Jožef Stefan Institute

Probabilistic grammars for equation discovery

Jure Brence, Nina Omejc, Boštjan Gec,
Ljupčo Todorovski, Sašo Džeroski

AAAI 2023 Spring Symposium Series

Equation discovery

(symbolic regression)

Jožef Stefan Institute

Data

v_1	v_2	$\ldots$	v_n
$v_{1,1}$	$v_{2,1}$	$\ldots$	$v_{n,1}$
$v_{1,2}$	$v_{2,2}$	$\ldots$	$v_{n,2}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$v_{1,m}$	$v_{2,m}$	$\ldots$	$v_{n,m}$

Equation discovery algorithm

Closed-form equations(s)

$$x_i = f(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_n)$$

Background knowledge

Two components:

Structure identification
- Background knowledge
- Expression generation
- Search method

Parameter estimation
- Numeric optimization
- Error-of-fit
- Well studied

Equation discovery at JSI

Jožef Stefan Institute


Grammars



Probabilistic grammars

1990 2000 2020

Process-based models

Autoencoders




Jožef Stefan Institute

Types of equations

Algebraic equations

$$y = f(x_1, x_2, \dots, x_m)$$

Newton's 2nd law

$$F = m \cdot a$$

Relativistic momentum

$$p = \frac{m_0 v}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Gaussian function

$$f = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\theta - \theta_0)^2}{2\,\sigma^2}}$$

Systems of differential equations

$$\dot{x_1} = f(x_1, x_2, \dots, x_m)$$

…

$$\dot{x_m} = f(x_1, x_2, \dots, x_m)$$

Lorenz system

$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = x(\rho - z) - y$$
$$\dot{z} = xy - \beta z$$

Glider system

$$\dot{x} = -\sin y - Dx^2$$
$$\dot{y} = -\frac{\cos x}{x} + x$$

Van der Pol oscillator

$$\ddot{x} = -x - \mu(x^2 - 1)\dot{x}$$

Integer sequences

$$a_n = f(n, a_{n-1}, \dots, a_{n-m})$$
$$n, a_{n-1}, \dots, a_{n-m} \in \mathbb{Z}$$

Fibonacci sequence:

$$a_n = a_{n-1} + a_{n-2}$$

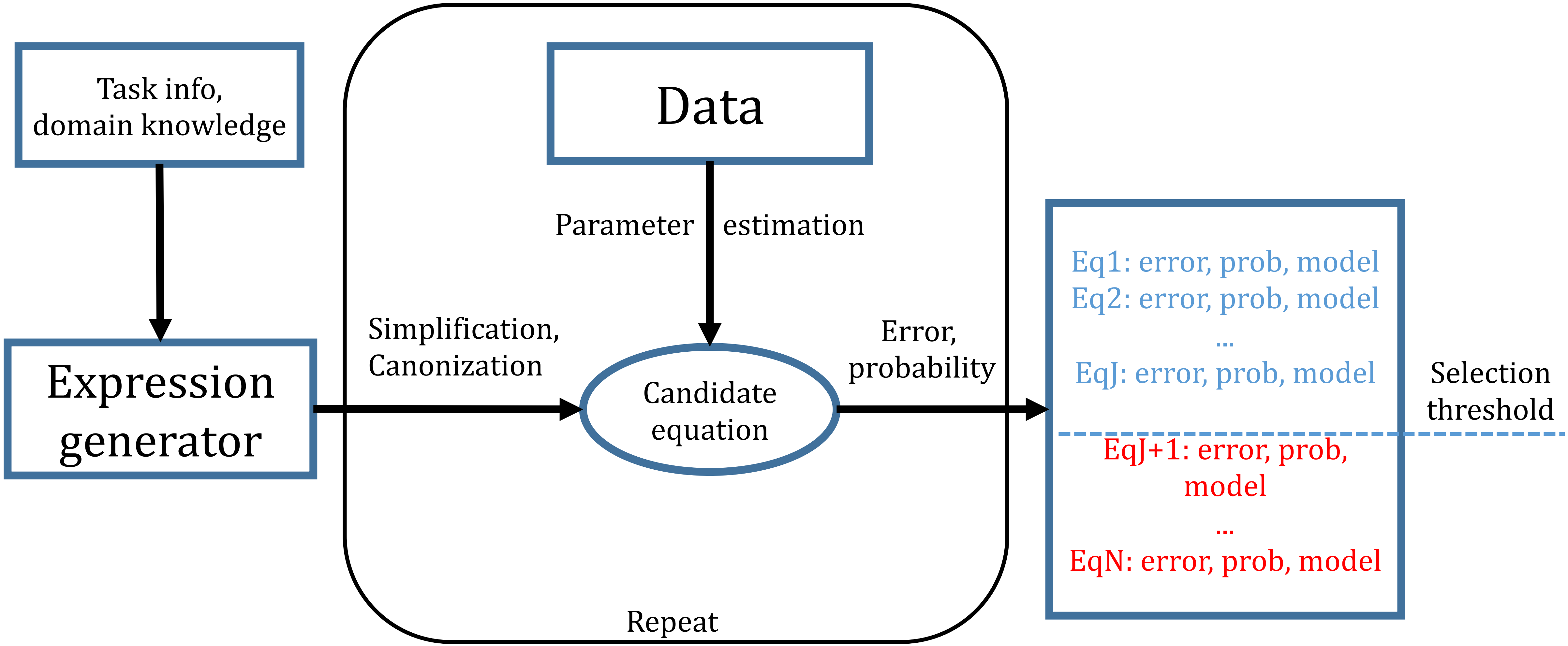
Tetrahedral numbers:

