De novo Prediction of Cell-Drug Sensitivities Using Deep Learning-based Graph Regularized Matrix Factorization

Shuangxia Ren[1], Yifeng Tao[2], Ke Yu[1], Yifan Xue[3], Russell Schwartz[2], Xinghua Lu[1,3]
[1]Intelligent Systems Program, and [3]Department of Biomedical Informatics, University of Pittsburgh
[2]Computational Biology Department, Carnegie Mellon University

University of Pittsburgh

Overview

- The success of **precision oncology** requires the capability to accurately predict the drug sensitivity of:
  1) new patient's cancer cell lines to **existing** anticancer drugs
  2) new patient's cancer cell lines to **new** anticancer drugs

- We demonstrate that our model (DeepGRMF)
  1) shows its superiority in **predicting drug sensitivity** in GDSC and CCLE dataset compared with competing models
  2) could **predict effectiveness of a chemotherapy regimen on patien**t outcomes for the lung cancer patients in TCGA dataset

Key Components:

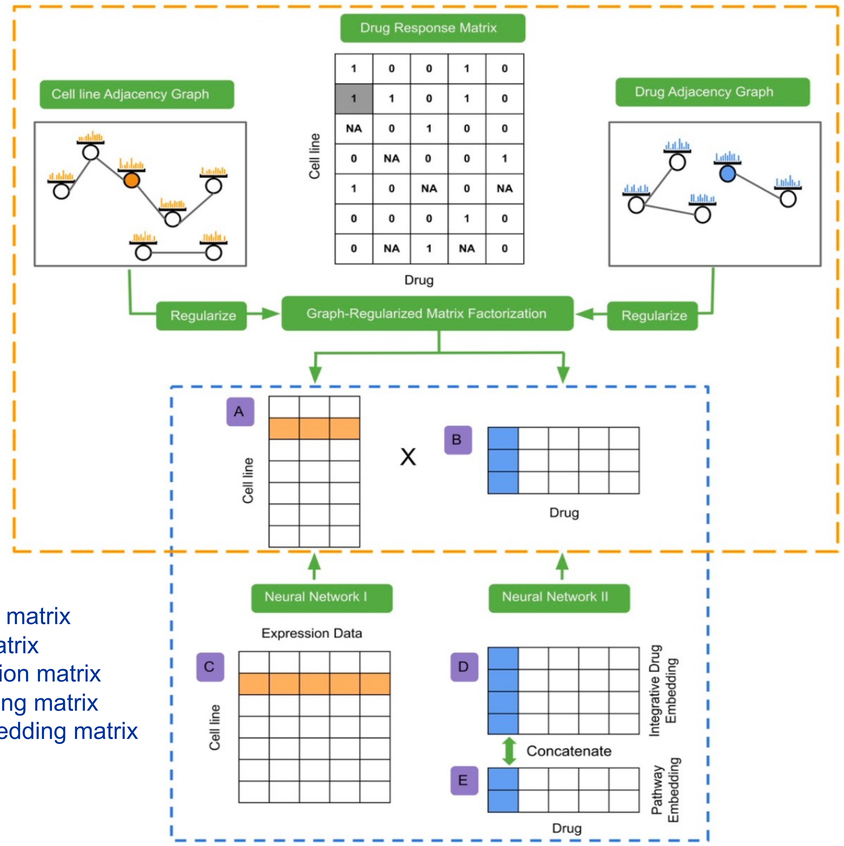
1) **Integrative drug embedding** that incorporates drug chemical structures, mechanisms of action (MOAs) and pathway information;

2) **Matrix-factorization-based collaborative filtering** that captures characteristic interactions between a set of similar cell lines and a set of similar drugs.

3) **Graph-based regularization** that encodes the similarity of cells and drugs in original input space.

4) **Neural networks** that accurately **map a new input (a new cell or a drug) to factor space**, thus performing prediction of responses between a pair of previously unseen cell and drug.

