Artificial Intelligence

A course about foundations

First-Order Logic

Marco Piastra

UNIVERSITÀ DI PAVIA

Propositional possible worlds

Each possible world is a structure $<\{0,1\}, \Sigma, v>$

- $\{0,1\}$ are the *truth values*
- Σ is the ***signature*** of the formal language: a set of propositional symbols
- v is a *function* $: \Sigma \rightarrow \{0,1\}$ assigning truth values to the symbols in Σ

Propositional symbols (*signature*)

- Each symbol in Σ stands for an actual *proposition* (in natural language)
  - In the simple convention, we use the symbols $A, B, C, D, \ldots$
  - Caution: Σ is not necessarily *finite*

Possible worlds

- The class of structures contains all possible worlds:

  $<\{0,1\}, \Sigma, v>$
  $<\{0,1\}, \Sigma, v'>$
  $<\{0,1\}, \Sigma, v''>$
  ...

  - Each class of structure shares Σ and $\{0,1\}$
    - The functions v are different: the assignment of truth values varies, depending on the possible world
    - If $\boldsymbol{P}$ is finite, there are only *finitely* many distinct possible worlds (actually $2^{|\boldsymbol{P}|}$)

Entering _extensional_ semantics: tuples, relations and functions

- **Tuple**

  Consider a generic set of objects $\mathbf{U}$

  An example *set* of objects from $\mathbf{U}$ is denoted as $\{u_1, u_2\}$, where $u_1, u_2 \in \mathbf{U}$

  In a set, the order of elements is not relevant

  An example of *tuple* of objects from $\mathbf{U}$ is denoted as $\langle u_1, u_2 \rangle$, where $u_1, u_2 \in \mathbf{U}$

  In a tuple, the order is relevant, i.e. $\langle u_1, u_2 \rangle \neq \langle u_2, u_1 \rangle$

- **Cartesian product**

  The cartesian product $\mathbf{U} \times \mathbf{U} =: \mathbf{U}^2$ is the set of all tuples $\langle u_1, u_2 \rangle, \; u_1, u_2 \in \mathbf{U}$

  Analogously, $\mathbf{U}^3$ is the set of all tuples $\langle u_1, u_2, u_3 \rangle, \; u_1, u_2, u_3 \in \mathbf{U}$
  $\mathbf{U}^4$ is the set of all tuples $\langle u_1, u_2, u_3, u_4 \rangle, \; u_1, u_2, u_3, u_4 \in \mathbf{U}$ and so on …

- **Relation**

  arity is always an integer

  A relation of *arity* n is a subset of $\mathbf{U}^n$

- **Function**

  A function of type $\mathbf{U}^n \to \mathbf{U}$ is a relation of arity $n+1$ such that each tuple is constructed by associating each tuple of $\mathbf{U}^n$ with exactly one object from $\mathbf{U}$

First-order possible worlds

Possible worlds made of objects, functions and relations

Each possible world is a structure <**U**, Σ, *v*>

U is a set of object, called ***domain*** (also *universe of discourse*)

Σ is a set of symbols, called ***signature***

v is a *function* that gives a *meaning* to the symbols in Σ with respect to **U**

Signature Σ

- *individual constants*: *a*, *b*, *c*, *d*, ...
- *function symbols (with arity)*: *f* /*n*, *g* /*p*, *h* /*q*, ...
- *predicate symbols* (with arity): *P* /*k*, *Q* /*l*, *R* /*m*, ...

