Massively Multitask Networks for Drug Discovery

Ramsundar et al. (2015)

What is Drug Discovery?

1. Hit finding: screen drug-like compounds in an effort to find a few attractive molecules for further optimization
2. ML goal: predict interactions between targets and small molecules

Dataset	Actives	Inactives	Target Class	Target
dude-igf1r	148	9298	other receptor	Insulin-like growth factor I receptor
dude-inha	43	2300	other enzyme	Enoyl-[acyl-carrier-protein] reductase
dude-ital	138	8498	miscellaneous	Leukocyte adhesion glycoprotein LFA-1 alpha

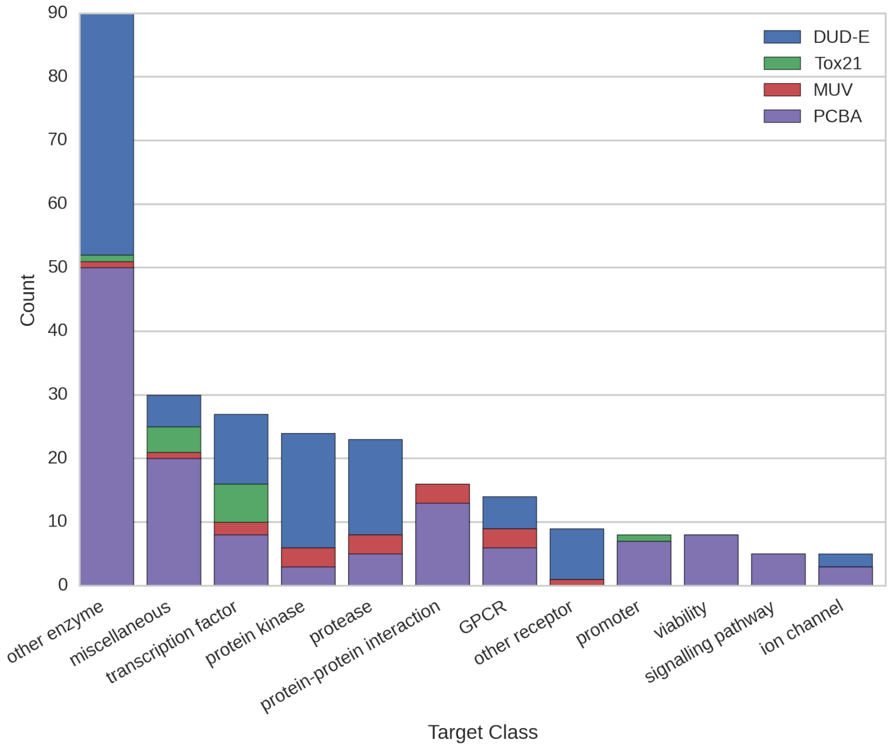
Motivation & Problem

1. Highly imbalanced datasets
   a. 1-2% of screened compounds are active against a given target

Group	Datasets	Data Points / ea.	% Active
PCBA	128	282K (122K)	1.8 (3.8)
DUD-E	102	14K (11K)	1.6 (0.2)
MUV	17	15K (1)	0.2 (0)
Tox21	12	6K (500)	7.8 (4.7)

