ModuLM:Enabling Modular and Multimodal Molecular Relational Learning with Large Language Models

数据智能实验室 Data Intelligence Lab

Zhuo Chen[1,2], Yizhen Zheng[4], Huan Yee Koh[4], Hongxin Xiang[3], Linjiang Chen[5], Wenjie Du[1,2,5] †, Yang Wang[1,2,5]†

1,University of Science and Technology of China, China
2,Suzhou Institute for Advanced Research, USTC, China
3,Hunan University, China 4,Monash University, Australia
5,State Key Laboratory of Precision and Intelligent Chemistry, USTC, China

Contact me: czchenzhuo@mail.ustc.edu.cn

Introduction

Lack of diverse input support:

- Molecules can be represented in 1D, 2D, and 3D forms, most MRL models support only a single input modality, limiting structural information utilization.

Lack of Flexible Architectures:

- Current LLM-based MRL models are limited by rigid encoders and weak interaction modeling, highlighting the need for a flexible, integrated framework.

The work proposed in this study

To address these challenges, reduce redundant coding, and ensure fair model comparison, we propose **ModuLM**, a framework designed to support flexible LLM-based model construction and diverse molecular representations. ModuLM provides a rich suite of modular components, including 8 types of 2D molecular graph encoders, 11 types of 3D molecular conformation encoders, 7 types of interaction layers, and 7 mainstream LLM backbones. Owing to its highly flexible model assembly mechanism, ModuLM enables the dynamic construction of over **50,000** distinct modelconfigurations. In addition, we provide comprehensive results to demonstrate the effectiveness of ModuLM in supporting LLM-based MRL tasks.

Methond

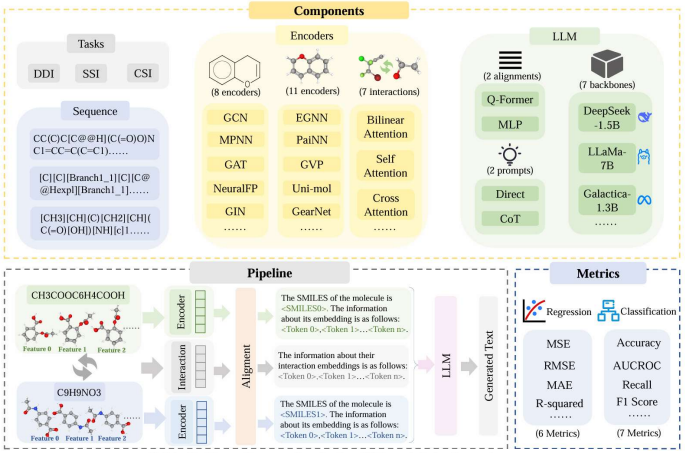


Table 2: Encoding methods for different data formats

Encoder Type	Methods
2D Graph	GCN [32], MPNN [20], GAT [66], NeuralFP [11], AttentiveFP [71], GIN [72], GraphSAGE[23], CoATGIN[78]
3D Conformation	EGNN [56], 3D-GeoFormer [84], SE3Transformer [17], PaiNN [57], GVP [28], GearNet [79], DimeNet++ [19], SchNet [58], SphereNet [42], G-SphereNet[44], Uni-mol [83]
Interaction	Bilinear Attention[1], Self Attention[65], Cross Attention[52], Highway[85], Gated Fusion[45], Bilinear Fusion[39], Mean

Experiment

We conducted validation experiments on MRL tasks using ModuLM to demonstrate the framework's ability to support a wide range of experiments, comparisons, and analyses. The following sections present the results for DDI and SSI tasks, highlighting the impact of different inputs and encoders on LLM performance in MRL. This also validates the effectiveness of our framework.

