5 March 2025 · CC-BY 4.0 Qeios

Peer Review

Review of: "Machine Learning of Slow Collective Variables and Enhanced Sampling via Spatial Techniques"

Hiroshi Fujisaki[1]

1. Nippon Medical School, Japan

Review of "Machine Learning of Slow Collective Variables and Enhanced Sampling via Spatial Techniques"

Summary

The paper "Machine Learning of Slow Collective Variables and Enhanced Sampling via Spatial Techniques" by Tuğçe Gökdemir and Jakub Rydzewski provides an insightful and timely review of spatial learning techniques for identifying slow collective variables (CVs) in molecular dynamics (MD) simulations. The authors focus on methods that extract CVs using spatial characteristics rather than temporal trajectory-based approaches, addressing a fundamental challenge in enhanced sampling and free-energy estimation.

The key contributions of the paper include:

- A review of unsupervised machine learning methods for determining slow CVs, with an emphasis on spectral methods and deep learning techniques.
- A discussion of diffusion maps and anisotropic kernels as foundational tools for CV discovery.
- A detailed examination of **Spectral Map**, a novel technique designed to maximize timescale separation in the learned CV space.
- A comparison between **Reweighted Stochastic Embedding (RSE)** and Spectral Map, highlighting their methodological distinctions and respective advantages.
- The role of enhanced sampling methods in constructing high-quality training datasets for learning CVs.

 doi.org/10.32388/NA3WCX

Points for Improvement

While the paper offers a thorough and technically detailed discussion, several areas could be improved or expanded:

1. Clearer Comparison Between Spectral Map and SGOOP

Spectral Map is presented as an improvement over other spectral-based techniques for learning CVs, but the differences from **Spectral Gap Optimization of Order Parameters (SGOOP)** by Tiwary et al. (2016) should be more explicitly articulated. SGOOP also maximizes the spectral gap, but it does so within the maximum caliber framework, incorporating time-dependent information. Spectral Map, on the other hand, operates entirely within the spatial domain. A more in-depth discussion on:

- Whether Spectral Map can capture similar kinetic properties as SGOOP without requiring temporal information.
- The extent to which Spectral Map maintains dynamical self-consistency compared to SGOOP.
- Scenarios where one method outperforms the other.

would be beneficial for clarity.

2. Distinguishing Spectral Map from Reweighted Stochastic Embedding (RSE)

The paper presents both **Spectral Map** and **Reweighted Stochastic Embedding (RSE)** as techniques for learning slow CVs, but their relative advantages and limitations remain somewhat ambiguous. Specifically, it would be useful to address:

- **Are these two methods complementary, or are they competing approaches?**
- **Does one method provide better interpretability, while the other offers superior accuracy?**
- **Are they suited for different types of systems (e.g., biomolecular folding vs. chemical reactions)?**
- **Can they be combined in a hybrid approach to leverage the benefits of both?**

Given the importance of unbiased transition probabilities in enhanced sampling, a clearer discussion on which method is more robust to sampling biases and which provides better predictive power for long-timescale dynamics would enhance the impact of the review.

3. The Relationship Between Spatial Techniques and Temporal Methods (tICA and Diffusion Maps)

While the paper focuses on **spatial** techniques, it does not sufficiently address the well-established numerical relationships between spatial and temporal methods. Notably, the connection between **time-structure independent component analysis (tICA) and diffusion maps** has been quantitatively explored (J. Chem. Phys. 2018, 149(13):134112, doi: 10.1063/1.5049420).

The review should acknowledge:

- **Theoretical and empirical links between tICA and diffusion maps.** Diffusion maps are mathematically related to tICA, as they both approximate the eigenfunctions of the transfer operator but use different formulations.
- **How the omission of temporal information in purely spatial methods may affect the quality of slow CVs.**
- **Whether spatial learning techniques can fully replace temporal methods or whether a hybrid approach is more effective.**

By incorporating these points, the review would provide a more balanced perspective on the strengths and limitations of spatial learning techniques relative to existing temporal approaches.

Conclusion and Suggested Improvements

The paper presents a comprehensive and up-to-date review of spatial learning techniques for identifying slow CVs, with particular emphasis on Spectral Map and RSE. However, to strengthen its impact, the authors should:

1. Provide a more explicit comparison between Spectral Map and **SGOOP**, detailing the advantages and disadvantages of using spectral gap maximization without temporal data.
2. Clearly articulate the differences between **Spectral Map and RSE**, specifying their ideal use cases and whether one method is superior in certain scenarios.
3. Discuss the **numerical connection between diffusion maps and tICA**, and clarify whether spatial techniques can entirely replace temporal methods for learning CVs.

Addressing these points would enhance the clarity of the review and provide the reader with a more nuanced understanding of how spatial techniques fit into the broader landscape of machine learning for enhanced sampling in molecular dynamics.

Declarations

Potential competing interests: No potential competing interests to declare.

doi.org/10.32388/NA3WCX