Loss-Balanced Task Weighting to Reduce Negative Transfer in Multi-Task Learning

Shengchao Liu[1,2] and **Yingyu Liang**[1] and **Anthony Gitter**[1,2,3]
[1]Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI
[2]Morgridge Institute for Research, Madison, WI
[3]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI
{shengchao, yliang}@cs.wisc.edu, gitter@biostat.wisc.edu

Abstract

In settings with related prediction tasks, integrated multi-task learning models can often improve performance relative to independent single-task models. However, even when the average task performance improves, individual tasks may experience **negative transfer** in which the multi-task model's predictions are worse than the single-task model's. We show the prevalence of negative transfer in a computational chemistry case study with 128 tasks and introduce a framework that provides a foundation for reducing negative transfer in multi-task models. Our Loss-Balanced Task Weighting approach dynamically updates task weights during model training to control the influence of individual tasks.

Introduction

Multi-task learning aims to exploit information from related tasks to improve the generalization performance of all the tasks jointly. Deep learning-based multi-task learning has been successfully applied in chemical screening, genomics, object detection, natural language processing, and other domains. Shared hidden layers in a neural network can transfer knowledge among related tasks, which may reduce overfitting and improve learned latent representations, especially when the task-specific training data is limited. However, when the tasks considered are not sufficiently related, the multi-task setting can be detrimental to performance.

Although the performance may improve on average over all tasks in the multi-task setting, for some specific tasks, the multi-task performance can be worse than a single-task model. This decrease in performance is known as **negative transfer**. Negative transfer occurs naturally in real scenarios and is especially problematic if a subset of tasks is of primary interest and the others are used only to improve the representation learning. We have two conjectures for why negative transfer may happen. (1) All tasks are diverse and unrelated to each other; there is no suitable common latent representation so multi-task learning produces poor representations. (2) One group of related tasks dominates the training process. The performance for those tasks improves as more related tasks are added, but tasks outside the dominant group suffer.

Despite abundant approaches for multi-task learning, few methods aim to improve average task performance while simultaneously minimizing negative transfer. Our contributions are (1) demonstrating the presence of negative transfer in a chemistry dataset and (2) proposing a preliminary algorithm intended to reduce negative transfer by learning task-specific weights. On our computational chemistry case study, this algoritm has the best average performance and fewest tasks with negative transfer.

Methods

Our goal is to design a multi-task learning framework that reduces negative transfer while still improving the average task performance. We consider five neural network-based transfer learning algorithms and a single-task baseline.

Single-Task Learning (STL) and **Multi-Task Learning (MTL)** are fully-connected neural networks, where MTL has shared hidden layers and separate outputs for T tasks in the last layer. **Fine Tuning** is another transfer learning strategy that first trains a MTL model on $T-1$ tasks and then initializes a STL model for the final task with the MTL weights. This strategy transfers the latent representation from the larger multi-task dataset to the single-task dataset so as to alleviate data insufficiency.

Another approach is to apply a task-specific weight vector. **Reinforced Multi-Task Learning (RMTL)** (Liu 2018) uses cosine similarity of the gradients between $T-1$ tasks and a single emphasized task as task weights. **GradNorm** (Chen et al. 2018) assumes that tasks with larger loss dominate the training and therefore learns a balanced global task weight. In GradNorm the task weight is static because it is identical among all batches, whereas RMTL assumes the task weight should be dynamic. Dynamic means that the task weight differs given different inputs. We propose that the challenging tasks can be identified by their loss, and the loss dynamically changes for different batches of data. Therefore, we introduce **Loss-Balanced Task Weighting (LBTW)**, which combines and expands upon ideas from RMTL and GradNorm.

LBTW follows the RMTL framework with dynamic task weights. However, LBTW assumes that the task-specific loss is informative for balancing different tasks. For each task and batch, LBTW considers the loss ratio between the current loss and the initial loss, which is a proxy for how

Table 1: Mean PR AUC on all 128 tasks and the number of tasks with negative transfer based on PR AUC.

Evaluation Metric	STL	MTL	FineTuning	GradNorm ($\alpha = 0.1$)	GradNorm ($\alpha = 0.5$)	RMTL	LBTW ($\alpha = 0.1$)	LBTW ($\alpha = 0.5$)
Mean PR AUC	0.232	0.241	0.239	0.189	0.181	0.238	0.247	0.253
# Negative Transfer	-	48	46	98	103	47	45	42

well the model has trained for that task. Poorly trained tasks have ratios close to 1 and contribute more to the overall loss and gradient. A hyperparameter α balances the influence of the task-specific weights, and LBTW approaches standard MTL as α goes to 0 (Algorithm 1). Implementation details and model hyperparameters are provided in the Appendix.

