using background knowledge in differential equations discovery

symposium on computational discovery of communicable knowledge

ljupčo todorovski
jožef stefan institute, ljubljana, slovenia

stanford university, march 2001

differential equations discovery: problem definition

- GIVEN:
  - a set of system variables $V = \{v_1, v_2, \ldots v_n\}$
  - a table of measurements (observations) of the system variables

time	v_1	v_2	$\ldots$	v_n
t_0	$v_{1,0}$	$v_{2,0}$	$\ldots$	$v_{n,0}$
t_1	$v_{1,1}$	$v_{2,1}$	$\ldots$	$v_{n,1}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	
t_m	$v_{1,m}$	$v_{2,m}$	$\ldots$	$v_{n,m}$

- FIND differential equation(s) that summarize the observations:

$$F(v_1, v_2, \ldots, v_n, \dot{v}_1, \dot{v}_2, \ldots, \dot{v}_n) = 0$$

background knowledge: building blocks

- genetic programming approach (Džeroski and Petrovski 1994):
  - user specified building blocks
  - example of expression used for modeling the process of saturation in mathematical models of population dynamics:

$$\frac{v}{v + const}$$

- problem: increased complexity of the space of possible equations

background knowledge: a declarative bias approach

- space of possible equations (language bias):
  - user defined (declarative)
  - based on the background knowledge from the domain of use
- declarative bias formalism used in lagramge: context free grammar
  - prescribes the form of the expression on the right-hand side
  - generates legal expressions in the C programming language
  - user defined functions based on the background knowledge
- discovered equations are of the form $\dot{v_d} = E$,
  where E is an expression derived using the given cfg

common language biases used in equation discovery

- universal grammar (of arithmetical expressions):

$$
\begin{aligned}
E &\rightarrow E + F \mid E - F \mid F \\
F &\rightarrow F * T \mid F/T \mid T \\
T &\rightarrow const \mid v \mid (E)
\end{aligned}
$$

- multivariate polynomials:

$$
\begin{aligned}
E &\rightarrow const \mid const \cdot F \mid E + const \cdot F \\
F &\rightarrow v \mid v \cdot F
\end{aligned}
$$

population dynamics example: context free grammar

- context free grammar used for modeling population dynamics:

$$E \rightarrow const \mid const \cdot F \mid E + const \cdot F$$
$$F \rightarrow v \mid Y \mid v \cdot Y$$
$$Y \rightarrow \texttt{monod}(const, v)$$

$$\frac{v}{v + const}$$

- user defined function `monod`:

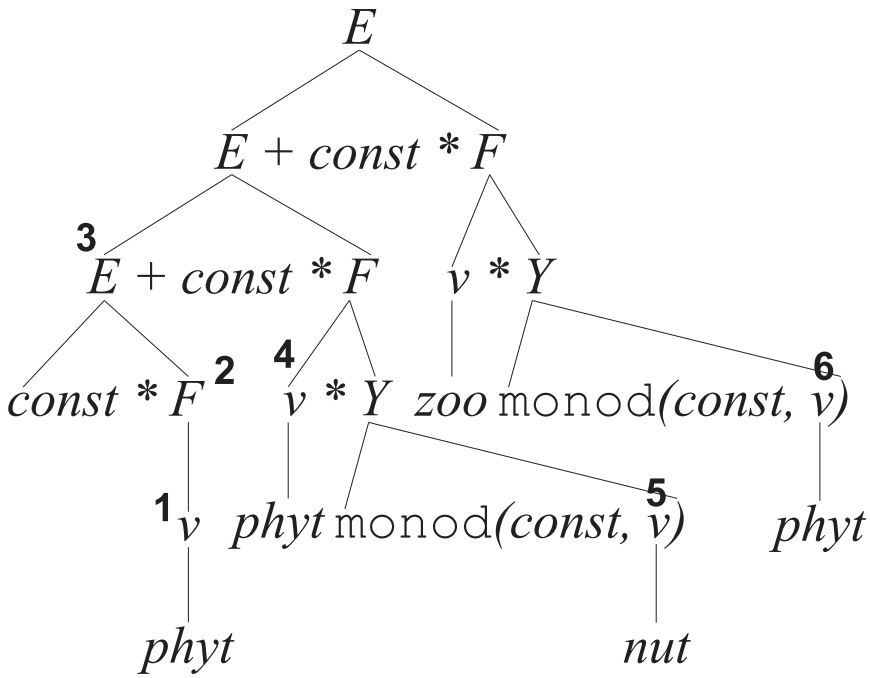
```
double monod(double c, double v) {
   return(v / (v + c));
}
```