Arity is an integer that describes the expected number of arguments

First-order possible worlds

Possible worlds made of objects, functions and relations

Each possible world is a structure <**U**, Σ, *v*>

U is a set of object, called ***domain*** (also *universe of discourse*)

Σ is a set of symbols, called ***signature***

v is a *function* that gives a *meaning* to the symbols in Σ with respect to **U**

Term

A single *individual constant* is a **term**

If $f\,/n$ is a *functional symbol* (with arity n) and $t_1, \ldots, t_n$ are **terms**,
then $f(t_1, \ldots, t_n)$ is a **term**

Atom

If $P\,/n$ is a *predicate symbol* (with arity n) and $t_1, \ldots, t_n$ are **terms**,
then $P(t_1, \ldots, t_n)$ is an **atom** (i.e a first-order well-formed formula – wff)

First-order possible worlds

Possible worlds made of objects, functions and relations

Each possible world is a structure $<\mathbf{U}, \Sigma, v>$

$\mathbf{U}$ is a set of object, called ***domain*** (also *universe of discourse*)

Σ is a set of symbols, called ***signature***

v is a *function* that gives a *meaning* to the symbols in Σ with respect to $\mathbf{U}$

Function v (***interpretation***)

- v assigns each *individual constant* to an *object* in $\mathbf{U}$

  $v(a) \in \mathbf{U}$ (a individual constant)

- v assigns each *functional symbol* a *function* defined on $\mathbf{U}$

  $v(f/n) : \mathbf{U}^n \to \mathbf{U}$ (f/n functional symbol)

- v assigns each *predicate symbol* a *relation* defined on $\mathbf{U}$

  $v(P/m) \subseteq \mathbf{U}^m$ (P/m predicate symbol)

First-order possible worlds

Possible worlds made of objects, functions and relations

Each possible world is a structure $<\mathbf{U}, \Sigma, v>$

$\mathbf{U}$ is a set of object, called ***domain*** (also *universe of discourse*)

Σ is a set of symbols, called ***signature***

v is a *function* that gives a *meaning* to the symbols in Σ with respect to $\mathbf{U}$

Term

A single *individual constant* is a **term**

If f/n is a *functional symbol* (with arity n) and $t_1, \dots, t_n$ are **terms**, then $f(t_1, \dots, t_n)$ is a **term**

The semantics of a **term** $f(t_1, \dots, t_n)$ is

$$v(f/n)(<v(t_1), \dots, v(t_n)>) \in \mathbf{U}$$

that is, the result of applying the function that v associates to f/n to the tuple of objects in $\mathbf{U}$ created from the semantics of $t_1, \dots, t_n$

It is yet an object in $\mathbf{U}$

First-order language (*without variables*)

- Well-formed formulae (wff)

  All symbols in the *signature* Σ (*constants*, *function* and *predicate symbols*)

  Two (primary) ***logical connectives***: $\neg$, $\rightarrow$

  Three (derived) ***logical connectives***: $\wedge$, $\vee$, $\leftrightarrow$

  Parenthesis: (,) (there are no *precedence rules* in this language)

  The definition of *terms* and *atoms* (see before)

  A set of syntactic rules

  The set of all the **wff** of L_{FO} is denoted as $\mathrm{wff}(L_{FO})$

  φ is an *atom* $\Rightarrow \varphi \in \mathrm{wff}(L_{FO})$

  $\varphi \in \mathrm{wff}(L_{FO}) \Rightarrow (\neg\varphi) \in \mathrm{wff}(L_{FO})$

  $\varphi, \psi \in \mathrm{wff}(L_{FO}) \Rightarrow (\varphi \rightarrow \psi) \in \mathrm{wff}(L_{FO})$

  $\varphi, \psi \in \mathrm{wff}(L_{FO}) \Rightarrow (\varphi \vee \psi) \in \mathrm{wff}(L_{FO}), \quad (\varphi \vee \psi) \Leftrightarrow ((\neg\varphi) \rightarrow \psi)$

  $\varphi, \psi \in \mathrm{wff}(L_{FO}) \Rightarrow (\varphi \wedge \psi) \in \mathrm{wff}(L_{FO}), \quad (\varphi \wedge \psi) \Leftrightarrow (\neg(\varphi \rightarrow (\neg\psi)))$

  $\varphi, \psi \in \mathrm{wff}(L_{FO}) \Rightarrow (\varphi \leftrightarrow \psi) \in \mathrm{wff}(L_{FO}), \quad (\varphi \leftrightarrow \psi) \Leftrightarrow ((\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi))$

These rules are identical to the propositional ones!

