Molecular Property Prediction using Graph Neural Networks

Beatriz Braga, João Miranda, Mafalda Pires and Riccardo Conte

Motivation

The accurate prediction of chemical properties is critical to advancements in fields like drug discovery, materials science and chemical engineering. Traditional computational methods, while precise, often involve complex calculations that can be prohibitively slow, especially for large molecules or extensive datasets. One promising alternative is Graph Neural Networks (GNNs).

Molecules can be represented as graphs of atoms and bonds. These structures can be effectively processed by GNNs, which preserve molecular properties and structures. However, given that standard message-passing formulations do not consider rotationally equivariant representations, in [1] a different approach was proposed: Polarizable Atom Interaction Neural Network (PaiNN). This approach captures geometric information and physical interactions, as it takes rotational equivariance of molecules into account.

Introduction

We considered the **QM9** dataset, which includes geometric, energetic, electronic and thermodynamic properties of 134k normal size stable organic molecules made up of CHONF [2], [3], [4]. The main goal of this project is to implement PaiNN and apply it to predict the **Internal Energy at 0K** of different molecules based on nuclear charges ($Z_i \in \mathbb{N}$) and atom positions ($r_i \in \mathbb{R}^3$).

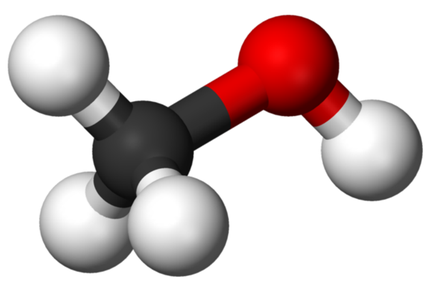

Figure 1: Methanol Molecule

Pre-processing

Preprocessing of the QM9 dataset is done inside PaiNN. This step takes as input the graph, with atoms' spatial information, and returns the Neighborhood of each atom i $(\mathcal{N}_i)$, consisting of the atoms that are close enough to influence each other. The cutoff hyperparameter defines the maximum distance such that two atoms can be considered neighbors. Additionally, preprocessing calculates the relative distances between atoms (r_{ij}). The new molecular data can be interpreted and used by PaiNN in the architecture shown below.

PaiNN model

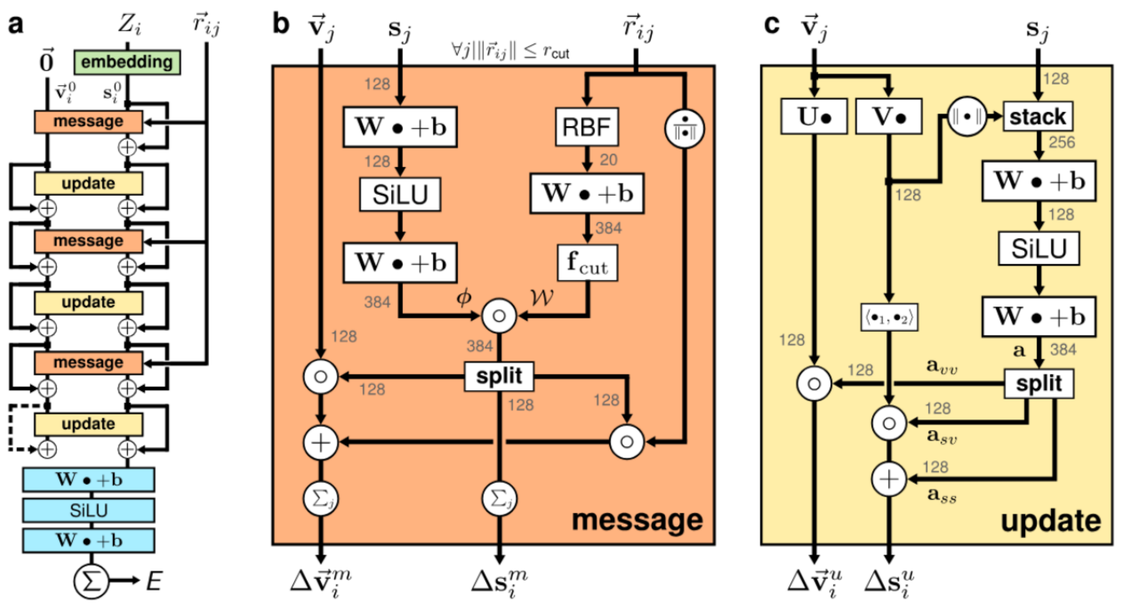

Figure 2: PaiNN model architecture representation.

Model Architecture

- Inputs: Atomic types (Z_i), $\mathcal{N}_i$, and 3D coordinates of each atom i (r_i)
- Outputs: Atomic contributions for each target property
- High-Level Architecture (a): The model uses a residual structure of interchanging message and update blocks
- Message Layer (b):
  - Aggregates information on the neighborhood of i $(\mathcal{N}_i)$ to capture local molecular structure.
  - Takes as inputs the scalars s_j (invariant atom representations), vectors v_j (equivariant atom representations), and relative positions r_{ij} for all $j \in \mathcal{N}_i$.
- Update Layer (c):
  - Combines contributions from scalar features of neighbors and interactions with vector features.
  - Refines vector features by applying transformations that preserve geometric directionality.
  - Balances scalar invariance (important for chemical features) with vector equivariance (essential for geometry-dependent tasks)

Postprocessing

The post-processing step regarding the target scalar feature takes atomic-level predictions from PaiNN, normalizes them using dataset-specific mean and standard deviation, and adds baseline contributions for each atom type. It then sums these processed atomic contributions to compute the final graph-level molecular property, the Internal Energy at 0K, for each molecule in the dataset.

