Electronic Supplementary Material (ESI) for Journal of Materials Chemistry A.

Supplementary Information for:

Artificial neural network using multi-head intermolecular attention for predicting chemical reactivity of organic materials

Jaekyun Yoo[1], ByungHoon Kim[1], Byungju Lee[2], Jun-hyuk Song[1] and Kisuk Kang*[1,3]

[1] Department of Materials Science and Engineering, Research Institute of Advanced Materials (RIAM), Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea

2 Center for Computational Science, Korea Institute of Science and Technology (KIST), Hwarangno 14-gil 5, Seongbuk-gu, Seoul 136-791, Korea

3 Center for Nanoparticle Research at Institute for Basic Science (IBS), 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

*e-mail: matlgen1@snu.ac.kr

Table S1. 40 examples of reaction rate constant fabricated from the Mayr's database. Those examples consist of 8 nucleophile and 5 electrophile molecules. Reaction rate constant was calculated from nucleophilicity *N*, S_N parameter and electrophilicity *E* by Mayr's formula $\log_{10} k_{20°\mathrm{C}} = S_N(N + E)$.

N_SMILES	E_SMILES	N	S_N	E	$k_{20°C}$
OCC(F)(F)F	[CH+]1SCCCS1	1.11	0.96	-2.14	1.03×10^{-1}
OCC(F)(F)F	C(/C=C/c1ccccc1)=[N+]1CCCC1	1.11	0.96	-9.8	4.55×10^{-9}
OCC(F)(F)F	C[CH+]c1c(C)cc(C)cc1C	1.11	0.96	6.04	7.31×10^{6}
OCC(F)(F)F	C=CC(=O)OC(C)(C)C	1.11	0.96	-20.22	4.51×10^{-19}
OCC(F)(F)F	Clc1ccc([C+]2C=CCC2)cc1	1.11	0.96	3.2	1.37×10^{-4}
O=N[O-]	[CH+]1SCCCS2	17.20	0.72	-2.14	6.97×10^{10}
O=N[O-]	C(/C=C/c1ccccc1)=[N+]1CCCC2	17.20	0.72	-9.8	2.13×10^{4}
O=N[O-]	C[CH+]c1c(C)cc(C)cc2C	17.20	0.72	6.04	5.41×10^{16}
O=N[O-]	C=CC(=O)OC(C)(C)C	17.20	0.72	-20.22	6.69×10^{-3}
O=N[O-]	Clc1ccc([C+]2C=CCC2)cc2	17.20	0.72	3.2	4.48×10^{14}
CC(C)=C(C)C	[CH+]1SCCCS3	-1.00	1.40	-2.14	4.02×10^{-5}
CC(C)=C(C)C	C(/C=C/c1ccccc1)=[N+]1CCCC3	-1.00	1.40	-9.8	7.59×10^{-16}
CC(C)=C(C)C	C[CH+]c1c(C)cc(C)cc3C	-1.00	1.40	6.04	1.14×10^{7}
CC(C)=C(C)C	C=CC(=O)OC(C)(C)C	-1.00	1.40	-20.22	1.96×10^{-30}
CC(C)=C(C)C	Clc1ccc([C+]2C=CCC2)cc3	-1.00	1.40	3.2	1.20×10^{3}
O=C([O-])C[S-]	[CH+]1SCCCS4	22.62	0.43	-2.14	6.40×10^{8}
O=C([O-])C[S-]	C(/C=C/c1ccccc1)=[N+]1CCCC4	22.62	0.43	-9.8	3.26×10^{5}
O=C([O-])C[S-]	C[CH+]c1c(C)cc(C)cc4C	22.62	0.43	6.04	2.11×10^{12}
O=C([O-])C[S-]	C=CC(=O)OC(C)(C)C	22.62	0.43	-20.22	1.08×10^{1}
O=C([O-])C[S-]	Clc1ccc([C+]2C=CCC2)cc4	22.62	0.43	3.2	1.27×10^{11}
c1ccc2cccc-2cc1	[CH+]1SCCCS5	6.66	1.02	-2.14	4.08×10^{4}
c1ccc2cccc-2cc2	C(/C=C/c1ccccc1)=[N+]1CCCC5	6.66	1.02	-9.8	6.27×10^{-4}
c1ccc2cccc-2cc3	C[CH+]c1c(C)cc(C)cc5C	6.66	1.02	6.04	8.99×10^{12}
c1ccc2cccc-2cc4	C=CC(=O)OC(C)(C)C	6.66	1.02	-20.22	1.48×10^{-14}

