NATIONAL RESEARCH
UNIVERSITY

Project Proposal

Machine Learning Approach
to Search for Patterns in Physical Phenomena

Mikhail Zybin

National Research University
Higher School of Economics (Moscow)
Faculty of Mathematics

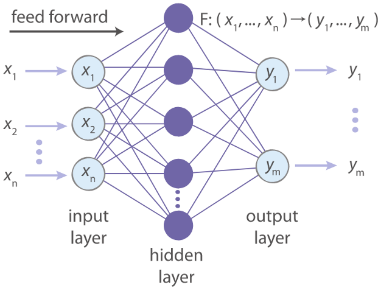
April 21, 2020

This work is an overview of several existing machine learning methods that aim at extracting knowledge from empirical data, and an attempt to create a new method.

Machine Learning algorithm

$f(x, \theta) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$

Multilayer perceptron
$f_n \circ f_{n-1} \circ \ldots f_1(x)$,
f_k - a layer, composition of nonlinearity
(e. g. $\frac{1}{1+e^{-x}}$) and affine transformation



Introduction

Definitions

This work is an overview of several existing machine learning methods that aim at extracting knowledge from empirical data, and an attempt to create a new method.

A *regression problem* $\{(X_i, y_i)\}, i = 1..N, X_i \in \mathbb{R}^n, y_i \in \mathbb{R}$

the goal: to predict values of y that correspond to values of X that are not present in the dataset.

A *symbolic regression problem*

explicitly understand the formula that maps X to y

In 1601, Johannes Kepler got access to the world's bestdata tables on planetary orbits, and after 4 years andabout 40 failed attempts to fit the Mars data to variousovoid shapes, he launched a scientific revolution by dis-covering that Mars' orbit was an ellipse.

goal - how can Machine Learning help scientists? can Artificial Intelligence become capable of investigating the laws of nature?

here I will cover some humble steps in this direction that were made by me and other researchers

The idea of this algorithm is to provide a general model for many possible scientific laws and recover them using the optimization of the model parameters.

$\varphi_0(x_1, \dots x_n) = \lambda_0$ for some $\lambda_0 \in \mathbb{R}$

$$\begin{aligned}\varphi_{k+1}(x_1, \dots x_n) = \psi(x_1, \dots x_n) + \sum_{i=1}^{n} \Big(& \lambda_i x_i^{\alpha_i} \varphi_i(x_1, \dots x_n) + \\ & \mu_i \sin\left(\chi_i(x_1, \dots x_n)\right) + \\ & \nu_i \cos\left(\psi_i(x_1, \dots x_n)\right) \Big)\end{aligned}$$

where $\varphi_i, \chi_i, \psi_i, \psi$ are nested formulas of depth k, $\lambda_i, \mu_i, \nu_i \in \mathbb{R}$,

Nested Formulas

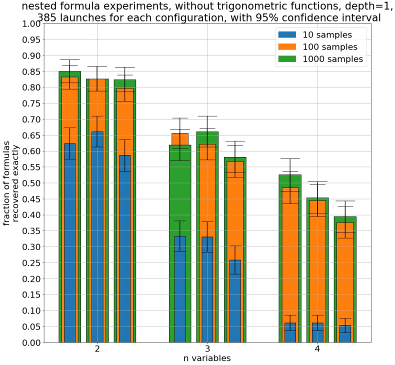
Examples

$Gm_1m_2r^{-2} + kq_1q_2r^{-2} = F$ (depth 3, universal gravity + Coulomb's law)

$T - \frac{2\pi}{\sqrt{G}}RM^{-0.5}l^{0.5} - \frac{2\pi}{\sqrt{G}}RM^{-0.5}hl^{0.5}$ (depth 2, pendulum + universal gravity)

$qvB\sin(\alpha) = F$ (depth 4, Ampère's force law)

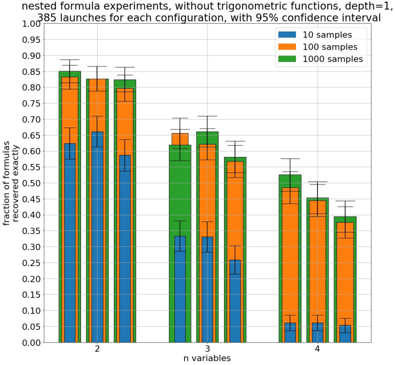
Nested Formulas
Results



$\varphi = \lambda_0 + \sum_{k=1}^{n} \lambda_k x_k^{\alpha_k}$
exact recovery - mean squared error between recovered parameters and true parameters is less than 10^{-4}

Nested Formulas

Results



α_i selection methods

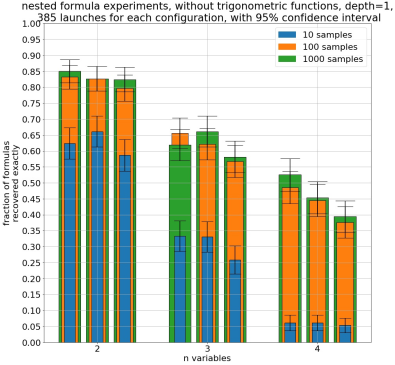
Uniform$(\{1, 2, 3, 4, 5, 6\})$

Uniform$(\{\frac{1}{2}, 1, \frac{3}{2}, 2, \ldots, \frac{11}{2}, 6\})$

Uniform$(\{\frac{k}{12}\}_{k=1}^{72})$

Nested Formulas

Results



$$2 \cdot 1.96\sqrt{\frac{\hat{p_n}(1-\hat{p_n})}{n}} < 0.1$$

$$\hat{p_n}(1-\hat{p_n}) \leq 0.25$$

$\lceil 19.6^2 \rceil = 385$ experiments

77 hours

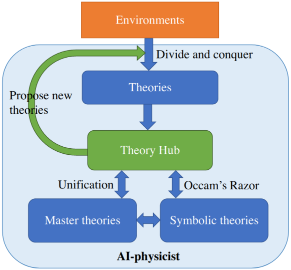
- measure the recovery rate with different number of variables, samples, and ways of selecting powers
- measure recovery rate for deeper formulas
- implement the possibility of adding trigonometric and other functions
- simplification (0.5 instead of 0.49999342)

