Supplementary file for *Realistic material property prediction using domain adaptation based machine learning*

Jeffrey Hu, David Liu, Nihang Fu, Rongzhi Dong

December 5, 2023

1 Dataset preparation

We downloaded two datasets (matbench_expt_gap and matbench_glass) from the Matbench site including one classification problem for glass materials, and one regression problem related to band gap prediction. We then prepare five DA test dataset using five different approaches including leave-one-cluster out (LOCO), 50 sparse samples in the X input space (SparseXSingle), 50 sparse samples in the Y (property) space (SparseYSingle), 50 sparse clusters each with 11 samples in the X input space (SparseXCluster), 50 sparse clusterse each with 11 samples in the Y (property) space (SparseYCluster). In total, we have 10 datasets for DA algorithm evaluations, including bandgap-LOCO, bandgap-SparseX, bandgap-SparseXCluster, bandgap-SparseY, and bandgap-SparseYCluster for regression and glass-LOCO, glass-SparseX, glass-SparseXCluster, glass-SparseY, and glass-SparseYCluster for classification. The test cluster statistics are shown in Table S1.

Table S1: 10 datasets for domain adaptation model evaluations

	bandgap					glass				
Cluster ID	LOCO	SparseX Cluster	SparseY Cluster	SparseX Single	SparseY Single	LOCO	SparseX Cluster	SparseY Cluster	SparseX Single	SparseY Single
0	78	11	11	1	1	27	11	11	1	1
1	77	11	11	1	1	224	11	11	1	1
2	110	11	11	1	1	213	11	11	1	1
3	59	11	11	1	1	245	11	11	1	1
4	113	11	11	1	1	149	11	11	1	1
5	117	11	11	1	1	121	11	11	1	1
6	63	11	11	1	1	184	11	11	1	1
7	73	11	11	1	1	125	11	11	1	1
8	96	11	11	1	1	126	11	11	1	1
9	75	11	11	1	1	107	11	11	1	1
10	40	11	11	1	1	230	11	11	1	1
11	134	11	11	1	1	107	11	11	1	1
12	104	11	11	1	1	97	11	11	1	1
13	199	11	11	1	1	202	11	11	1	1
14	40	11	11	1	1	146	11	11	1	1
15	124	11	11	1	1	95	11	11	1	1
16	42	11	11	1	1	77	11	11	1	1
17	127	11	11	1	1	88	11	11	1	1
18	99	11	11	1	1	110	11	11	1	1
19	138	11	11	1	1	108	11	11	1	1
20	170	11	11	1	1	84	11	11	1	1
21	27	11	11	1	1	39	11	11	1	1
22	57	11	11	1	1	87	11	11	1	1
23	70	11	11	1	1	117	11	11	1	1
24	172	11	11	1	1	125	11	11	1	1
25	175	11	11	1	1	118	11	11	1	1
26	81	11	11	1	1	264	11	11	1	1
27	47	11	11	1	1	96	11	11	1	1
28	232	11	11	1	1	71	11	11	1	1
29	79	11	11	1	1	19	11	11	1	1
30	96	11	11	1	1	189	11	11	1	1
31	118	11	11	1	1	91	11	11	1	1
32	130	11	11	1	1	321	11	11	1	1
33	84	11	11	1	1	92	11	11	1	1
34	97	11	11	1	1	276	11	11	1	1
35	25	11	11	1	1	26	11	11	1	1
36	106	11	11	1	1	118	11	11	1	1
37	76	11	11	1	1	25	11	11	1	1
38	35	11	11	1	1	60	11	11	1	1
39	16	11	11	1	1	61	11	11	1	1
40	205	11	11	1	1	76	11	11	1	1
41	122	11	11	1	1	79	11	11	1	1
42	28	11	11	1	1	61	11	11	1	1
43	69	11	11	1	1	55	11	11	1	1
44	57	11	11	1	1	84	11	11	1	1
45	98	11	11	1	1	9	11	11	1	1
46	52	11	11	1	1	170	11	11	1	1
47	22	11	11	1	1	20	11	11	1	1
48	104	11	11	1	1	21	11	11	1	1
49	46	11	11	1	1	45	11	11	1	1
avg	92.08	11	11	1	1	113.6	11	11	1	1
min	16	11	11	1	1	9	11	11	1	1
max	232	11	11	1	1	321	11	11	1	1

2 Figures

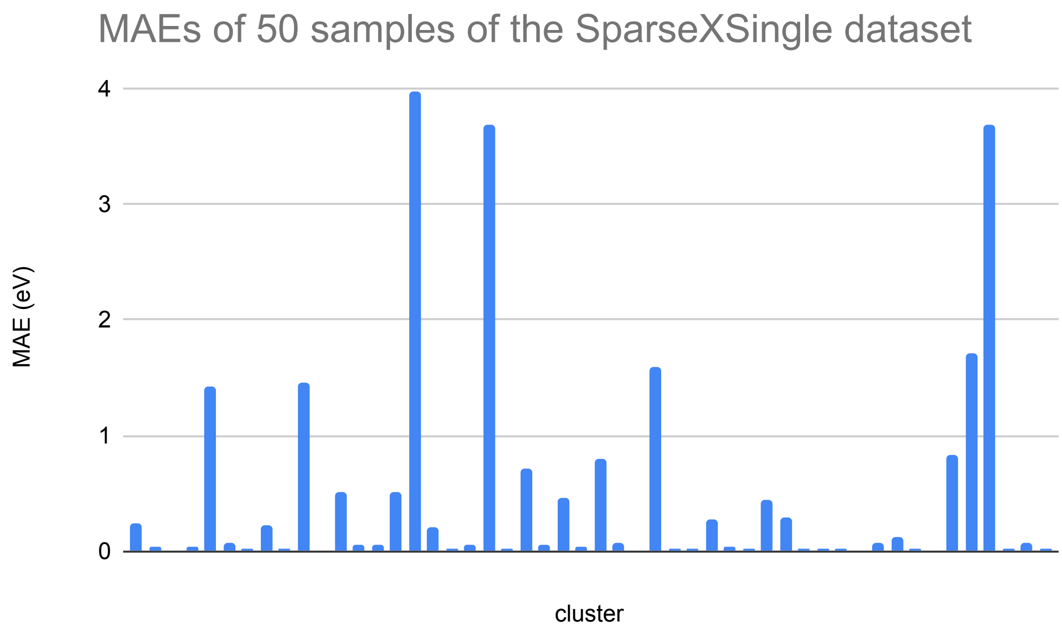

Figure S1: Distribution of the MAEs for the 50 single-sample clusters by the Roost algorithm. It shows that a few clusters/samples are extremely difficult to predict with more than 1.0 eV MAEs, which leads to high variation of the model predictions.

3 Tables

Table S2: How the number of fine-tuning samples affects the DA performance

fine-tune sample #	metrics	BW	TransferTreeClf	RegTransferLC	TrAdaBoost	LinInt
3 labeled samples	baseline avg. mae	0.652	0.644	0.597	0.658	0.656
	adapt avg. mae	0.704	0.556	0.527	0.633	0.533
	performance gain %	8.03%	-13.64%	-11.78%	-3.70%	-18.74%
50% of test set	baseline avg. mae	0.652	0.632	0.597	0.658	0.650
	adapt avg. mae	0.742	0.645	0.640	0.725	0.638
	performance gain %	13.86%	2.18%	7.12%	10.20%	-1.73%