$$a_n = \frac{n(n+1)}{2}$$

Jacobsthal sequence:

$$a_n = 2^n - \mathrm{a}_{n-1}$$

Algorithm: Monte-Carlo sampling



Jožef Stefan Institute

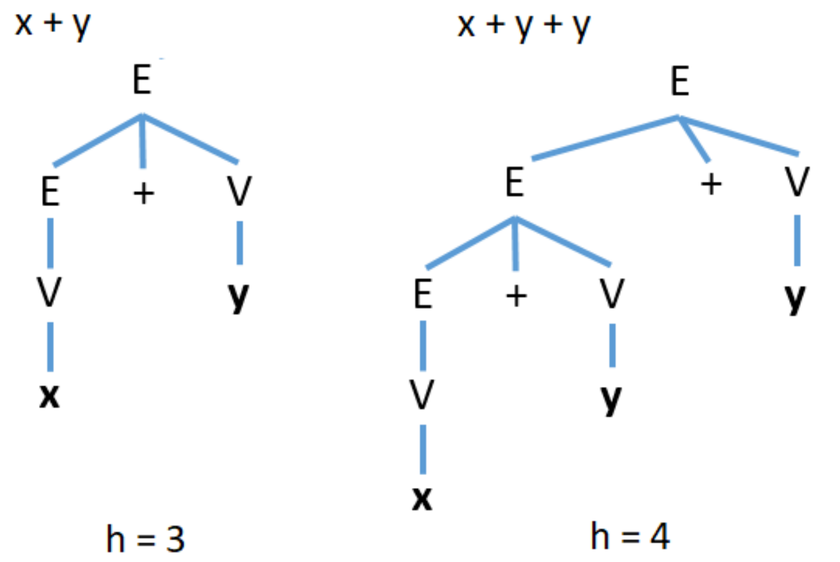
Probabilistic grammars - PCFG

Traditionally: a formal specification of a language

In our case: - the specification of the search space

- the generator of candidate expressions
- defines distribution over equations

$$
\begin{aligned}
E &\rightarrow E + V \; [p] \\
E &\rightarrow V \; [1 - p] \\
V &\rightarrow x \; [q] \\
V &\rightarrow y \; [1 - q]
\end{aligned}
$$



Jožef Stefan Institute

Parsimony in equation discovery

- Parsimony principle:
  simpler explanations are more likely to be correct.
- Model selection: error vs. complexity
- Common solution: regularization term
- Inherent in PCFGs:

$$P(\psi) = \prod_{(A \to \alpha) \in \mathcal{R}} P(A \to \alpha)^{f(A \to \alpha, \psi)}$$

recursive production

p governs parsimony

$$\boldsymbol{E} \to \boldsymbol{E} + \boldsymbol{V}\ [\boldsymbol{p}]$$
$$E \to V\ [1-p]$$
$$V \to x\ [q]$$
$$V \to y\ [1-q]$$

J. Brence et al. (2021). "Probabilistic grammars for equation discovery". Knowledge-Based Systems 224: 107077.

Results – Feynman database*

- Popular benchmark for equation discovery / symbolic regression
- 100 algebraic equations from physics texbooks by R. Feynman
- Used generator: universal arithmetic grammar

```
exp(-theta**2/2)/sqrt(2*pi)
exp(-(theta/sigma)**2/2)/(sqrt(2*pi)*sigma)
exp(-((theta-theta1)/sigma)**2/2)/(sqrt(2*pi)*sigma)
sqrt((x2-x1)**2+(y2-y1)**2)
G*m1*m2/((x2-x1)**2+(y2-y1)**2+(z2-z1)**2)
m_0/sqrt(1-v**2/c**2)
x1*y1+x2*y2+x3*y3
mu*Nn
q1*q2*r/(4*pi*epsilon*r**3)
q1*r/(4*pi*epsilon*r**3)
```

Successfully reconstructed equations: 36

-> limits of Monte-Carlo

-> unconstrained space, no background knowledge

* Udrescu, S. M., & Tegmark, M. (2020). AI Feynman: A physics-inspired method for symbolic regression. Science Advances, 6(16), eaay2631.

