Hypothesis Generation in Linear Temporal Logic for Clauses in a Restricted Syntactic Form

Marta Cialdea

Università di Roma Tre

2016

The story

Comme tu sais nous travaillons sur la modélisation logique des systèmes biologiques, dans ce contexte nous avons besoins de l'abduction. [...]

J'ai besoins de faire de l'abduction sur une logique propositionnelle avec la possibilité de l'expression du temps. [...]

Pour la modélisation nous avons besoins du temps, puisque pour chaque réaction nous devons changer d'état. [...]

Un abrazo Luis

When Andreas, David and Pedro asked for contributions for Luis's celebration, I decided to get down to my homework.

Modelling biological systems [Demolombe, Fariñas, Obeid]

Cellular and molecular interactions in a biological system are often modelled by diagrams (**molecular interaction maps**) representing causal relationships between different kinds of proteins.

To the aim of **reasoning** about such networks, they can be given a **logical model**, which can be used for

- query answering (deduction)
- abductive reasoning: find out what could explain some particular behaviour, i.e. **hypothesis generation** for a given fact in the context of a background theory
  (which proteins should be activated or inhibited in order to obtain a given effect, such as the death of a cancer cell?)

Modeling the causality relationship between components of the system requires to take into account **temporal aspects**.

Temporal reasoning

(Thanks to Robert Demolombe)

A protein of type A can activate a protein of type B:

- If A is activated, then B is activated.

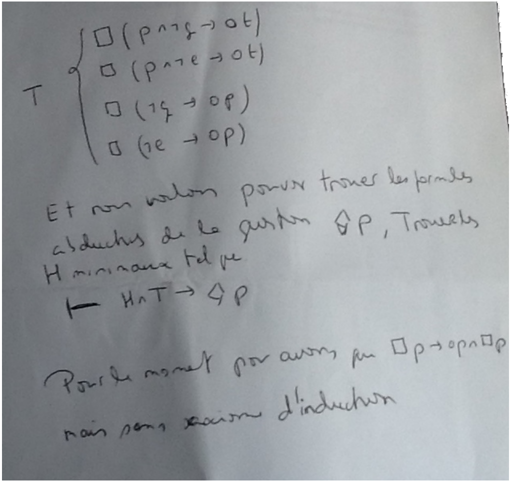
Protein B can inhibit protein A:

- if B is activated, then A is inhibited.

But A cannot be both activated and inhibited...

The language needed to model biological systems

Voici un exemple. Nous avons le langage avec next et nécessaire, sans axiome d'induction et les règles sont de la forme comme dans la photo.



Hypothesis generation

What explains a fact F in the context of a background theory T?
(assuming $T \not\models F$)

- Find E such that $T \cup \{E\} \models F$

But also:

- E must be **minimal** w.r.t. logical consequence: every other explanation E' such that $E \models E'$ (E' is not stronger than E) is logically equivalent to E

  (E is a **relevant explanation**).
- E must **not** be a **trivial** explanation:
  - E must be consistent with T ($T \not\models \neg E$)
  - E must not be equivalent to F
    ($E \not\models F$ suffices, assuming that E is a minimal explanation of F, since F is an explanation of F)

Hypothesis generation and consequence finding

Hypothesis generation can be reduced (to some extent) to **consequence finding**:

- Find C such that $T \models C$

And also: C must be **maximal** w.r.t. logical consequence: every other consequence C' such that $C' \models C$ is logically equivalent to C.

Obviously: Find E such that $T \cup \{E\} \models F$
$\equiv$ Find E such that $T \cup \{\neg F\} \models \neg E$

So we may look for consequences C of $T \cup \{\neg F\}$ such that:

- C is maximal among the C' such that $T \cup \{\neg F\} \models C'$
- $T \not\models C$: T alone is not enough to derive C.
- $\neg C \not\models F$, i.e. $\neg F \not\models C$: $\neg F$ is not enough to derive C.

For any of such C, $\neg C$ is a relevant and non-trivial explanation of F.

The consequence generation method for formulae in restricted form

Using either resolution or tableaux methods for LTL to generate consequences, like in classical logic, is hard: such methods are quite complex.

But syntactical restrictions may help to make things easier (even if not so easy as I expected!)

Here: consequence generation method based on a resolution system for LTL defined by **Cavalli and Fariñas** in 1984, restricted to

flat clauses

- **always clauses:** $\Box(L_1 \vee \cdots \vee L_k)$
- **initial clauses:** $L_1 \vee \cdots \vee L_k$ (harmless, up to a certain point)

where $k \geq 1$ and $L_1, \ldots, L_k$ are **modal literals**, of the form $\bigcirc^n \ell$, for $n \geq 0$ and ℓ a classical literal.

