Dataset of energies and forces on organic molecules for training HIPNN+SEQM

Ben Nebgen, Guoqing Zhou

LA-UR-22-25419

This is a dataset of quantum mechanical calculations performed on atomic configurations of organic molecules through active learning. The primary purpose of this dataset is as a supplemental file to the publication "Deep learning of dynamically responsive chemical Hamiltonians with semiempirical quantum mechanics" submitted to the journal *Proceedings of the National Academy of Sciences*. The data was generated with the Gaussian software package, for which LANL is a licensed user. This dataset can be used to train the dynamically reparametrized Hamiltonian model constructed from the interface of HIPPYNN (https://github.com/lanl/hippynn) and PYSEQM (https://github.com/lanl/PYSEQM/). The numpy files can easily be read by the data loader functionality contained with HIPPYNN. This dataset will be useful for fundamental science applications such as the creation of a tin interatomic potential.

data-AL-atom5f18-GZ-1-EtEi.npy: Atomization energies for each configuration in eV.

data-AL-atom5f18-GZ-1-Gradient_ev.npy: Atomic gradient information for each configuration in eV/Å.

data-AL-atom5f18-GZ-1-R.npy: Atomic location information in Å.

data-AL-atom5f18-GZ-1-Z_long.npy: Atomic species labels.

Element 0:

Z_long:

8, 8, 7, 6, 6, 6, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0

R:

1.5308	-1.1421	0.1542
1.3173	1.0929	-0.3993
-1.1032	1.2409	0.1729
-0.5428	-0.0748	0.4187
0.8747	-0.1420	0.1464
-1.4180	-1.1342	-0.3530
-1.6811	1.5448	0.8967
-1.6825	1.1370	-0.6968
-0.6347	-0.3747	1.4671
-1.2536	-2.1657	-0.0411
-1.3208	-1.0793	-1.4452
-2.4721	-0.8954	-0.3040
0.4995	1.6467	-0.3982

EtEi: -68.1530003

Gradient_ev:

-0.4644	0.1019	-0.5708
-0.0854	2.3377	-1.0075
-0.1329	0.4251	2.7864
3.6191	1.6074	-0.9030
-2.2216	-2.5652	3.4072
-0.6657	-0.1973	0.5749
1.2134	-0.6483	-1.4527
-0.5761	-0.6189	-1.2372
0.2164	-0.3202	-0.2692
-0.1280	-0.0196	-0.0396
-0.9267	0.0323	0.0324
0.4998	-0.0954	-1.0354
-0.3479	-0.0395	-0.2855

Element 1000:

Z_long:

8, 6, 6, 6, 6, 6, 6, 6, 6, 6, 1, 1, 1, 1, 1, 1, 1, 1

R:

-0.8672	-1.5210	-0.1976
-1.2969	1.3588	0.1717
1.2604	1.4790	0.0039
0.1031	-0.7450	-0.1483
2.4942	-0.6715	0.1033

2.4239	0.8106	0.0000
-2.8603	-0.3572	0.3426
-2.2936	0.4994	-0.4521
-0.0083	0.7887	-0.0615
1.4388	-1.4017	0.0601
1.5491	-2.5024	0.4552
3.4998	-1.1115	0.2212
3.3342	1.4256	-0.1553
1.1073	2.5329	-0.2072
-3.4045	-1.2760	0.2564
-3.0393	-0.0712	1.4623
-2.3288	0.4103	-1.4642
-1.3476	2.1946	0.8945

EtEi: -105.906457

Gradient_ev:

-1.5857	-0.1794	0.4589
-3.6604	1.4031	0.6613
1.0699	0.1286	0.3915
1.2113	1.4386	-0.3862
-4.0529	-3.4188	0.0994
-1.0694	-0.3396	0.4316
3.1421	-0.1618	-4.9332
0.6515	-2.0028	-2.3691
1.5614	0.6346	-0.5078
2.8954	2.8645	-1.6010
0.9200	-2.0480	1.5548
0.4723	-0.2229	-0.0947
0.3388	0.7509	-0.3011
-0.8621	-0.3188	-0.4037
0.1786	0.4322	1.4102
-1.2464	0.3424	1.7299
-0.5164	0.5534	3.0772
0.5522	0.1438	0.7819