Experiment No.	Backbone	Encoder	Interaction	Input Feature
1.1	Galactica-1.3B	-	-	m_s
1.2	Galactica-1.3B	GIN	-	$m_s + m_g$
1.3	Galactica-1.3B	GIN	Cross Attention	$m_s + m_g$
1.4	Galactica-1.3B	Uni-mol	-	$m_s + m_c$
1.5	Galactica-6.7B	MPNN	Gated Fusion	$m_s + m_g$
1.6	DeepSeek-1.5B	-	-	m_s
1.7	DeepSeek-1.5B	GIN	-	$m_s + m_g$
1.8	DeepSeek-1.5B	Uni-mol	-	$m_s + m_c$
1.9	DeepSeek-7B	3D-GeoFormer	Highway	$m_s + m_c$
1.10	DeepSeek-14B	Uni-mol	-	$m_s + m_c$
1.11	DeepSeek-14B	GAT	Self Attention	$m_s + m_g$
1.12	LLaMA-1B	-	-	m_s
1.13	LLaMA-1B	CoATGIN	-	$m_s + m_g$
1.14	LLaMA-1B	EGNN	Gated Fusion	$m_s + m_c$
1.15	LLaMA-13B	SchNet	Bilinear Attention	$m_s + m_c$

Note: m_s = molecular sequence, m_g = molecular graph, m_c = molecular conformation. '-' indicates that no method is applied.

Performance on DDI Datasets

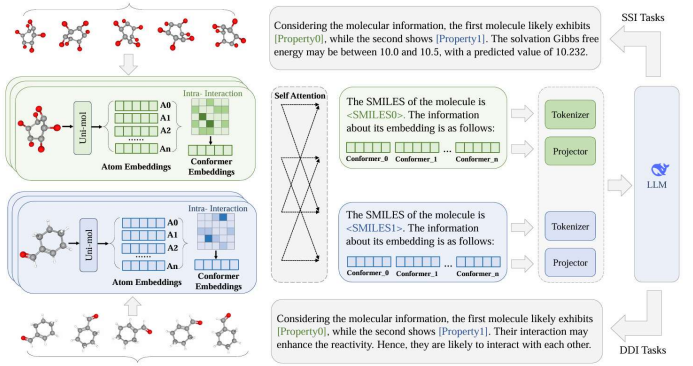
Experiment	AUC-ROC (ChChMiner)	Accuracy (ChChMiner)	AUC-ROC (ZhangDDI)	Accuracy (ZhangDDI)	AUC-ROC (DeepDDI)	Accuracy (DeepDDI)
Chem T5[5]	0.867 ± 0.012	0.814 ± 0.009	0.889 ± 0.017	0.751 ± 0.021	0.856 ± 0.012	0.784 ± 0.013
MolCA[43]	0.924 ± 0.006	0.901 ± 0.009	0.895 ± 0.006	0.745 ± 0.010	0.878 ± 0.014	0.841 ± 0.015
MolT5[13]	0.914 ± 0.019	0.862 ± 0.022	0.901 ± 0.011	0.802 ± 0.015	0.907 ± 0.014	0.870 ± 0.016
MolTC[15]	0.964 ± 0.008	0.957 ± 0.006	**0.941 ± 0.006**	0.896 ± 0.008	**0.977 ± 0.013**	0.956 ± 0.011
1.1	0.933 ± 0.011	0.924 ± 0.009	0.912 ± 0.008	0.854 ± 0.004	0.899 ± 0.010	0.855 ± 0.009
1.2	0.956 ± 0.008	0.943 ± 0.009	0.930 ± 0.006	0.872 ± 0.009	0.924 ± 0.008	0.887 ± 0.008
1.3	0.960 ± 0.010	0.954 ± 0.006	0.933 ± 0.007	0.891 ± 0.004	0.939 ± 0.007	0.904 ± 0.008
1.4	0.955 ± 0.005	0.949 ± 0.008	0.936 ± 0.014	0.901 ± 0.010	0.956 ± 0.008	0.919 ± 0.007
1.5	0.940 ± 0.009	0.932 ± 0.008	0.921 ± 0.008	0.866 ± 0.005	0.948 ± 0.009	0.912 ± 0.006
1.6	0.936 ± 0.010	0.930 ± 0.011	0.920 ± 0.009	0.860 ± 0.008	0.906 ± 0.008	0.872 ± 0.008
1.7	0.957 ± 0.008	0.953 ± 0.010	0.934 ± 0.006	0.889 ± 0.004	0.958 ± 0.007	0.942 ± 0.007
1.8	**0.966 ± 0.007**	**0.964 ± 0.005**	0.938 ± 0.005	**0.907 ± 0.006**	0.972 ± 0.009	**0.959 ± 0.010**
1.9	0.944 ± 0.010	0.935 ± 0.009	0.925 ± 0.005	0.870 ± 0.003	0.955 ± 0.008	0.930 ± 0.007
1.10	0.931 ± 0.012	0.918 ± 0.010	0.916 ± 0.009	0.861 ± 0.011	0.943 ± 0.010	0.915 ± 0.008
1.11	0.935 ± 0.007	0.921 ± 0.012	0.906 ± 0.009	0.855 ± 0.008	0.936 ± 0.009	0.908 ± 0.007
1.12	0.925 ± 0.013	0.911 ± 0.011	0.901 ± 0.008	0.852 ± 0.006	0.895 ± 0.010	0.850 ± 0.009
1.13	0.945 ± 0.009	0.937 ± 0.008	0.925 ± 0.007	0.870 ± 0.006	0.935 ± 0.008	0.904 ± 0.008
1.14	0.951 ± 0.007	0.946 ± 0.011	0.928 ± 0.004	0.875 ± 0.005	0.946 ± 0.007	0.918 ± 0.006
1.15	0.915 ± 0.016	0.896 ± 0.013	0.913 ± 0.008	0.860 ± 0.002	0.928 ± 0.011	0.897 ± 0.008