Algorithm 1: Loss-Balanced Task Weighting

Given T tasks and parameter α.
Initialize neural network weights W.
for each epoch i **do**
  for each batch of data B **do**
    Get the loss on each task $\ell_B \in \mathbb{R}^T$.
    Store the first batch loss as $\ell_{(0,i)} \in \mathbb{R}^T$.
    for each task t **do**
      Set the task weight $w_t = \left(\frac{\ell_{(B,t)}}{\ell_{(0,i,t)}}\right)^{\alpha}$.
      Update weighted loss $\ell_{(B,t)} = \ell_{(B,t)} \times w_t$.
    end for
    Update W with respect to ℓ_B.
  end for
end for

Results

We test LBTW on the PubChem BioAssay chemistry dataset. It includes 128 tasks and approximately 440,000 chemicals, where each task is a binary classification problem on whether the chemical affects a biological target. The data processing follows (Liu et al. 2018).

We compute the difference in predictive performance for each transfer learning model versus the STL baseline using PR AUC (Figure 1 and Table 1) or ROC AUC (Appendix) for evaluation. In this domain, there are far more inactive than active chemicals, so PR AUC is more meaningful than ROC AUC (Liu et al. 2018). All five approaches improve mean ROC AUC, but on average the GradNorm PR AUC is worse than STL (Table 1). No method eliminates negative transfer or even reduces it substantially for PR AUC. However, LBTW has the best mean PR AUC overall and the fewest tasks with negative transfer, slightly fewer than MTL, Fine Tuning, and RMTL.

Conclusion

Although the preliminary version of LBTW provides only a minor reduction in the number of tasks with negative transfer, the LBTW framework provides flexibility to tune task weights. Currently, LBTW uses uniform task weights when making predictions, but future versions could apply gradient-based task weighting during prediction as well. In settings where there is no shared optimal latent representation for all tasks, there will be natural tradeoffs between average performance and instances of negative transfer. LBTW provides a platform to continue to explore and tune that tradeoff.

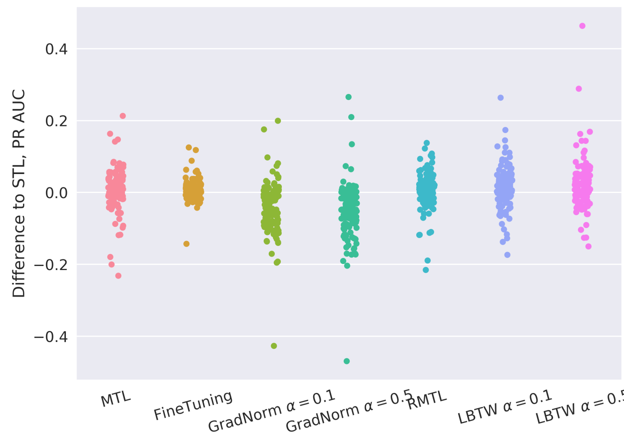

Figure 1: Distribution of the change in PR AUC relative to STL for 128 tasks. Values below 0 indicate tasks with negative transfer. Details and ROC AUC results are in the Appendix.

Acknowledgements

The authors acknowledge computing resources from the University of Wisconsin-Madison Center for High Throughput Computing and GPU hardware from NVIDIA. This work was also supported in part by FA9550-18-1-0166, the Morgridge Institute for Research, and the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. The authors thank Spencer S. Ericksen for transfer learning discussions.

References

Chen, Z.; Badrinarayanan, V.; Lee, C.-Y.; and Rabinovich, A. 2018. GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *International Conference on Machine Learning*, 793–802.

Liu, S.; Alnammi, M.; Ericksen, S. S.; Voter, A. F.; Ananiev, G. E.; Keck, J. L.; Hoffmann, F. M.; Wildman, S. A.; and Gitter, A. 2018. Practical model selection for prospective virtual screening. *bioRxiv* 337956.

Liu, S. 2018. Exploration on deep drug discovery: Representation and learning. *Master's Thesis* TR1854.

Appendix: Loss-Balanced Task Weighting to Reduce Negative Transfer in Multi-Task Learning

Shengchao Liu[1,2] and **Yingyu Liang**[1] and **Anthony Gitter**[1,2,3]
[1]Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI
[2]Morgridge Institute for Research, Madison, WI
[3]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI
{shengchao, yliang}@cs.wisc.edu, gitter@biostat.wisc.edu

Implementation

We tested six algorithms on the PubChem BioAssay (PCBA) dataset (Wang et al. 2017). Data processing followed (Liu et al. 2018).

Hyperparameters were guided by prior work on chemical screening (Liu et al. 2018), and all algorithms used the same network architecture. We used two layers with 2000 hidden units each, the first with ReLU activation functions and the second with sigmoid activation functions. The output layer used sigmoid activation functions, one per task in multi-task models. We used a dropout rate of 0.25 and trained with the Adam optimizer. We fixed the epoch number as 100 and the batch size as 1024 for all algorithms. For Single-Task Learning (STL), Fine Tuning, and Reinforced Multi-Task Learning (RMTL), we ran 128 jobs corresponding to each PCBA task. Multi-Task Learning (MTL), GradNorm, and Loss-Balanced Task Weighting (LBTW) made predictions for all 128 tasks in a single run. For GradNorm and LBTW, we tested multiple values of α, which controls the task weights. Code implementing LBTW is available at (Liu 2018).

Results

Figure S1 visualizes the difference in ROC AUC between STL and each transfer learning approach, complimenting the PR AUC results in the main text. Table S1 summarizes the mean ROC AUC for each method and the number of tasks with negative transfer when evaluated with ROC AUC. The performance on each of the 128 tasks is listed in Table S2 (PR ROC) and Table S3 (ROC AUC).

