# INFO8010: Reading Assignment - Discovering Symbolic Models from Deep Learning with Inductive Biases

**Maxime Amodei**[1] and **Cédric Siboyabasore**[2]
[1] *maxime.amodei@student.uliege.be (s171832)*
[2] *c.siboyabasore@student.uliege.be (s175202)*

## I. INTRODUCTION

Finding a mathematical formula to characterize a physical phenomenon is highly laborious. Machine learning techniques such as symbolic regression exist to automate this task but can become intractable[1]. Deep learning methods are difficult to interpret and much more complex than the laws they approximate. In this paper [2], the authors try to get the best of both worlds by first applying a Graph Neural Network (GNN) to the data because of its strong physical inductive biases then a symbolic regression to fit internal components of the GNN hoping to obtain a simple expression. They first apply this framework to recover Newton and Hamilton laws and by the end, they hope to improve generalization by discovering new symbolic expressions for non-trivial astrophysical datasets.

## II. INSPIRATION

As bachelor students in engineering, how long it took for renowned physicians to derive the scientific formulas we used in class is a question that crossed our minds more than once. Now, as master students in artificial intelligence, it is natural for us to wonder how to automate such a tedious task with deep learning.
Our interest was also nurtured during a Deep Learning class in which we learned about [3] where a physics simulator is approximated by using a Graph Network that rediscovers the laws of fluid dynamics. The work we here summarize is an extended version of that paper.
We believe this work is relevant as our daily life physics include particles interactions.

## III. BACKGROUND

A good portion of the authors of the analyzed paper [2] worked on [4] in which they managed to show that they could train graph-based models on the physical data that generalize well. Based on these GNNs they fitted symbolic expressions for the messages of the graph and showed that this method generalizes much better than the symbolic expression learned on the raw data (without neural network). By their words, the new paper is a generalized and extended version of [4].

Graph Networks use three distinct and interpretable functions: the edge model $\phi^e$ maps each connected pair of nodes in a graph to a message vector. These messages vectors are summed element-wise for each receiving node. Given the summed messages vectors, the node model $\phi^v$ updates the receiving nodes. The global model $\phi^u$ aggregates all messages and all updated nodes and computes a global property. After training the Graph Network, the authors perform symbolic regression to fit closed-form expressions using basic mathematical operations and functions to $\phi^e$, $\phi^v$ and $\phi^u$ independently by using symbolic regression package *eureqa* [1]. The authors then refit the symbolic model parameters a second time to the data to avoid an accumulation of the approximated error.

## IV. METHODOLOGY

The first case study on which the authors apply their framework is on Newton dynamics data, as there exist analogies between dynamics of particles and the messages exchanged in a Graph Network : forces defined on pairs of particles are analogous to the message function $\phi^e$ in the Graph network. The Graph Network is trained on a dataset of N-body particle simulations undergoing and exercising multiple interaction forces on each other in $D = 2$ or $D = 3$ dimensions and must predict the instantaneous acceleration of every particle. Out of 4 different strategies (standard, bottleneck, L1 regularization and KL regularization) for Graph Network training, the authors find that the model with the L1 regularization has the greatest prediction performance and its $D$ message features with the highest variance have the highest correlation with a linear combination of the force components. The authors then apply symbolic regression to fit the most significant messages $\phi_i^e$ of the graph to physical input variables by minimizing the mean absolute error. They then succeed in recovering analytical expressions corresponding to the simulated force laws.

For Hamilton dynamics, which characterizes a system of particles by their summed energies, the authors train a Flattened Hamiltonian Graph Network [5] on the same dataset they used for Newton dynamics. Symbolic regression applied to the message function learned by the model allows the authors to successfully extract potential energies, rather than forces like in Newton dynamics.

The authors then apply their framework on a static dataset of N-body dark matter simulations from [6] for which there exist no known laws, though a hand-designed estimator was made by [7]. The dataset contains dark matter halos which the authors connect together to form a graph. Then, they feed the graph to a Graph Network to predict the smoothed overdensity $\delta_i$ at the location of the center of each halo. The Graph Network using L1 regularization together with the symbolic regression yield a new formula which makes physical sense and which achieves a lower loss than the formula hand-designed by scientists in [7].