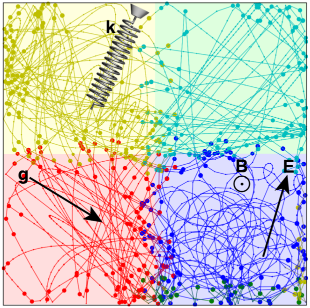
The algorithm is implemented using Python3 programming language and PyTorch framework.
The code is available at
https://github.com/mishazybin/NestedFormulas

scientific theory - a formula that can be used for prediction, and an area of space where this formula is applicable
given movements of a particle in the environment and learn the physical laws that govern the environment
$f : \mathbb{R}^{n \times m} \to \mathbb{R}^n$ maps m previously observed coordinates of the particle to its coordinate in the next time step
$c : \mathbb{R}^n \to [0, 1]$ - is theory applicable here?

Literature Overview

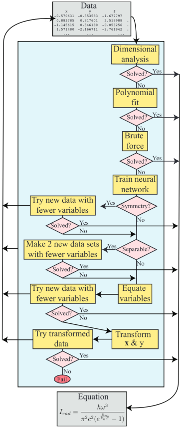
Toward an AI Physicist for Unsupervised Learning





Literature Overview

AI Feynman: a Physics-Inspired Method for Symbolic Regression



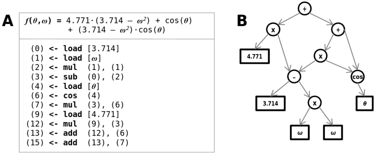
Literature Overview

AI Feynman: a Physics-Inspired Method for Symbolic Regression

- *Units*: f and the variables upon which it depends have known physical units
- *Low-order polynomial*: f(or part thereof) is a polynomial of low degree
- *Compositionality*: f is a composition of a small set of elementary functions, each typically taking no more than two arguments
- *Smoothness*: f is continuous and perhaps even analytic in its domain
- *Symmetry*: f exhibits translational, rotational or scaling symmetry with respect to some of its variables
- *Separability*: f can be written as a sum or product of two parts with no variables in common

Literature Overview

Genetic Algorithm



repeat until reasonable quality is reached:

1. Random changes and/or merging
2. Selection of the best models using the following function:

$$-\frac{1}{N}\sum_{i=1}^{N}\log\left(1+abs\left(\frac{\Delta x_i}{\Delta y_i}-\frac{\delta x_i}{\delta y_i}\right)\right),$$

where $\frac{\Delta x}{\Delta y} \approx \frac{dx}{dt}/\frac{dy}{dt}$, $\frac{\delta x}{\delta y} = \frac{\delta f}{\delta y}/\frac{\delta f}{\delta x}$

Conclusion

The field of Machine Learning is rapidly growing and evolving. As Max Tegrark put it in his paper, We look forward to the day when, for the first time in the historyof physics, a computer, just like Kepler, discovers a usefula nd hitherto unknown physics formula through symbolic regression!
There are many opportunities and possible future work directions, that I have not coreved in this proposal:

- creating an agent that would interact with the environment and make experiments
- hamiltonian neural networks
- compact latent representations

Acknowledgments

This work is a part of a project held at the Laboratory of Methods for Big Data Analysis, Faculty of CS, HSE.
The team working on this project includes

- Vladislav Belavin, Research Assistant, Postgraduate Student at the Laboratory of Methods for Big Data Analysis, Faculty of Computer Science, HSE), de-facto project supervisor
- Julia Semavina, 3rd-year graduate student at Faculty of CS, HSE
- Andrey Ustyuzhanin, PhD, Head of the Laboratory of Methods for Big Data Analysis, Associate Professor at the Faculty of CS, HSE
- Denis Derkach, PhD, Senior Research Fellow at the Laboratory of Methods for Big Data Analysis, Assistant Professor at the Faculty of CS, HSE

Bibliography

- Tailin Wu, Max Tegmark *Toward an AI Physicist for Unsupervised Learning*, 2019
  https://arxiv.org/pdf/1810.10525
- Michael Schmidt, Hod Lipson *Distilling Free-Form Natural Laws from Experimental Data*, 2009
  https://www.uvm.edu/~cdanfort/courses/237/schmidt-lipson-2009.pdf
- Sam Greydanus, Misko Dzamba, Jason Yosinski, *Hamiltonian Neural Networks*, 2019
  https://arxiv.org/pdf/1906.01563.pdf
- Peter Tot, Danilo J. Rezend, Andrew Jaegle, Sébastien Racanière, Aleksandar Botev, Irina Higgins, *Hamiltonian Generative Networks*, 2019
  https://arxiv.org/pdf/1909.13789.pdf

Bibliography

- Raban Iten, Tony Metger, Henrik Wilming, Lidia del Rio, and Renato Renner, *Discovering physical concepts with neural networks*, 2020
  https://arxiv.org/pdf/1807.10300.pdf
- Silviu-Marian Udrescu, Max Tegmark *AI Feynman: a Physics-Inspired Method for Symbolic Regression*, 2019
  https://arxiv.org/pdf/1905.11481.pdf

Thank you for your attention!

Project Proposal

Mikhail Zybin

National Research University
Higher School of Economics (Moscow)
Faculty of Mathematics

April 21, 2020