Although all of the transfer learning strategies have widespread negative transfer, LBTW has the most robust results relative to the other methods. It has the highest mean PR AUC and ROC AUC and the fewest tasks with negative transfer. RMTL achieves comparatively few instances of negative transfer, tying LBTW for ROC AUC, but its average performance is worse than LBTW and MTL. GradNorm struggles in this setting with the PR AUC evaluation metric. Its average performance is worse than STL, and the majority of the tasks have negative transfer.

References

Liu, S.; Alnammi, M.; Ericksen, S. S.; Voter, A. F.; Ananiev, G. E.; Keck, J. L.; Hoffmann, F. M.; Wildman, S. A.; and Gitter, A. 2018. Practical model selection for prospective virtual screening. *bioRxiv* 337956.

Liu, S. 2018. Loss-balanced task weighting. `https://github.com/chao1224/Loss-Balanced-Task-Weighting`.

Wang, Y.; Bryant, S. H.; Cheng, T.; Wang, J.; Gindulyte, A.; Shoemaker, B. A.; Thiessen, P. A.; He, S.; and Zhang, J. 2017. PubChem BioAssay: 2017 update. *Nucleic Acids Research* 45(D):D955–D963.

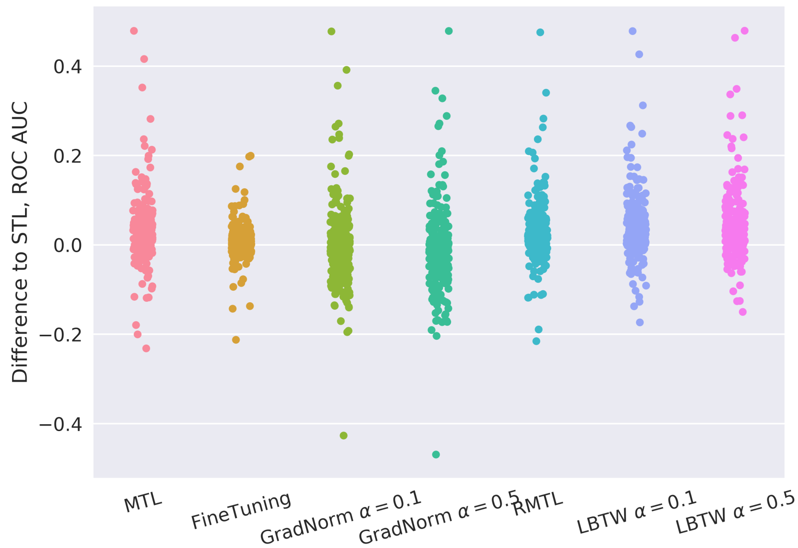

Figure S1: Distribution of the change in ROC AUC relative to STL for 128 tasks. Values below 0 indicate tasks with negative transfer.

Table S1: Mean ROC AUC on all 128 tasks and the number of tasks with negative transfer based on ROC AUC.

Evaluation Metric	STL	MTL	FineTuning	GradNorm ($\alpha = 0.1$)	GradNorm ($\alpha = 0.5$)	RMTL	LBTW ($\alpha = 0.1$)	LBTW ($\alpha = 0.5$)
Mean ROC AUC	0.799	0.857	0.806	0.840	0.833	0.852	0.859	0.863
# Negative Transfer	-	13	50	44	46	11	13	11

Table S2: PR AUC on all 128 PCBA tasks.

Task id	STL	MTL	FineTuning	GradNorm ($\alpha = 0.1$)	GradNorm ($\alpha = 0.5$)	RMTL	LBTW ($\alpha = 0.1$)	LBTW ($\alpha = 0.5$)
PCBA-1030	0.436	0.429	0.457	0.296	0.325	0.432	0.430	0.425
PCBA-1379	0.218	0.239	0.209	0.196	0.147	0.180	0.241	0.245
PCBA-1452	0.122	0.111	0.121	0.132	0.117	0.096	0.089	0.099
PCBA-1454	0.185	0.143	0.203	0.078	0.056	0.142	0.184	0.162
PCBA-1457	0.062	0.064	0.052	0.029	0.014	0.046	0.048	0.073
PCBA-1458	0.363	0.433	0.370	0.320	0.328	0.423	0.455	0.427
PCBA-1460	0.605	0.591	0.598	0.469	0.477	0.595	0.604	0.578
PCBA-1461	0.131	0.124	0.113	0.074	0.077	0.115	0.125	0.112
PCBA-1468	0.293	0.235	0.353	0.156	0.102	0.320	0.296	0.285
PCBA-1469	0.244	0.229	0.216	0.151	0.073	0.214	0.240	0.270
PCBA-1471	0.002	0.003	0.006	0.003	0.002	0.003	0.007	0.003
PCBA-1479	0.069	0.074	0.069	0.048	0.038	0.075	0.064	0.069
PCBA-1631	0.096	0.088	0.133	0.020	0.012	0.067	0.058	0.053
PCBA-1634	0.104	0.121	0.086	0.132	0.043	0.166	0.118	0.235
PCBA-1688	0.118	0.151	0.119	0.100	0.109	0.149	0.159	0.155
PCBA-1721	0.316	0.305	0.321	0.248	0.163	0.297	0.321	0.307
PCBA-2100	0.248	0.201	0.238	0.124	0.075	0.228	0.185	0.210
PCBA-2101	0.182	0.167	0.174	0.090	0.009	0.122	0.123	0.158
PCBA-2147	0.386	0.383	0.378	0.322	0.308	0.392	0.387	0.375
PCBA-2242	0.589	0.389	0.596	0.163	0.120	0.471	0.416	0.439
PCBA-2326	0.029	0.017	0.036	0.014	0.016	0.019	0.017	0.023
PCBA-2451	0.275	0.233	0.281	0.137	0.145	0.229	0.224	0.227
PCBA-2517	0.198	0.283	0.209	0.238	0.218	0.292	0.279	0.280

