. Simplify expression using equivalence rules

     $(P \vee Q \vee \neg R \vee S) \wedge (\neg Q \vee \neg P \vee \neg R \vee S) \wedge (\neg Q \vee \neg P \vee \neg R)$

Example: Conversion to CNF, II

- **Consider the proposition**

  $((P \Leftrightarrow Q) \Rightarrow (\neg R \wedge S)) \wedge (Q \Rightarrow \neg (P \wedge R))$

- Convert the proposition to **equivalent CNF**

  1. Eliminate implication symbols

     $(\neg ((\neg P \vee Q) \wedge (\neg Q \vee P)) \vee (\neg R \wedge S)) \wedge (\neg Q \vee \neg (P \wedge R))$

  2. Reduce the scope of $\neg$

     $((\neg (\neg P \vee Q) \vee \neg (\neg Q \vee P)) \vee (\neg R \wedge S)) \wedge (\neg Q \vee \neg P \vee \neg R)$

     $((\neg\neg P \wedge \neg Q) \vee (\neg\neg Q \wedge \neg P)) \vee (\neg R \wedge S)) \wedge (\neg Q \vee \neg P \vee \neg R)$

     $(((P \wedge \neg Q) \vee (Q \wedge \neg P)) \vee (\neg R \wedge S)) \wedge (\neg Q \vee \neg P \vee \neg R)$

  3. Apply associative / distributive laws

     $((P \wedge \neg Q) \vee (Q \wedge \neg P) \vee (\neg R \wedge S)) \wedge (\neg Q \vee \neg P \vee \neg R)$

P	$\neg Q$	(generate from this matrix 8 conjunctions of
Q	$\neg P$	disjunctions, where each disjunction has
$\neg R$	S	one element from each row)

$(P \vee Q \vee \neg R) \wedge (P \vee Q \vee S) \wedge (P \vee \neg P \vee \neg R) \wedge (P \vee \neg P \vee S) \wedge$
$(\neg Q \vee Q \vee \neg R) \wedge (\neg Q \vee Q \vee S) \wedge (\neg Q \vee \neg P \vee \neg R) \wedge (\neg Q \vee \neg P \vee S) \wedge$
$(\neg Q \vee \neg P \vee \neg R)$

4. Simplify expression

   $(P \vee Q \vee \neg R) \wedge (P \vee Q \vee S) \wedge (\neg Q \vee \neg P \vee \neg R) \wedge (\neg Q \vee \neg P \vee S)$

Entailment

- **Entailment:** If a wff `w` has value *True* under all interpretations for which all wffs of Δ have value *True*, then
  - » Δ logically entails `w`

    $$\Delta \models w$$
  - » `w` is a logical consequence of Δ

    (also `w` follows from Δ)

E.g. $\{P\} \models P$

$\{P, P \Rightarrow Q\} \models Q$

$F \models w$ (w any wff)

Soundness and completeness

- **Soundness (or truth preserving):**
  If, for any set of wffs Δ, and a wff w
  $$\Delta \vdash_R w \text{ implies } \Delta \models w$$
  then, the **set of rules *R*** is said to be **sound**.

- **Completeness:**
  If, for any set of wffs Δ, and a wff w
  $$\Delta \models w \text{ implies } \Delta \vdash_R w$$
  then, the **set of rules *R*** is said to be **complete**.

- If we find ***R* sound and complete,** we can determine whether **w follows from** Δ ($\Delta \models w$) by either
  - » Model checking: Constructing **truth tables** for all possible interpretations (computationally expensive: $O(2^n)$, n = number of atoms)
  - » **Searching for a proof** using a complete search procedure
    - – The inference rules are the operators that generate successors of the current search state.
    - – Usually a transformation to a normal form is needed.

Metalinguistics

Do not confuse

- **linguistic symbols** that form sentences in propositional logic:
  T, F, P, Q, P1, P2…, $\vee$, $\wedge$, $\neg$, $\Rightarrow$, $\Leftrightarrow$
- **metalinguistic symbols**, used to make statements about propositional logic:
  $\models$ (entailment), $\vdash$ (inference)

Do not confuse

- **Theorems in propositional logic**, produced by the application of rules of inference on sets of wffs.
- **Metatheorems:** Theorems about propositional logic.

Q: Is the following (correct) statement
"Equivalence rules are sound inference rules"
a theorem or a metatheorem?

Our goal: A knowledge-based agent

Knowledge base: Set of sentences in a formal language, each of which represents an assertion about the world.

- » The agent can represent knowledge about the world by including **atoms** in its **knowledge base (KB)**. The explicit occurrence of an atom in the agent's knowledge base means that the agent regards the associated proposition to be ***True*** in its world.

- **A knowledge-based agent should have:**
  - » **Mechanisms for grounding its knowledge:**
    - – 'Innate' **background knowledge.**
    - – A **sensory apparatus** to determine **directly** the **truth or falsity of propositions** about its **world**.
  - » **A query mechanism** that allows to access the knowledge stored in KB.
  - » **A sound (truth preserving) inference mechanism** that allows it to deduce new sentences and to add them to its knowledge base.

    If KB has the value *True* in the real world, then any sentence derived from KB by a sound inference procedure has also the value *True* in the real world.
  - » **Completeness** of the inference mechanism is also desirable (Gödel says: not always possible).

Exercises (from Nilsson, ch. 13)

- Show by means of a truth table the equivalence between $\neg(P \wedge Q) \equiv (\neg P \vee \neg Q)$
- Prove that if a set of wffs is inconsistent, then $\Delta \models w$, for any wff w.
- How would you use a truth table to demonstrate that modus ponens is sound?