MFPS XXVI

Special Session on Systems Biology

Internal coarse-graining of molecular systems*

[PNAS’09]

Jérôme Feret
Laboratoire d'Informatique de l'École Normale Supérieure
INRIA, ÉNS, CNRS

* This research has been done during my Post-Doc at Harvard Medical School

Mai 8, 2010
Joint-work with...

Walter Fontana  
Harvard Medical School

Vincent Danos  
Edinburgh

Russ Harmer  
Paris VII  
(Harvard Medical School)

Jean Krivine  
Paris VII
Overview

1. Context and motivations
2. Handmade ODEs
3. Abstract interpretation framework
4. Instantiation
5. Conclusion
Signalling Pathways

EGF, TGF-alpha, etc

EGFR

PI3-K

AKT

mTOR

STAT

GRB2

SOS

RAS

RAF

MEK

ERK

Gene transcription
Cell cycle progression

Cell proliferation
Inhibition of apoptosis
Angiogenesis
Migration, Adhesion, Invasion

Eikuch, 2007
A gap between two worlds

Two levels of description:

1. Databases of proteins interactions in natural language
   + documented and detailed description
   + transparent description
   − cannot be interpreted

2. ODE-based models
   + can be integrated
   − opaque modelling process, models can hardly be modified
   − there are also some scalability issues.
We use site graph rewrite systems

1. The description level matches with both
   - the observation level
   - and the intervention level
   of the biologist.
   We can tune the model easily.

2. Model description is very compact.

3. Quantitative semantics can be defined.
A breach in the combinatorial wall?

Jérôme Feret
7 Mai 8, 2010


Overview

1. Context and motivations
2. Handmade ODEs
3. Abstract interpretation framework
4. Instantiation
5. Conclusion
Overview

1. Context and motivations
2. Handmade ODEs
3. Abstract interpretation framework
4. Instantiation
5. Conclusion
Case study: A protein with a switch
Case study: A protein with a switch

\[
\begin{align*}
(u,u,u) & \rightarrow (u,p,u) \quad \kappa^c \\
(u,p,u) & \rightarrow (p,p,u) \quad \kappa^l \\
(u,p,p) & \rightarrow (p,p,p) \quad \kappa^l \\
(u,p,u) & \rightarrow (u,p,p) \quad \kappa^r \\
(p,p,u) & \rightarrow (p,p,p) \quad \kappa^r
\end{align*}
\]
Case study: A protein with a switch

\[
\begin{align*}
\frac{d[(u,u,u)]}{dt} &= -k^c [(u,u,u)] \\
\frac{d[(u,p,u)]}{dt} &= -k^l [(u,p,u)] + k^c [(u,u,u)] - k^r [(u,p,u)] \\
\frac{d[(u,p,p)]}{dt} &= -k^l [(u,p,p)] + k^r [(u,p,u)] \\
\frac{d[(p,p,u)]}{dt} &= k^l [(u,p,u)] - k^r [(p,p,u)] \\
\frac{d[(p,p,p)]}{dt} &= k^l [(u,p,p)] + k^r [(p,p,u)]
\end{align*}
\]
Case study: Two subsystems
Case study: Two subsystems
Case study: Two subsystems

\[ [(u,p,?)] \triangleq [(u,p,u)] + [(u,p,p)] \]
\[ [(p,p,?)] \triangleq [(p,p,u)] + [(p,p,p)] \]

\[
\begin{align*}
\frac{d[(u,u,u)]}{dt} &= -k^c \cdot [(u,u,u)] \\
\frac{d[(u,p,?)]}{dt} &= -k^l \cdot [(u,p,?)] + k^c \cdot [(u,u,u)] \\
\frac{d[(p,p,?)]}{dt} &= k^l \cdot [(u,p,?)]
\end{align*}
\]

\[
\begin{align*}
\frac{d[(u,u,u)]}{dt} &= -k^c \cdot [(u,u,u)] \\
\frac{d[(? ,p, u)]}{dt} &= -k^r \cdot [(?,p,u)] + k^c \cdot [(u,u,u)] \\
\frac{d[(? ,p, p)]}{dt} &= k^r \cdot [(?,p,u)]
\end{align*}
\]
Case study: Dependence index

