Computational Bioscience

Antonis C. Kakas
Department of Computer Science
University of Cyprus

2 July, 2007

Summer School,
Nicosia, Cyprus

MOTIVATION and AIMS

- **Develop a new approach to Computational Science based on high-level (direct) declarative scientific modelling**
  - **Rich expressive power (close to human one)**
  - **Computational and Formal foundations**
- **Automated help in the process of Scientific Theory development**
  - **Analysis of data, generation and evaluation of hypotheses.**

Underlying THESIS of Approach

- Development of scientific theories is incremental where we need to exploit fully the knowledge acquired so far
  - Declarative Modelling can facilitate this.
- Declarative modelling is well-suited for Bioscience (e.g. Functional Genomics)
  - nature of existing biological knowledge is largely descriptive,
  - exploit the large and varied corpus of knowledge accumulated so far.
- Compare with "conventional bioinformatics".

BACKGROUND

- Computational Modelling in Systems Biology
  - Mathematical/statistical modelling approaches for Metabolic and Genetic Networks
- Artificial Intelligence in Bioinformatics
- Qualitative Biochemistry

BACKGROUND (cnt.)

- Knowledge Representation and Reasoning
  - Medical Expert Systems
- Machine Learning & Data Mining
  - Medical applications e.g. protein folding
- Modelling and hypotheses formation
  - ALP and ILP representation frameworks
- Closed Loop Learning and Knowledge Development
  - R.D. King et al, “Functional Genomic hypothesis generation and experimentation by a robot scientist”, Nature Vol. 427, 15 Jan 2004, pp. 247-251.

CONTENTS

Scientific Modelling and Logic
(Logic for Declarative Modelling)

- Start with models that represent the prior knowledge of our domain in a logical form
  - Model cellular metabolism by capturing through logical formulae all the objects and key relationships between protein-coding sequences, enzymes and metabolites in known pathways:
    - Coding
    - Reactions
    - Transport
    - Feedback
- Compute and develop the models through logical reasoning.

7

EXAMPLE: A "socio-economic" model of Universities

- Language / Ontology of relations
  {sad / 1, overworked / 1, academic / 1, student / 1, lecturer / 1, poor / 1}
- Model & background knowledge
  sad(X) if overworked(X), poor(X)

  overworked(oliver)
  overworked(alex)
  overworked(krycia)
  lecturer(alex)
  lecturer(krycia)
  student(oliver)
  academic(alex), …
- We can deduce (compute) the information of sad(alex) but we can not deduce anything new for oliver or krycia.

8

Reasoning for Declarative Problem Solving

- **Deduction**: **analytic** reasoning, inferring a result from applying general rules (model) to particular cases, e.g.

  from A (case) and *B if A* (general rule)

  infer B (result)

- **Deduction**: produces **observable (phenotype)** information.
  - Example: **sad(ale)** can be **observed/tested.**
- But this information is already known in the model. How can we improve the model?

Scientific Modelling

- **Any scientific model can be (is) incomplete!**
- **Information is separated into three types:**
  - **Observable (phenotype) – obtained from experiments via observations**
  - **Theoretical (functional genotype) – underlying relations that cause the observable behaviour**
  - **Background – known relevant properties, e.g. structural or chemical information.**
- **Example: {sad/1, overworked/1, academic/1, student/1, lecturer/1, poor/1}**

Scientific Modelling (cnt.)

- The **incompleteness** of the model resides in the theoretical part (e.g poor / 1)
- The task is to **complete** the model by finding theoretical information and developing a theory for this.
  - HOW do we **synthesize** definitions for the unobserved theoretical relations?
  - ANSWER: Scientific theories are **"explanatory"** and **"unifying"**

11

Reasoning for Declarative Problem Solving

- **Deduction**: **analytic** reasoning, inferring a result from applying general rules to particular cases, e.g.

  from A (case) and *B if A* (general rule)
  infer B (result)
- **Abduction**: **synthetic** reasoning, inferring the case from the rule and a result, e.g.