Performance on SSI Datasets

Experiment	MAE (FreeSolv)	RMSE (FreeSolv)	MAE (CompSol)	RMSE (CompSol)	MAE (CombiSolv)	RMSE (CombiSolv)
Chem T5[5]	0.923 ± 0.022	1.511 ± 0.043	0.611 ± 0.017	0.766 ± 0.032	0.840 ± 0.040	1.294 ± 0.043
MolCA[43]	0.761 ± 0.034	1.303 ± 0.039	0.505 ± 0.036	0.726 ± 0.040	0.771 ± 0.033	1.130 ± 0.027
MolT5[13]	0.733 ± 0.047	1.135 ± 0.059	0.496 ± 0.028	0.708 ± 0.020	0.677 ± 0.024	1.066 ± 0.027
MolTC[15]	0.533 ± 0.018	0.726 ± 0.022	0.244 ± 0.018	0.356 ± 0.022	0.237 ± 0.019	0.465 ± 0.022
1.1	0.710 ± 0.021	1.120 ± 0.030	0.472 ± 0.024	0.665 ± 0.028	0.615 ± 0.026	0.984 ± 0.032
1.2	0.570 ± 0.020	0.910 ± 0.028	0.384 ± 0.022	0.540 ± 0.025	0.568 ± 0.023	0.930 ± 0.030
1.3	0.556 ± 0.018	0.840 ± 0.025	0.366 ± 0.021	0.522 ± 0.024	0.487 ± 0.021	0.820 ± 0.027
1.4	0.534 ± 0.017	0.808 ± 0.024	0.347 ± 0.020	0.501 ± 0.023	0.447 ± 0.020	0.780 ± 0.026
1.5	0.580 ± 0.016	0.972 ± 0.023	0.403 ± 0.019	0.575 ± 0.021	0.579 ± 0.019	0.898 ± 0.025
1.6	0.685 ± 0.015	1.086 ± 0.022	0.451 ± 0.018	0.643 ± 0.020	0.602 ± 0.018	0.945 ± 0.024
1.7	0.550 ± 0.014	0.749 ± 0.021	0.271 ± 0.017	0.415 ± 0.019	0.289 ± 0.017	0.515 ± 0.023
1.8	**0.510 ± 0.013**	**0.698 ± 0.020**	**0.191 ± 0.016**	**0.298 ± 0.018**	**0.190 ± 0.016**	**0.388 ± 0.022**
1.9	0.555 ± 0.014	0.825 ± 0.021	0.335 ± 0.017	0.490 ± 0.019	0.393 ± 0.017	0.595 ± 0.023
1.10	0.605 ± 0.015	0.850 ± 0.022	0.443 ± 0.018	0.598 ± 0.020	0.548 ± 0.018	0.864 ± 0.024
1.11	0.590 ± 0.016	0.876 ± 0.023	0.458 ± 0.019	0.601 ± 0.021	0.556 ± 0.019	0.839 ± 0.025
1.12	0.745 ± 0.017	1.091 ± 0.024	0.514 ± 0.020	0.692 ± 0.022	0.687 ± 0.020	1.008 ± 0.026
1.13	0.605 ± 0.016	0.880 ± 0.023	0.374 ± 0.019	0.530 ± 0.021	0.538 ± 0.019	0.820 ± 0.025
1.14	0.580 ± 0.015	0.887 ± 0.022	0.360 ± 0.018	0.515 ± 0.020	0.460 ± 0.018	0.790 ± 0.024
1.15	0.630 ± 0.018	0.910 ± 0.025	0.450 ± 0.021	0.633 ± 0.023	0.567 ± 0.021	0.892 ± 0.027