Table S2: PR AUC on all 128 PCBA tasks.

Task id	STL	MTL	FineTuning	GradNorm ($\alpha = 0.1$)	GradNorm ($\alpha = 0.5$)	RMTL	LBTW ($\alpha = 0.1$)	LBTW ($\alpha = 0.5$)
PCBA-2528	0.119	0.148	0.122	0.124	0.131	0.195	0.125	0.193
PCBA-2546	0.517	0.536	0.513	0.435	0.447	0.536	0.535	0.554
PCBA-2549	0.184	0.214	0.211	0.174	0.177	0.224	0.219	0.213
PCBA-2551	0.640	0.638	0.650	0.557	0.557	0.648	0.641	0.662
PCBA-2662	0.066	0.113	0.106	0.061	0.083	0.032	0.089	0.080
PCBA-2675	0.120	0.001	0.128	0.123	0.068	0.002	0.004	0.060
PCBA-2676	0.115	0.107	0.122	0.044	0.012	0.109	0.137	0.163
PCBA-411	0.323	0.470	0.364	0.373	0.387	0.460	0.496	0.491
PCBA-463254	0.003	0.021	0.009	0.028	0.007	0.054	0.032	0.022
PCBA-485281	0.010	0.008	0.006	0.004	0.002	0.003	0.012	0.033
PCBA-485290	0.095	0.151	0.096	0.082	0.079	0.133	0.130	0.137
PCBA-485294	0.105	0.047	0.222	0.185	0.033	0.144	0.042	0.215
PCBA-485297	0.483	0.527	0.482	0.373	0.393	0.515	0.530	0.527
PCBA-485313	0.459	0.496	0.466	0.346	0.358	0.494	0.507	0.502
PCBA-485314	0.268	0.329	0.305	0.259	0.234	0.308	0.319	0.315
PCBA-485341	0.028	0.026	0.026	0.025	0.025	0.034	0.031	0.033
PCBA-485349	0.009	0.033	0.007	0.011	0.011	0.062	0.028	0.026
PCBA-485353	0.109	0.071	0.099	0.099	0.067	0.098	0.064	0.098
PCBA-485360	0.195	0.276	0.215	0.218	0.205	0.250	0.262	0.280
PCBA-485364	0.450	0.464	0.467	0.336	0.332	0.460	0.454	0.454
PCBA-485367	0.146	0.113	0.183	0.057	0.025	0.172	0.137	0.108
PCBA-492947	0.000	0.003	0.023	0.012	0.008	0.004	0.004	0.013
PCBA-493208	0.250	0.241	0.244	0.227	0.213	0.236	0.147	0.214
PCBA-504327	0.108	0.159	0.116	0.140	0.117	0.126	0.165	0.160
PCBA-504332	0.526	0.527	0.533	0.427	0.439	0.533	0.524	0.525
PCBA-504333	0.534	0.570	0.555	0.438	0.465	0.556	0.562	0.562
PCBA-504339	0.542	0.555	0.550	0.464	0.469	0.552	0.557	0.557
PCBA-504444	0.235	0.244	0.243	0.145	0.151	0.243	0.260	0.248
PCBA-504466	0.276	0.353	0.280	0.217	0.228	0.321	0.348	0.347
PCBA-504467	0.296	0.311	0.291	0.203	0.224	0.313	0.298	0.298
PCBA-504706	0.002	0.017	0.017	0.017	0.011	0.020	0.016	0.014
PCBA-504842	0.013	0.030	0.101	0.072	0.147	0.007	0.009	0.157
PCBA-504845	0.004	0.023	0.010	0.009	0.034	0.020	0.023	0.024
PCBA-504847	0.287	0.312	0.301	0.158	0.153	0.310	0.312	0.302
PCBA-504891	0.000	0.163	0.125	0.097	0.265	0.064	0.111	0.463
PCBA-540276	0.230	0.312	0.238	0.217	0.213	0.278	0.320	0.308
PCBA-540317	0.202	0.255	0.222	0.210	0.184	0.250	0.251	0.271
PCBA-588342	0.686	0.667	0.689	0.515	0.558	0.666	0.665	0.674
PCBA-588453	0.372	0.419	0.372	0.266	0.230	0.411	0.388	0.393
PCBA-588456	0.099	0.311	0.108	0.298	0.308	0.207	0.362	0.387
PCBA-588579	0.321	0.353	0.304	0.315	0.272	0.367	0.364	0.357
PCBA-588590	0.270	0.302	0.289	0.261	0.241	0.294	0.299	0.296
PCBA-588591	0.372	0.380	0.378	0.314	0.295	0.386	0.381	0.369
PCBA-588795	0.261	0.332	0.254	0.266	0.198	0.292	0.336	0.335
PCBA-588855	0.256	0.302	0.265	0.219	0.205	0.293	0.294	0.306
PCBA-602179	0.003	0.010	0.006	0.007	0.007	0.020	0.022	0.026
PCBA-602233	0.031	0.063	0.083	0.105	0.104	0.133	0.159	0.127
PCBA-602310	0.037	0.069	0.031	0.022	0.010	0.040	0.057	0.049
PCBA-602313	0.047	0.079	0.056	0.062	0.023	0.067	0.086	0.100
PCBA-602332	0.000	0.002	0.000	0.001	0.002	0.001	0.001	0.002
PCBA-624170	0.035	0.069	0.048	0.050	0.042	0.062	0.075	0.043
PCBA-624171	0.103	0.102	0.096	0.046	0.035	0.114	0.088	0.106
PCBA-624173	0.367	0.363	0.390	0.378	0.304	0.393	0.429	0.424
PCBA-624202	0.083	0.116	0.080	0.074	0.072	0.102	0.115	0.114
PCBA-624246	0.053	0.066	0.053	0.228	0.060	0.054	0.197	0.169