population dynamics example: parse tree

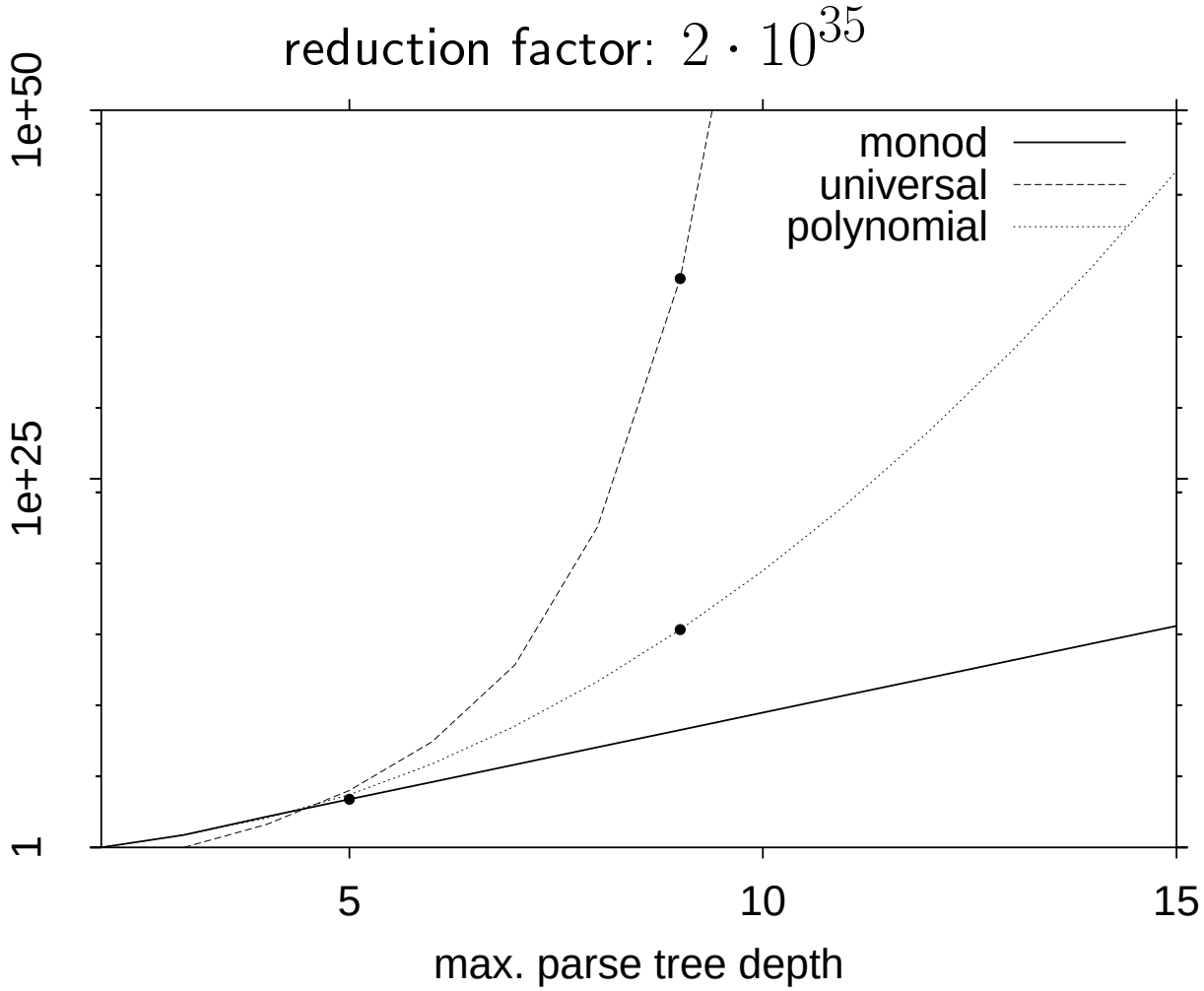
- equation:

$$\dot{phyt} = const \cdot phyt + const \cdot phyt \cdot \frac{nut}{nut+const} + const \cdot zoo \cdot \frac{phyt}{phyt+const}$$

- the parse tree:



equation space complexity: number of parse trees



lagramge: outline of the algorithm

- search space: parse trees with limited depth
- refinement operator defined on parse trees
  - orders the search space
  - allows the use of different search strategies
- two search strategies implemented:
  - beam search
  - exhaustive search
- nonlinear optimization (LSE) method for constant parameters fitting

lagramge: summary

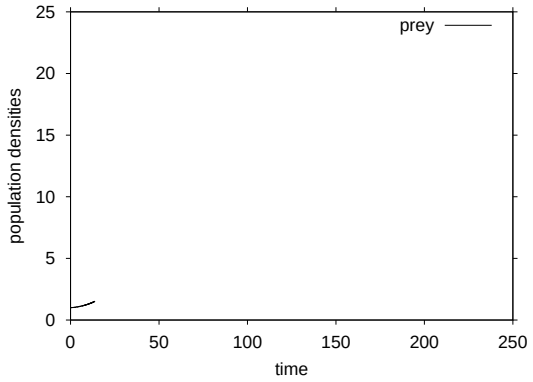
- benefits of using declarative bias in equation discovery:
  - incorporating background knowledge from the domain of use
  - compensate incomplete (and noisy) data
  - discovery of (relatively) complex models
- successfully used for modeling:
  - two poles on cart (from simulated data)
  - phyotoplankton growth in lake Glumsoe (from sparse measurements)
- how to encode a context free grammar from the domain knowledge?

population dynamics background knowledge

- collected from textbooks in mathematical modeling of biological systems
- population: group of individuals of the same species
  inhabiting the same area
- population dynamics: change of the population density
- population dynamics processes:
  - single population (natality and mortality)
  - interactions between populations
    predator-prey, symbiosis, parasytism and competitive exclusion

simple predator-prey interaction model

- Lotka-Volterra population dynamics model:
  - N is prey population density (hares)
  - P is predator population density (lynx)

$$\dot{N} = a \cdot N - b \cdot NP$$
$$\dot{P} = c \cdot NP - d \cdot P$$

- oscilatorry behavior of the Lotka-Volterra model: