EXPLOITING MACHINE LEARNING IN ENHANCED SAMPLING CALCULATIONS

GARETH TRIBELLO

Many chemical processes take place over timescales that are considerably longer than the timescales that can be accessed during conventional molecular dynamics simulations. Consequentially, to examine these sorts of problems the simulator must make use of enhanced sampling techniques such as metadynamics [1]. In using these methods one must typically first select a small number of collective variables (cvs). A bias is then added on these variables so that the system explores configurations with a wide range of cv values. Typically, one would like to select cvs that describe the process of interest. However, whereas this is straightforward for simple processes such as ion pairing in solution, for more complex systems the selection of appropriate collective variables represents the most difficult portion of any investigation.

An obvious solution to the problem of selecting collective variables is to simply increase the number of biased degrees of freedom. However, this is difficult in part because the cost of the calculation increases exponentially with the number of collective variables but also because of the difficulty associated with understanding the resulting high dimensionality free energy surfaces. Furthermore, it may even be the wrong approach as there is growing evidence [2] that, even for complex reactions, the dimensionality of the accessible portion of phase space is quite low. Therefore, an attractive idea is to use ideas from the dimensionality reduction and unsupervised learning communities to find the directions that span the important regions of phase space directly from the simulation trajectory.

I will present our ideas in this area. In particular I will present a new algorithm we have developed, reconnaissance metadynamics [3], which works by fitting the trajectory in a high dimensional cv space using a Gaussian mixture model. Results obtained from this fitting are then used to inform the biassing strategy. I will show how this algorithm can be used to enhance phase space exploration in a number of systems. Furthermore, we will further show how dimensionality reduction algorithms can be used to analyze the resulting trajectories to obtain small numbers of collective coordinates, which can be used in subsequent metadynamics simulations to obtain quantitative, low-dimensional free energy data [4, 5].

References

[1] A. Laio and M. Parrinello, *Proc. Natl. Acad Sci U.S.A.*, 2002, **99**, 12562.
[2] S. Piana and A. Laio, *Physical Review Letters*, 2008, **101**(20), 208101.
[3] G. A. Tribello, M. Ceriotti, and M. Parrinello, *Proc. Natl. Acad Sci U.S.A.*, 2010, **107**, 17509.
[4] M. Ceriotti, G. A. Tribello, and M. Parrinello, *Proceedings of the National Academy of Sciences*, 2011, **108**(32), 13023–13028.
[5] G. A. Tribello, M. Ceriotti, and M. Parrinello, *Proceedings of the National Academy of Sciences*, 2012, **109**(14), 5196–5201.