Benchmarking KANG on Molecular Graph Learning Tasks

Sahar Khanlari, Marco Natale

1 Introduction

We evaluate the performance of the KANG model on various molecular property prediction tasks, both regression and classification, using datasets from MoleculeNet. The implementation relies on PyTorch Geometric and includes custom dataset processing pipelines for HIV, ToxCast, QM8, and QM9. We perform Optuna-based hyperparameter tuning for each task. The model uses only the 2D molecular structure and does not incorporate 3D conformations.

2 Graph Classification

2.1 HIV

The HIV dataset contains a binary label for HIV replication inhibition. As it contains only one classification task, we evaluate KANG using ROC-AUC. The best published model on this dataset is CIN++ [1].

2.2 ToxCast

ToxCast comprises 617 binary classification tasks. We select 5 representative tasks and treat them as single-task binary classification problems for evaluation. The selected tasks are:

- **TOX21_AhR_LUC_Agonist**: measures the activation of the Aryl hydrocarbon Receptor pathway.
- **TOX21_Aromatase_Inhibition**: assesses inhibition of the aromatase enzyme, important in hormone biosynthesis.
- **TOX21_AutoFluor_HEK293_Cell_blue**: a control task to detect autofluorescence in HEK293 cells under blue channel excitation.
- **TOX21_p53_BLA_p3_ch1**: measures activation of the tumor suppressor protein p53 using a reporter assay.
- **TOX21_p53_BLA_p4_ratio**: an additional reporter-based p53 activation task capturing response ratios.

The best reported model on this dataset is DumplingGNN [2].

3 Graph Regression

3.1 QM8

QM8 provides quantum mechanical properties and includes 12 regression tasks derived from TD-DFT and CC2 calculations. We evaluate the KANG model on all 12 tasks and report the average MAE across them. The state-of-the-art baseline for QM8 is the D-MPNN model [3].

3.2 QM9

QM9 provides quantum chemical properties for small organic molecules and includes 12 regression tasks covering electronic, thermodynamic, and vibrational properties. We evaluate the KANG model on all tasks and report the average MAE. D-MPNN is also the best known model on this dataset [3].

4 Experimental Setup

Experiments are run using custom PyTorch Geometric code for both classification and regression. All molecules are represented using SMILES strings and converted to graph structures using a custom preprocessing pipeline. Each molecule is transformed into a graph where atoms are nodes and bonds are edges.

The atomic features used in our preprocessing are:

- Atomic number (as integer)
- Atom degree (number of directly bonded atoms)
- Formal charge
- Aromaticity flag (boolean)
- Total number of hydrogens (including neighbors)

Bond features used are:

- One-hot encoding of bond type: single, double, triple, aromatic
- Conjugation flag (boolean)
- Ring membership flag (boolean)
- One-hot encoding of stereochemistry: none, Z, E, any

Hyperparameters are optimized using Optuna. Each experiment logs the best validation performance and stores the corresponding test result.

5 Comparison with State-of-the-Art

Table 1 compares KANG with SoTA models.

Dataset	Metric	SoTA Model	SoTA Value	KANG Value
QM8	MAE	D-MPNN	**0.0190**	0.0221
QM9	MAE	D-MPNN	**0.00814**	7.1407]
HIV	ROC-AUC	CIN++	**0.8063**	0.6997
ToxCast	ROC-AUC	DumplingGNN	0.782	0.7896[1]

Table 1: Comparison of KANG model performance against SoTA baselines.

6 Conclusions

The KANG model shows mixed performance across different tasks. It performs comparably to state-of-the-art models on the QM8 dataset, despite using only 2D input features. Notably, the ToxCast evaluation was conducted on only 5 of the 617 available tasks, providing a limited view of its overall potential on this dataset. However, KANG significantly underperforms on the full QM9 regression benchmark, where its average MAE is substantially higher than the SoTA. On the HIV dataset, it also falls short. These results suggest that while KANG can be effective for certain property prediction tasks it may require architectural improvements or additional features (e.g., 3D structure, attention mechanisms) to generalize across more complex or high-resolution regression problems.

[1] Average ROC-AUC over 5 tasks: $(0.8689 + 0.7194 + 0.9581 + 0.7019 + 0.6999)/5 = 0.7896$

References

[1] L. Giusti, T. Reu, F. Ceccarelli, C. Bodnar, and P. Liò, "Cin++: Enhancing topological message passing," *arXiv preprint arXiv:2306.03561*, 2023.

[2] S. Xu and L. Xie, "Dumplinggnn: Hybrid gnn enables better adc payload activity prediction based on chemical structure," *arXiv preprint arXiv:2410.05278*, 2024.

[3] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, *et al.*, "Analyzing learned molecular representations for property prediction," *arXiv preprint arXiv:1904.01561*, 2019.