  from B (result) and *B if A* (general rule)
  infer A (case)
- **Induction**: **synthetic** reasoning, but inferring the rule from the case and the result

12

Synthetic Reasoning for Declarative Problem Solving

- **Deduction**: concerned with PREDICTION of phenotype from a given model
  - T |= Obs
- **Abduction**: concerned with EXPLANATION according to a given model: produces genotype information
  - T U H |= Obs   H specific (genotype) information
- **Induction**: concerned with GENERALIZATION of information outside the observed situations
  - T U H |= Obs   H general (genotype) information

13

EXAMPLE: A "socio-economic" model of Universities (Cnt.)

- Model & background knowledge

  sad(X) if overworked(X), poor(X)

  Abducible

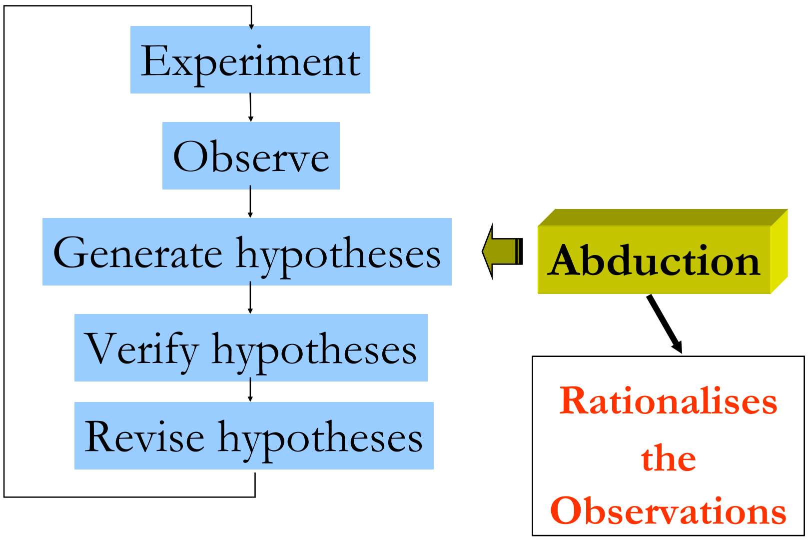
  overworked(oliver)
  overworked(alex)
  overworked(krycia)
  lecturer(alex)
  lecturer(krycia)
  student(oliver)
  academic(alex), ...

- Observations = {sad(alex), sad(krycia), not sad(oli)}

- Abductive Explanation =
  {poor(ale), poor(krycia), not poor(oli)}

14

Abduction for Scientific Analysis



15

Evaluating Abductive Explanations

- **Predictive Accuracy**
  - **On known and unknown cases**
- **Simplicity**
  - **Minimality**
- **Domain specific (preference) criteria**
  - **Language Bias**
- **Computational Complexity**
  - **E.g. Localized effects**
- **Generalizability**
  - **LINK WITH INDUCTION**

16

Integration of Abduction & Induction

- **Abduction**: problem solving given an adequate but incomplete model of the problem domain
  - Generates explanations: specific hypotheses on the incomplete part of the model
  - Rationalizes/normalizes the observations.
- **Integration with Induction**
  - Feed explanations to Induction.
  - **Induction**: development of the model by a (partial) theory for its incomplete part.
- **Tight Integration**
  - Explanations and their Generalization are evaluated as a whole.

17

EXAMPLE: A "socio-economic" model of Universities (Cnt.)

