Executable Knowledge

for rule-based modelling

of cellular signalling networks

Russ Harmer (CNRS & ENS Lyon)

+ Adrien Basso-Blandin (ENSL) & Walter Fontana (HMS)

+ John Bachman (HMS) & Pierre Boutillier (HMS) & Lucian Galescu (IHMC) & Ben Gyori (HMS)
Context
Cellular signalling

• Decentralized coordination of tissue formation and maintenance

  • extra-cellular ligands trigger intra-cellular signalling pathways to control cell growth, death, division, …

• Perturbed in disease states, e.g. cancer, diabetes, …

  • kinetic perturbations: over-expression, knock-outs, …

  • causal perturbations: mutations, truncations, …
Rule-based modelling

- **Formal** representation for the (10s? of 1000s of) protein-protein interactions (**PPIs**) in signalling
  - **graph rewriting** formalism
  - **scalable** stochastic simulation
  - pathways as **causal traces**
- Handles **kinetic**, but not **causal**, perturbations
Serendipity
The cognitive barrier

- Have to read many papers to find various fragments of knowledge about a single PPI
  - many different ‘puzzle pieces’, at varying levels of detail, that must be assembled into rules
  - the effects of causal perturbations must be hard-wired by enumerating all cases
  - not scalable for a human curator [believe me, I’ve done it]
Big Mechanism

• Seeks causal explanations of complex system behaviour [not ‘just’ correlations]

• Machine reading of papers, automatic assembly into models that yield causal explanations …

• The chosen use case: signalling pathways in cancer!
Breaching the cognitive barrier
Assembly

• Big Mechanism makes reading scalable and RBM provides causal explanations — once your PPIs have been formalized as rules

• The hard problem is assembly

  • combining fragments of knowledge into rules …

  • … in such a way that (apparently) conflicting information can be accommodated …

  • … and the effects of causal perturbations emerge
KAMI
knowledge aggregator & model instantiator

• A graph-based representation of PPIs

• A graph with two directed edge structures, respecting a meta-model:

• uses graph rewriting to update and aggregate PPIs
KAMI
in BigM

PubMed

Reading
Deep

KAMI

Instantiation

RBM

Annotation
“Grb2’s SH2-domain binds phosphorylated EGFR”

thanks to Lucian Galescu et alia!
KAMI

input

thanks to Ben Gyori & John Bachman!
“Grb2 binds EGFR phosphorylated on Y1092”
I already know something about this interaction…

“Grb2’s SH2-domain binds [phosphorylated] EGFR phosphorylated on Y1092”

this is not yet fully automated: requires a semantic layer
“Grb2’s SH2-domain binds [phosphorylated] EGFR phosphorylated on Y1092”

“this is a step of graph rewriting”
“Grb2’s SH2-domain binds [phosphorylated] EGFR phosphorylated on Y1092”

“this is another step of graph rewriting”
“Grb2’s SH2-domain binds phosphorylated Shc”
Grb2($SH2e$), EGFR($g,Y1092-p$) $\rightarrow$ Grb2($SH2e!1$), EGFR($g!1,Y1092-p$)
Grb2($SH2s$), Shc($g,shc-p$) $\rightarrow$ Grb2($SH2s!1$), Shc($g!1,shc-p$)
independent!
These interactions use the same mechanism!

KAMI
aggregate

PubMed
Reading
Instantiation
RBM
Deep
Annotation

this is not yet fully automated: requires a semantic layer

This is not yet fully automated: requires a semantic layer
KAMI
aggregate

this is a step of graph rewriting

merged (reversibly)
KAMI

instantiate

\[ \text{KAMI} \]

\[ \text{Reading} \rightarrow \text{KAMI} \]

\[ \text{Instantiation} \rightarrow \text{RBM} \]

\[ \text{PubMed} \leftarrow \text{KAMI} \]

\[ \text{Deep} \leftarrow \text{KAMI} \]

\[ \text{Annotation} \leftarrow \text{KAMI} \]

\[ \text{conflict!} \]

\[ \text{Grb2}(\text{SH2}), \text{EGFR}(g,Y1092-p) \rightarrow \text{Grb2}(\text{SH2}!1), \text{EGFR}(g!1,Y1092-p) \]

\[ \text{Grb2}(\text{SH2}), \text{Shc}(g,\text{shc}-p) \rightarrow \text{Grb2}(\text{SH2}!1), \text{Shc}(g!1,\text{shc}-p) \]
“Grb2-S90D does not bind EGFR”
only one rule for Grb2_D90

Grb2_S90(SH2e), EGFR(g,Y1092-p) -> Grb2_S90(SH2!1), EGFR(g!1,Y1092-p)

Grb2_S90(SH2s), Shc(g,shc-p) -> Grb2_S90(SH2!1), Shc(g!1,shc-p)

Grb2_D90(SH2s), Shc(g,shc-p) -> Grb2_D90(SH2!1), Shc(g!1,shc-p)
“Grb2-S90D does not bind EGFR”

KAMI

negation

only aa:S can bind

aa:S
loc:90

Grb2

BND

EGFR

aa can be S or D

this automatically propagates to the interaction with Shc too

PubMed

Reading

Deep

KAMI

Instantiation

Annotation

RBM
KAMI

enumration

no rules for Grb2_D90

Grb2_S90(SH2), EGFR(g,Y1092-p) -> Grb2_S90(SH2!1), EGFR(g!1,Y1092-p)
Grb2_S90(SH2), Shc(g,shc~p) -> Grb2_S90(SH2!1), Shc(g!1,shc~p)

Publication

 Instantiation

Deep Annotation

automatic enumeration of rules
Wrapping up
Summary

• A purely formal graph rewriting foundation
  • represents knowledge and [revokable] hypotheses using formal operations of (update and) aggregation
  
• Model instantiation into RBM
  • automatically maintains desired [conflict] invariants and handles the effects of mutations because all enumeration is done by the machine
Work in progress
with Christine Froidevaux

- A signalling-specific semantic layer
  - semantic checking: so we only “write what a (careful) expert would write”
  - automatically update/aggregate in common cases, i.e. “when an expert would”
  - this recapitulates the by similarity reasoning used (informally) by biologists all the time
Work in progress
with Jean Yang (CMU)

• ‘Cleaning up’ databases (e.g. NCI/Nature)
  • automatic detection of subsumption and aggregation
    • remove duplicates, obtain most detailed versions of rules
  • detection of causal chains of rules [pathways]
    • resolve action-at-a-distance into local interactions