nature portfolio

https://doi.org/10.1038/s42256-022-00501-8

Supplementary information

An adaptive graph learning method for automated molecular interactions and properties predictions

In the format provided by the authors and unedited

Supplementary Information for "An adaptive graph learning method for automated molecular interactions and properties predictions"

Yuquan Li [1,2,†], Chang-Yu Hsieh [2,†], Ruiqiang Lu [1], Xiaoqing Gong [1], Xiaorui Wang [3], Pengyong Li [4], Shuo Liu [5], Yanan Tian [5], Dejun Jiang [6], Jiaxian Yan [7], Qifeng Bai [8], Huanxiang Liu[5], Shengyu Zhang [2], & Xiaojun Yao [1,*]

[1] College of chemistry and chemical engineering, Lanzhou University, Lanzhou, China.

[2] Tencent Quantum Laboratory, Tencent, Shenzhen, China.

[3] State Key Laboratory of Quality Research in Chinese Medicines, Macau University of Science and Technology, Macau, China.

[4] School of computer science and technology, Xidian University, Xian, China.

[5] School of pharmacy, Lanzhou University, Lanzhou, China.

[6] College of Computer Science and Technology, Zhejiang University, Hangzhou, China.

[7] School of Data Science, University of Science and Technology of China, Hefei, China.

[8] School of Basic Medical Sciences, Lanzhou University, Lanzhou, China.

[†]These authors contributed equally to this work.

[*]Corresponding authors. E-mail: xjyao@lzu.edu.cn

Supplementary Table 1. Performance comparison on drug-drug interaction datasets

Dataset	DrugBank	
Metrics	**ROC-AUC**	**PR-AUC**
LR	0.774(±0.003)	0.745(±0.005)
Nat.Prot	0.786(±0.003)	0.753(±0.003)
Mol2Vec	0.849(±0.004)	0.828(±0.006)
MOLVAE	0.852(±0.006)	0.828(±0.009)
DeepDDI	0.844(±0.003)	0.828(±0.002)
CASTER	0.861(±0.005)	0.829(±0.003)
GLAM	**0.911(±0.017)**	**0.899(±0.019)**

Note: (a) In the drug-drug interaction task, scores of all baselines are taken from CASTER[1]. (b) the highlight score with bold-black style means this method got the highest score. (c) all scores of GLAM are run under the same default configuration setting. (d) The ensemble size used of GLAM in this table is 3.

Supplementary Table 2. Ablation study on time consumption

Dataset	Dataset Size	Method	Params	Time	Device	Detail
ESOL	Train: 904 Valid: 112 Test:112	GCN	16,225	0.9 min	1*V100	200 epochs
		GAT	63,297	1.1 mins	1*V100	200 epochs
		MPNN	27,681	1.2 mins	1*V100	200 epochs
		GLAM	-	34.3 mins	4*V100	100 low- and 3 high-fidelity training
BACE	Train: 1633 Valid: 203 Test:203	GCN	63,169	1.7 mins	1*V100	200 epochs
		GAT	63,297	2.1 mins	1*V100	200 epochs
		MPNN	108,609	2.5 mins	1*V100	200 epochs
		GLAM	-	41.1 mins	4*V100	100 low- and 3 high-fidelity training
KAT2A	Train:183089 Valid:78468 Test:87185	Glide (SP)	-	10.3 hrs	40 CPU	-
		Glide (XP)	-	120.3 hrs	40 CPU	-
		DGraphDTA	1,344,920	3.4 hrs	1*V100	200 epochs
		TransformerCPI	449,737	5.1 hrs	1*V100	200 epochs
		GLAM	-	48.5 hrs	4*V100	100 low- and 3 high-fidelity training

Note: (a) All the time consumptions are averaged on three independent runs. (b) On ESOL[2] and BACE [2] datasets, "100 low- and 3 high-fidelity training" means that GLAM performs 100 low-fidelity training process (30 epochs * 3 run) and then performs 3 high-fidelity training process (2000 epochs with early stopping * 3 run). (c) On KAT2A[3], "100 low- and 3 high-fidelity training" here means that GLAM performs 100 low-fidelity training process (20 epochs * 1 run) and then performs 3 high-fidelity training process (2000 epochs with early stopping * 2 run).