We introduce:

\[((?,p,?) \triangleq [(?,p,u)] + [(?,p,p)]\]

The states of left site and right site would be independent if, and only if:

\[\frac{[(p,p,p)]}{[(p,p,?)]} = \frac{[(?,p,p)]}{[(?,p,?)]}\]

Thus we define the dependence index as follows:

\[X \triangleq [(p,p,p) \cdot [(?,p,?)] - [(?,p,p) \cdot (p,p,?)].\]

We have (after a short computation):

\[
\frac{dX}{dt} = -X \cdot (k^l + k^r) + k^c \cdot [(p,p,p)] \cdot [(u,u,u)].
\]

As a consequence, the property \[X = 0\] is not an invariant.
Case study: Erroneous recombination

Concentrations evolution with respect to time \( \left( \left[ (u, u, u) \right](0) = 100 \right) \).

\( \left[ (p, p, p) \right] \) and \( 25 \cdot \frac{\left[ (p, p, ?) \right] \cdot \left[ (? , p , p) \right]}{\left[ (? , p , ?) \right]} \)
Conclusion
self-consistency

− some information is abstracted away
  we cannot recover the concentration of any species;

+ it is a weak property
  which is easy to ensure,
  which is easy to propagate;
+ it captures the essence of the kinetics of systems.

We are going to track the correlations that are read by the system.
Overview

1. Context and motivations
2. Handmade ODEs
3. Abstract interpretation framework
4. Instantiation
5. Conclusion
Continuous differential semantics

Given $\mathcal{V}$, a finite set of variables; and $F$, a $C^\infty$ mapping from $\mathcal{V} \rightarrow \mathbb{R}^+$ into $\mathcal{V} \rightarrow \mathbb{R}$.

as for instance,

- $\mathcal{V} \overset{\Delta}{=} \{[(u,u,u)], [(u,p,u)], [(p,p,u)], [(u,p,p)], [(p,p,p)]\},$

$$
\begin{align*}
\mathcal{V}(\rho) &\overset{\Delta}{=} \\
[(u,u,u)] &\mapsto -k^c \cdot \rho([(u,u,u)]) \\
[(u,p,u)] &\mapsto -k^l \cdot \rho([(u,p,u)]) + k^c \cdot \rho([(u,u,u)]) - k^r \cdot \rho([(u,p,u)]) \\
[(p,p,u)] &\mapsto k^l \cdot \rho([(u,p,u)]) - k^r \cdot \rho([(p,p,u)]) \\
[(p,p,p)] &\mapsto k^l \cdot \rho([(u,p,p)]) + k^r \cdot \rho([(p,p,u)])
\end{align*}
$$

we can define the continuous differential semantics as follows:

$$
X_c : \left\{ \begin{array}{c}
(\mathcal{V} \rightarrow \mathbb{R}^+) \times \mathbb{R}^+ \rightarrow (\mathcal{V} \rightarrow \mathbb{R}^+) \\
(X_0, T) \mapsto X_0 + \int_{t=0}^{T} F_c(X_c(X_0, t)) \cdot dt.
\end{array} \right.
$$
Abstraction

An abstraction \((\mathcal{V}^\#, \psi, F^\#)\) is given by:

- \(\mathcal{V}^\#\): a finite set of observables,
- \(\psi\): a mapping from \(\mathcal{V} \rightarrow \mathbb{R}\) into \(\mathcal{V}^\# \rightarrow \mathbb{R}\),
- \(F^\#\): a \(C^\infty\) mapping from \(\mathcal{V}^\# \rightarrow \mathbb{R}^+\) into \(\mathcal{V}^\# \rightarrow \mathbb{R}\);

such that:

- \(\psi\) is linear with positive coefficients,
- \(F^\#\) is \(\psi\)-complete
  i.e. the following diagram commutes:

\[
\begin{array}{ccc}
\mathcal{V} & \rightarrow & \mathbb{R}^+ \\
\psi & \downarrow & \psi \\
\mathcal{V}^\# & \rightarrow & \mathcal{V} \\
\downarrow & & \downarrow \\
\mathcal{V}^\# & \rightarrow & \mathbb{R} \\
\end{array}
\]

i.e. \(\psi \circ F = F^\# \circ \psi\).
Abstraction example

- $\mathcal{V} \triangleq \{(u,u,u), (u,p,u), (p,p,u), (u,p,p), (p,p,p)\}$

- $\mathcal{F}(\rho) \triangleq \begin{cases} 
(u,u,u) \mapsto -k^c \cdot \rho([u,u,u]) \\
(u,p,u) \mapsto -k^l \cdot \rho([u,p,u]) + k^c \cdot \rho([u,u,u]) - k^r \cdot \rho([u,p,u]) \\
(u,p,p) \mapsto -k^l \cdot \rho([u,p,p]) + k^r \cdot \rho([u,p,u]) \\
\end{cases}$