Generating "relevant" consequences

- The **background theory** T is made of **flat clauses**.
- The negation of the **fact** F to be explained is a flat clause, hence F has the form $\boldsymbol{\Diamond(L_1 \wedge \cdots \wedge L_k)}$, where $L_1, \ldots, L_k$ are modal literals.
- Consequences of $T \cup \{\neg F\}$ are flat clauses, therefore **explanations** (negations of consequences) have the same form ($\boldsymbol{\Diamond(L_1 \wedge \cdots \wedge L_k)}$).
- **Relevant consequences** must be **maximal** w.r.t. logical consequence.
- Relevant consequences **must depend**, in the derivation, both from $\neg F$ and from some clause in T.

Maximality, minimality and subsumption

Subsumption in classical FOL:

a clause C subsumes a clause D ($\boldsymbol{C \sqsubseteq D}$)
if there exists a substitution θ such that $C\theta \subseteq D$.

If $C \sqsubseteq D$ then $\forall C \models \forall D$, i.e. C is "stronger" than D: minimality w.r.t. subsumption reflects maximality w.r.t. logical consequence.

Relevant consequences are minimal w.r.t. subsumption.

Subsumption for flat clauses

$C \sqsubseteq D$ iff one of the following cases holds (treating disjunctions as sets):

- $C = \Box(L_1 \vee \cdots \vee L_k)$ and $D = \Box(\bigcirc^m L_1 \vee \cdots \vee \bigcirc^m L_k \vee M_1 \vee \cdots \vee M_n)$, for some $m, n \geq 0$.

  Example: $\Box(p \vee \bigcirc\neg p) \sqsubseteq \Box(\bigcirc^2 p \vee \bigcirc^3 \neg p \vee q)$ $\quad (m = 2)$

- $C = \Box(L_1 \vee \cdots \vee L_k)$ and $D = \bigcirc^m L_1 \vee \cdots \vee \bigcirc^m L_k \vee M_1 \vee \cdots \vee M_n$, for some $m, n \geq 0$.

  Example: $\Box(p \vee \bigcirc\neg p) \sqsubseteq \bigcirc p \vee \bigcirc^2 \neg p \vee q$ $\quad (m = 1)$

- $C = L_1 \vee \cdots \vee L_k$ and $D = L_1 \vee \cdots \vee L_k \vee M_1 \vee \cdots \vee M_n$, for $n \geq 0$.

  Example: $\bigcirc^2 \neg p \vee \bigcirc p \sqsubseteq \bigcirc p \vee \bigcirc^2 \neg p \vee q$

If $C \sqsubseteq D$ then $C \models D$

The resolution rules

Simplification of the rules in Cavalli & Fariñas (1984), restricted to flat clauses

$$\frac{\Box(\, L \vee L_1 \vee \cdots \vee L_n) \quad \Box(\, \bigcirc^k \sim L \vee M_1 \vee \cdots \vee M_m)}{\Box(\bigcirc^k L_1 \vee \cdots \vee \bigcirc^k L_n \vee M_1 \vee \cdots \vee M_m)} \; (R1)$$

$$\frac{\Box(\, L \vee L_1 \vee \cdots \vee L_n) \quad \bigcirc^k \sim L \vee M_1 \vee \cdots \vee M_m}{\bigcirc^k L_1 \vee \cdots \vee \bigcirc^k L_n \vee M_1 \vee \cdots \vee M_m} \; (R2)$$

$$\frac{L \vee L_1 \vee \cdots \vee L_n \quad \sim L \vee M_1 \vee \cdots \vee M_m}{L_1 \vee \cdots \vee L_n \vee M_1 \vee \cdots \vee M_m} \; (R3)$$

$\sim L$ is the **complement** of the modal literal L:

$$\begin{array}{rcl} \sim\bigcirc^n p & = & \bigcirc^n \neg p \\ \sim\bigcirc^n \neg p & = & \bigcirc^n p \end{array}$$

(+ Simplification rules)

Example (from Luis's photo)

$$\{\Box(\neg p \vee q \vee t), \Box(\neg p \vee e \vee t), \Box(q \vee p), \Box(e \vee p)\} \cup \{\ ??\ \} \models \Diamond p$$

1)	$\Box(\neg p \vee q \vee \bigcirc t)$	(in T)
2)	$\Box(\neg p \vee e \vee \bigcirc t)$	(in T)
3)	$\Box(q \vee \bigcirc p)$	(in T)
4)	$\Box(e \vee \bigcirc p)$	(in T)
5)	$\Box \neg p$	(negation of the *explanandum*)
6)	$\Box q$	(from 3 and 5)
7)	$\Box e$	(from 4 and 5)
8)	$\Box(e \vee \bigcirc q \vee \bigcirc^2 t)$	(from 1 and 4)
9)	$\Box(e \vee \bigcirc e \vee \bigcirc^2 t)$	(from 2 and 4)
10)	$\Box(q \vee \bigcirc q \vee \bigcirc^2 t)$	(from 1 and 3)
11)	$\Box(q \vee \bigcirc e \vee \bigcirc^2 t)$	(from 2 and 3)

$6 \sqsubseteq 8$, $6 \sqsubseteq 10$, $6 \sqsubseteq 11$, $7 \sqsubseteq 8$, $7 \sqsubseteq 9$, $7 \sqsubseteq 11$.

