

Supplementary information

CataCon: a contrastive graph representation learning framework for catalyst prediction

Hua Shi^{1#}, Yuzhe Wang^{2#}, Shouzhen Song¹, Kefei Li², Wenjia Gao², Shuan Liu², Jianbo Qiao^{2*}, Leyi Wei^{3, 4*}

¹ School of Opto-electronic and Communication Engineering, Xiamen University of Technology, Xiamen, China

² School of Software, Shandong University, Jinan, China

³ Faculty of Applied Sciences, Macao Polytechnic University, Macao 999078, China

⁴ Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR), Shandong University, Jinan, China

#These authors contributed equally to this work as the first authors

*These authors jointly supervised this work

* Correspondence:

Jianbo Qiao: jamble@mail.sdu.edu.cn

Leyi Wei: weileyi@sdu.edu.cn

Supplementary Table S1. Best- and worst-performing catalyst classes under the product-scaffold split

Best-performing catalyst classes	Test support	Train support	Top-1 recall	Catalyst SMILES
Label 55	14	230	1	<chem>CCN(CC)CC.Cc1ccccc1.ClCCl</chem>
Label 67	31	325	0.9677	<chem>[C-]#N.[C-]#N.[Zn+2].c1ccc([P](c2ccccc2)(c2ccccc2)[Pd]([P](c2ccccc2)(c2ccccc2)c2ccccc2)([P](c2ccccc2)(c2ccccc2)c2ccccc2)[P](c2ccccc2)(c2ccccc2)c2ccccc2)cc1</chem>
Label 76	18	215	0.9444	<chem>C1CCOC1.C[P+](c1ccccc1)(c1ccccc1)c1ccccc1.[Br-]</chem>
Worst-performing catalyst classes	Test support	Train support	Top-1 recall	Catalyst SMILES
Label 29	83	619	0	<chem>CO.[C].[Pd]</chem>
Label 64	68	389	0	<chem>CCO.[OH-].[OH-].[Pd+2]</chem>
Label 44	52	484	0	<chem>CCO.[C].[Pd]</chem>

Supplementary Table S2. Most frequent confusion pairs under the product-scaffold split

Count	True label	Predicted label	True catalyst SMILES	Predicted catalyst SMILES
460	2	1	<chem>CCO.[Pd]</chem>	<chem>CO.[Pd]</chem>
150	13	1	<chem>CCOC(C)=O.[Pd]</chem>	<chem>CO.[Pd]</chem>
90	4	1	<chem>[Pd]</chem>	<chem>CO.[Pd]</chem>
86	19	1	<chem>CO.[OH-].[OH-].[Pd+2]</chem>	<chem>CO.[Pd]</chem>
73	26	1	<chem>C1CCOC1.[Pd]</chem>	<chem>CO.[Pd]</chem>

Supplementary Table S3. Best- and worst-performing product scaffolds under the product-scaffold split

Best-performing product scaffolds	Test support	Top-1	Top-5	Top-10
<chem>C(=Cn1c2c(c3cccc31)CCNC2)c1cccc1</chem>	20	1	1	1
<chem>O=c1c(Cc2ccc(-c3cccc3)cc2)cncn1-c1cccc1</chem>	63	0.8254	1	1
<chem>c1cncc(C2CC2)c1</chem>	35	0.6286	0.7429	0.9429
Worst-performing product scaffolds	Test support	Top-1	Top-5	Top-10
<chem>c1ccc(-c2cnc3ccccn23)cc1</chem>	22	0.1364	0.4091	0.6364
<chem>c1ccc(-n2cccn2)nc1</chem>	32	0.1562	0.5	0.8438
<chem>c1ccc(-c2[nH]c3cccc3c2C2CCCC2)cc1</chem>	25	0.16	0.44	0.6