c1ccc2cccc-2cc5	Clc1ccc([C+]2C=CCC2)cc5	6.66	1.02	3.2	1.14×10^{10}
CC(=O)C=[N+]=[N-]	[CH+]1SCCCS6	3.96	0.91	-2.14	4.53×10^{1}
CC(=O)C=[N+]=[N-]	C(/C=C/c1ccccc1)=[N+]1CCCC6	3.96	0.91	-9.8	4.85×10^{-6}
CC(=O)C=[N+]=[N-]	C[CH+]c1c(C)cc(C)cc6C	3.96	0.91	6.04	1.26×10^{9}
CC(=O)C=[N+]=[N-]	C=CC(=O)OC(C)(C)C	3.96	0.91	-20.22	1.6×10^{-15}
CC(=O)C=[N+]=[N-]	Clc1ccc([C+]2C=CCC2)cc6	3.96	0.91	3.2	3.28×10^{6}
C=CC[Si](C)(C)Cl	[CH+]1SCCCS7	-0.57	1.06	-2.14	1.34×10^{-3}
C=CC[Si](C)(C)Cl	C(/C=C/c1ccccc1)=[N+]1CCCC7	-0.57	1.06	-9.8	1.02×10^{-11}
C=CC[Si](C)(C)Cl	C[CH+]c1c(C)cc(C)cc7C	-0.57	1.06	6.04	6.28×10^{5}
C=CC[Si](C)(C)Cl	C=CC(=O)OC(C)(C)C	-0.57	1.06	-20.22	9.17×10^{-23}
C=CC[Si](C)(C)Cl	Clc1ccc([C+]2C=CCC2)cc7	-0.57	1.06	3.2	6.13×10^{2}
C=CC[B-]1(c2ccccc2)OC(C)(C)C(C)(C)O1.[Li+]	[CH+]1SCCCS8	11.20	0.64	-2.14	6.29×10^{5}
C=CC[B-]1(c2ccccc2)OC(C)(C)C(C)(C)O1.[Li+]	C(/C=C/c1ccccc1)=[N+]1CCCC8	11.20	0.64	-9.8	7.87
C=CC[B-]1(c2ccccc2)OC(C)(C)C(C)(C)O1.[Li+]	C[CH+]c1c(C)cc(C)cc8C	11.20	0.64	6.04	1.08×10^{11}
C=CC[B-]1(c2ccccc2)OC(C)(C)C(C)(C)O1.[Li+]	C=CC(=O)OC(C)(C)C	11.20	0.64	-20.22	1.69×10^{-6}
C=CC[B-]1(c2ccccc2)OC(C)(C)C(C)(C)O1.[Li+]	Clc1ccc([C+]2C=CCC2)cc8	11.20	0.64	3.2	1.64×10^{9}

Table S2. Performance of ImRRNet trained with split nucleophile data based on K-means clustering and the number of molecules in each nucleophile group. MAE and RMSE value of one group denotes the values when that group is used as validation set, whereas rest of dataset is used as training set.

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Average
The number of molecules (The number of corresponding dataset)	353 (76601)	29 (6293)	26 (5642)	139 (30163)	50 (10850)	128 (27776)	50 (10850)	36 (7812)	-
MAE	3.332	3.795	4.492	2.034	2.284	2.499	2.407	2.302	2.893
RMSE	5.259	5.352	6.685	3.214	3.994	4.533	4.263	4.072	4.672

Table S3. Performance of ImRRNet trained with split electrophile data based on K-means clustering and the number of molecules in each electrophile group. MAE and RMSE value of one group denotes the values when that group is used as validation set, whereas rest of dataset is used as training set.

	Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**	**Group 7**	**Group 8**	**Average**
The number of molecules (The number of corresponding dataset)	39 (31629)	60 (48660)	15 (12165)	20 (16220)	15 (12165)	38 (30818)	11 (8921)	19 (15409)	-
MAE	3.957	3.553	4.266	1.871	2.532	1.723	4.409	1.482	2.974
RMSE	5.736	5.581	6.004	3.458	4.130	3.266	6.005	3.122	4.663

Table S4. Hyperparameter sets which were used during hyperparameter optimization process for ImRRNet model and resulting performance. Only the number of heads and dimension of final dense layers were optimized. Num_heads, Dense_dim1 and Dense_dim2 denotes for the number of heads, dimension of the first dense layer and dimension of the second dense layer at final dense layers, respectively. The model with lowest MAE was selected for final hyperparameter set which is denoted by bold letters.

Num_heads	Dense_dim1	Dense_dim2	Average of MAE	Average of RMSE
10	**400**	**200**	**1.5596**	**2.0461**
2	400	200	1.5703	2.0864
6	600	300	1.5751	2.0564
5	600	300	1.5803	2.0492
2	800	400	1.5818	2.1403
12	200	100	1.6017	2.0644
2	200	100	1.6042	2.1015
1	400	200	1.6273	2.1369
6	200	100	1.6375	2.1054
12	1000	500	1.6525	2.1633
5	400	200	1.6587	2.1666
3	400	200	1.6696	2.1822
4	600	0	1.6758	2.2159
3	800	400	1.6792	2.2495
12	800	400	1.6908	2.1974
4	400	200	1.7028	2.2219
4	600	300	1.7042	2.2283
1	600	0	1.7221	2.2628
1	400	0	1.7250	2.2815
2	800	0	1.7613	2.3193
2	200	0	1.8701	2.3913
5	400	0	1.8840	2.4603

12	600	0	1.9160	2.5173
10	1000	0	1.9542	2.5612
12	800	0	1.9732	2.5559
12	1000	0	2.0115	2.6111

Table S5. Hyperparameter sets and resulting performance which were used during hyperparameter optimization process for GRU, LSTM amd Delfos-base model. The number of Final dense layers and dimension of dense layers were optimized. The GRU and LSTM model with lowest MAE was selected for final hyperparameter set. For the case of Delfos-base model, MAE and RMSE were unexpectedly high at hyperparameter optimization stage, we trained and tested again with the entire training and test sets, which recorded the lowest MAE with model whose dimension of final dense layers was 400/200. Final hyperparameter sets were denoted by bold letters.

Model type	Dense_dim1	Dense_dim2	Dense_dim3	Average of MAE	Average of RMSE
GRU model	**600**	**300**		**1.5605**	**2.7310**
GRU model	400	200		1.6005	2.7639
GRU model	200	100		1.6143	2.7703
GRU model	400	200	100	1.6294	2.8568
LSTM model	**200**	**100**		**1.7697**	**2.8794**
LSTM model	600	300		1.7844	2.9334
LSTM model	400	200		1.8456	2.9500
LSTM model	400	200	100	1.8480	3.0537
Delfos-base model	600	300		5.1989	8.5807
Delfos-base model	**400**	**200**		**5.2151**	**8.6694**
Delfos-base model	400	200	100	5.3060	9.1801
Delfos-base model	200	100		5.3409	8.5185