

仮説発見システムに関する研究 Research on Hypothesis Finding Systems

井上 克巳
Katsumi Inoue

アンドレイ・ドンチェスク
Andrei Doncescu (LAAS-CNRS)

何がわかる？

Intelligent machines ---
Thinking like human being.
Automated discovery of scientific knowledge,
in particular biological knowledge.

どんな研究？

Automated hypothesis-finding through
deductively complete methods.
Induction of causal laws in action theories,
and applications to systems biology.
Web-based ILP system.

Background

How Human Beings Think ?



Deduction
Abduction
Induction

Limited by
computational
and memory
capacities



The genius people are able to
mix these three fundamental
modes of reasoning.

How Intelligent Machines Think ?



Induction
Abduction
Deduction

Diagnosis
Design
Characterization
Discovery
Verification

Discovery from
huge data



Combination of
Induction and
Abduction



One of the most powerful theoretical
answers for the next generation of
Intelligent Machine (Inoue 2001,2004)

Logic and Computation

Abduction and Induction: Logic

Input:

B : background theory
 E : examples / observations

Output:

H : hypothesis satisfying that

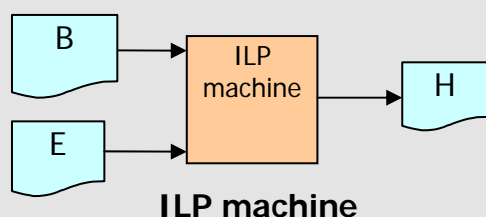
1. $B \wedge H \models E$,
2. $B \wedge H$ is consistent.

Inverse Entailment (IE)

Computing a hypothesis H can be done **deductively** by:

$$B \wedge \neg E \models \neg H$$

We have good tools for this inverse computation.



IE for Abduction --- SOLAR

(Nabeshima, Iwanuma & Inoue 2003)

B : full clausal theory
 E : conjunction of literals ($\neg E$ is a clause)
 H : conjunctions of literals ($\neg H$ is a clause)

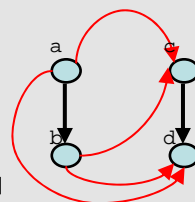
Example: graph completion problem – pathway finding

Find an arc which enables a path from a to d.

Axioms: [-node(X), -node(Y), -arc(X,Y), path(X,Y)].
[-node(X), -node(Y), -node(Z), -arc(X,Y), -path(Y,Z), path(X,Z)].
[node(a)]. [node(b)]. [node(c)]. [node(d)].
[arc(a,b)]. [arc(c,d)].

Negated Observation: [-path(a,d)].
Production_field: [-arc(.,.)].

SOLAR outputs four consequences:
[-arc(a, d)], [-arc(a, c)], [-arc(b, d)], [-arc(b, c)]



IE for Induction --- CF-induction (Inoue 2004)

B, E, H : full clausal theory

Note: CF-induction is the only existing ILP system that is complete for full clausal theories. *c.f.* Progol (Muggleton 1995) can deal with Horn clauses only.

システム生物学における知的発見 Knowledge-based Discovery in Systems Biology

井上 克巳
Katsumi Inoue

アンドレイ・ドンチェスク
Andrei Doncescu (LAAS-CNRS)

何がわかる?

Discover hidden rules in systems biology.

Explain the relationships between causes and effects from genotype to phenotype.

Use generic models in biology, *Saccharomyces Cerevisiae* and *E-coli*.

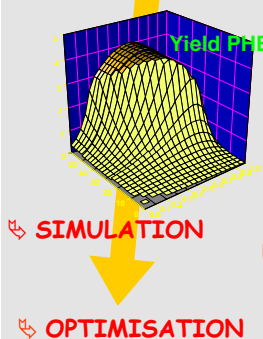
どんな研究?

Use Inductive Logic Programming (ILP).

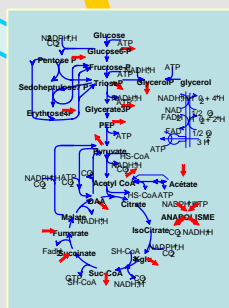
Bridge between biologists and computer scientists, due to the possibility to represent biologist knowledge in the form of logical formulas.

Modeling

Phenomenological models



METABOLIC Analyzer



Metabolic Constraints

Stoichiometric Modeling

INTRACELLULARE pools

GENOMIC RESPONSE

Structural Models

INDUSTRIAL RELATIONSHIP



Roots : basic knowledge

« Japanese bonsai »



Approach

Experts' Knowledge

The different situations are:

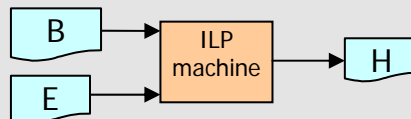
The expert can express only a partial knowledge from processes. She/he does not know the existence of several physiological states, and does not have a clear idea on which state to recognize.

Conclusion

The expert is not able to draw conclusion from the analysis of multiple signals between which there exist some true relationships.

Why Inductive Logic Programming ?

ILP can overcome the two main limitations of other machine learning techniques: (1) the use of a limited knowledge representation formalism (essentially propositional logic); (2) the difficulties to use substantial background knowledge in the learning process.



In many induced rules, the next one could be explained by biochemical experts.

```
to_state(E,E,A,B,C,T) :-
  derive(p1,A,T), derive(p2,B,T), derive(p3,C,T),
  positive(p1,T), positive(p2,T), positive(p3,T).
```

This rule indicates that there is no evolution of the metabolism state (the bio-reactor remains in the same state) when the parameters have an increasing slope but that we do not encounter maxima or minima.



Signals containing transitory events, which are impossible to predict even in a statistical form using the memory of signals, are obtained from bioreactor's sensors.