Satisfaction (*without* *variables*)

- Given a possible world $<\mathbf{U}, \Sigma, v>$

If φ is an *atom* (i.e. φ has the form $P(t_1, \ldots, t_n)$)

$<\mathbf{U}, \Sigma, v> \models \varphi$ iff $<v(t_1), \ldots, v(t_n)> \in v(P/n)$

If φ e ψ are wffs

$<\mathbf{U}, \Sigma, v> \models (\neg\varphi)$ iff $<\mathbf{U}, \Sigma, v> \not\models \varphi$

$<\mathbf{U}, \Sigma, v> \models (\varphi \rightarrow \psi)$ iff NOT $<\mathbf{U}, \Sigma, v> \models \varphi$ OR $<\mathbf{U}, \Sigma, v> \models \psi$

$<\mathbf{U}, \Sigma, v> \models (\varphi \wedge \psi)$ iff $<\mathbf{U}, \Sigma, v> \models \varphi$ AND $<\mathbf{U}, \Sigma, v> [s] \models \psi$

$<\mathbf{U}, \Sigma, v> \models (\varphi \vee \psi)$ iff $<\mathbf{U}, \Sigma, v> \models \varphi$ OR $<\mathbf{U}, \Sigma, v> [s] \models \psi$

What is *true*?

- **A world of *cats***

Likes	Tom	Spot	Kitty	Felix
Tom	x			
Spot	x		x	
Kitty		x	x	
Felix			x	

translates into $\langle \mathbf{U}, \Sigma, v \rangle$

- ***Universe***

  $\mathbf{U} := \{\underline{\text{tom}}, \underline{\text{spot}}, \underline{\text{kitty}}, \underline{\text{felix}}\}$ ← *Could not put real cats in* $\mathbf{U}$ *:)*
  underlined names here stand for objects

- ***Signature***

  $\Sigma := \{tom, spot, kitty, felix, Likes/2\}$ *four constants and one predicate symbol*

- ***Interpretation***

  $v(tom) = \underline{\text{tom}}$, $v(spot) = \underline{\text{spot}}$, $v(kitty) = \underline{\text{kitty}}$, $v(felix) = \underline{\text{felix}}$,

  $v(Likes/2) =$ *a subset of* $\mathbf{U} \times \mathbf{U}$
  $\{\langle \underline{\text{tom}}, \underline{\text{tom}} \rangle, \langle \underline{\text{spot}}, \underline{\text{tom}} \rangle, \langle \underline{\text{spot}}, \underline{\text{kitty}} \rangle, \langle \underline{\text{kitty}}, \underline{\text{spot}} \rangle, \langle \underline{\text{kitty}}, \underline{\text{kitty}} \rangle, \langle \underline{\text{felix}}, \underline{\text{kitty}} \rangle\}$

What is *true*?

- **A world of *cats***

Likes	Tom	Spot	Kitty	Felix
Tom	x			
Spot	x		x	
Kitty		x	x	
Felix			x	

translates into $\langle \mathbf{U}, \Sigma, v \rangle$

- ***Sentences***

$\langle \mathbf{U}, \Sigma, v \rangle \models Likes(spot, kitty)$ because $\langle v(spot), v(kitty) \rangle \in v(Likes/2)$

$\langle \mathbf{U}, \Sigma, v \rangle \models Likes(tom, tom)$ because $\langle v(tom), v(tom) \rangle \in v(Likes/2)$

$\langle \mathbf{U}, \Sigma, v \rangle \models \neg Likes(kitty, felix)$ because $\langle v(kitty), v(felix) \rangle \notin v(Likes/2)$

$\langle \mathbf{U}, \Sigma, v \rangle \not\models Likes(tom, kitty)$ because $\langle v(tom), v(kitty) \rangle \notin v(Likes/2)$

$\langle \mathbf{U}, \Sigma, v \rangle \not\models \neg Likes(felix, kitty)$ because $\langle v(felix), v(kitty) \rangle \in v(Likes/2)$

What is *true*?