Motivation & Problem

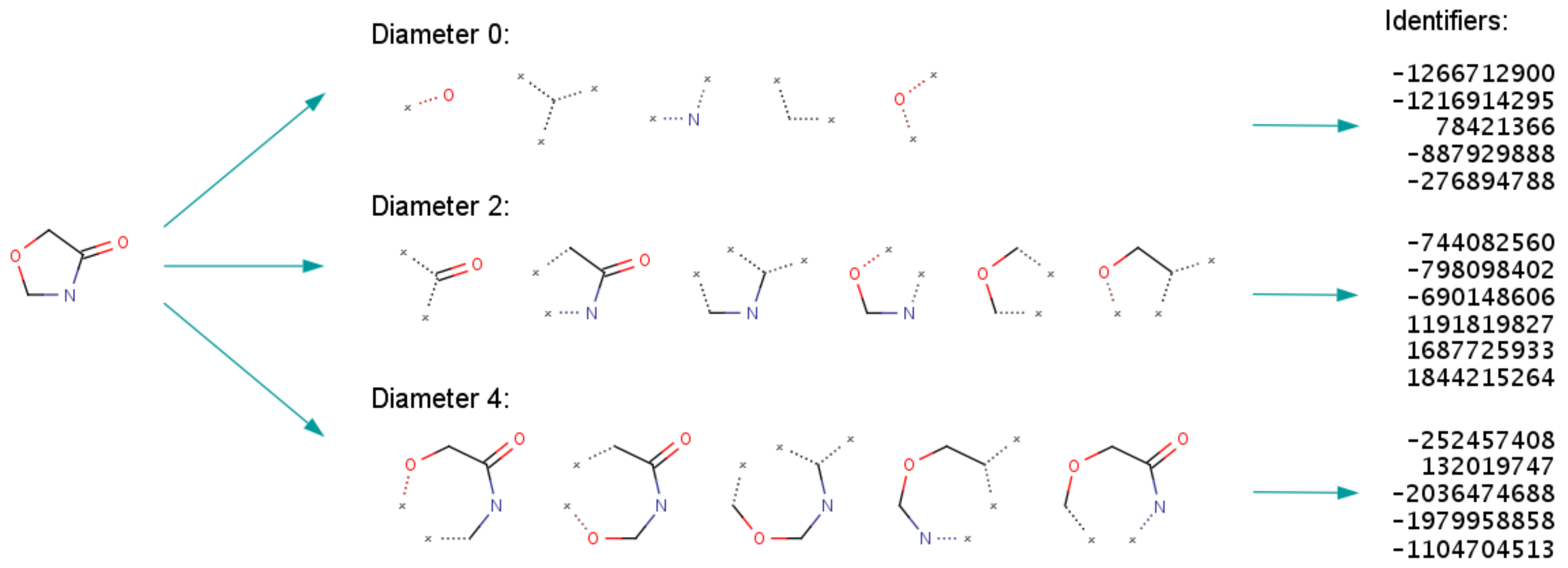
1. Highly imbalanced datasets
    a. 1-2% of screened compounds are active against a given target
2. Disparate sources of experimental data across multiple targets
    a. 259 datasets
    b. 37.8M experimental data points
    c. 1.6M compounds
    d. 249 tasks



Motivation & Problem

1. Highly imbalanced datasets
   a. 1-2% of screened compounds are active against a given target
2. Disparate sources of experimental data across multiple targets
   a. 259 datasets
   b. 37.8M experimental data points
   c. 1.6M compounds
   d. 249 tasks
3. Prior work unclear whether multitask learning beneficial in drug discovery
   a. Dahl (2012), Lowe (2012): Too small sample size and gains in predictive accuracy too small to justify increase in complexity
   b. Unterthiner et al.: Performance gains due to multitask networks
   c. Erhan et al. (2006): Multitask networks did not consistently outperform singletask networks

Method Overview



Method Overview

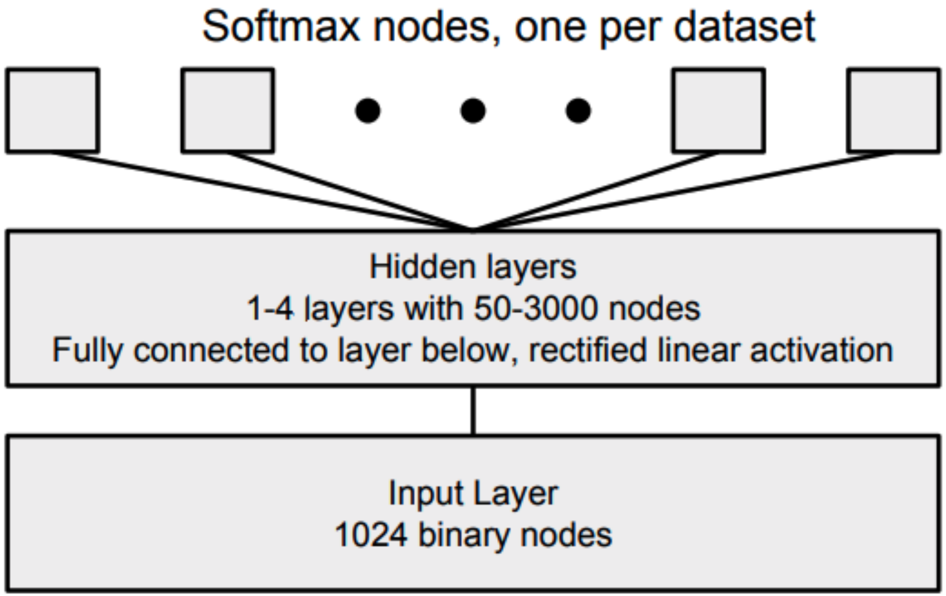


Figure 1. Multitask neural network.

