Use of Autoencoders to speed up molecular dynamic simulations

Alessio Gagliardi

Technische Universität München

With the advent of nanotechnology and the perspective of atomic-precision engineering the holy grail of material scientists has become the ability to design, in silico, materials with target properties for functionality objectives. Recently, the improvement in deep learning and generative methods has open the possibility to boost the discovery of new materials [1]. This is possible by a direct structure-to-property prediction, by coupling with experimental characterization data or by improving numerical models making them more efficient, especially when they involve expensive dynamical simulations. When the dynamics of the interested phenomenon is in some sense collective, the biasing can be performed along the slowest modes of this collective dynamics, the modes termed as collective variables (CVs). There are several methods developed over the past decades in this direction in the framework of molecular dynamic simulations [2, 3]. Despite the availability of this variety of methods, the prerequisite for efficient exploration of the configurational space is the availability of good CVs [4]. To select a number of hand-picked CVs require deep chemical intuition about the system dynamics and becomes increasingly difficult for complex systems [5]. One the other hand, one can choose as many general coordinates of the system as necessary and consider the collective variables a function of them. The task then is to come up with the transformation between the general coordinates and the collective variables. Recently, the focus of research has been either development of powerful Bayesian framework methods [6] or using a variation of an auto-associative artificial neural network or Autoencoder [7]. An Autoencoder is a nonlinear compression-decompression architecture which aims to embed input data to a relatively low dimensional space. This property is associated with a class of mathematical models called latent space models. Recent application of deep Autoencoder networks to different types of data demonstrate their ability as dimensionality reduction tool [8]. Recently, we used this method and demonstrated its utility on a large molecular system to study binding modes for drug-target interactions [9].

References:

[1] F Mayr et al.; The Journal of Physical Chemistry Letters 13 (8), 1940-1951 (2022).
[2] Invernizzi, M. and M. Parrinello, Rethinking Metadynamics: From Bias Potentials to Probability Distributions. J. Phys. Chem. Lett., 2020. 11(7): p. 2731–2736.
[3] Hénin, J., et al., Enhanced sampling methods for molecular dynamics simulations. arXiv preprint arXiv:2202.04164, 2022.
[4] Chen, M., Collective variable-based enhanced sampling and machine learning. Eur Phys J B, 2021. 94(10): p. 211.
[5] Bhakat, S., Collective variable discovery in the age of machine learning: reality, hype and everything in between. RSC Adv, 2022. 12(38): p. 25010-25024.
[6] Wu, H. and F. Noé, Variational Approach for Learning Markov Processes from Time Series Data. Journal of Nonlinear Science, 2020. 30(1): p. 23-66.
[7] Bengio, Y., A. Courville, and P. Vincent, Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013. 35(8): p. 1798-1828.
[8] Rifai, S., et al., Contractive Auto-Encoders: Explicit Invariance During Feature Extraction. p. 8.

[9] Siddiqui, G.A., et al., Application of Machine Learning Algorithms to Metadynamics for the Elucidation of the Binding Modes and Free Energy Landscape of Drug/Target Interactions: a Case Study. Chemistry, 2023. 29(62): p. e202302375.