Supplementary Information for Powerful, transferable representations for molecules through intelligent task selection in deep multitask networks

Clyde Fare, Lukas Turcani, and Edward O. Pyzer-Knapp

Table 1: Task list

Id	Task	Dataset	Code	Reference
0	log solubilities	Huuskonen	huusk_logs	[1]
1	melting points	Karthikeyan	karth_mp	[2]
2	anti HIV activities	AIDS	aids_anti_HIV	[3]
3	blood-brain-barrier activities	BBB	bbb_bbb	[4]
4	photo conversion efficiency	CEP	cep_pce	[5]
5	heat capacity (C_v)	DSDB9	qm9_cv	[6]
6	formation energy (G @ 298.15K)	DSDB9	qm9_G	[6]
7	formation energy (H @ 298.15K)	DSDB9	qm9_H	[6]
8	formation energy (U @ 0K)	DSDB9	qm9_U2	[6]
9	formation energy (U @ 298.15K)	DSDB9	qm9_U1	[6]
10	isotropic polarizability	DSDB9	qm9_alpha	[6]
11	band gap	DSDB9	qm9_gap	[6]
12	highest occupied molecular orbital	DSDB9	qm9_homo	[6]
13	lowest unoccupied molecular orbital	DSDB9	qm9_lumo	[6]
14	total dipole moment	DSDB9	qm9_mu	[6]
15	electronic spatial extent	DSDB9	qm9_r2	[6]
16	zero point energy	DSDB9	qm9_zpve	[6]
17	electronic band gap	HOPV15	hopv15_egap	[7]
18	fill factor	HOPV15	hopcv15_ff	[7]
19	highest occupied molecular orbital	HOPV15	hopv15_homo	[7]
20	model short circuit current density	HOPV15	hopv15_jsc	[7]
21	lowest unoccupied molecular orbital	HOPV15	hopv15_lumo	[7]
22	optical band gap	HOPV15	hopv15_ogap	[7]
23	photo conversion efficiency	HOPV15	hopv15_pce	[7]
24	anti malarial activity	MMV3	mal_am	[8]
25	raw ec50	MMV3	mal_ec50	[8]
26	log octanol-water partition coefficient	NCI2000_r2v2	nci2000_klogp	[9]
27	hlms stability	DTP2010_r4	nci2010_hlms	[10]
28	ionisation energy (adiabatic)	SRD101	srd101_EI1A	[11]
29	ionisation energy (vertical)	SRD101	srd101_EI1V	[11]
30	formation energy (H @ 0K)	SRD101	srd101_H2	[11]
31	formation energy (H @ 298.15K)	SRD101	srd101_H1	[11]
32	rotational constant (R_A)	SRD101	srd101_RA	[11]
33	rotational constant (R_C)	SRD101	srd101_RC	[11]
34	Entropy (S @ 298.15K)	SRD101	srd101_S	[11]
35	dipole moment	SRD101	srd101_mu	[11]
36	aryl hydrocarbon receptor activity	TOX21	tox21_ahr	[12]
37	androgen receptor full activity	TOX21	tox21_arf	[12]
38	androgen receptor LBD activity	TOX21	tox21_arlbd	[12]
39	aromatase activity	TOX21	tox21_arom	[12]
40	atad5 activity	TOX21	tox21_atad5	[12]
41	estrogen receptor alpha full activity	TOX21	tox21_erf	[12]
42	estrogen receptor alpha LBD activity	TOX21	tox21_erldb	[12]
43	heat shock factor response element activity	TOX21	tox21_hse	[12]
44	mitochondrial membrane potential activity	ToX21	tox21_mmp	[12]
45	nrf2 activity responsive element activity	TOX21	tox21_nrf2a	[12]
46	p53 activity	TOX21	tox21_p53	[12]
47	ppa receptor gamma activity	TOX21	tox21_pparg	[12]

References

[1] J. Huuskonen, *Journal of Chemical Information and Computer Sciences*, 2000.

[2] M. Karthikeyan, R. C. Glen and A. Bender, *Journal of Chemical Information and Modeling*, 2005.

[3] NCI, *AIDS Antiviral Screen Data*, 2015, `https://wiki.nci.nih.gov/display/NCIDTPdata/AIDS+Antiviral+Screen+Data`.

[4] H. Li, C. W. Yap, C. Y. Ung, Y. Xue, Z. W. Cao and Y. Z. Chen, *Journal of Chemical Information and Modeling*, 2005, **45**, 1376–1384.

[5] J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway and A. Aspuru-Guzik, *Journal of Physical Chemistry Letters*, 2011.

[6] R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. von Lilienfeld, *Scientific Data*, 2014, **1**, 1–7.

[7] S. A. Lopez, E. O. Pyzer-Knapp, G. N. Simm, T. Lutzow, K. Li, L. R. Seress, J. Hachmann and A. Aspuru-Guzik, *Scientific Data*, 2016.

[8] A. Gaulton, A. Hersey, M. L. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magariños, J. P. Overington, G. Papadatos, I. Smit, A. R. Leach, A. Patricia Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrian-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magarinos, J. P. Overington, G. Papadatos, I. Smit and A. R. Leach, *Nucleic Acids Research*, 2017, **45**, D945–D954.

[9] NCI, *NCI Open Database Release 2*, 2000, `https://cactus.nci.nih.gov/download/nci/{\#}release-2`.

[10] NCI, *GUSAR Human Liver Microsoman Stability Prediction Data*, 2012, `https://cactus.nci.nih.gov/download/nci/{\#}release-4`.

[11] R. D. Johnson III, *NIST Standard Reference, Database Number 101*, 2015.

[12] T. Unterthiner, A. Mayr, G. Klambauer and S. Hochreiter, *frontiers in environmental science*, 2015, **3**, 10.