- **A world of *cats***

Likes	Tom	Spot	Kitty	Felix
Tom	x			
Spot	x		x	
Kitty		x	x	
Felix			x	

translates into $<\mathbf{U}, \Sigma, v>$

- ***Sentences***

$<\mathbf{U}, \Sigma, v> \models (Likes(spot, kitty) \wedge Likes(felix, kitty))$

$<\mathbf{U}, \Sigma, v> \models (Likes(tom, kitty) \vee Likes(tom, tom))$

$<\mathbf{U}, \Sigma, v> \models (Likes(spot, tom) \vee \neg Likes(spot, tom))$

is satisfied in this possible world but also in <u>any</u> possible world

Variables & Quantifiers

- Well-formed formulae (wff)

  All symbols in the *signature* Σ (i.e. *constants*, *function* and *predicate* *symbols*)
  A set of **variables**: x, y, z
  Two (primary) ***logical connectives***: $\neg, \rightarrow$
  Three (derived) ***logical connectives***: $\wedge, \vee, \leftrightarrow$
  Two **quantifiers**: $\forall, \exists$
  Parentheses: (,) (there are no *precedence rules* in this language)

An extended definition of *terms* and *atoms*

Term

A single *individual constant* or a variable is a **term**

If f/n is a *functional symbol* (with arity n) and $t_1, \dots, t_n$ are **terms**,
then $f(t_1, \dots, t_n)$ is a **term**

Atom

If P/n is a *predicate symbol* (with arity n) and $t_1, \dots, t_n$ are **terms**,
then $P(t_1, \dots, t_n)$ is an **atom** (i.e a first-order well-formed formula – wff)

Variables & Quantifiers

- Well-formed formulae (wff)

  All symbols in the *signature* Σ (i.e. *constants*, *function* and *predicate symbols*)

  A set of **variables**: x, y, z

  Two (primary) ***logical connectives***: $\neg$, $\rightarrow$

  Three (derived) ***logical connectives***: $\wedge$, $\vee$, $\leftrightarrow$

  Two **quantifiers**: $\forall$, $\exists$

  Parentheses: (,) (there are no *precedence rules* in this language)

  An extended definition of *terms* and *atoms (see before)*

  A set of syntactic rules

  φ is an *atom* $\Rightarrow$ $\varphi \in \mathrm{wff}(L_{FO})$

  $\varphi \in \mathrm{wff}(L_{FO}) \Rightarrow (\neg\varphi) \in \mathrm{wff}(L_{FO})$

  $\varphi, \psi \in \mathrm{wff}(L_{FO}) \Rightarrow (\varphi \rightarrow \psi) \in \mathrm{wff}(L_{FO})$

  $\varphi, \psi \in \mathrm{wff}(L_{FO}) \Rightarrow (\varphi \vee \psi) \in \mathrm{wff}(L_{FO})$, $(\varphi \vee \psi) \Leftrightarrow ((\neg\varphi) \rightarrow \psi)$

  $\varphi, \psi \in \mathrm{wff}(L_{FO}) \Rightarrow (\varphi \wedge \psi) \in \mathrm{wff}(L_{FO})$, $(\varphi \wedge \psi) \Leftrightarrow (\neg(\varphi \rightarrow (\neg\psi)))$

  $\varphi, \psi \in \mathrm{wff}(L_{FO}) \Rightarrow (\varphi \leftrightarrow \psi) \in \mathrm{wff}(L_{FO})$, $(\varphi \leftrightarrow \psi) \Leftrightarrow ((\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi))$

  $\varphi \in \mathrm{wff}(L_{FO}) \Rightarrow (\forall x\, \varphi) \in \mathrm{wff}(L_{FO})$ ← x *can be any variable*

  $\varphi \in \mathrm{wff}(L_{FO}) \Rightarrow (\exists x\, \varphi) \in \mathrm{wff}(L_{FO})$

Satisfaction

- Given a possible world $<\mathbf{U}, \Sigma, v>$ and a valuation s (on that world)

