Chemical Property Prediction via Graph Knowledge Transfer

Zhenbang Wu[12*], Haonan Wang[2*], Yizhou Sun[3]
[1] Zhejiang University
[2] University of Illinois at Urbana-Champaign
[3] University of California, Los Angeles
* Equal Contribution

MOTIVATION

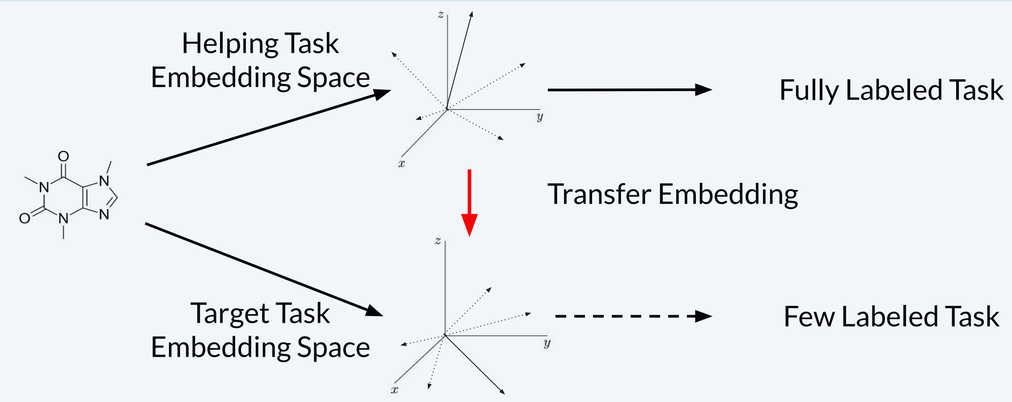
- Current deep learning frameworks require large amounts of data.
- In chemical and biomedical domain, the cost of fully labeling a dataset is usually unaffordable [1].
- Because measurements of some properties are more expensive than the others, label ratio between properties is usually imbalanced [2].

Insight: Leveraging chemical graph structure, knowledge extracted from fully labeled properties can be efficiently transferred to enhance the representation learning of properties with few labels.

ToxCast dataset. x: no label.

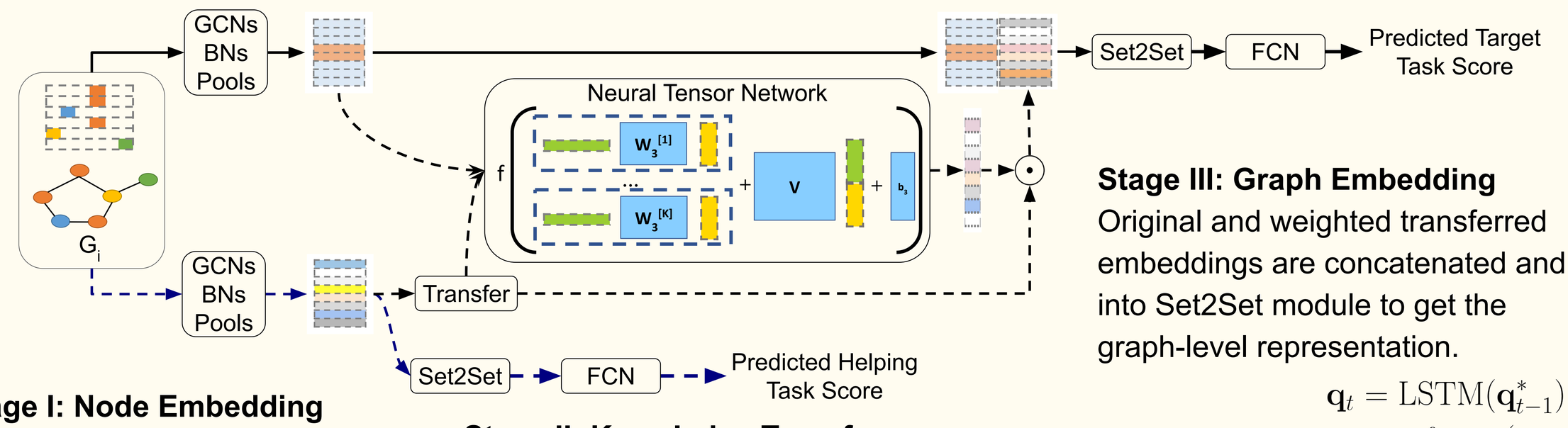
Property	ESRE BLA	APR HepG2
Mol 1	1	x
Mol 2	0	0
Mol 3	0	x
Mol 4	1	1
Mol 5	1	x
…	…	…
Mol 8597	0	x
Mol 8598	1	x
Label Ratio	0.84	0.12