Experiments

1) How do multitask neural nets perform relative to baselines?
2) How does adding more tasks effect accuracy?
3) Would we rather have more tasks or more examples?
4) How does adding more tasks effect pre-training accuracy?
5) When do datasets benefit from multitask training?

Experiment 1: How do multitask neural nets perform relative to baselines?

Model	PCBA ($n = 128$)	MUV ($n = 17$)	Tox21 ($n = 12$)
Logistic Regression (LR)	.801	.752	.738
Random Forest (RF)	.800	.774	.790
Single-Task Neural Net (STNN)	.795	.732	.714
Pyramidal (2000, 100) STNN (PSTNN)	.809	.745	.740
Max{LR, RF, STNN, PSTNN}	.824	.781	.790
1-Hidden (1200) Layer Multitask Neural Net (MTNN)	.842	.797	.785
Pyramidal (2000, 100) Multitask Neural Net (PMTNN)	**.873**	**.841**	**.818**

Experiment 1: How do multitask neural nets perform relative to baselines?

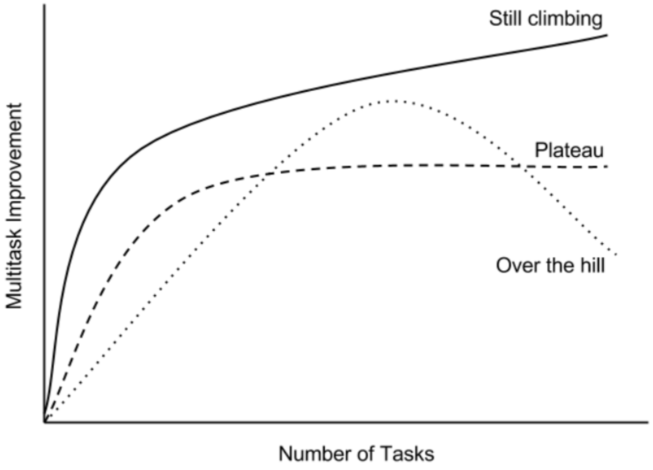
Model	PCBA ($n = 128$)	MUV ($n = 17$)	Tox21 ($n = 12$)
Logistic Regression (LR)	.801	.752	.738
Random Forest (RF)	.800	.774	.790
Single-Task Neural Net (STNN)	.795	.732	.714
Pyramidal (2000, 100) STNN (PSTNN)	.809	.745	.740
Max{LR, RF, STNN, PSTNN}	.824	.781	.790
1-Hidden (1200) Layer Multitask Neural Net (MTNN)	.842	.797	.785
Pyramidal (2000, 100) Multitask Neural Net (PMTNN)	**.873**	**.841**	**.818**

Experiment 1: How do multitask neural nets perform relative to baselines?

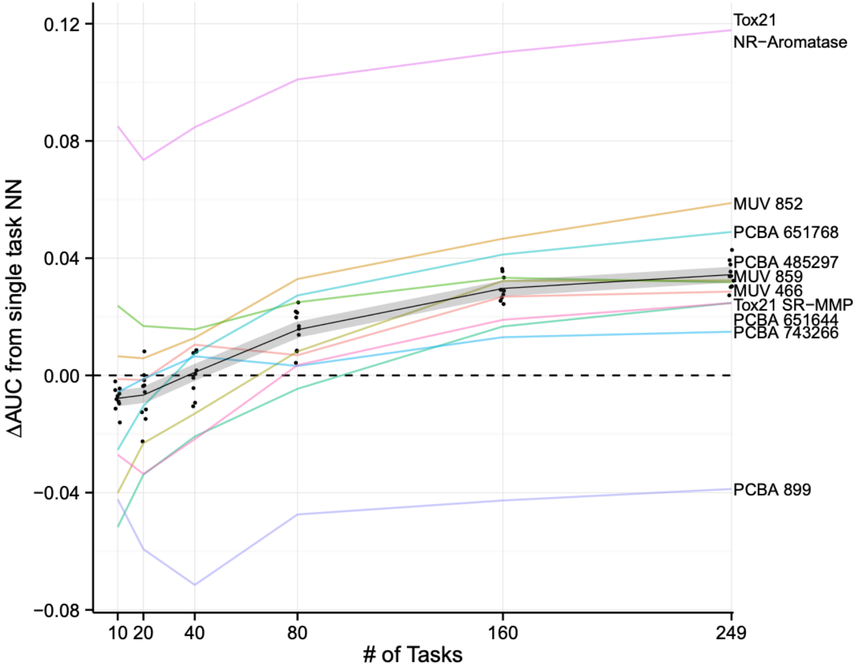
Model	PCBA ($n = 128$)	MUV ($n = 17$)	Tox21 ($n = 12$)
Logistic Regression (LR)	.801	.752	.738
Random Forest (RF)	.800	.774	.790
Single-Task Neural Net (STNN)	.795	.732	.714
Pyramidal (2000, 100) STNN (PSTNN)	.809	.745	.740
Max{LR, RF, STNN, PSTNN}	.824	.781	.790
1-Hidden (1200) Layer Multitask Neural Net (MTNN)	.842	.797	.785
Pyramidal (2000, 100) Multitask Neural Net (PMTNN)	**.873**	**.841**	**.818**