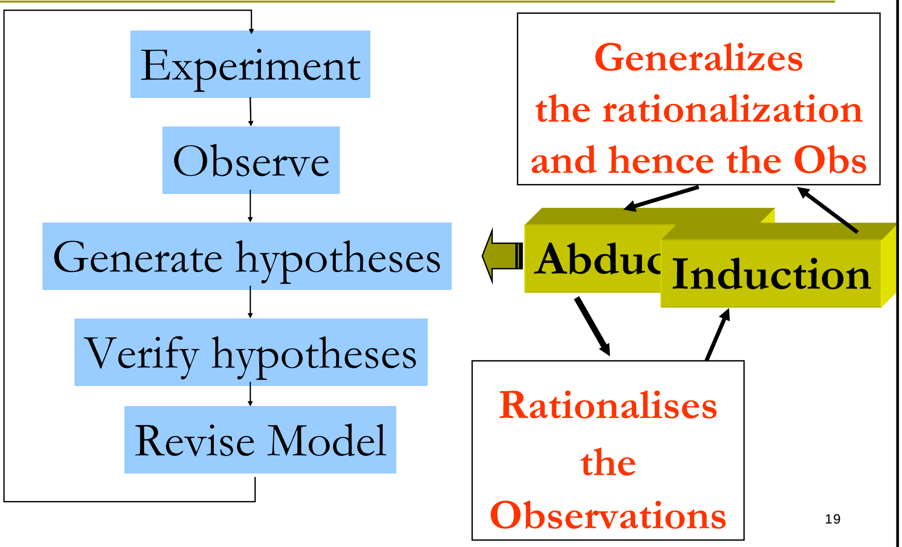
sad(X) if overworked(X), poor(X)

overworked(oliver)
overworked(alex)
overworked(krycia)
lecturer(alex)
lecturer(krycia)
student(oliver)
academic(alex), …

- Observations = {sad(alex), sad(krycia), not sad(oli)}
- Abductive Explanation =
  {poor(ale), poor(krycia), not poor(oli)}
- Inductive Hypotheses: poor(X) if lecturer(X)

18

Logical Reasoning for Scientific Analysis



19

EXAMPLE: Searching for Hypotheses

- **poor(X) if female(X)**
  - **Not as accurate (as poor(X) if lecturer(X)) as we only know female(krycia).**
- **poor(X) if academic(X)**
- **poor(X) if young(X)**
  - **Not as accurate as each would imply sad(oliver).**
- **poor(X) if academic(X), lecturer(X)**
- **poor(X) if young(X), lecturer(X)**
  - **Same accuracy but not as compressive.**

20

EXAMPLE: Verifying Hypotheses

- poor(X) if lecturer(X)
- The socio-economy specialist is sceptical as it knows that students are poor.
  - The hypotheses is rejected and new partial information is now given:

    poor(X) if student(X)
- Model now is refined to:

sad(X) if not student(X),overworked(X),poor(X)
sad(X) if student(X), alone(X)
poor(X) if student(X)

EXAMPLE: Verifying Hypotheses

- Model now is refined to:

sad(X) if not student(X),overworked(X),poor(X)
sad(X) if student(X), alone(X)
poor(X) if student(X)

- Observations = {sad(alex), sad(krycia), not sad(oli)}
- Abductive Explanation =

  {poor(ale), poor(krycia), not alone(oli)}
- Inductive Hypotheses: poor(X) if young(X), lecturer(X)

(Note: "poor(X) if young(X)" is rejected because we know young(bill), rich(bill).)

Uncertainty in the Model/Hypotheses

- Rules are not absolute!
- Non-monotonic Logic
  - Rules and Exceptions
  - Rules and Integrity Constraints
  - Preference Reasoning
- Accuracy is typically not 100%
  - Statistical methods in search for hypotheses
  - Predictive accuracy improvement.
- No time to go into the details.

Frameworks and Systems

- Deduction: Logic Programming & PROLOG (1970-)
- Abduction: Abductive Logic Programming (1990-)
  - ACLP & A-system
  - IFF
  - SLDNFA
- Induction: Inductive Logic Programming (1990-)
  - FOIL
  - Progol
- Integrated Abduction and Induction (2000-)
  - Progol
  - HAIL

Applications to Bioscience

- **Gene Regulatory Pathways**
- **Inhibition in Metabolic Networks**
- **HIV Drug Resistance**