Constraints on the search space

- Search space is generally infinite → need for constraints
- Background knowledge → constraints

$$
\begin{aligned}
E &\rightarrow E + V \ [0.4] \\
E &\rightarrow V \ [0.6] \\
V &\rightarrow x \ [0.75] \\
V &\rightarrow y \ [0.25]
\end{aligned}
$$

Production rules
- **hard** constraints

Rule probabilities
- **soft** constraints

Domain-specific knowledge

Jožef Stefan Institute

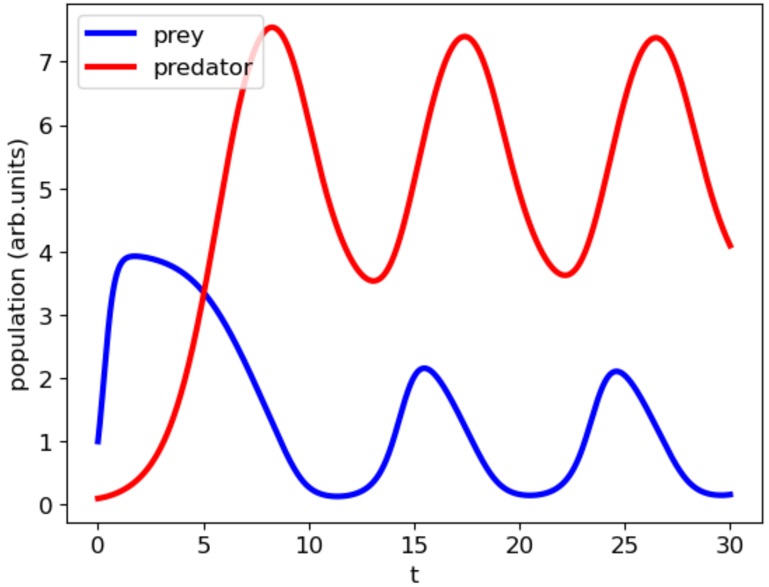
State oscillators

- Polynomials
- Population models: the monod function $\frac{v}{v+c}$

$$E \rightarrow E + T \mid T$$
$$T \rightarrow T * V \mid T * M \mid c$$
$$M \rightarrow \frac{x}{x+c} \mid \frac{y}{y+c}$$
$$V \rightarrow x \mid y$$

Example – predator-prey

$$\dot{x} = x\left(b - x - \frac{y}{y+1}\right)$$
$$\dot{y} = y\left(\frac{x}{x+1} - ay\right)$$



Domain-specific knowledge

Phase oscillators:

- Linear combination of terms
- Terms: sin and cos of variables
- Arguments: linear functions

$$E \rightarrow E + c * T \mid c * T \mid \mathrm{c}$$

$$T \rightarrow \sin(L) \mid \cos(\mathrm{L})$$

$$L \rightarrow L + c * V \mid c$$
$$V \rightarrow x \mid y$$

Example - bar magnets

$$\dot{x} = K\sin(x-y) - \sin(x)$$
$$\dot{y} = K\sin(y-x) - \sin(y)$$

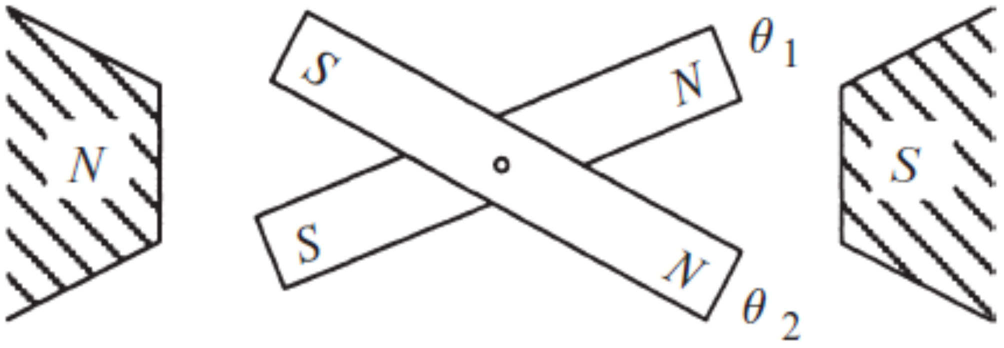
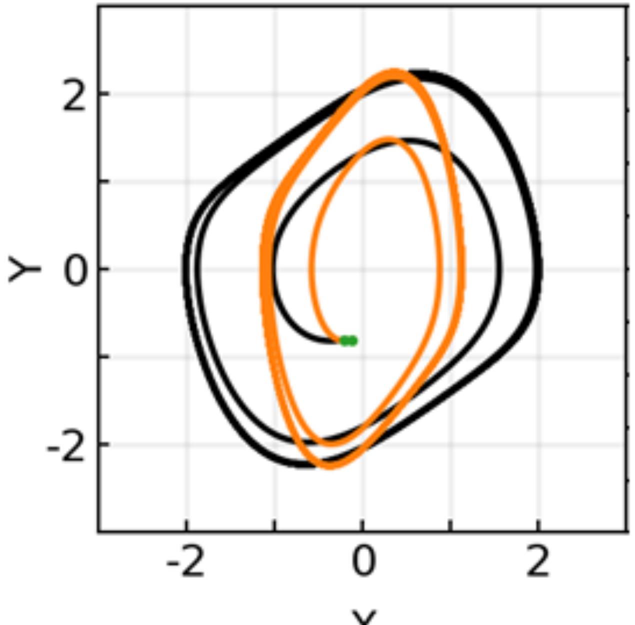


Image from Steven H. Strogatz, "Nonlinear Dynamics and Chaos with Student Solutions Manual", 2015.

