Electronic Supplementary Information

Delfos: Deep Learning Model for Prediction of Solvation Free Energies in Generic Organic Solvents

Hyuntae Lim and YounJoon Jung

July 29, 2019

Benchmark for the FreeSolv Database

The present ESI provides a benchmark for the FreeSolv database, which contains hydration free energies for 642 small molecules[1, 2]. Delfos results (green box in Fig. 1) are calculated using the BiLSTM encoder model with the attention mechanism, while results of other ML models (blue boxes in Fig. 1) are taken from the MoleculeNet[4]. We use the default 80:10:10 dataset split method that is implemented in MoleculeNet submodule of DeepChem library[3]. Calculated RMSEs of our model were 1.16 ± 0.03 kcal/mol for the validation set and 1.19 ± 0.08 kcal/mol for the test set.

References

[1] David L. Mobley and J. Peter Guthrie. "FreeSolv: a database of experimental and calculated hydration free energies, with input files". In: *Journal of Computer-Aided Molecular Design* 28.7 (July 2014), pp. 711–720. ISSN: 1573-4951. DOI: 10.1007/s10822-014-9747-x. URL: https://doi.org/10.1007/s10822-014-9747-x.

[2] David L. Mobley et al. *MobleyLab/FreeSolv: Version 0.52*. Jan. 2018. DOI: 10.5281/zenodo.1161245. URL: https://doi.org/10.5281/zenodo.1161245.

[3] Bharath Ramsundar et al. *Deep Learning for the Life Sciences*. https://www.amazon.com/Deep-Learning-Life-Sciences-Microscopy/dp/1492039837. O'Reilly Media, 2019.

[4] Zhenqin Wu et al. "MoleculeNet: A benchmark for molecular machine learning". In: *Chemical Science* 9.2 (2018), pp. 513–530. ISSN: 20416539. DOI: 10.1039/c7sc02664a.

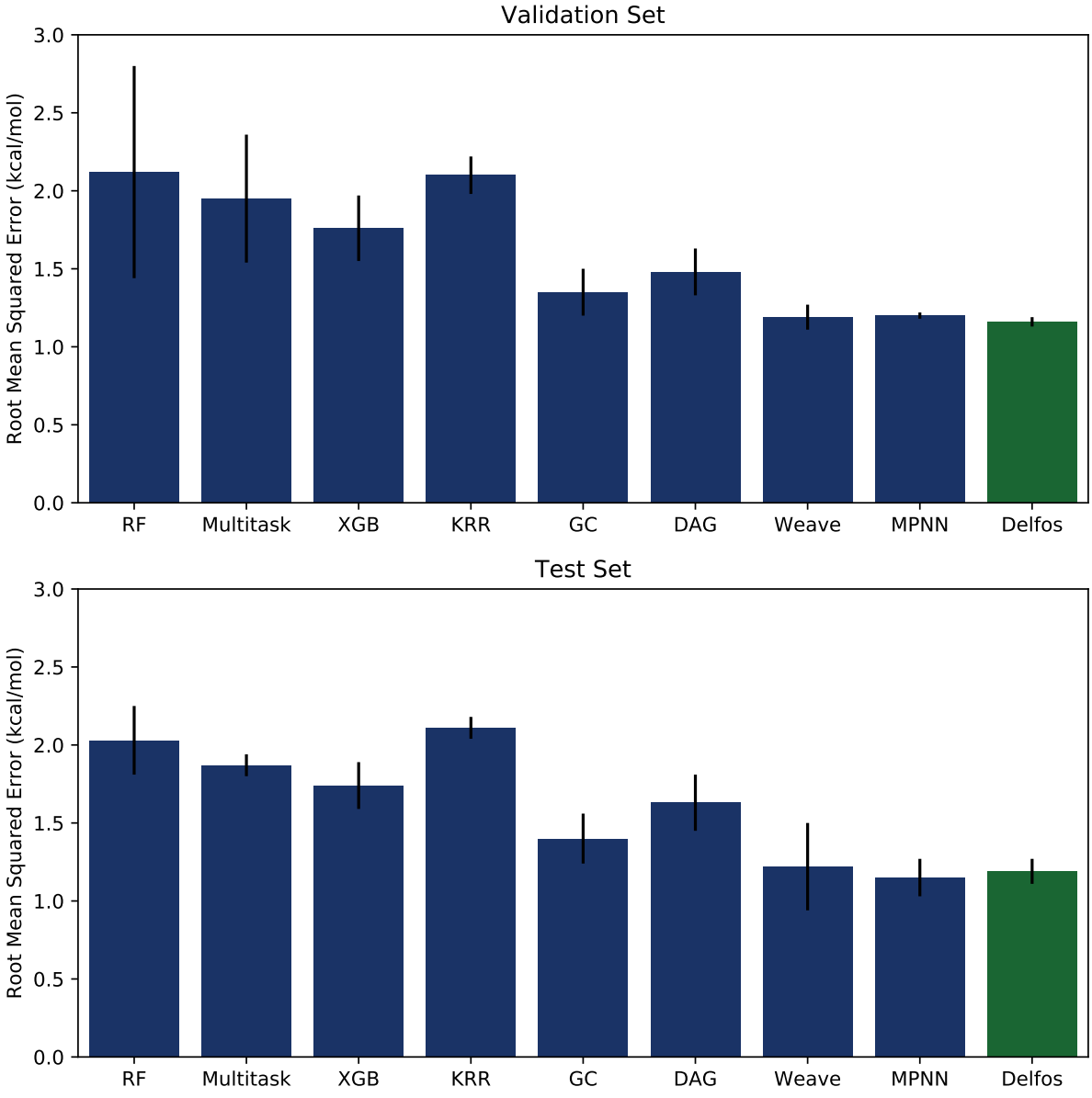

Figure 1: Benchmark data for FreeSolv dataset.