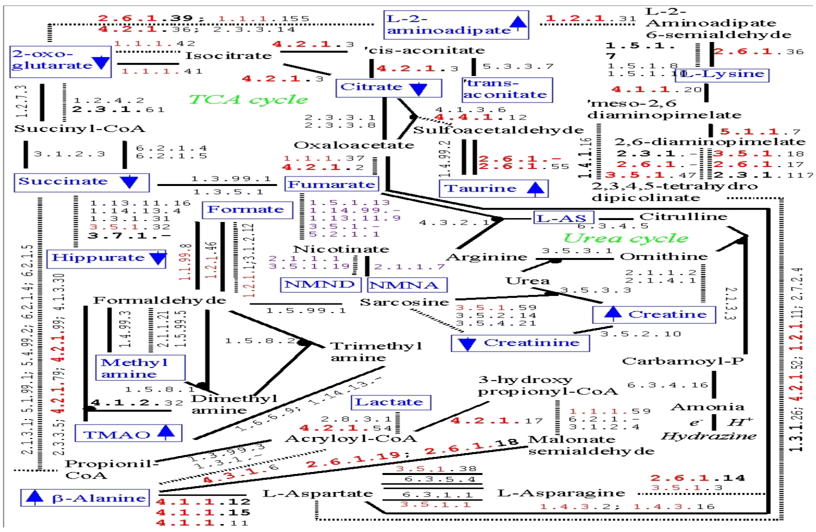
CONCLUSIONS

- **Use of logical inference can be powerful method for "rationalization" of scientific phenomena.**
  - **But it needs a good problem domain model with:**
    - **Clear hypotheses of the basic model**
    - **Parameters of variation of analysis of the phenomena**
    - **Well-informed strategy of use and feedback analysis**
- **Offers a systematic methodology for modelling and analysis under known/engineered hypotheses.**

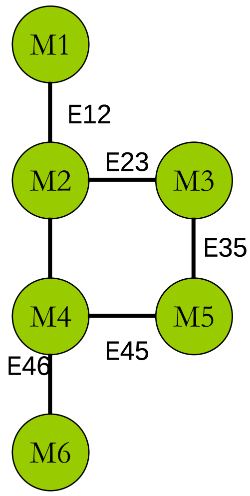
CONCLUSIONS (again)

- **In the complex bioscience domain this analysis can result in hypotheses to be tested.**
  - **Start with a good biological model and go one step further**
  - **Type of hypotheses can be controlled by "biological expertise"**
  - **Declarative and modular logical representations facilitate the modelling and analysis.**

Inhibition in Metabolic Pathways METALOG Project



Metabolic pathways ALP Model



PROGRAM T:

down(M):-
    reaction(E1,M,M1),
    inhibited(E1),
    not compensated(M,E1).

compensated(M,E1):-
    reaction(E2,M,M2),
    diff(E2,E1),
    increased(E2).

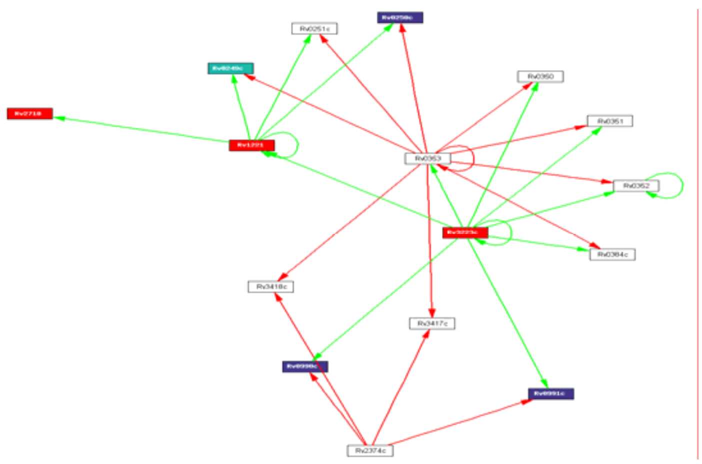
INTEGRITY CONSTRAINTS IC:

false if inhibited(E), increased(E).

¬(inhibited(E), increased(E))

29

Microarray Gene Mutation Experiments at CMMI



30