Campuswire
Campuswire
https://campuswire.com/c/GC36E5829/feed
Class feed
MSE598ML: Machine Learnin
All categories
This week
Notifications
DMs
Calendar
Search
Class feed
Rooms
Files
Grades
Settings
Collapse
Wenwen
Active
Events
See all
Today May, 8th
No events today
Your day is all clear!

Jahid's question in Machine Learning for MatSE

Jahid Emon 1524 asked a question 3 months ago
Visible to: Everyone

Resolved

Pseudo Force Field in MD Simulation #6

General

In molecular dynamics simulations, we typically use integration to calculate forces from atomic positions. What if we use machine learning to predict forces directly from atomic positions, bypassing integration? Would this speed up the simulation?

2 0 236 46

Answers Comments

Answer this question

Instructor answers

Best

0 Andre Schleife ⭐ answered this question 3 months ago

You're absolutely on the right track. I don't think you would skip integration, but people absolutely use ML to predict forces based on atomic positions. The advantage is that ML gives you relatively accurate forces much faster than comparable quantum-mechanical methods (like DFT).

Reply

Student answers

Best

1 Sheikh Mohd Ta-Seen Afrid 306 answered this question 3 months ago

Nice question! I actually have some experience working on this particular topic, and it's fascinating to explore the potential of machine learning in molecular dynamics. While bypassing integration to predict forces directly from atomic positions can indeed speed up simulations, the success of this approach depends heavily on the quality of the training data, the model's ability to generalize across different systems, and the computational cost of evaluating the machine learning model itself. If these challenges are addressed, it could revolutionize the field by enabling faster exploration of complex molecular systems.

Reply

1 Wenwen LENG 59 answered this question 3 months ago

Clarification on "Integration" in MD Simulations

I am not entirely sure whether I fully understand your confusion regarding **integration** in MD. In MD simulations, **integration typically refers to updating atomic positions and velocities over time based on forces**, not computing forces themselves.

- **Force Calculation**: In classical MD, forces are derived from empirical potentials (e.g., Lennard-Jones), whereas in **Ab Initio Molecular Dynamics (AIMD)**, they are computed from **first-principles (DFT-based) electronic structure calculations**, which is computationally expensive (scaling as **$O(N^3–N^4)$** with system size).
- **Time Integration**: Once the forces are known, numerical integration methods (e.g., **Verlet, Velocity-Verlet**) are used to propagate the atomic motion.

Thus, if your question concerns **bypassing force computation**, **machine learning methods can replace this step**. However, **time integration cannot be skipped**, as it is necessary to evolve atomic trajectories.

What Happens If We Use Machine Learning to Predict Forces?

the current machine learning, like **Deep Potential Molecular Dynamics** or **Neuroevolution Potential **, allows us to **directly predict forces from atomic positions without computing them from explicit potential functions**. This means:

- **Bypass the expensive force calculation step**, such as solving Schrödinger's equation in AIMD.
- **NOT bypass integration**, as numerical integration (like Verlet or Velocity-Verlet) is still needed to evolve atomic trajectories over time.

(*I am unsure if your question about integration pertains to force computation or numerical time integration in MD.*)

**Would This Speed Up the Simulation? **

✅ **Yes, it can significantly speed up simulations**, especially compared to AIMD, which require solving quantum mechanical equations for each step. Because:

- **DFT-based AIMD**: Computational cost scales as **$O(N^3–N^4)$** with the number of atoms.
- **ML-based MD (e.g., DeePMD)**: Once trained, the force evaluation scales **linearly** with system size, **O(N)**, similar to classical force fields.

🚀 **If your goal is to accelerate MD simulations by replacing traditional force fields with ML-based force predictions, then ML can help drastically reduce computational cost while maintaining near-DFT accuracy.**

2 replies Reply

Jahid Emon 1524 3 months ago • (edited)

My mistake; now I can recall the underlying principle of MD simulation, ie. solving Newton's 2nd law, to determine the trajectory of each atom using numerical integration. I appreciate your time and effort to give such an elaborate response. However, recent studies suggest that Machine Learning Interatomic Force Fields (MLIFs) are often limited to specific problems and require substantial computational resources for training. Given this high training cost, do MLIFs still outperform traditional force fields in terms of overall efficiency and accuracy?

1 reply Reply

James Tallman 59 3 months ago

Someone in my lab actually does a lot of work on this (statt lab). In their work, they are trying to get "effective interactions" of differently shaped colloids (as typically MD only solves for particles, but often ellipsoids, superballs, and other shapes are of interest).

https://ui.adsabs.harvard.edu/abs/2024JChPh.160x4901A/abstract

Is this a specific problem that required a lot of resources for training? Yes. But, it is still a very general solution to a wide number of problems. Also, the speed-up is hardware dependent but can be between 2x and 25x speed-up (comparing a naive implementation of the MLIFs vs a highly optimized MD simulation). This implies the possibility of significant more speed-up, while the accuracy is there.

I'm not sure if this completely answers your question, but I thought I could highlight work going on in my lab that was relevant.

Reply

Wenwen LENG 59 3 months ago

In most materials science studies, the training process for MLIFs does not necessarily impose an high computational cost. The crucial factor is not just the training itself but rather the preparation of a high-quality and accurate dataset for training the MLIFs. The most challenging aspect lies in generating reliable raw data, typically from DFT calculations, to ensure the robustness of the trained model.

One of the key advantages and diffuults of MLIFs over traditional force fields is their potential for transferability to unforeseen scenarios—those that are either difficult or entirely inaccessible to DFT calculations. While ML models can accurately reproduce known data, their true challenge lies in extrapolating beyond previously seen configurations, making generalization a critical research focus.

For instance, if one aims to develop a reliable interatomic potential for a **multicomponent alloy** in **solid-state electrolytes (SSEs)**, a straightforward approach would be to include data for all possible alloy compositions in the training dataset. In theory, this would allow the machine learning model to learn a comprehensive representation of the interatomic interactions. However, the challenge lies in the fact that constructing such a dataset is **both computationally expensive and technically difficult**. If such an extensive dataset is available, then the problem is largely solved.

The reality, however, is that for **simple systems**, obtaining sufficient high-quality data through first-principles calculations is relatively easy. But for **complex multicomponent systems**, we often have **only a limited amount of data**, which is far from sufficient to develop a robust force field. Ideally, if machine learning models could **leverage information from elemental systems and binary alloys**, and then **extrapolate or transfer knowledge** to predict the behavior of ternary, quaternary, and even more complex systems, that would be the most desirable outcome.

While **fitting data is straightforward**, addressing **the challenge of sparse data in complex systems**—even with **active learning techniques**—remains highly **computationally expensive**. This **cost** is particularly reflected in the **iterative active learning process**, which requires continuous exploration and refinement. However, despite these challenges, **this approach is not just theoretically feasible—it has already been successfully demonstrated in practical applications**.

Reply

0 Enock Bunan 16 answered this question a month ago

I think the answer is yes, but for the accuracy of the results, the ML must be trained on high-quality QM or experimental data.

Reply