METHOD



Our model **transfers** and **fuses** the knowledge between tasks to enhance the performance on target task.
Chemical graph will first be mapped to task-specific embedding space separately. Then, node embeddings of helping task will be transferred by a powerful transfer layer (red arrow) to the target task embedding space.

MODEL DETAIL



Stage I: Node Embedding

$$\mathrm{Conv}(A, X) = \hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}X\Theta$$

$$\mathrm{BN}(X) = \frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}[X] + \epsilon}} * \gamma + \beta$$

$$\mathrm{Pool}(v) = \max\{\max_{(u,v)\in E}\{u, v\}\}$$

Stage II: Knowledge Transfer

Two linear layers with ReLU activation transfer the node embeddings from helping task to target task.
Neural Tensor Network models the node-level interaction and decides the transfer weights.

Stage III: Graph Embedding

Original and weighted transferred embeddings are concatenated and fed into Set2Set module to get the graph-level representation.

$$\mathbf{q}_t = \mathrm{LSTM}(\mathbf{q}^*_{t-1})$$

$$\alpha_{i,t} = \mathrm{softmax}(\mathbf{x}_i \cdot \mathbf{q}_t)$$

$$\mathbf{r}_t = \sum_{i=1}^{N} \alpha_{i,t}\mathbf{x}_i$$

$$\mathbf{q}^*_t = \mathbf{q}_t \,\|\, \mathbf{r}_t,$$

RESULTS

Model	TOX21 Target Task (10% Training)	TOX21 Helping Task (90% Training)	SIDER Target Task (10% Training)	SIDER Helping Task (90% Training)
Single-task Model				
GCN [3]	0.6776	0.8638	0.5938	0.6266
MoleculeNet [4]	0.7156	0.8315	0.6189	0.6294
Our	0.7385	0.9096	0.6266	**0.8212**
Multi-task Model				
MoleculeNet [4]	0.7298	0.8382	0.6315	0.6503
Our	**0.7762**	**0.9233**	**0.6569**	0.8037

We select SR-ARE (TOX21) and Investigations (SIDER) as target tasks; SR-MMP (TOX21) and Vascular Disorders (SIDER) as helping tasks.

CONCLUSION

In this work, we make the following contributions:

- Identify imbalanced labeling issues on chemical and biomedical tasks.
- Introduce novel graph neural network to extract, transfer and fuse knowledge between tasks.
- Improve AUC-ROC score by 6.9% without fine tuning.

ACKNOWLEDGEMENTS

We would like to thank Yunsheng Bai for helpful discussions and useful poster templates.

REFERENCE

[1] H. Altae-Tran, B. Ramsundar, A. S. Pappu, V. Pande. Low Data Drug Discovery with One-Shot Learning. ACS Cent. Sci. 2017, 3 (4), 283−293.
[2] W. Lin, D. Xu, Imbalanced multi-label learning for identifying antimicrobial peptides and their functional types. Bioinformatics, Volume 32, Issue 24, 15 December 2016, Pages 3745.
[3] Kipf, T. N., and Welling, M. Semi-supervised classification with graph convolutional networks, 2016. arXiv preprint arXiv:1609.02907.
[4] Z. Wu, B. Ramsundar, E. N. Feinberg; J. Gomes,C. Geniesse, A. S. Pappu, K. Leswing, V. Pande. MoleculeNet: A Benchmark for Molecular Machine Learning, 2017. https://arxiv.org/abs/1703.00564.