- $\mathcal{V}^{\sharp} \triangleq \{(u,u,u), (?,p,u), (?,p,p), (u,p,?) , (p,p,?)\}$

- $\psi(\rho) \triangleq \begin{cases} 
(u,u,u) \mapsto \rho([u,u,u]) \\
(?,p,u) \mapsto \rho([u,p,u]) + \rho([p,p,u]) \\
(?,p,p) \mapsto \rho([u,p,p]) + \rho([p,p,p]) \\
\end{cases}$

- $\mathcal{F}^{\sharp}(\rho^{\sharp}) \triangleq \begin{cases} 
(u,u,u) \mapsto -k^c \cdot \rho^{\sharp}([u,u,u]) \\
(?,p,u) \mapsto -k^r \cdot \rho^{\sharp}([?,p,u]) + k^c \cdot \rho^{\sharp}([u,u,u]) \\
(?,p,p) \mapsto k^r \cdot \rho^{\sharp}([?,p,u]) \\
\end{cases}$

(Completeness can be checked analytically.)
Abstract continuous trajectories

Given an abstraction \((\mathcal{V}^\#, \psi, \mathbb{F}^\#)\), we have:

\[
\begin{align*}
X_c(X_0, T) &= X_0 + \int_{t=0}^{T} \mathbb{F} (X_c(X_0, t)) \cdot dt \\
\psi (X_c(X_0, T)) &= \psi \left( X_0 + \int_{t=0}^{T} \mathbb{F} (X_c(X_0, t)) \cdot dt \right) \\
\psi (X_c(X_0, T)) &= \psi(X_0) + \int_{t=0}^{T} \left[ \psi \circ \mathbb{F} \right] (X_c(X_0, t)) \cdot dt \quad (\psi \text{ is linear}) \\
\psi (X_c(X_0, T)) &= \psi(X_0) + \int_{t=0}^{T} \mathbb{F}^\# (\psi (X_c(X_0, t))) \cdot dt \quad (\mathbb{F}^\# \text{ is } \psi\text{-complete})
\end{align*}
\]

We set \(Y_0 \triangleq \psi(X_0)\) and \(Y_c \triangleq \psi \circ X_c\).

Then we have:

\[
Y_c(X_0, T) = Y_0 + \int_{t=0}^{T} \mathbb{F}^\# (Y_c(X_0, t)) \cdot dt
\]
Fluid trajectories

\[ Y(t) \]
Fluid trajectories
Overview

1. Context and motivations
2. Handmade ODEs
3. Abstract interpretation framework
4. Instantiation
5. Conclusion
Differential system

Let $\mathcal{R}$ be an over-approximation of the set of reachable species. Let us consider a rule $lhs \rightarrow rhs \quad k$.

1. We write $lhs$ as a multi-set $\{C_i\}$ of non empty connected components.
2. A ground instantiation of the rule rule is defined by a tuple $(r_i, \Phi_i)$ such that $\forall i, r_i \in \mathcal{R}$ and $C_i \prec_\Phi r_i$.
3. The ground instantiation can be written as follows:
   \[
   r_1, \ldots, r_m \rightarrow p_1, \ldots, p_n \quad k.
   \]
4. The activity of a ground instantiation is defined as:
   \[
   act_{(r_i, \Phi_i)} = \frac{k \cdot \prod [r_i]}{\#\{\Phi \mid lhs \prec_\Phi lhs\}}.
   \]
5. Each ground instantiation induces the following contributions:
   \[
   \frac{d[r_i]}{dt} \Rightarrow -act_{(r_i, \Phi_i)}, \quad \frac{d[p_i]}{dt} \Rightarrow act_{(r_i, \Phi_i)}.
   \]
Abstract domain

We are looking for suitable pair \((\mathcal{V}^#, \psi)\) (such that \(F^#\) exists)

The set of linear variable changement is too big to be explored.