  A valuation is a function s : *Variables* $\rightarrow \mathbf{U}$

  In a sense, a valuation s transforms all variables into constants

  If φ is an *atom* (i.e. φ has the form $P(t_1, \ldots, t_n)$)
  $<\mathbf{U}, \Sigma, v> [s] \models \varphi$ iff $<v(t_1)\,[s], \ldots, v(t_n)\,[s]> \in v(P)\,[s]$

  If φ e ψ are wffs
  $<\mathbf{U}, \Sigma, v> [s] \models (\neg\varphi)$ iff $<\mathbf{U}, \Sigma, v> [s] \not\models \varphi$
  $<\mathbf{U}, \Sigma, v> [s] \models (\varphi \wedge \psi)$ iff $<\mathbf{U}, \Sigma, v> [s] \models \varphi$ AND $<\mathbf{U}, \Sigma, v> [s] \models \psi$
  $<\mathbf{U}, \Sigma, v> [s] \models (\varphi \vee \psi)$ iff $<\mathbf{U}, \Sigma, v> [s] \models \varphi$ OR $<\mathbf{U}, \Sigma, v> [s] \models \psi$
  $<\mathbf{U}, \Sigma, v> [s] \models (\varphi \rightarrow \psi)$ iff NOT $<\mathbf{U}, \Sigma, v> [s] \models \varphi$ OR $<\mathbf{U}, \Sigma, v> [s] \models \psi$

  Quantified formulae
  $<\mathbf{U}, \Sigma, v> [s] \models \forall x\, \varphi$ iff FORALL $\underline{d} \in \mathbf{U}$ we have $<\mathbf{U}, \Sigma, v> [s](x{:}\underline{d}) \models \varphi$
  $<\mathbf{U}, \Sigma, v> [s] \models \exists x\, \varphi$ iff it EXISTS $\underline{d} \in \mathbf{U}$ such that $<\mathbf{U}, \Sigma, v> [s](x{:}\underline{d}) \models \varphi$
  Where $[s](x{:}\underline{d})$ is the *variant* of function s that assigns $\underline{d}$ to x and remains unaltered for any other variables.

What is *true*?

- **A world of *cats***

Likes	Tom	Spot	Kitty	Felix
Tom	x			
Spot	x		x	
Kitty		x	x	
Felix			x	

translates into $<\mathbf{U}, \Sigma, v>$

- ***Sentences***

$<\mathbf{U}, \Sigma, v>$ [s] $\models$ $(\forall x\ (\exists y\ Likes(x, y)))$ because

FORALL <u>cat1</u> $\in$ $\mathbf{U}$, $<\mathbf{U}, \Sigma, v>$ [s](x:<u>cat1</u>) $\models$ $(\exists y\ Likes(x, y))$ because

it EXISTS <u>cat2</u> $\in$ $\mathbf{U}$, $<\mathbf{U}, \Sigma, v>$ ([s](x:<u>cat1</u>))(y:<u>cat2</u>) $\models$ $Likes(x, y)$

What is *true*?

- **A world of *cats***

Likes	Tom	Spot	Kitty	Felix
Tom	x			
Spot	x		x	
Kitty		x	x	
Felix			x	

translates into $<\mathbf{U}, \Sigma, v>$

- ***Sentences***

$<\mathbf{U}, \Sigma, v>$ [s] $\not\models$ $(\exists x\ (\forall y\ Likes(x, y)))$ because

it EXISTS $\underline{\text{cat1}} \in \mathbf{U}$, $<\mathbf{U}, \Sigma, v>$ $[s](x{:}\underline{\text{cat1}}) \not\models (\forall y\ Likes(x, y))$ because

NOT FORALL $\underline{\text{cat2}} \in \mathbf{U}$, $<\mathbf{U}, \Sigma, v>$ $([s](x{:}\underline{\text{cat1}}))(y{:}\underline{\text{cat2}}) \models Likes(x, y)$

Variables & Quantifiers: further examples

- "Being brothers means being relatives"

  $\forall x \forall y\ (Brother(x, y) \rightarrow Relative(x, y))$

- "Being relative is a symmetric relation"