Explanations: $\Diamond \neg q$ and $\Diamond \neg e$

Example 2

$$\{\Box(\neg p \vee q \vee \bigcirc r), \Box(\neg s \vee \bigcirc p)\} \cup \{\ ??\ \} \models \Diamond r$$

1)	$\Box(\neg p \vee q \vee \bigcirc r)$	(in T)
2)	$\Box(\neg s \vee \bigcirc p)$	(in T)
3)	$\Box \neg r$	(negation of the *explanandum*)
4)	$\boldsymbol{\Box(\neg p \vee q)}$	(from 1 and 3)
5)	$\Box(\neg s \vee \bigcirc q \vee \bigcirc^2 r)$	(from 1 and 2)
6)	$\boldsymbol{\Box(\neg s \vee \bigcirc q)}$	(from either 2 and 4, or 3 and 5)

5 is a consequence of T alone

Explanations: $\boldsymbol{\Diamond(\neg q \wedge p)}$ and $\boldsymbol{\Diamond(s \wedge \bigcirc \neg q)}$

Refutational completeness

The calculus R1-R3 is equivalent to the calculus **CF** defined by Cavalli & Fariñas, when restricted to flat clauses:

Theorem

If $S \cup \{C\}$ is a set of flat clauses, then:

1. *if $S \vdash_{CF} C$, then $S \vdash_{R1-R3} C'$ for some flat clause C' that is logically equivalent to C;*
2. *if $S \vdash_{R1-R3} C$, then $S \vdash_{CF} C'$ for some clause C' that is logically equivalent to C.*

As a consequence:

Corollary

The resolution system consisting of the rules R1-R3 is sound and refutationally complete for flat clauses.

A weak form of implicational completeness

A mapping τ from flat clauses to classical first order clauses is defined:

$$\begin{array}{lcl} \tau(\bigcirc p \vee q \vee \bigcirc\bigcirc\neg p)) & = & p(f(a)) \vee q(a) \vee \neg p(f(f(a))) \\ \tau(\Box(\bigcirc p \vee q \vee \bigcirc\bigcirc\neg p))) & = & p(f(x)) \vee q(x) \vee \neg p(f(f(x))) \end{array}$$

- Subsumption for flat clauses corresponds to subsumption for their translations
- Derivability for flat clauses by R1-R3 corresponds to derivability for their translation by classical resolution

Exploiting the implicational completeness of classical resolution [Lee 1967]:

Theorem

If

- $C_1, \ldots, C_n, C$ *are flat clauses,*
- C *is not valid and*
- $\tau(C_1), \ldots, \tau(C_n) \models_{FOL} \tau(C)$,

then there exists a clause C' *subsuming* C *such that* $C_1, \ldots, C_n \vdash_{R1-R3} C'$.

Aprés voir si nous pouvons traiter le même problème mais cette fois avec une axiome d'induction

The calculus R1-R3 is implicationally incomplete:

$$p, \Box(\neg p \vee \bigcirc p) \models \Box p$$

but $\Box p$ (that is a flat clause) cannot be derived from $\{p, \Box(\neg p \vee \bigcirc p)\}$ by R1-R3.

This does not contradict the refutational completeness of the calculus, since the negation of the induction axiom, $\{A, \Box(\neg A \vee \bigcirc A), \Diamond\neg A\}$, is not a set of flat clauses.

Conjecture

Full implicational completeness (for flat clauses) can be achieved by adding the rule

$$\frac{L_1 \vee \cdots \vee L_k \quad \Box(\sim L_1 \vee \bigcirc L_1 \vee \cdots \vee \bigcirc L_k) \quad \ldots \quad \Box(\sim L_k \vee \bigcirc L_1 \vee \cdots \vee \bigcirc L_k)}{\Box(L_1 \vee \cdots \vee L_k)} \ (\mathrm{Ind})$$

Non termination

$$\Box(\neg p \vee \bigcirc p) \models \Box(\neg p \vee \bigcirc^n p) \quad \text{for all } n \geq 1$$

$$\dfrac{\Box(\neg p \vee \bigcirc p) \qquad \dfrac{\Box(\neg p \vee \bigcirc p) \qquad \dfrac{\Box(\neg p \vee \bigcirc p) \quad \Box(\neg p \vee \bigcirc p)}{\Box(\neg p \vee \bigcirc^2 p)}\ (R1)}{\Box(\neg p \vee \bigcirc^3 p)}\ (R1)}{\dfrac{\Box(\neg p \vee \bigcirc^4 p)}{\vdots}}\ (R1)$$

None of the consequences is subsumed by the others....

They are all implied by $\Box(\neg p \vee \Box p)$, but this is not a flat clause

Implementation: termination is forced by setting a bound on the maximal length of allowed sequences of the $\bigcirc$ operator.

Future work

On the application side

- Check the method on significative problems on biological systems and design a general methodology to encode them

On the technical side

- Prove/disprove the conjecture on the implicational completeness
- Study possible strategies/refinements of the resolution method

Thanks to Andi, David, Pedro, Robert, Serena and above all Luis!