We introduce a specific shape on \((\mathcal{V}^#, \psi)\) so as:

- restrict the exploration;
- drive the intuition;
- having efficient way to find suitable abstractions \((\mathcal{V}^#, \psi)\)
  and to compute \(F^#\).

Our choice might be not optimal, but we can live with that.
Contact map

G

E

b

r

a

l

b

r

Sh

Y_7

pi

Y_{68}

Y_{48}

So

d

Jérôme Feret

Mai 8, 2010
Annotated contact map
A fragment
Basic properties

The set of fragments enjoys two convenient properties:

1. Closure with respect to the operational semantics:
   When we apply a rule with a tuple of fragments, we get a tuple of fragments.

2. Subfragments:
   We can express the concentration of any sub-fragment as a linear combination of the concentration of some fragments.

Which other properties do we need so that the function $F$ can be defined?
Can we express the amount (per time unit) of this fragment (bellow) concentration that is consumed by this rule (above)?
No, because we have abstracted away the correlation between the state of the site \( r \) and the state of the site \( l \).
Fragments consumption
Proper intersection

Whenever a fragment intersects a connected component of a lhs on a modified site, then the connected component must be embedded in the fragment!
We reflect each path that stems from a modified site (in the lhs of a rule) into the annotated contact map.
Connected components

We need to express the “concentration” of any connected component of a lhs with respect to the “concentration” of fragments.
Each connected component of a LHS must be a sub-fragment.
Each connected component of a lhs must be a sub-fragment.
Fragment properties

If:

- an annotated contact map satisfies the syntactic criteria,
- fragments are defined by this annotated contact map,
- we know the concentration of fragments;

then:

- we can express the concentration of any connected component occurring in lhss,
- we can express fragment proper consumption,
- we can express fragment proper production (eg. see the LICS’2010 paper),
- WE HAVE A CONSTRUCTIVE DEFINITION FOR $\mathbb{F}^\dagger$. 

Jérôme Feret

34

Mai 8, 2010
Overview

1. Context and motivations
2. Handmade ODEs
3. Abstract interpretation framework
4. Instantiation
5. Conclusion
Experimental results

On early egfr, 356 species are simplified into 38 fragments:

/home/feret/MFPS/demo/egfr-compressed.ka

(reduced) [EGFR(Y48=0), SHC(Y71=1,p=0), GRB2(a=1,b=2), SOS(d=2)]
(reduced) [EGFR(Y68=0), GRB2(a=0,b=1), SOS(d=1)]
(ground) [EGFR(Y48=0), SHC(Y71=1,p=0), GRB2(a=1,b=2), SOS(d=2)]
(ground) [EGFR(Y68=0), GRB2(a=0,b=1), SOS(d=1)]

Superposition of the ground and the abstract differential semantics.

On a bigger example, $\approx 2 \cdot 10^{19}$ species are simplified into $\approx 2 \cdot 10^{5}$ fragments.
Related issues I: **Semantics comparisons**

Species–based semantics  Rule–based semantics  Abstract semantics

C  T  M  C

refinements

⊂

limit

O  D  E

refinements

⊂

limit

♯
1. ODE approximations:
   - Because of the use of annotated contact map, fragments have a homogeneous structure (or signature).
     Can we design and use heterogeneous fragments?
     
     Joint work with Ferdinanda Camporesi (Bologne)

2. Stochastic semantics approximations:
   - Can we design abstraction?
   - Find the adequate soundness criteria.
   
     Joint work with Tatjana Petrov and Heinz Koeppl (EPFL)
Announcements

• Call for candidates:
  If you are interested in (at least one) of these issues, there are open positions (Internships, PhD students or Post-doc fellows)...

  **ANR-Chair of Excellence: AbstractCell**
  [http://www.di.ens.fr/~feret/abstractcell](http://www.di.ens.fr/~feret/abstractcell)

• Call for paper/participation:

  **First Workshop on Static Analysis and Systems Biology (SASB 2010)**
  (co-chaired with Andre Levchenko)
  13th Sept 2010, Perpignan