MATDAT18: Materials and Data Science Hackathon
MATERIALS SCIENCE TEAM APPLICATION FORM

Complete and return via email to brian_reich@ncsu.edu by 15 January 2018

Team Composition (2 people max.)

Name	Department	Institution	Email
Anand Chandrasekaran	Materials Science	University of Connecticut	anand.chandrasekaran@uconn.edu
Prasanna V. Balachandran	Materials Science and Engineering	University of Virginia	pvb5e@virginia.edu

Project Title

A scalable deep representation learning framework to accelerate *ab initio* molecular dynamics simulations

Project Synopsis (approx. 100 words)

We motivate the problem of learning dynamics from *ab initio* molecular dynamics (AIMD) simulations within the framework of deep learning. One of the key challenges is how to optimally *represent* the atom coordinates for deep learning, because they must follow strict symmetry constraints (e.g translational, rotational and permutational symmetry). This is a non-trivial problem, because of the inherent complexity in capturing the many-body character of the atomic interactions and several physics-inspired strategies have been explored with varying degrees of success. We propose the problem of exploring various representation learning schemes (graphs, networks etc) to map the coordinates and utilize deep neural networks to model the AIMD dynamics.

Identified Data-Science Collaborative Need (approx. 100 words)

Representation learning is a rapidly growing field in data science, but has not been fully explored in the domain of AIMD. Traditionally, materials scientists have relied on their expertise in fingerprinting the intrinsic structure of the data. Although significant advancements have been made, it is often unclear whether some of the physics-inspired representations are indeed optimal. On the other hand, representing data in the form of networks and graphs (that are also differentiable) have been used in computer science for relational reasoning and modeling of complex and non-linear systems [1]. Desired is a principled approach towards representation learning for AIMD trajectories and, in turn, explore its potential in modeling the energies within the paradigm of a scalable deep neural network.

Data Origin and Accesss

We intend to use open source computational tools (e.g GPAW, Quantum ESPRESSO) to generate the required data for the proposed problem. We will share our data with the data science team.

Project Description (approx. 1.5 pages, plus figures and references; please describe data size, form, dimensionality, uncertainties, number of examples, etc.)

Molecular dynamics is an integral tool in materials science through which we gain an understanding of various atomistic phenomena that span across multiple length and time-scales. The accuracies of these simulations are strongly dependent on the quality of force-fields that

describe the local interaction between different atoms. Further, they are built for specific materials systems and therefore, are not transferable. Although density functional theory (DFT) provides a higher level of accuracy on the forces and total energies, it is computationally expensive for large system sizes. There is a need to develop novel strategies that has the potential to learn the dynamics from high fidelity AIMD simulations and then apply it to rapidly predict the dynamics of realistic materials systems with significantly large system sizes.

In this project, we propose to capture the complexities of the many-body interaction in AIMD simulations using a novel and scalable deep representation learning framework. Unlike earlier approaches [2-4], where the forces/energies are learned for static configurations, we propose to capture directly the dynamics the of the system, as depicted in Figure 1. Given the positions and velocities of N-atoms at a time-step 't', the model should predict positions and velocities at the next time-step 't+1'. The prediction is only likely possible if the model captures correctly the complex many-body interactions between an atom and its local environment.

We intend to generate large number of AIMD runs for clusters of elemental systems (for e.g Cu, Al, C, Ti, Fe) and intermetallics. The size of the clusters would be between 50 to 100 atoms. Randomly sampled pairs of sequential configurations (for time-step 't' and 't+1') would be then split into train, validation and test sets. The input would consist of the x,y and z coordinates of the atoms at a particular time step, atomic energies, and also the x,y and z components of the velocity for every atom at that time-step. The target would be the same set of properties but for the subsequent time-step. In order to reduce the size of the model, one must also take advantage of the locality of atomic environment [7-10].