Custom Model

In addition to supporting existing model configurations, our ModuLM framework allows for custom model design, which can be combined with existing configurations to conduct relevant experiments.



We conducted further experimental comparisons between the best-performing model configuration from previous experiments and our custom-designed model.

Performance of Custom Model on DDI and SSI Datasets

Experiment	Accuracy (ChChMiner)	Accuracy (ZhangDDI)	Accuracy (DeepDDI)	RMSE (FreeSolv)	RMSE (CompSol)	RMSE (CombiSolv)
1.7	0.953 ± 0.010	0.889 ± 0.004	0.942 ± 0.007	0.749 ± 0.021	0.415 ± 0.019	0.515 ± 0.023
1.8	0.964 ± 0.005	0.907 ± 0.006	0.959 ± 0.010	0.698 ± 0.020	0.298 ± 0.018	0.388 ± 0.022
Custom Model	**0.968 ± 0.006**	**0.911 ± 0.006**	**0.964 ± 0.008**	**0.680 ± 0.019**	**0.288 ± 0.013**	**0.359 ± 0.013**

Results of the Ablation Study on DDI and SSI Datasets

Experiment	Accuracy (ChChMiner)	Accuracy (ZhangDDI)	Accuracy (DeepDDI)	RMSE (FreeSolv)	RMSE (CompSol)	RMSE (CombiSolv)
w/o M-Encoder	0.962 ± 0.007	0.905 ± 0.005	0.959 ± 0.007	0.700 ± 0.019	0.299 ± 0.017	0.370 ± 0.020
w/o Interaction	0.955 ± 0.008	0.896 ± 0.006	0.950 ± 0.007	0.705 ± 0.018	0.317 ± 0.016	0.388 ± 0.019
w/o CoT	0.959 ± 0.006	0.901 ± 0.006	0.956 ± 0.008	0.732 ± 0.020	0.335 ± 0.018	0.382 ± 0.021
Full Model	**0.968 ± 0.006**	**0.911 ± 0.006**	**0.964 ± 0.008**	**0.680 ± 0.019**	**0.288 ± 0.013**	**0.359 ± 0.013**

The above experiments demonstrate ModuLM's strong capability in supporting user-defined models. Following our provided protocol, users can create custom encoders and flexibly integrate them with other encoders and interaction layers as modular components. These custom models can involve more complex configurations, including user-defined blocks and additional operations such as stacking and flattening the outputs of ModuLM components. Furthermore, ModuLM's dynamic model construction mechanism allows users to easily adjust the model architecture to perform ablation studies.

Future Work

In the future, we will continue to expand and maintain ModuLM by adding new encoders, interaction layers, evaluation metrics, and broader LLM support to enhance its flexibility, generalizability, and scalability. We will also deepen its application in molecular relational learning, especially in drug discovery and protein interaction analysis, aiming to make ModuLM a powerful, flexible, and efficient tool for advancing interdisciplinary research.