Qeios

Peer Review

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

Eric Beyerle[1]

1. Department of Biology, University of Copenhagen, Denmark

In this short review (which is now available in published form in *Chem. Phys. Rev.* (link), the authors summarize recent (and not-so-recent) advances in spatial learning of slow collective variables (CVs) from computationally generated data in molecular systems.

I think the review covers the high points in the field and makes the key point that, in a specific Markovian limit, the slow CVs learned using these types of spatial learning methods allow for an approximation of the true slow processes *without requiring any temporal information*. To my knowledge, this connection is made explicit in the commute maps paper from Noé, Banisch, and Clementi; this reference is cited in the review but probably deserves more coverage given that what most researchers in the field are interested in is the slow dynamics of the molecular systems under study; otherwise, we would do a principal component analysis in this space, then model with a Markov state model, and call things done, but few people perform that process nowadays as a final analysis. Without thinking about it more, this type of approach should get around the timescale problem (since spatial distance is equal to kinetic distance, we should be able to perform much faster Langevin simulations in these reduced spaces without running any or much less *a priori* molecular dynamics).

The authors also discuss in some detail techniques for learning slow CVs using machine learning approaches. However, for these spatial-based methods, it seems like it might be useful to somehow machine-learn one or more of the 1) ϵ scale parameter, 2) the functional form of the kernel (e.g., use the network to extend beyond the Gaussian approximation from the traditional diffusion maps method), or 3) the anisotropic diffusion constant α in the 'traditional' diffusion map approaches. That is, instead of learning the slow CV space using machine learning, can the slow CVs be fixed using, e.g., time-lagged

 doi.org/10.32388/VO4N2F

independent component analysis, Markov state models, VAMPnets, or a related method, and the parameters of the diffusion (or Mahalanobis) map be learned using a neural network?

Some other minor points:

- I think it would be worthwhile to change the notation from ‘slow CVs’ to reaction coordinates…there are cases where the slow degrees of freedom are, in fact, not collective variables (e.g., interatomic distance).
- I think there are some methods out there for generating latent spaces obeying Langevin dynamics directly from molecular dynamics data, e.g., this one. Of course, this type of method is relying on temporal information, but it also obeys the objective of learning a low-dimensional reaction coordinate space where the dynamics are Markov, which is also an assumption required to relate the slow CVs learned from the spatial learning methods to temporal dynamics and time correlation functions.
- The authors might want to discuss whether or not there should, in general, be a Jacobian term in eq. 4 to account for nonlinearities in the mapping from the configurational space to CV space.
- It seems like the authors might benefit from citing some of the Markov state model clustering literature (e.g., Perron cluster-cluster) when discussing spectral clustering and the interpretation of the eigenfunctions of the transition matrix as defining the metastable states.

Declarations

Potential competing interests: No potential competing interests to declare.