Experiment 2: How does adding more tasks effect accuracy?

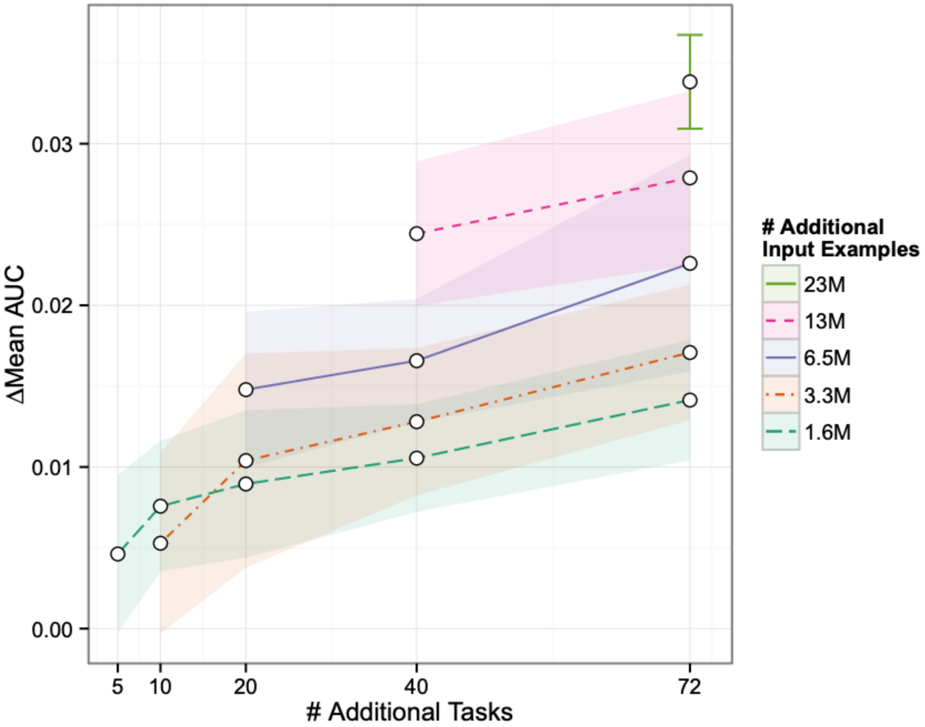
- Train models for 10 “held-in” tasks and variable number of additional randomly sampled tasks
- Observe accuracy as function of number of additional tasks
- Three possibilities