  $\forall x \forall y\ (Relative(x, y) \leftrightarrow Relative(y, x))$

- "By definition, being mother is being parent and female"

  $\forall x\ (Mother(x) \leftrightarrow (\exists y\ Parent(x, y) \land Female(x)))$

- "A cousin is a son of either a brother or a sister of either parents"

  $$\forall x \forall y\ (Cousin(x,y) \\ \leftrightarrow \exists z \exists w\ (Parent(z, x) \land Parent(w, y) \land (Brother(z, w) \lor Sister(z, w))))$$

- "Everyone has a mother"

  $\forall x \exists y\ Mother(y, x)$

  BE CAREFUL about the order of quantifiers, in fact:

  $\exists y \forall x\ Mother(y, x)$

  "There is one (common) mother to everyone"

Open formulae, Sentences

- ***Bound*** and ***free*** variables

  The occurrence of a *variable* in a wff is ***bound*** if it lies in the scope of a quantifier (for that *variable*)

  The occurrence of a *variable* in a wff is ***free*** if it is not *bound*

  Examples of bound variables:
  $\forall x\ P(x)$
  $\exists x\ (P(x) \rightarrow (A(x) \wedge B(x))$

  Examples of free variables:
  $P(x)$
  $\exists y\ (P(y) \rightarrow (A(x,y) \wedge B(y)))$

- ***Open*** and ***closed*** formulae: ***sentences***

  A wff is ***open*** if there is at least one free occurrence of a variable

  Otherwise, the wff is ***closed*** (also called ***sentence***)

  Only *closed* wffs, i.e. *sentences*, have a truth value (see after)

Models

- **Validity** in a possible world, **model**

  A wff φ such that $<\mathbf{U}, \Sigma, v>$ $[s] \models \varphi$ for any *valuation* s is **valid** in $<\mathbf{U}, \Sigma, v>$

  $<\mathbf{U}, \Sigma, v>$ is also a **model** of φ

  and we write $<\mathbf{U}, \Sigma, v> \models \varphi$ (i.e. the reference to s can be omitted)

  A possible world $<\mathbf{U}, \Sigma, v>$ is a **model** of a *set* of wff Γ iff it is a model for all the wffs in Γ

  and we write $<\mathbf{U}, \Sigma, v> \models \Gamma$

- **Truth**

  A **sentence** ψ such that $<\mathbf{U}, \Sigma, v>$ $[s] \models \psi$ for one valuation s is **valid** in $<\mathbf{U}, \Sigma, v>$

  If the sentence is true for one valuation s *, then is true for all valuations*

  A **sentence** ψ is **true** in $<\mathbf{U}, \Sigma, v>$ if it is **valid** in $<\mathbf{U}, \Sigma, v>$

Validity in general

- Validity and logical truth

  A wff (either open or closed) is **valid** (also **logically valid**)
  if it is **valid** in any possible world $<\mathbf{U}, \Sigma, v>$

  Example:

  $(P(x) \vee \neg P(x))$

  A sentence ψ is a **logical truth**
  if it is **true** in any possible world $<\mathbf{U}, \Sigma, v>$

  we write then $\models \psi$ (i.e. no reference to $<\mathbf{U}, \Sigma, v>$)

  Examples:

  $\forall x\, (P(x) \vee \neg P(x))$

  $\forall x \forall y\, (G(x,y) \rightarrow (H(x,y) \rightarrow G(x,y)))$

- Inconsistence

  A wff (either open or closed) is **inconsistent** if its not *satisfiable*

  Example:

  $\forall x\, (P(x) \wedge \neg P(x))$

Entailment

- Definition

  Given a set of wffs Γ and one wff φ , we have

  $$\Gamma \models \varphi$$

  iff all possible worlds $<\mathbf{U}, \Sigma, v> [s]$ satisfying Γ also satisfy φ

  This definition embraces all possible combinations $<\mathbf{U}, \Sigma, v> [s]$
  The only thing that does not vary is the language Σ

  Is this problem decidable?

  In general, in first-order logic, a direct calculus of entailment is impossible…