TRANSFER-LEARNING-BASED COARSE-GRAINING METHOD FOR SIMPLE FLUIDS: TOWARD DEEP INVERSE LIQUID-STATE THEORY

Allocation: Illinois/377.8 Knh
PI: Narayana R. Aluru[1]

[1]University of Illinois at Urbana–Champaign

EXECUTIVE SUMMARY

Machine learning is an intriguing method to circumvent difficulties faced with the development and optimization of force field parameters. In this project, a deep neural network (DNN) was used to solve the inverse problem of the liquid-state theory—particularly, to find the relationship between the radial distribution function (RDF) and the Lennard–Jones (LJ) parameters at various thermodynamic states. Using molecular dynamics (MD), many observables, including RDF, are uniquely determined by specifying the interatomic potentials. However, the inverse problem (*e.g.*, determining the potential using a specific RDF) is not feasible through MD simulations unless it is combined with a complementary method.

In this project, the PI developed a framework integrating DNN with 1.5-terabyte MD trajectories (26,000 distinct systems and a total simulation time of 52 microseconds) to predict LJ parameters. The results show that DNN was successful not only in parameterization of the atomic LJ liquids but also in parameterizing the potentials for coarse-grained models of multiatom molecules.

RESEARCH CHALLENGE

The Lennard–Jones (LJ) potential form is one of the widely used interatomic potentials to study atomistic systems [1]. By specifying the potential parameters and the thermodynamic state, MD can compute various quantities of interest such as the radial distribution function (RDF). However, given a specific RDF, MD cannot directly predict the potential parameters; *i.e.*, the inverse problem is considered a difficult task [2]. As per Henderson's theorem [3], the relationship between the pair potential and RDF is unique at a given thermodynamic state, implying that the potential parameters can be determined uniquely based on the RDF. In this work, the PI explored the feasibility of force field development based on Henderson's theorem with a data-driven and deep learning-based approach.

This inverse problem can also be viewed as a coarse-graining problem where the objective is to develop a pair-potential between coarse-grained particles such that the RDF of the original system is reproduced. While different frameworks such as the fundamental measure theory and integral equations have been developed to address this problem, both accuracy and generalizability to more complex systems are still issues. Alternatively, reproducing a given RDF relies on MD to refine the potential parameters. For example, MD data are either integrated with a theoretical framework (*e.g.*, iterative Boltzmann inversion) or used to optimize statistical errors with given RDFs (*e.g.*, simplex and relative entropy). These approaches require thousands of simulations for a specific system and the data are often not reused. The main bottleneck in reusing data to parameterize a new system originates from the complexity of physics-based model development and long-term storage of the data. To overcome these issues, recent data-driven approaches, known as the physics of big data [4], use surrogate models.

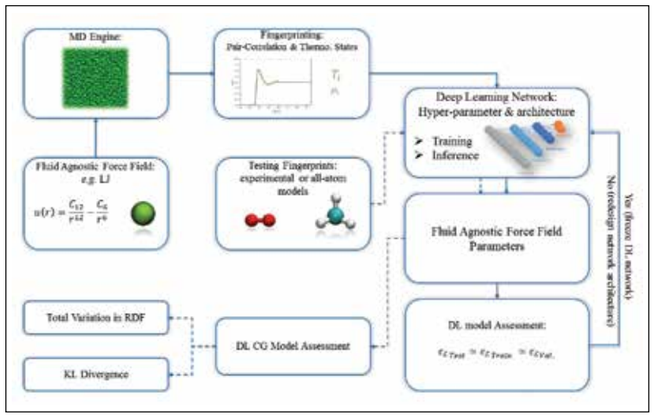

Figure 1: Workflow of the current study. An MD engine (GROMACS) generated big data that is used in the training of a deep learning model.