Table S2: PR AUC on all 128 PCBA tasks.

Task id	STL	MTL	FineTuning	GradNorm ($\alpha = 0.1$)	GradNorm ($\alpha = 0.5$)	RMTL	LBTW ($\alpha = 0.1$)	LBTW ($\alpha = 0.5$)
PCBA-624287	0.013	0.017	0.008	0.011	0.002	0.009	0.014	0.042
PCBA-624288	0.025	0.039	0.022	0.019	0.019	0.049	0.038	0.036
PCBA-624291	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
PCBA-624296	0.247	0.305	0.258	0.193	0.207	0.279	0.295	0.291
PCBA-624297	0.214	0.256	0.218	0.135	0.147	0.228	0.255	0.223
PCBA-624417	0.175	0.172	0.183	0.112	0.120	0.181	0.199	0.179
PCBA-651635	0.332	0.386	0.341	0.244	0.244	0.363	0.370	0.383
PCBA-651644	0.230	0.231	0.229	0.118	0.075	0.205	0.176	0.185
PCBA-651768	0.249	0.288	0.206	0.243	0.212	0.268	0.293	0.295
PCBA-651965	0.187	0.159	0.170	0.071	0.070	0.169	0.147	0.144
PCBA-652025	0.038	0.069	0.021	0.013	0.006	0.036	0.051	0.091
PCBA-652104	0.127	0.128	0.133	0.073	0.063	0.130	0.124	0.111
PCBA-652105	0.369	0.361	0.391	0.287	0.271	0.362	0.364	0.346
PCBA-652106	0.025	0.024	0.053	0.008	0.007	0.014	0.017	0.026
PCBA-686970	0.241	0.298	0.252	0.203	0.202	0.270	0.319	0.314
PCBA-686978	0.713	0.734	0.721	0.602	0.644	0.724	0.732	0.734
PCBA-686979	0.642	0.683	0.661	0.543	0.584	0.671	0.681	0.680
PCBA-720504	0.257	0.265	0.254	0.169	0.170	0.263	0.248	0.255
PCBA-720532	0.538	0.570	0.576	0.490	0.490	0.539	0.566	0.545
PCBA-720542	0.221	0.258	0.191	0.152	0.099	0.243	0.260	0.261
PCBA-720551	0.080	0.095	0.076	0.035	0.029	0.086	0.113	0.114
PCBA-720553	0.133	0.155	0.125	0.082	0.059	0.124	0.158	0.162
PCBA-720579	0.197	0.216	0.200	0.162	0.144	0.206	0.227	0.218
PCBA-720580	0.245	0.261	0.233	0.172	0.149	0.260	0.279	0.268
PCBA-720707	0.086	0.069	0.058	0.061	0.012	0.169	0.124	0.084
PCBA-720708	0.036	0.041	0.038	0.025	0.022	0.032	0.054	0.046
PCBA-720709	0.066	0.123	0.074	0.109	0.081	0.087	0.165	0.136
PCBA-720711	0.065	0.122	0.061	0.080	0.014	0.132	0.152	0.134
PCBA-743255	0.286	0.276	0.309	0.198	0.129	0.295	0.256	0.247
PCBA-743266	0.003	0.001	0.004	0.001	0.001	0.001	0.001	0.001
PCBA-875	0.447	0.215	0.478	0.339	0.429	0.257	0.309	0.343
PCBA-881	0.383	0.290	0.370	0.291	0.212	0.323	0.295	0.292
PCBA-883	0.639	0.641	0.665	0.553	0.588	0.611	0.636	0.618
PCBA-884	0.860	0.830	0.873	0.792	0.824	0.847	0.828	0.826
PCBA-885	0.366	0.248	0.362	0.352	0.354	0.256	0.419	0.388
PCBA-887	0.196	0.178	0.193	0.136	0.139	0.170	0.177	0.165
PCBA-891	0.673	0.664	0.710	0.606	0.675	0.675	0.675	0.678
PCBA-899	0.711	0.692	0.721	0.666	0.698	0.734	0.695	0.716
PCBA-902	0.249	0.314	0.238	0.192	0.232	0.317	0.312	0.305
PCBA-903	0.292	0.369	0.350	0.237	0.272	0.390	0.403	0.455
PCBA-904	0.203	0.221	0.171	0.112	0.136	0.138	0.249	0.196
PCBA-912	0.106	0.143	0.108	0.101	0.123	0.085	0.102	0.127
PCBA-914	0.662	0.575	0.693	0.467	0.458	0.550	0.589	0.537
PCBA-915	0.372	0.274	0.435	0.287	0.301	0.301	0.349	0.317
PCBA-924	0.246	0.387	0.262	0.223	0.265	0.368	0.372	0.390
PCBA-925	0.220	0.041	0.077	0.028	0.073	0.005	0.093	0.094
PCBA-926	0.040	0.015	0.029	0.018	0.013	0.014	0.011	0.021
PCBA-927	0.056	0.013	0.029	0.010	0.024	0.008	0.014	0.009
PCBA-938	0.249	0.175	0.250	0.153	0.124	0.206	0.183	0.188
PCBA-995	0.054	0.070	0.075	0.058	0.056	0.062	0.088	0.081