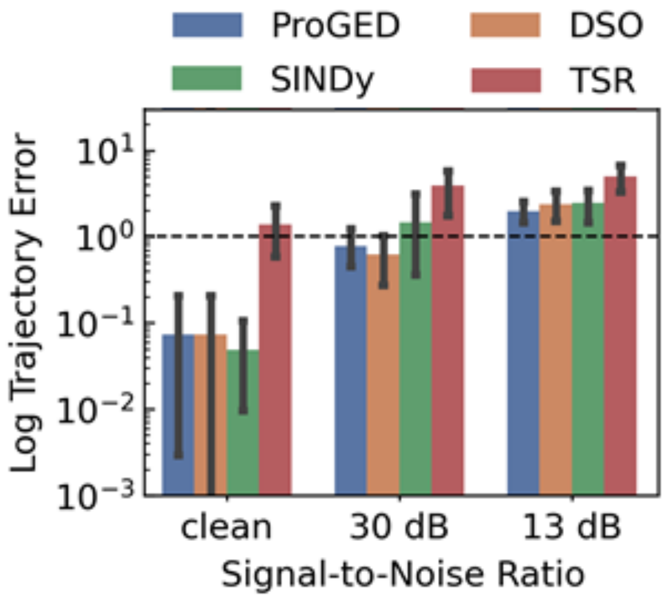
Results – dynamical systems

Jožef Stefan Institute

- **Competitive results** for fully-observed systems
- Can handle **partially-observed** systems well



Extended Strogatz benchmark:
10 dynamical systems



N. Omejc, et al. (2023). "Probabilistic grammars for modeling dynamical systems from coarse, noisy, and partial data". In review.

Learning from equation corpora

- Grammar probabilities – “soft” encoding of knowledge

Initial grammar

$E \rightarrow E + F\ [0.2]\ |\ E - F\ [0.2]\ |\ F\ [0.6]$

$F \rightarrow F * T\ [0.2]\ |\ F / T\ [0.2]\ |\ T\ [0.6]$

$T \rightarrow R\ [0.2]\ |\ V\ [0.4]\ |\ c\ [0.4]$

$R \rightarrow (E)\ [0.6]\ |\ \sin(\mathrm{E})\ [0.1]\ |\ \cos(\mathrm{E})\ [0.1]$

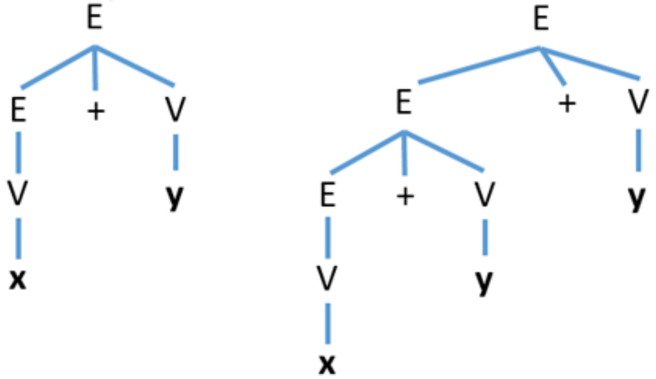
$\rightarrow \sqrt{E}\ [0.1]\ |\ \exp(E)\ [0.1]$

Parse

Corpus of equations

n_rho*mom*tanh(mom*B/(kb*T))
mom*H/(kb*T)+(mom*alpha)/(epsilon*c**2*kb*T)*M
mom*(1+chi)*B
Y*A*x/d
Y/(2*(1+sigma))
1/(exp((h/(2*pi))*omega/(kb*T))-1)
(h/(2*pi))*omega/(exp((h/(2*pi))*omeg

Parse trees



Production frequencies

Updated grammar

$E \rightarrow E + F\ [0.10]\ |\ E - F\ [0.15]\ |\ F\ [0.75]$

$F \rightarrow F * T\ [0.36]\ |\ F / T\ [0.24]\ |\ T\ [0.40]$

$T \rightarrow R\ [0.15]\ |\ V\ [0.72]\ |\ c\ [0.13]$

$R \rightarrow (E)\ [0.56]\ |\ \sin(\mathrm{E})\ [0.12]\ |\ \cos(\mathrm{E})\ [0.09]$

$\rightarrow \sqrt{E}\ [0.14]\ |\ \exp(E)\ [0.09]$

General knowledge: dimensions

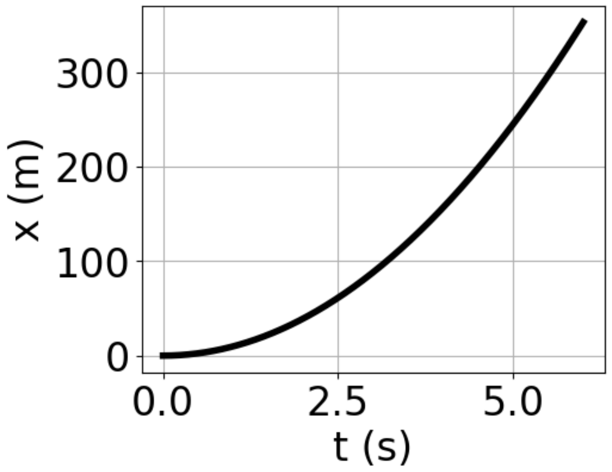
- Units of measurement
- Related: Buckingham PI theorem, dimensional analysis
- Impose constraints on the structure of expressions