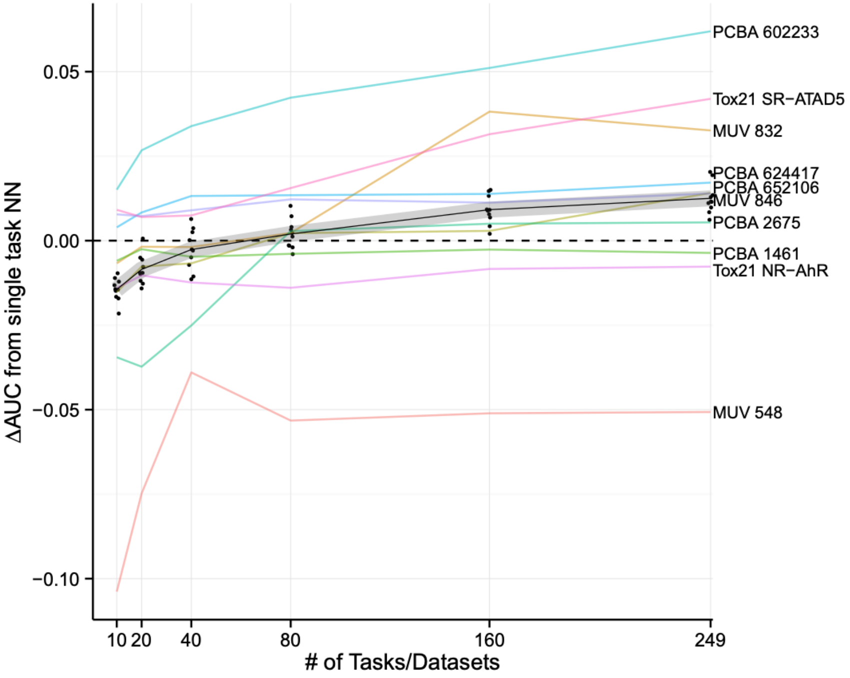
Experiment 2: How does adding more tasks effect accuracy?



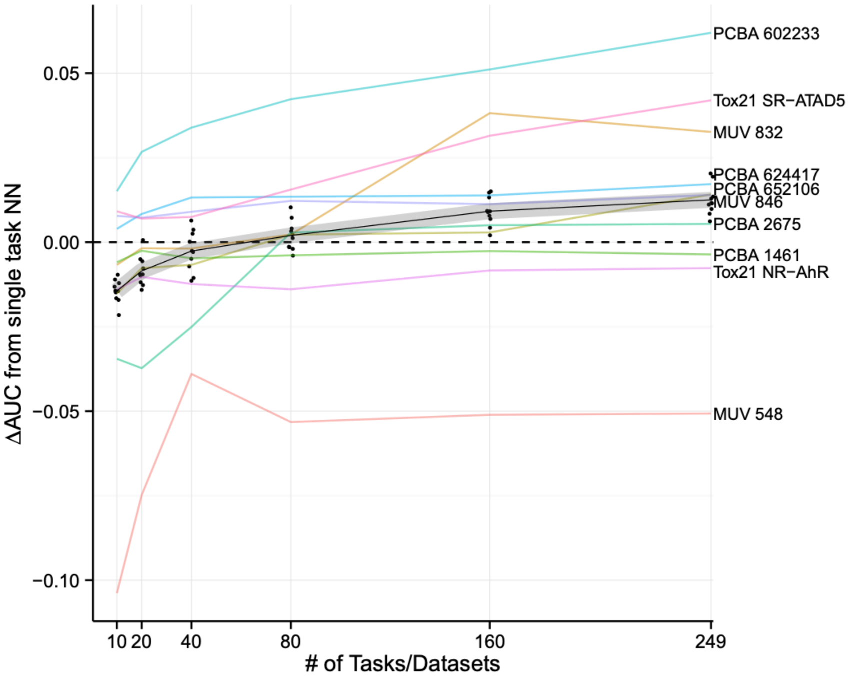
Experiment 3: Would we rather have more tasks or more examples?



Experiment 4: How does adding more tasks effect pre-training accuracy?



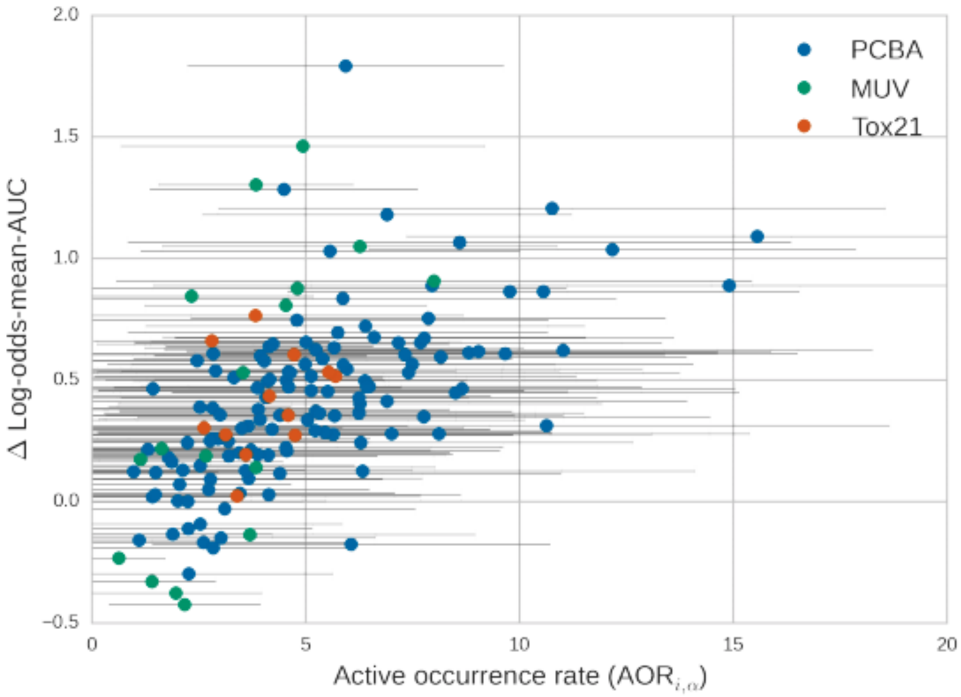
Experiment 4: How does adding more tasks effect pre-training accuracy?