Table S3: ROC AUC on all 128 PCBA tasks

Task id	STL	MTL	FineTuning	GradNorm ($\alpha = 0.1$)	GradNorm ($\alpha = 0.5$)	RMTL	LBTW ($\alpha = 0.1$)	LBTW ($\alpha = 0.5$)
PCBA-1030	0.801	0.799	0.806	0.731	0.755	0.803	0.798	0.792
PCBA-1379	0.893	0.951	0.893	0.944	0.930	0.944	0.954	0.958
PCBA-1452	0.805	0.908	0.750	0.837	0.873	0.863	0.902	0.908
PCBA-1454	0.774	0.898	0.765	0.869	0.883	0.903	0.901	0.900
PCBA-1457	0.800	0.848	0.813	0.853	0.786	0.852	0.880	0.856
PCBA-1458	0.884	0.906	0.880	0.877	0.883	0.902	0.908	0.907
PCBA-1460	0.937	0.941	0.939	0.920	0.922	0.944	0.947	0.940
PCBA-1461	0.816	0.856	0.802	0.833	0.837	0.847	0.867	0.858
PCBA-1468	0.886	0.906	0.904	0.886	0.855	0.912	0.903	0.912
PCBA-1469	0.852	0.936	0.915	0.956	0.944	0.962	0.945	0.953
PCBA-1471	0.524	0.544	0.470	0.581	0.560	0.595	0.599	0.553
PCBA-1479	0.710	0.683	0.698	0.706	0.699	0.717	0.697	0.730
PCBA-1631	0.824	0.824	0.830	0.800	0.769	0.853	0.834	0.810
PCBA-1634	0.910	0.942	0.895	0.947	0.950	0.942	0.935	0.958
PCBA-1688	0.733	0.810	0.720	0.758	0.768	0.792	0.809	0.804
PCBA-1721	0.914	0.954	0.924	0.945	0.939	0.947	0.947	0.951
PCBA-2100	0.880	0.880	0.885	0.858	0.835	0.888	0.890	0.884
PCBA-2101	0.782	0.879	0.778	0.872	0.806	0.878	0.859	0.875
PCBA-2147	0.846	0.870	0.854	0.836	0.842	0.872	0.869	0.865
PCBA-2242	0.955	0.958	0.933	0.934	0.907	0.967	0.963	0.965
PCBA-2326	0.728	0.751	0.746	0.747	0.742	0.752	0.749	0.753
PCBA-2451	0.882	0.894	0.876	0.867	0.878	0.902	0.900	0.893
PCBA-2517	0.846	0.890	0.867	0.868	0.868	0.888	0.892	0.887
PCBA-2528	0.855	0.923	0.853	0.920	0.892	0.921	0.904	0.934
PCBA-2546	0.882	0.904	0.884	0.871	0.884	0.905	0.912	0.914
PCBA-2549	0.804	0.865	0.810	0.848	0.846	0.865	0.868	0.872
PCBA-2551	0.913	0.920	0.915	0.893	0.899	0.921	0.922	0.922
PCBA-2662	0.686	0.907	0.712	0.921	0.866	0.895	0.880	0.856
PCBA-2675	0.633	0.687	0.608	0.737	0.741	0.634	0.763	0.827
PCBA-2676	0.836	0.872	0.839	0.837	0.766	0.892	0.887	0.890
PCBA-411	0.882	0.933	0.876	0.905	0.916	0.922	0.938	0.934
PCBA-463254	0.566	0.693	0.590	0.805	0.724	0.719	0.720	0.803
PCBA-485281	0.705	0.742	0.676	0.698	0.670	0.707	0.721	0.726
PCBA-485290	0.781	0.833	0.776	0.821	0.810	0.812	0.841	0.836
PCBA-485294	0.578	0.859	0.777	0.842	0.866	0.860	0.890	0.868
PCBA-485297	0.932	0.944	0.929	0.914	0.921	0.941	0.944	0.948
PCBA-485313	0.930	0.947	0.930	0.918	0.922	0.944	0.946	0.950
PCBA-485314	0.851	0.883	0.866	0.850	0.851	0.880	0.874	0.874
PCBA-485341	0.725	0.730	0.728	0.742	0.701	0.746	0.740	0.753
PCBA-485349	0.655	0.749	0.578	0.741	0.744	0.749	0.751	0.747
PCBA-485353	0.713	0.774	0.726	0.790	0.774	0.764	0.777	0.771
PCBA-485360	0.829	0.905	0.858	0.875	0.870	0.893	0.899	0.908
PCBA-485364	0.887	0.885	0.889	0.847	0.851	0.889	0.882	0.884
PCBA-485367	0.847	0.835	0.844	0.829	0.736	0.876	0.843	0.850
PCBA-492947	0.463	0.879	0.660	0.854	0.808	0.803	0.889	0.799
PCBA-493208	0.733	0.781	0.778	0.795	0.777	0.808	0.778	0.775
PCBA-504327	0.715	0.844	0.761	0.828	0.820	0.854	0.868	0.849
PCBA-504332	0.842	0.842	0.845	0.790	0.803	0.845	0.842	0.843
PCBA-504333	0.892	0.909	0.900	0.869	0.878	0.910	0.908	0.910
PCBA-504339	0.870	0.885	0.875	0.854	0.858	0.884	0.888	0.885
PCBA-504444	0.805	0.854	0.804	0.798	0.807	0.845	0.856	0.855
PCBA-504466	0.895	0.927	0.894	0.898	0.899	0.918	0.923	0.930
PCBA-504467	0.851	0.876	0.847	0.834	0.843	0.875	0.872	0.870
PCBA-504706	0.687	0.923	0.702	0.852	0.873	0.880	0.911	0.902
PCBA-504842	0.604	0.777	0.779	0.875	0.875	0.775	0.800	0.824