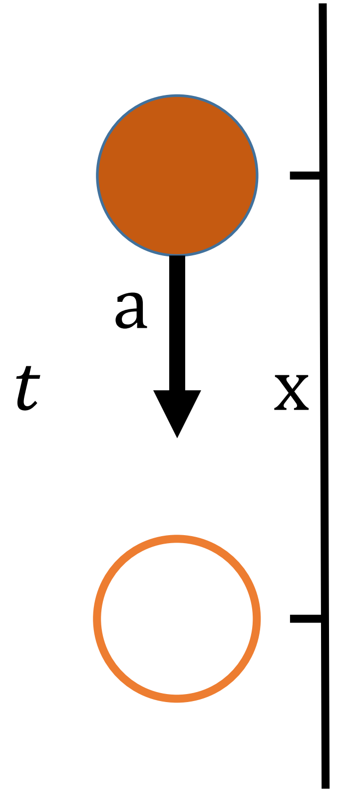
$$u(x) = \boldsymbol{m} = (1,0)$$
$$u(t) = \boldsymbol{s} = (0,1)$$
$$u(a) = \frac{\boldsymbol{m}}{\boldsymbol{s}^2} = (1,-2)$$

$$u(v_1 \pm v_2) = u(v_1) = u(v_2)$$

$$u(v_1 * v_2) = u(v_1) + u(v_2)$$
$$u(v_1 / v_2) = u(v_1) - u(v_2)$$

$$\boldsymbol{x = at^2}$$





General knowledge: dimensions

Attribute grammars

- Nonterminals can have *attributes*
- *Attribute rules* encode knowledge

Dimensional attribute grammar

$$
\begin{array}{lll}
P & \rightarrow P + c * M & \{P1.u = P2.u = M.u\} \\
 & \rightarrow c * M & \{P.u = M.u\} \\
M & \rightarrow M * V & \{M1.u = M2.u + V.u\} \\
 & \rightarrow V & \{M.u = V.u\} \\
V & \rightarrow a & \{V.u = a.u\} \\
 & \rightarrow t & \{V.u = t.u\}
\end{array}
$$

transform ⇒

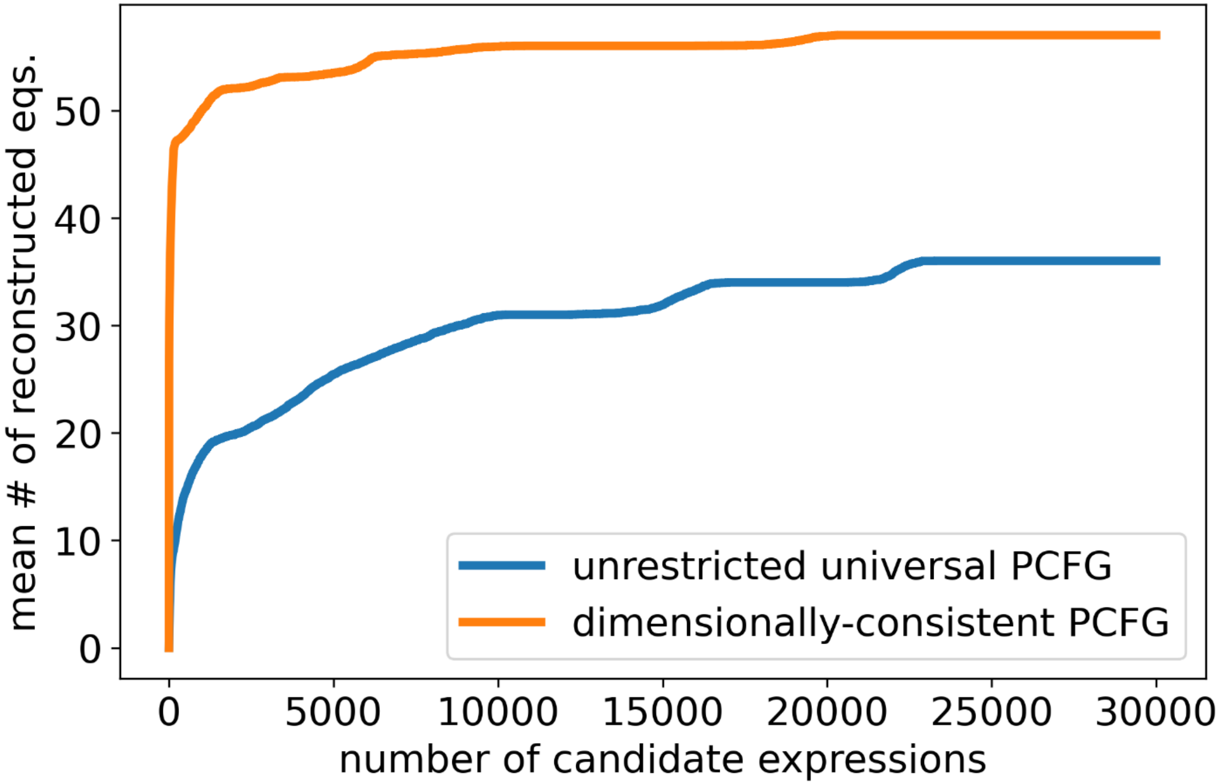
Dimensionally-consistent context-free grammar

