Final Project Proposal: Connecting AI X Physics and Comptuational mechanics via rate distortion theory, symbolic regression and Bayesian inference

Omar Aguilar
University of California Santa Cruz

(Dated: May 2, 2022)

I. GOAL

My primary goal is to establish connections between AI X Physics (physics-inspired algorithms for inferring physical laws) and computational mechanics as both of these lines of research share the goal of automated theory building albeit from different angles. Towards this end, I would like to deepen my understanding of automated theory building from the computational mechanics perspective by learning about rate distortion theory and Bayesian structural inference from hidden processes. Here, I am including summaries of a few of the advances in AI X Physics that seem to be connected to computational mechanics so you can advocate for me better.

A. AI Physicist + Rate Distortion Theory

The AI Physicist seeks to discover "theories" and the region in which they are valid by combining 4 "mental" strategies that physicists use on a daily basis: Divide-and-Conquer, Occam's Razor, Unification and Lifelong Learning. It seems to me that the first two strategies can be connected to computational mechanics via rate distortion theory.

1. Divide-and-Conquer

Consider an object moving in an environment divided in regions that are each of them governed by different physical laws. Conventional machine learning algorithms would fit a function to the entire environment, which leads to complicated models that lack interpretability and fleixibility. In contrast, the AI Physicist, employs a Divide-and-Conquer approach that discovers many theories that specialize on different regions. A theory is denoted by $\mathcal{T}_i = (\mathbf{f_i}, c)$ where f_i is the prediction function and c is the region where f_i is valid.

Mathematically, the Divide-and-conquer approach consists of minimizing the following novel generalized-mean loss:

$$\mathcal{L}_\gamma = \sum_t \Big(\frac{1}{M}\sum_{i=1}^{M} l[\mathbf{f_i}(\mathbf{x_t}), \mathbf{y_t}]^\gamma\Big)^{\frac{1}{\gamma}}$$

where $f_i(x_t)$ is the predicted function, y is the target, l is some non-negative distance function quantifying how far each prediction is from the target, M is the number of theories and $\gamma = -1, 0, 1$ or $-\infty$

I think Rate Distortion theory (RDT) could potentially benefit from incorporating the Divide-and-Conquer technique.

2. Occam's Razor

The AI Physicist uses Occam's razor to turn theories into compact symbolic theories. For example, the program converts $y = 0.1 + x^{1.9999998}$ into $y = x^2$. More specifically, the AI Physicist mathematically implements Occam's razor by applying the minimum description length (MDL) principle. This procedure consists of minimizing the description length DL of some theory $\mathcal{T}$ on a dataset D, which is given by

$$\mathrm{DL}(\mathcal{T}, D) = \mathrm{DL}(\mathcal{T}) + \sum_t \mathrm{DL}(u_t)$$

where $\mathrm{DL}(\mathcal{T})$ is the sum of the DLs of the numbers that specify the theory $\mathcal{T}$ and $\sum_t \mathrm{DL}(u_t)$ is the sum of the prediction errors of our theory at time step t.

The AI Physicist provides no justification for the description lengths it assigns to natural and rational numbers, which seems to suggest that this approach lacks a strong theoretical footing. Since RDT and MDL have analogous concepts and equations, RDT could potentially fill the AI Physicist's theoretical gap.

B. AI Feynman + Rate Distortion Theory

The AI Feynman (the follow-up of the AI Physicist) discovers equations using a symbolic regression approach that performs better than the state of the art, by exploiting common properties of physics equations: symmetries, separability, compositionality, low-order polynomial, etc. Moreover, the AI Feynman 2.0 discovers generalized symmetries (arbitrary modularity in the computational graph of a formula) from gradient properties of a neural network fitting. This method employs a Pareto-frontier (of description-length complexity versus inaccuracy) that provides an approximated equation at different levels of complexity (ex. classical and relativistic kinetic energy). The main limitation of the AI Feynman 2.0 is that it sometimes fails to discover the correct expression

because of noise in the data or inaccuracies introduced by the neural network fitting. Since RDT was developed with the intent of differentiating structure from noise, it could be a good candidate for correcting the limitations in the AI Feynman and further generalize symbolic regression.

C. Bayesian Symbolic framework + Bayesian Structural Inference

Inspired by the highly data-efficient ability of humans to learn and reason about their physical environment with incomplete information, the Bayesian-symbolic physics (BSP) model combines Bayesian inference of latent properties (mass, charge, etc.) and symbolic regression to learn explicit force expressions, which are functions of the latent properties. However, like the AI Feynman, BSP assumes the form of symbolic equations it will search over. Thus, it wourd interesting to see how the data-efficient Bayesian structural inference generalize BSP and strengthen its theoretical foundations.

II. SYSTEM

The state space I will consider for all of these approaches is the position-momentum state space. If time allows, I will consider other traditionally physically relevant state spaces such as current-voltage. Furthermore, the dynamic will be the predicted function that governs the system's behavior and is non-linear. This system is interesting because it is dynamic yet tractable enough to infer a set of approximated models and their corresponding symbolic equations (some of them which might be non-linear).

III. DYNAMICAL PROPERTIES

I am interested in inferring equations from noisy sparse data at different levels of approximation and if time allows, provide a computational mechanics account (as opposed to a computational cognitive science one) of the process of human discovery. For this, I will need to use the toolkit from the first part of the first course (identify the invariant sets of my system, compute Lyapunov coefficients, etc.).

IV. INTRINSIC COMPUTATION PROPERTIES

I will use the information toolkit we learned in class. In particular, I will measure processess' randomness using Shannon entropy rate and their structure using statistical complexity.

V. METHODS

Besides the toolkit learned in class, I will use rate distortion theory, symbolic regression and Bayesian structural inference.

VI. HYPOTHESIS

A. AI Physicist + Rate Distortion Theory

1. Divide-and-Conquer + Rate Distortion Theory

My hypothesis is that rate distortion theory can reproduce the results of the AI Physicist by incorporating Divide-and-Conquer strategy. Furthermore, rate distortion theory might produce more accurate results as it is more adept at dealing with realistic noisy environment.

2. Occam's Razor/MDL + Rate Distortion Theory

My hypothesis is that Rate Distortion Theory can provide a series of approximated models for our data satisfying different modelers' interests. Some modelers might want a simple linear or power law and other modelers might be interested in the actual non-linear set of equations describing the data. Computational mechanics may provide a better way to estimate the complexity of "theories", rather than relying on on arbitrary description lengths assigned to integers and rational numbers.

B. AI Feynman + Rate Distortion Theory + Symbolic Regression

My hypothesis is that rate distortion could generalize the AI Physicist's symbolic regression approach by grounding its theoretical foundation in computational mechanics.

C. Bayesian Symbolic Physics framework + Bayesian Structural Inference

My hypothesis is that Bayesian structural inference could generalize the Bayesian symbolic physics framework by grounding its theoretical foundation in computational mechanics.

VII. STEPS/TIME

1. Read literature: I have finished reading the literature from the AI X Physics community that pertains to this project. Now, I am reading the rele-

vant literature from the computational mechanics reader: 3-5 days

2. Do mathematical analysis: Establish mathematical connections between AI X Physics and computational mechanics communities and compute dynamical and intrinsic computation properties corresponding to rate distortion theory and Bayesian structural inference: 13-18 days

3. Code: Develop code of the connections between AI X Physics and computational mechanics: 7-10 days

4. Write up a report: 1-2 days (I will write as the project develops)

I believe I can complete this project in one month. I plan to continue working on this project over the summer to broaden my understanding of automated theory building and to publish a paper that can connect two communities that share a common goal.