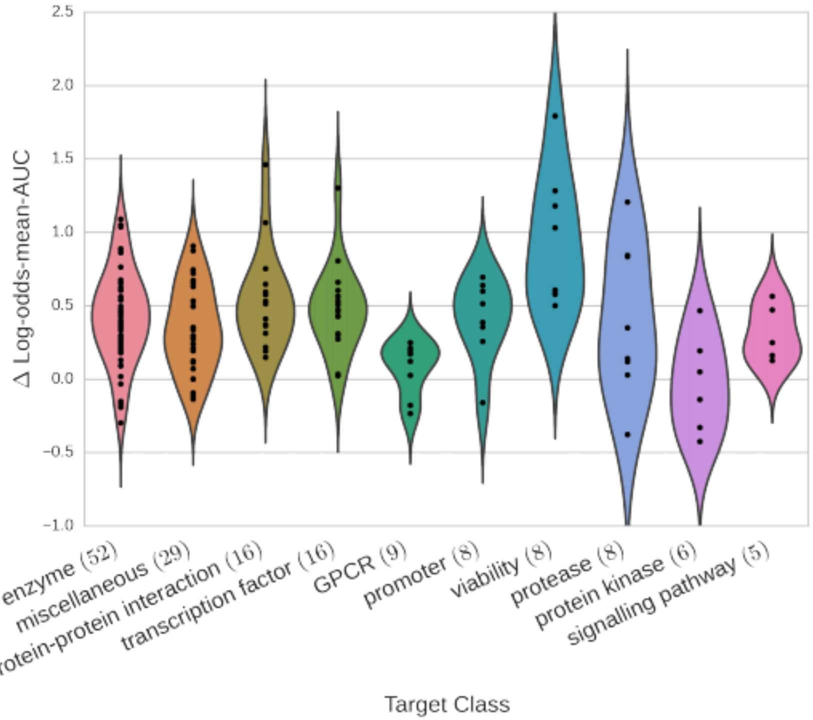
stronger as multitask networks were trained on more data. Large multitask networks exhibited better transferability, but the average effect even with 249 datasets was only $\sim$ .01 AUC. We hypothesize that the extent of this generalizability is determined by the presence or absence of relevant data in the multitask training set.

Experiment 5: When do datasets benefit from multitask training?

$$\mathrm{AOR}_{i,\alpha} = \sum_{d \neq i} \mathbb{1}\left(\alpha \in \mathrm{Actives}\left(D_d\right)\right)$$



Experiment 5: When do datasets benefit from multitask training?



Strengths

1. Empirical analysis on real world data
2. Challenging problem with extreme data skew (1-2% of screened compounds are activate against a given target)
3. Simple network for simple analysis
4. Exploring under what conditions multitask learning produces positive and negative results
5. Achieve results outperforming other approaches to the task

Weaknesses

1. Confound between data size and number of tasks
2. No clear analysis of when not to use multitask learning
3. Could have explored other architectures

Potential Improvements

1. More theoretical results on task overlap, covariance analysis
2. Comparison of models trained on related categories of tasks vs all tasks
3. Control training set size vs. number of tasks
4. Compare different architectures
5. Have benchmark comparisons against models from related papers

Takeaways

1. Multitask learning can yield superior results to singletask learning
2. Limited transferability to tasks not contained in training set
3. Multitask effect stronger for some datasets than others
4. Presence of shared active compounds moderately correlated with multitask improvement
5. Efficacy of multitask learning directly related to availability of relevant data

Questions?