Extensions

We implemented two different extensions of the model to improve its performances.

1. **Layer Optimization:** We investigate the impact of varying the number of message-passing layers in PaiNN: 1, 3, 5, and 10 layers.
2. **Bayesian Modelling:** Implement SWA [5] and SWAG [6] to PaiNN.
   - ▷ **SWA (Stochastic Weight Averaging):** SWA is a technique used in training deep neural networks. It involves averaging the weights of a model over several iterations towards the end of training. This can lead to better generalization and more robust models by stabilizing the final weight values.
   - ▷ **SWAG (Stochastic Weight Averaging-Gaussian):** SWAG extends SWA by also modeling the uncertainty in weights as a Gaussian distribution. It approximates the posterior distribution of model weights, which can be useful for uncertainty estimation and Bayesian deep learning.

Model performance

- Increasing the number of layers improves the resulting MAE, but with diminishing returns.
- SWA and SWAG extensions improved the original model, even if by little
- The best model is PaiNN10, however, the training is twice as computationally expensive time-wise than PaiNN3

Table 1: MAE achieved by different PaiNN architectures in the test set.

Method	MAE (meV)
PaiNN3 (original model)	7.9850
PaiNN1	13.2553
PaiNN5	7.5428
PaiNN10	6.9043
PaiNN3-SWA (lr=5e-4)	7.8244
PaiNN3-SWAG (lr=1e-5)	7.8393

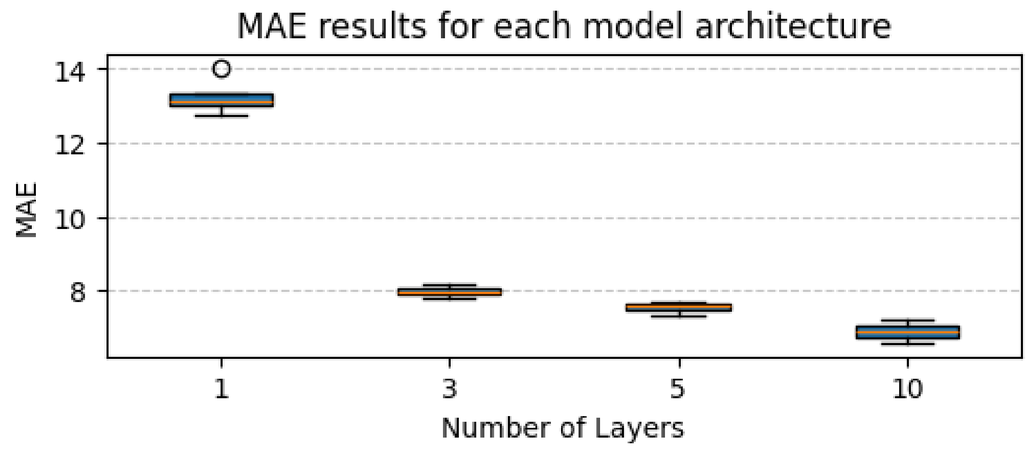

Figure 3: MAE results relative to the model Architecture

Conclusions

Performances: the PaiNN model successfully predicted molecular Internal Energy at 0K with low MAE. Increasing the number of layers improved performance, with the best results achieved using PaiNN10. The Stochastic Weight Averaging (SWA) and SWAG extensions further enhanced the original PaiNN3 model performances, by 2% and 1.8% respectively. Compared to the original paper, our implementation achieved comparable results, with a reported MAE of 5.85 meV.
Oversmoothing: this common issue in deeper Graph Neural Networks, was not observed in our implementations. The shortcut connections in PaiNN mitigated this problem, enabling the use of up to 10 layers without performance degradation.
Future work: exploring alternative implementations of PaiNN to further improve performances and generalizing the model to predict multiple molecular properties.

References

[1] Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pages 9377–9388. PMLR, 2021.

[2] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

[3] Lars Ruddigkeit, Ruud Van Deursen, Lorenz C Blum, and Jean-Louis Reymond. Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17. *Journal of chemical information and modeling*, 52(11):2864–2875, 2012.

[4] Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay S. Pande. Moleculenet: A benchmark for molecular machine learning. *CoRR*, abs/1703.00564, 2017.

[5] Pavel Izmailov, Dmitrii Podoprikhin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. Averaging weights leads to wider optima in deep learning. *arXiv preprint arXiv:1803.05407*, 1803:05407, 2018.

[6] Timur D Garipov, Pavel Izmailov, Dmitrii Podoprikhin, Dmitry P Vetrov, and Andrew Gordon Wilson. Loss surfaces, mode connectivity, and fast ensembling of dnns. *Advances in Neural Information Processing Systems*, 31, 2018.