Supplementary Table 3. Ablation study on preferences of GLAM

Config Items	Options	Physical chemistry ESOL(1128)			Pharmacokinetics BBBP(2053)			Toxicity Tox21(8014)		
		Top-10	Top-20	Top-50	Top-10	Top-20	Top-50	Top-10	Top-20	Top-50
Pooling method(readout)	Set2Set	0.0%	0.0%	4.7%	16.7%	20.0%	29.3%	6.7%	18.3%	34.0%
	Global Pool	100.0%	100.0%	92.7%	70.0%	63.3%	52.0%	90.0%	76.7%	47.3%
	Attention Pool	0.0%	0.0%	2.7%	13.3%	16.7%	18.7%	3.3%	5.0%	18.7%
Optimizer	Adam	66.7%	70.0%	61.3%	46.7%	56.7%	61.3%	76.7%	68.3%	60.7%
	Ranger	33.3%	30.0%	38.7%	53.3%	43.3%	38.7%	23.3%	31.7%	39.3%
Message passing core	GCN	10.0%	11.7%	18.7%	10.0%	8.3%	19.3%	13.3%	20.0%	13.3%
	GAT	6.7%	23.3%	22.7%	6.7%	15.0%	16.0%	0.0%	1.7%	10.7%
	MPN	40.0%	28.3%	24.7%	36.7%	33.3%	26.0%	10.0%	15.0%	20.7%
	Tri-MPN	23.3%	20.0%	20.7%	16.7%	15.0%	14.0%	20.0%	13.3%	22.0%
	Light Tri-MPN	20.0%	16.7%	13.3%	30.0%	28.3%	24.7%	56.7%	50.0%	33.3%
Message passing step	1	23.3%	26.7%	26.7%	30.0%	33.3%	27.3%	13.3%	20.0%	24.7%
	2	26.7%	21.7%	22.0%	36.7%	33.3%	27.3%	3.3%	16.7%	27.3%
	3	30.0%	28.3%	29.3%	20.0%	15.0%	18.7%	46.7%	36.7%	27.3%
	6	20.0%	23.3%	22.0%	13.3%	18.3%	26.7%	36.7%	26.7%	20.7%
Learning rate	0.01	26.7%	35.0%	26.7%	16.7%	15.0%	16.0%	13.3%	18.3%	24.7%
	0.005	26.7%	28.3%	24.0%	36.7%	26.7%	24.0%	26.7%	20.0%	23.3%
	0.001	26.7%	18.3%	21.3%	20.0%	30.0%	25.3%	26.7%	35.0%	29.3%
	0.0005	20.0%	15.0%	15.3%	20.0%	20.0%	25.3%	33.3%	25.0%	17.3%
	0.0001	0.0%	3.3%	12.7%	6.7%	8.3%	9.3%	0.0%	1.7%	5.3%
Graph norm	None	66.7%	50.0%	45.3%	40.0%	41.7%	42.7%	60.0%	53.3%	49.3%
	Batch Norm	0.0%	5.0%	8.7%	10.0%	11.7%	9.3%	3.3%	6.7%	9.3%
	Layer Norm	20.0%	16.7%	19.3%	33.3%	23.3%	24.7%	10.0%	18.3%	20.7%
	Pair Norm	13.3%	28.3%	26.7%	16.7%	23.3%	23.3%	26.7%	21.7%	20.7%
End Dropout	None	23.3%	25.0%	21.3%	33.3%	25.0%	22.7%	30.0%	33.3%	30.0%
	0.1	26.7%	25.0%	30.0%	20.0%	21.7%	26.7%	13.3%	15.0%	26.7%
	0.2	26.7%	26.7%	25.3%	33.3%	31.7%	31.3%	20.0%	18.3%	21.3%
	0.5	23.3%	23.3%	23.3%	13.3%	21.7%	19.3%	36.7%	33.3%	22.0%

Note: (a) All the percents of preferences are averaged on three independent runs.

Supplementary Table 4. Ablation study on effect of ensemble size on performance

Method	ESOL	BBBP
GCN	1.032(0.091)	0.724(0.093)
3*GCN	0.954	0.766
GAT	1.012(0.061)	0.686(0.038)
3*GAT	0.942	0.696
GIN	0.732(0.073)	0.742(0.076)
3*GIN	0.690	0.814
MPNN	0.819(0.097)	0.887(0.029)
3*MPNN	0.720	0.927
GCN+GAT+GIN	0.841	0.843
GCN+GAT+MPNN	0.852	0.890
GCN+GIN+MPNN	0.686	0.907
GAT+GIN+MPNN	0.722	0.889
GLAM (best, n=1)	0.652(0.109)	0.901(0.031)
GLAM (n=2)	0.592(0.150)	0.918(0.016)
GLAM (n=3)	0.585(0.167)	0.931(0.021)
GLAM (n=5)	0.604(0.165)	0.935(0.022)
GLAM (n=7)	0.593(0.179)	0.934(0.021)

Note: (a) 3* GCN means the ensemble of three GCN predictors. (b) GCN+GAT+GIN means the ensemble of GCN, GAT and GIN predictors. (c) GLAM(n=3) means that the ensemble of top-3 predictors after low- and high-fidelity train process.

Supplementary Table 5. Performance comparison with Auto-sklearn and Auto-Gluon

Method	ESOL		BBBP		Device
	Performance(RMSE)	Time	Performance(AUC)	Time	
AutoGluon	0.630(0.004)	**9.4 mins**	0.722(0.005)	**8.0 mins**	20 CPU
Auto-sklearn	0.792(0.015)	60.1 mins	0.584(0.008)	60.1 mins	20 CPU
GLAM	0.592(0.036)	34.3 mins	0.932(0.015)	41.1 mins	4*V100

Note: (a) AutoGluon[4] and Auto-sklearn[5] use ECFP[6] fingerprints as input (calculated by rdkit [7]).

Supplementary references

1. Huang, K., Xiao, C., Hoang, T., Glass, L. & Sun, J. CASTER: Predicting Drug Interactions with Chemical Substructure Representation. *Proceedings of the AAAI Conference on Artificial Intelligence* **34**, 702–709 (2020).
2. Wu, Z. *et al.* MoleculeNet: A benchmark for molecular machine learning. *Chemical Science* **9**, 513–530 (2018).
3. Tran-Nguyen, V. K., Jacquemard, C. & Rognan, D. LIT-PCBA: An unbiased data set for machine learning and virtual screening. *Journal of Chemical Information and Modeling* **60**, 4263–4273 (2020).
4. Erickson, N. *et al.* AutoGluon-tabular: Robust and accurate AutoML for structured data. *arXiv* (2020).
5. Feurer, M., Eggensperger, K., Falkner, S., Lindauer, M. & Hutter, F. Auto-Sklearn 2.0: The Next Generation. *arXiv* (2020).
6. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling* **50**, (2010).
7. Landrum, G. RDKit : A software suite for cheminformatics , computational chemistry , and predictive modeling. https://www.rdkit.org/ (2011).