## V. CONTRIBUTIONS

By fitting the closed-form expressions to the functions $\phi^e$, $\phi^v$ and $\phi^u$ learned by Graph Networks which have well-motivated physical inductive biases, the authors manage to extend symbolic regression to high-dimensional datasets, something which was intractable in [1] where only symbolic regression was used. This framework and the more explicit substructure of the Graph Network also allowed the authors to interpret the trained weights of the network, something which is not always possible for other deep learning methods.

They recover the closed-form physical laws for Newtonian and Hamiltonian dynamics because the messages learnt by the Graph Network have a representation equivalent to a linear transformations of vector components of the true force or energy. The authors also discover a physically meaningful formula for a a non-trivial dataset with complex interactions which achieves a better fit than formulas hand-designed by scientists in [7]. After training the Graph Network on 80% of the data, the symbolic regression manages to extract the same symbolic expressions obtained when training the network on the entire dataset. It leads the authors to believe that the symbolic expression generalizes better than the graph network it was extracted from. By their work, the authors could help develop researches by discovering unknown formulas in other areas of physics.

The four training strategies (standard, bottleneck, L1 regularization and KL regularization) does not lead to the same kind of results regarding the extracted symbolic expression. The authors show that standard (i.e. no particular restriction on the message size) fails to recover the known laws by symbolic regression. Bottleneck (which limits the components of the message to mach the dimensionality of the problem) manages to recover all known laws tested in the paper but performs worse than L1 in terms of test prediction loss. The L1 regularization is applied on the message sent between all node pairs to limit the information sent. This weakens the inductive bias as the dimensionality of the messages (forces) is not respected compared to the bottleneck version. However the test errors is lower and this method recovers most laws also recovered by the bottleneck version which makes it a great candidates. The KL regularization is applied on a modified version of the GNN where each message is shaped as a Gaussian distribution rather than a single value. The messages are sent as pairs $(\mu, \log \sigma^2)$ which are regularized using the KL divergence with $\mathcal{N}(0,1)$. This regularization ensures that the messages are close to a standard normal distribution.

## VI. LIMITATIONS AND DISCUSSION

While the graph-base neural networks are likely to learn a good representation of the physical model and generalize well, there is no guarantee that any meaningful symbolic expression can be extracted from them. This point is acknowledged by the authors : some regularization techniques (L1, bottleneck) on the network allow a better extraction of the expression. This disparity of outcome seems hard to explain : while we can understand that models which are more constrained in the form of their messages are easier to interpret, not all formula can be recovered. The example with charged particles was only recovered by the bottleneck training and it is not very different from the gravitational example.

We could also argue on the complexity measures for mathematical expressions used in the paper. Measuring the complexity of a formula is challenging and it has some part of subjectivity to it. While the laws of motion recovered by the authors are relatively simple, the formula for overdensities in dark matter simulations uses the distances between the halos to a power ranging from 13 to 21 in the denominator. Although this formula certainly is better than the theoretical one it is compared to, it is also significantly more complex and trying to make theoretical generalization out of it seems impossible.

To continue on expression complexity, only one way of measuring the complexity was used, some other complexity models could provide more useful formulas obtained via symbolic regression. This isn't really a shortcoming of their work but rather an idea to extend on what was explored in the paper.

The approach taken by the authors gives an interesting solution to the shortcomings of symbolic regression which is often intractable on large datasets. By using their methods, the expression obtained for different problem remain remarkably simple while providing a great generalization.


[1] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.

[2] Miles Cranmer, Alvaro Sanchez-Gonzalez, Peter Battaglia, Rui Xu, Kyle Cranmer, David Spergel, and Shirley Ho. Discovering symbolic models from deep learning with inductive biases, 2020.

[3] Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter W. Battaglia. Learning to simulate complex physics with graph networks,

2020.
[4] Miles D. Cranmer, Rui Xu, Peter Battaglia, and Shirley Ho. Learning symbolic physics with graph networks, 2019.
[5] Alvaro Sanchez-Gonzalez, Victor Bapst, Kyle Cranmer, and Peter Battaglia. Hamiltonian graph networks with ode integrators, 2019.
[6] Francisco Villaescusa-Navarro, ChangHoon Hahn, Elena Massara, Arka Banerjee, Ana Maria Delgado, Doogesh Kodi Ramanah, Tom Charnock, Elena Giusarma, Yin Li, Erwan Allys, and et al. The quijote simulations, Aug 2020.
[7] Carlos S. Frenk, Simon D. M. White, Marc Davis, and George Efstathiou. The Formation of Dark Halos in a Universe Dominated by Cold Dark Matter, April 1988.