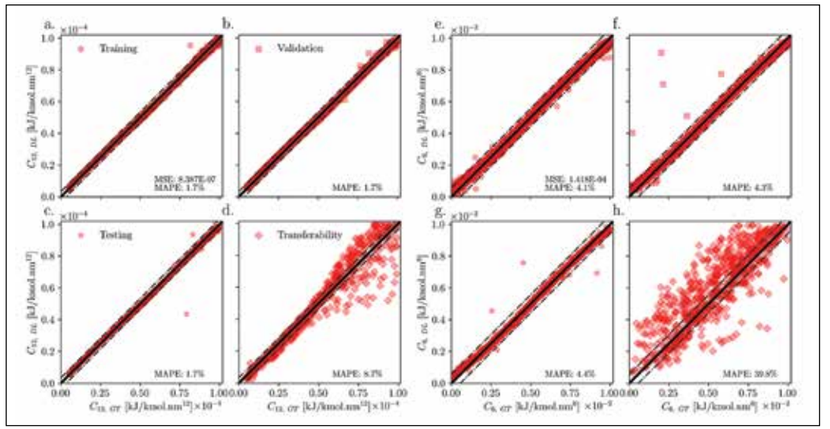

Figure 2: Predictive capability of the model.

METHODS & CODES

The PI performed MD simulations using the GROningen MAchine for Chemical Simulations package (GROMACS). GROMACS uses the well-known LJ potentials with the Verlet algorithm. The deep learning model development was performed using TensorFlow on Blue Waters. Fig. 1c shows the workflow of the data generation, training, inference, and assessment phases. molecules such as carbon monoxide, fluorine, and methane. The center of mass RDFs of these systems were first obtained using all-atom MD simulations. Then, the center of mass RDFs and thermodynamic states were fed into the DNN, as shown in Fig. 1. The results indicated that DNN force fields are indistinguishable from the other available methods. Considering that DNN is a single-shot method, its speed in deriving the CG force field is faster compared to the simplex and relative entropy methods.

RESULTS & IMPACT

The PI assessed the performance of DNN by considering two cases. First, he investigated the generalizability and transferability of the interatomic potential for LJ particles (development of atomistic force fields for single-atom particles). Second, he considered transfer learning for coarse-grained (CG) force-field development. Generalizability points to using DNN to predict potential parameters for LJ particles for a given thermodynamic data set. Transferability refers to using DNN to estimate potential parameters with thermodynamic states outside the training data set. Transfer learning refers to the use of DNN to estimate potential parameters for CG representation of simple multiatom molecules.

Fig. 2 compares the predicted DNN and the ground truth potential parameters used in the MD simulations. All the points are distributed almost uniformly around the one-to-one mapping line (the line on which the ground truth and DNN parameters match exactly). While the thermodynamic states and RDFs vary for each point, the DNN is able to relate them correctly to the underlying potential parameters. The DNN results for the LJ particles exhibit no more than 4.4% mean absolute percentile error over the data set, which implies that the DNN is an efficient approach to solve the inverse problem of the liquid-state theory for the LJ particles.

To investigate the transferability as a new coarse-graining route, the PI developed single-bead CG models of simple multiatom Following the procedure shown in Fig. 1, the PI assessed the accuracy of CG models with two additional metrics: total deviation in the RDFs and Kullback–Leibler (KL) divergence. Both error metrics show a small deviation from the DNN-predicted LJ parameters with a distance less than 0.1% of the maximum error. As DNN does not have prior knowledge about the information theory (KL metric) or the statistical mechanics metric (error in the total variation of the RDF), the PI concluded that deep learning is a good coarse-graining strategy as it performs well on both metrics.

WHY BLUE WATERS

This project involved 40,000 MD simulations of up to 10,000 atoms with a total simulation time of 52 microseconds. The feasibility of such computations is a result of the petascale resources and GROMACS' linear scaling on Blue Waters. Similarly, deep learning training using the TensorFlow module of Python relies heavily on access to the XK nodes on Blue Waters. Various neural network architectures were tried on Blue Waters (over 10 million training iterations) to select the optimized network and to avoid overfitting and underfitting.

PUBLICATIONS & DATA SETS

A. Moradzadeh and N. R. Aluru, "Transfer-learning-based coarse-graining method for simple fluids: toward deep inverse liquid-state theory," *J. Phys. Chem. Lett.*, vol. 10, no. 6, pp. 1242–1250, Feb. 2019.