Table S3: ROC AUC on all 128 PCBA tasks

Task id	STL	MTL	FineTuning	GradNorm ($\alpha = 0.1$)	GradNorm ($\alpha = 0.5$)	RMTL	LBTW ($\alpha = 0.1$)	LBTW ($\alpha = 0.5$)
PCBA-504845	0.883	0.926	0.905	0.919	0.903	0.863	0.899	0.918
PCBA-504847	0.855	0.884	0.863	0.844	0.837	0.873	0.886	0.890
PCBA-504891	0.520	0.999	0.507	0.998	0.999	0.996	0.998	0.999
PCBA-540276	0.811	0.880	0.813	0.837	0.847	0.865	0.887	0.880
PCBA-540317	0.905	0.933	0.899	0.917	0.920	0.938	0.937	0.937
PCBA-588342	0.933	0.930	0.936	0.889	0.903	0.932	0.929	0.931
PCBA-588453	0.894	0.922	0.896	0.888	0.885	0.921	0.917	0.915
PCBA-588456	0.715	0.914	0.767	0.962	0.871	0.951	0.963	0.955
PCBA-588579	0.913	0.931	0.908	0.921	0.911	0.926	0.931	0.930
PCBA-588590	0.843	0.880	0.845	0.853	0.843	0.868	0.875	0.872
PCBA-588591	0.916	0.923	0.915	0.905	0.896	0.927	0.924	0.925
PCBA-588795	0.933	0.948	0.919	0.921	0.912	0.942	0.939	0.946
PCBA-588855	0.852	0.901	0.853	0.869	0.876	0.903	0.904	0.906
PCBA-602179	0.612	0.804	0.649	0.815	0.812	0.819	0.824	0.858
PCBA-602233	0.826	0.965	0.888	0.951	0.960	0.937	0.973	0.961
PCBA-602310	0.837	0.904	0.700	0.886	0.893	0.905	0.891	0.889
PCBA-602313	0.860	0.904	0.838	0.886	0.866	0.908	0.917	0.909
PCBA-602332	0.587	0.642	0.493	0.612	0.718	0.585	0.654	0.715
PCBA-624170	0.742	0.884	0.816	0.836	0.835	0.869	0.857	0.876
PCBA-624171	0.906	0.911	0.900	0.884	0.877	0.918	0.910	0.916
PCBA-624173	0.876	0.925	0.888	0.925	0.928	0.929	0.931	0.944
PCBA-624202	0.853	0.894	0.849	0.861	0.868	0.891	0.893	0.896
PCBA-624246	0.431	0.783	0.407	0.787	0.758	0.693	0.697	0.779
PCBA-624287	0.769	0.793	0.737	0.740	0.600	0.773	0.786	0.787
PCBA-624288	0.699	0.797	0.705	0.761	0.762	0.785	0.794	0.784
PCBA-624291	0.567	0.497	0.597	0.536	0.569	0.512	0.505	0.519
PCBA-624296	0.808	0.845	0.810	0.811	0.816	0.837	0.845	0.845
PCBA-624297	0.771	0.820	0.787	0.785	0.792	0.812	0.823	0.825
PCBA-624417	0.841	0.867	0.852	0.828	0.834	0.861	0.877	0.866
PCBA-651635	0.882	0.919	0.877	0.897	0.897	0.915	0.921	0.924
PCBA-651644	0.886	0.945	0.890	0.928	0.928	0.943	0.936	0.947
PCBA-651768	0.874	0.929	0.883	0.914	0.905	0.919	0.932	0.928
PCBA-651965	0.781	0.766	0.778	0.704	0.706	0.773	0.762	0.762
PCBA-652025	0.742	0.893	0.803	0.900	0.861	0.870	0.916	0.879
PCBA-652104	0.825	0.828	0.822	0.790	0.778	0.837	0.835	0.824
PCBA-652105	0.882	0.897	0.890	0.871	0.870	0.897	0.902	0.893
