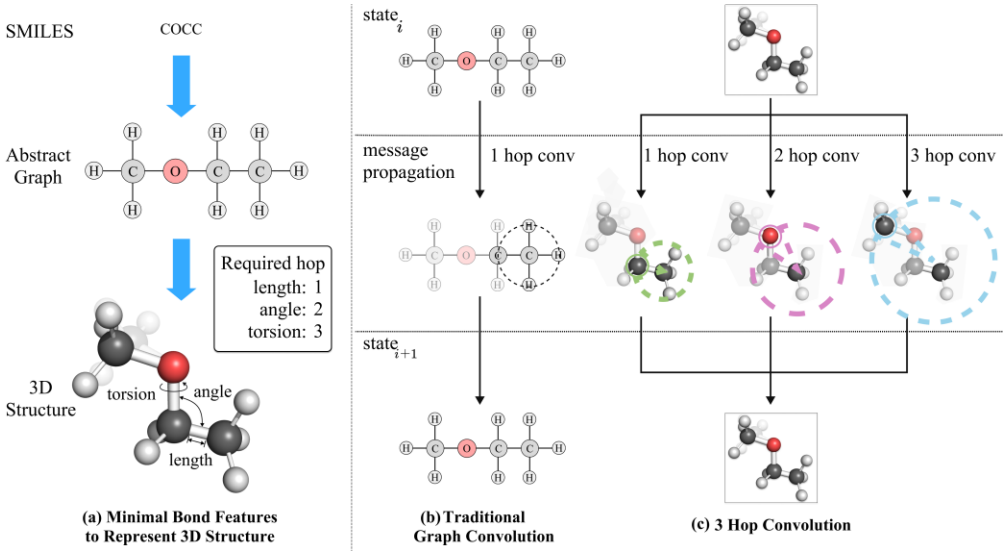
MetaGIN: A Lightweight Framework for Molecular Property Prediction

Xuan ZHANG, Cheng CHEN, Xiaoting WANG, Haitao JIANG, Wei ZHAO, Xuefeng CUI

Frontiers of Computer Science, DOI: 10.1007/s11704-024-3784-y

Problems & Ideas

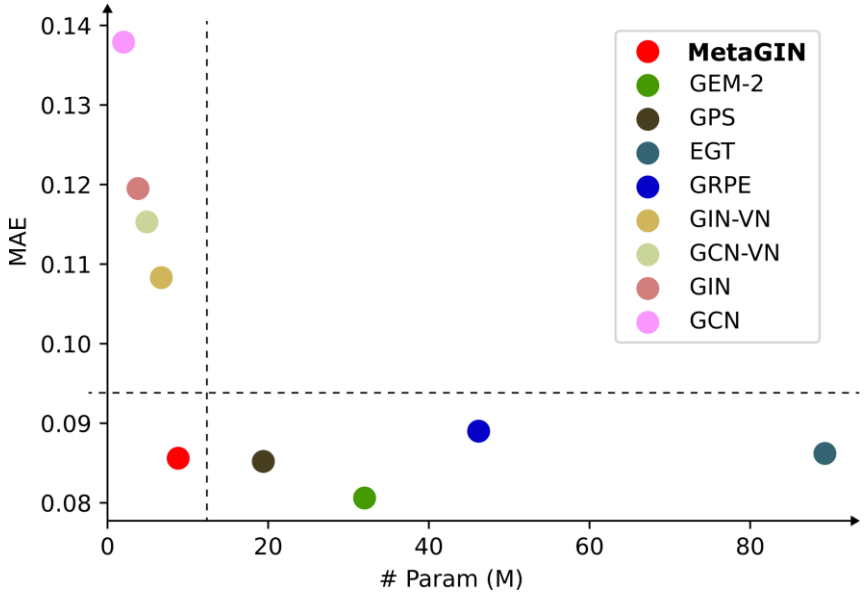
- Problems of conventional molecular property prediction approaches:
  - 1 hop convolution cannot describe the 3D conformation information.
  - Existing models face a trade-off between model size and performance.
- Ideas:
  - Use the "3-hop convolution" technique to capture spatial relationships.
  - Achieve an efficient model by using the MetaFormer-like architecture.



To accurately represent 3D structures, (a) it is necessary to incorporate at least 3-hop features, which include bond length, bond angle, and torsion angle. This means that (b) traditional Graph Convolutional Networks (GCNs) that only utilize 1-hop features fall short in representing 3D structures. (c) MetaGIN, on the other hand, efficiently uses 3-hop features, satisfying the bare minimum requirements for 3D structure representation.

Main Contributions

- Contributions:
  - Association of 3-Hop Convolution with 3D Molecular Structure : enables effective capture of complex molecular geometries without using explicit 3D structural data during training. And the superior 3D structure learning capability of 3-hop convolution are proved.
  - Balance of Performance and Efficiency: With fewer than 10 million parameters, MetaGIN achieves state-of-the-art performance across various molecular property prediction tasks, often surpassing more complex models.



	FreeSolv	ESOL	Lipophilicity
# Mols	642	1128	4200
ECFP [31]	5.275(0.751)	2.359(0.454)	1.188(0.061)
TF_Robust [32]	4.122(0.085)	1.722(0.038)	0.909(0.060)
GraphConv [15]	2.900(0.135)	1.068(0.050)	0.712(0.049)
Weave [33]	2.398(0.250)	1.158(0.055)	0.813(0.042)
SchNet [34]	3.215(0.755)	1.045(0.064)	0.909(0.098)
MGCN [35]	3.349(0.097)	1.266(0.147)	1.113(0.041)
AttentiveFP [36]	2.030(0.420)	0.853(0.060)	0.650(0.030)
TrimNet [37]	2.529(0.111)	1.282(0.029)	0.702(0.008)
MPNN [38]	2.185(0.952)	1.167(0.430)	0.885(0.030)
DMPNN [39]	2.177(0.914)	0.980(0.258)	0.653(0.046)
FunQG-MPNN [40]	1.542(0.460)	0.879(0.091)	0.638(0.020)
FunQG-DMPNN [40]	1.501(0.376)	0.818(0.047)	0.622(0.028)
MetaGIN	**1.397(0.062)**	**0.780(0.061)**	**0.532(0.013)**

Performance evaluation comparison. Left: the parameter and MAE relation on PCQM4Mv2 dataset; Right: regression tasks on MoleculeNet datasets.