$$
\begin{array}{ll}
P_{(1,0)} & \rightarrow P_{(1,0)} + c * M_{(1,0)} \\
 & \rightarrow c * M_{(1,0)} \\
M_{(1,0)} & \rightarrow M_{(1,-1)} * V_{(0,1)} \\
 & \rightarrow V_{(0,1)} \\
\ldots & \\
V_{(1,0)} & \rightarrow x \\
V_{(0,1)} & \rightarrow t
\end{array}
$$

Jožef Stefan Institute

Results – Feynman database, again

- Total equations: **100**
- Universal grammar: **36**
- Dimensionally-consistent universal grammar: **58**



Dimensional consistency enables

-> more successes

-> with fewer candidates

J. Brence, et al. (2023). "Dimensionally consistent equation discovery through probabilistic attribute grammars". Information Sciences.

Summary

Probabilistic grammars enable

- inherent and intuitive control over **parsimony**,
- the encoding of general and domain-specific **background knowledge,**
- both hard and **soft constraints** on the space of equations.

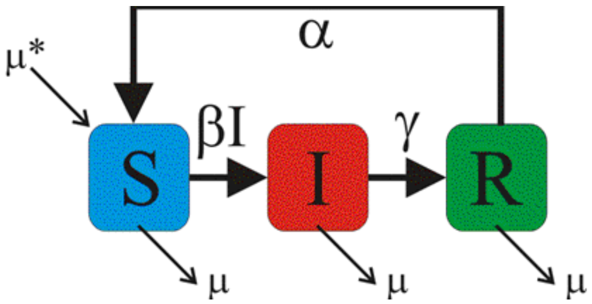
Attribute grammars are a promising framework for more complex types of background knowledge.