PCBA-652106	0.729	0.784	0.688	0.784	0.776	0.801	0.818	0.816
PCBA-686970	0.834	0.887	0.841	0.843	0.855	0.878	0.888	0.893
PCBA-686978	0.885	0.892	0.890	0.835	0.854	0.889	0.894	0.893
PCBA-686979	0.876	0.890	0.879	0.833	0.850	0.887	0.891	0.891
PCBA-720504	0.730	0.734	0.735	0.708	0.709	0.738	0.737	0.734
PCBA-720532	0.833	0.876	0.871	0.829	0.840	0.858	0.875	0.875
PCBA-720542	0.875	0.946	0.851	0.910	0.915	0.918	0.944	0.948
PCBA-720551	0.690	0.712	0.707	0.709	0.697	0.721	0.717	0.720
PCBA-720553	0.744	0.759	0.744	0.753	0.754	0.754	0.760	0.752
PCBA-720579	0.784	0.862	0.811	0.848	0.853	0.871	0.864	0.879
PCBA-720580	0.857	0.875	0.853	0.866	0.869	0.880	0.883	0.892
PCBA-720707	0.666	0.760	0.706	0.770	0.730	0.777	0.752	0.795
PCBA-720708	0.830	0.864	0.821	0.839	0.814	0.843	0.843	0.838
PCBA-720709	0.883	0.939	0.878	0.924	0.907	0.916	0.948	0.939
PCBA-720711	0.780	0.913	0.866	0.907	0.864	0.888	0.897	0.897
PCBA-743255	0.888	0.921	0.896	0.905	0.892	0.914	0.919	0.922
PCBA-743266	0.520	0.512	0.570	0.550	0.546	0.478	0.547	0.556
PCBA-875	0.800	0.936	0.900	0.921	0.870	0.894	0.914	0.891
PCBA-881	0.858	0.934	0.851	0.921	0.941	0.949	0.942	0.955

Table S3: ROC AUC on all 128 PCBA tasks

Task id	STL	MTL	FineTuning	GradNorm ($\alpha = 0.1$)	GradNorm ($\alpha = 0.5$)	RMTL	LBTW ($\alpha = 0.1$)	LBTW ($\alpha = 0.5$)
PCBA-883	0.835	0.851	0.856	0.791	0.807	0.832	0.834	0.832
PCBA-884	0.910	0.893	0.916	0.868	0.889	0.903	0.896	0.892
PCBA-885	0.903	0.850	0.932	0.906	0.816	0.827	0.895	0.872
PCBA-887	0.718	0.807	0.764	0.785	0.781	0.806	0.777	0.792
PCBA-891	0.830	0.849	0.854	0.831	0.858	0.851	0.853	0.856
PCBA-899	0.837	0.855	0.838	0.830	0.844	0.856	0.851	0.854
PCBA-902	0.826	0.899	0.819	0.867	0.874	0.892	0.893	0.898
PCBA-903	0.859	0.938	0.810	0.888	0.895	0.919	0.938	0.931
PCBA-904	0.783	0.897	0.747	0.827	0.853	0.854	0.892	0.889
PCBA-912	0.715	0.840	0.688	0.819	0.828	0.848	0.843	0.867
PCBA-914	0.879	0.902	0.903	0.888	0.865	0.893	0.925	0.912
PCBA-915	0.755	0.812	0.842	0.806	0.807	0.800	0.818	0.828
PCBA-924	0.785	0.921	0.876	0.895	0.906	0.924	0.930	0.936
PCBA-925	0.860	0.884	0.648	0.881	0.926	0.851	0.870	0.952
PCBA-926	0.666	0.723	0.695	0.726	0.699	0.726	0.662	0.715
PCBA-927	0.816	0.700	0.731	0.744	0.696	0.705	0.725	0.753
PCBA-938	0.791	0.810	0.792	0.783	0.770	0.804	0.795	0.797
PCBA-995	0.761	0.796	0.770	0.781	0.803	0.798	0.796	0.808