Our idea for a deep representation learning framework for AIMD was inspired from existing open-source projects [1,5,6], where graphs and network-like representations have been applied to *dynamically model* physical systems outside the materials science domain. For example, the "interaction network" approach of Battaglia *et al* [1], explicitly separates the reasoning of "relations" from the reasoning of "objects". In the context of AIMD, "objects" refer to individual atoms while "relations" refers to the interaction between two or more different atoms. This separation allows the model to generalize between variable numbers of arbitrarily ordered objects and relations. As shown in Figure. 2, this workflow has shown great promise in reproducing the dynamical behavior of various physical systems. In our opinion, this project enables multiple opportunities and challenges that require extensive collaboration between data scientists and materials researchers.

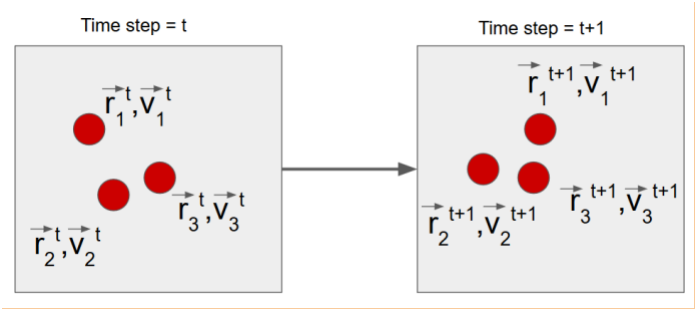

Figure 1. The objective of the model is to predict the positions (r) and velocities (v) of n-atoms at time-step 't+1' given the positions and velocities at time-step 't'

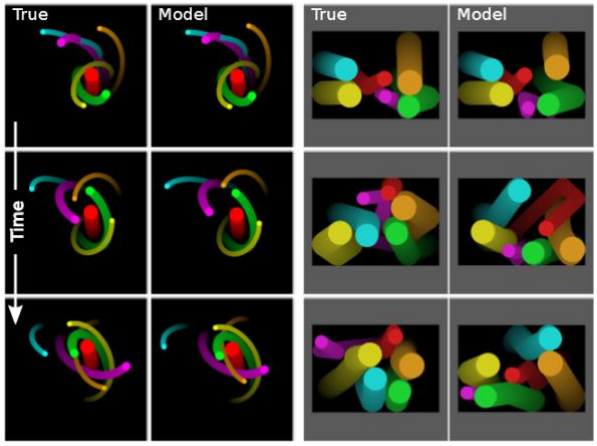

Figure 2. The interaction network has shown promise in predicting the dynamics of different physical systems such as a n-body gravitational system (Columns 1-2) and a n-bouncing-ball system (Columns 3-4). Figure taken from Reference [1]

References

[1] Battaglia, Peter, *et al*. "Interaction networks for learning about objects, relations and physics." *Advances in Neural Information Processing Systems*. 2016.

[2]Botu, Venkatesh, *et al*. "Machine learning force fields: construction, validation, and outlook." *The Journal of Physical Chemistry C* 121.1 (2016): 511-522.

[3]Behler, Jörg. "Atom-centered symmetry functions for constructing high-dimensional neural network potentials." *The Journal of chemical physics* 134.7 (2011): 074106.

[4]Bartók, Albert P., *et al*. "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons." *Physical review letters* 104.13 (2010): 136403.

[5] https://github.com/jaesik817/Interaction-networks_tensorflow

[6] https://github.com/higgsfield/interaction_network_pytorch

[7]Behler, Jörg. "Atom-centered symmetry functions for constructing high-dimensional neural network potentials." *The Journal of chemical physics* 134.7 (2011): 074106.

[8]De, Sandip, *et al*. "Comparing molecules and solids across structural and alchemical space." *Physical Chemistry Chemical Physics* 18.20 (2016): 13754-13769.

[9]Zhang, Linfeng *et al*. “Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics.” https://arxiv.org/pdf/1707.09571.pdf.

[10] Chmiela, Stefan, *et al*. “Machine learning of accurate energy-conserving molecular force fields.” *Science Advances* 3.5 (2017): e1603015. (DOI: 10.1126/sciadv.1603015).