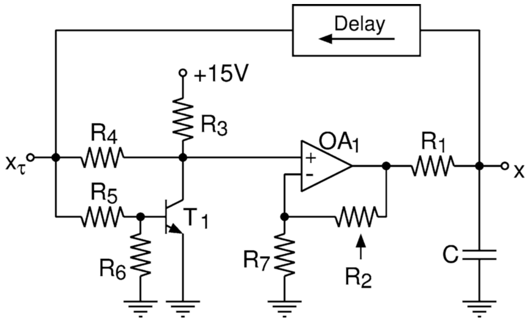
Jožef Stefan Institute

Further work

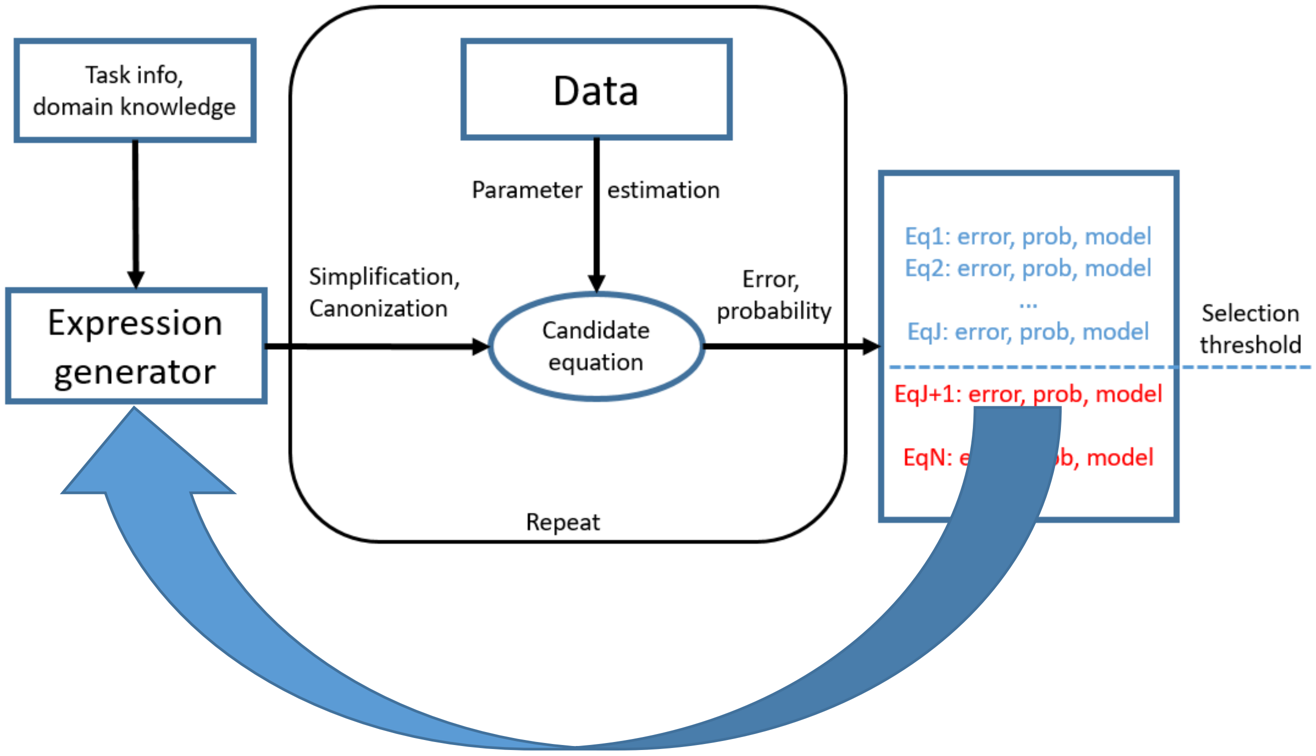
Attribute grammars

- Direct sampling
- Other types of knowledge:
  - Process-based models
  - Compartmental models
  - Electronic circuits
  - ...





Algorithmic improvement

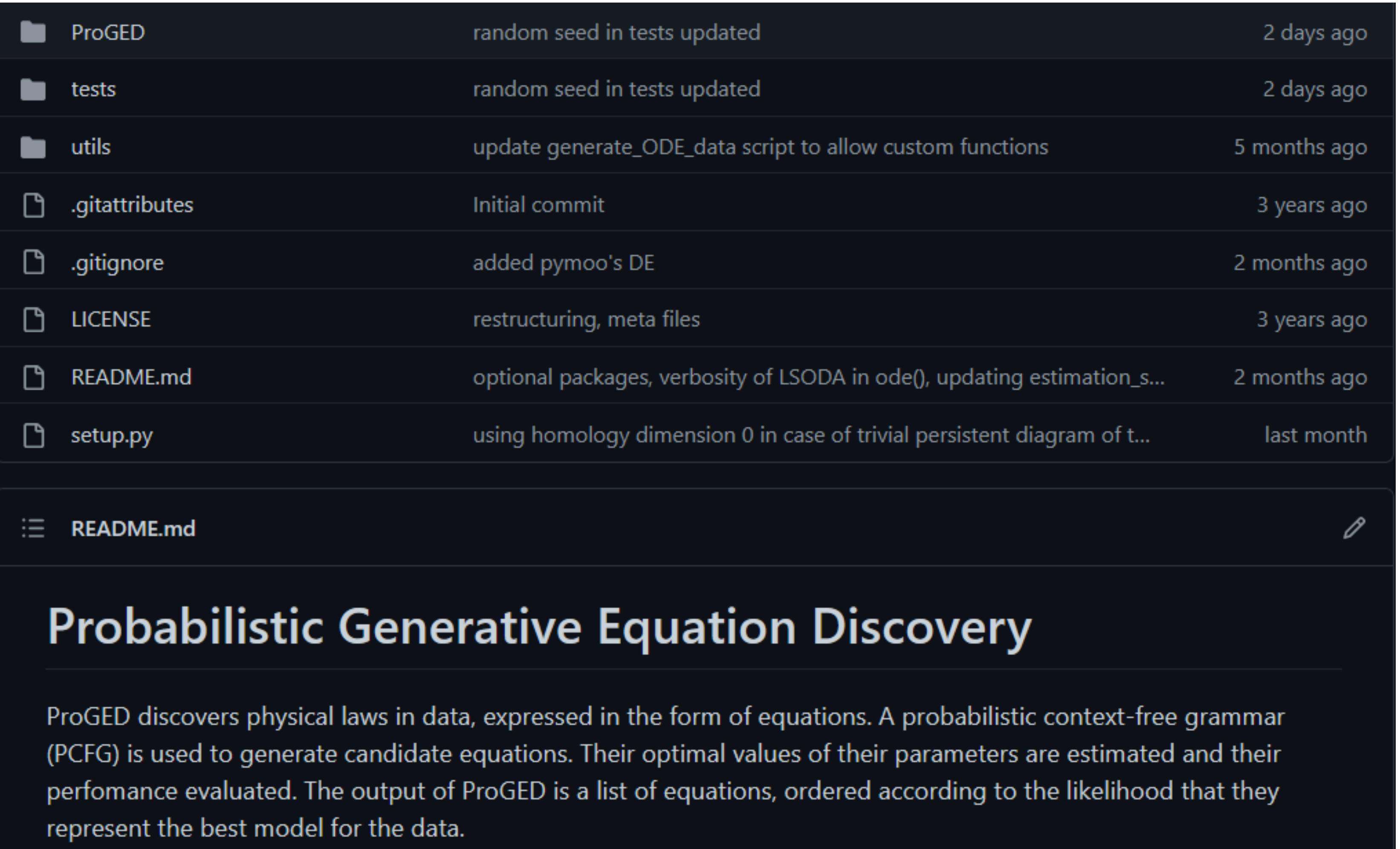


Iterative grammar updating

Jožef Stefan Institute

ProGED

https://github.com/brencej/ProGED



ProGED	random seed in tests updated	2 days ago
tests	random seed in tests updated	2 days ago
utils	update generate_ODE_data script to allow custom functions	5 months ago
.gitattributes	Initial commit	3 years ago
.gitignore	added pymoo's DE	2 months ago
LICENSE	restructuring, meta files	3 years ago
README.md	optional packages, verbosity of LSODA in ode(), updating estimation_s...	2 months ago
setup.py	using homology dimension 0 in case of trivial persistent diagram of t...	last month

README.md

Probabilistic Generative Equation Discovery

ProGED discovers physical laws in data, expressed in the form of equations. A probabilistic context-free grammar (PCFG) is used to generate candidate equations. Their optimal values of their parameters are estimated and their perfomance evaluated. The output of ProGED is a list of equations, ordered according to the likelihood that they represent the best model for the data.

- Python package
- Open source
- Active development

Thank you for your attention!

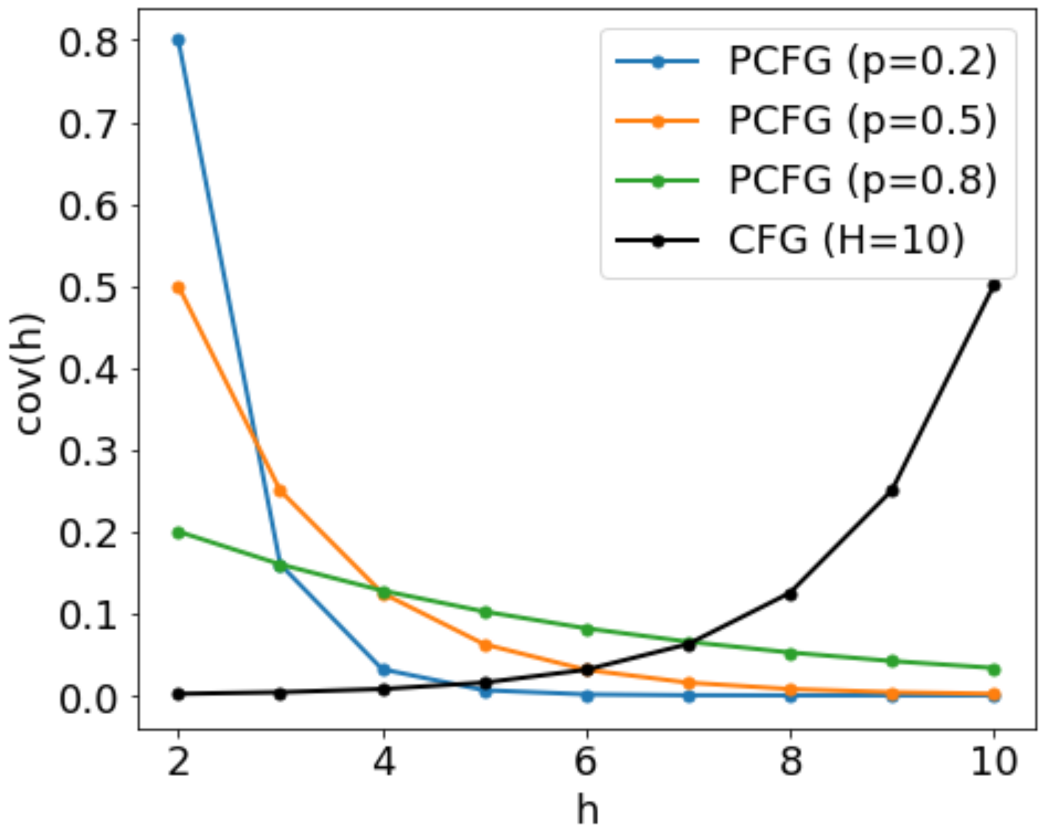
Extras

Jožef Stefan Institute

CFG vs PCFG

- **cov** - prob. of all parse trees with height h
- Number of parse trees increases with height.
- In CFGs, all parse trees are equally likely.

- PCFG: cov drops exponentially
- CFG: cov rises exponentially



Domain-specific knowledge

Dynamical systems:
- Systems of ODEs
- Linear combinations of a few terms:
  - Low order monomials
  - Simple rational functions
  - sin or cos functions

Example: glider

$$\dot{x} = -\sin(y) - Dx^2$$

$$\dot{y} = -\frac{\cos(y)}{x} + x$$

$$
\begin{aligned}
E &\rightarrow E + T \ [0.6] \quad | \ T/(D) \ [0.15] \quad | \ T \ [0.25] \\
D &\rightarrow D + T \ [0.5] \quad | \ T \ [0.5] \\
T &\rightarrow T * V \ [0.3] \quad | \ T * R \ [0.1] \quad | \ c \ [0.6] \\
R &\rightarrow \sin(M) \ [0.5] \quad | \ \cos(M) \ [0.5] \\
M &\rightarrow M * V \ [0.5] \quad | \ c \ [0.5] \\
V &\rightarrow x \ [0.5] \quad | \ y \ [0.5]
\end{aligned}
$$