Method



A: cell line factor matrix
B: drug factor matrix
C: gene expression matrix
D: drug embedding matrix
E: pathway embedding matrix

Module I: Graph-regularized matrix factorization (orange dotted box)

$$\mathcal{L}_{i,j}(A_i, B_j; y_{i,j}, \vec{W}_{\text{cell}}, \vec{W}_{\text{drug}}) = \text{CrossEnt}(\hat{y}_{i,j}, y_{i,j}) + \lambda_c \sum_{r=1}^{n} (\vec{W}_{\text{cell}})_{i,r} \cdot \|\vec{A}_i - \vec{A}_r\|^2 + \lambda_d \sum_{k=1}^{m} (\vec{W}_{\text{drug}})_{j,k} \cdot \|\vec{B}_j - \vec{B}_k\|^2,$$

Where $\vec{W}_{cell}$ is an adjacency matrix for cell lines and $\vec{W}_{drug}$ is an adjacency matrix for drugs

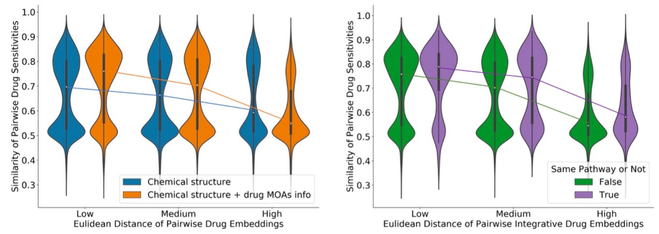
Module II: Using neural networks to learn mapping functions (blue dotted box)

$$\mathcal{L}(\theta; \vec{C}_i) = \sum_{i=1}^{n} \|f_\theta(\vec{C}_i) - \vec{A}_i\|^2 \qquad \text{(Neural Network I)}$$

$$\mathcal{L}(\gamma, \phi; \vec{D}_j, \vec{P}_j) = \sum_{j=1}^{m} \|f_\phi(\vec{D}_j \cup f_\gamma(\vec{P}_j)) - \vec{B}_j\|^2 \qquad \text{(Neural Network II)}$$

We adopted an embedding layer (denoted as fγ) to convert a drug's pathway information ($\vec{P}_j$) into its pathway embedding ($E_j = f\gamma(P_j)$).

Results



We observed a decreasing trend (the blue curve) of drug sensitivity similarity from the low to the high quantiles of Euclidean distance in the drug embedding space using only chemical structure information. This trend (the orange curve) becomes more evident by using the MOAs as a regularization, suggesting that the drug embedding is augmented by adding the MOAs information. In addition, including the pathway information into the IDE further boosted its quality.

Table 1: Performance of different models to predict drug response of new cell lines to existing drugs.

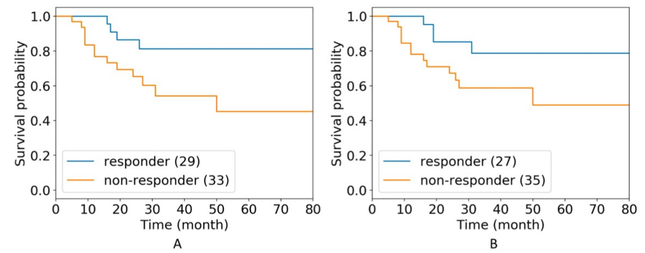
			Per Cell Line		Per Drug		Micro	
Train/Val Data	Test Data	Model	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
GDSC	GDSC	Lasso	79.1	53.8	67.1	38.2	79.3	55.4
		DeepDSC	80.0	54.8	67.7	38.8	79.9	56.4
		DeepGRMF	**83.2**	**60.1**	**70.9**	**41.8**	**83.1**	**62.0**
GDSC	CCLE	Lasso	79.2	67.5	66.2	38.2	74.1	50.5
		DeepDSC	80.0	68.3	67.0	40.5	75.1	51.5
		DeepGRMF	**82.0**	**70.9**	**67.9**	**41.6**	**76.0**	**53.7**

Table 2: Performance of different models to predict drug response of existing cell lines to new drugs.

			Per Cell Line		Per Drug		Micro	
Train/Val Data	Test Data	Model	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
GDSC	GDSC	DeepDSC	58.6	33.0	64.5	35.3	65.4	37.7
		DeepGRMF	**65.5**	**38.1**	**70.7**	**41.8**	**72.9**	**46.7**

Table 3: Performance of different models to predict drug response of new cell lines to new drugs.

			Per Cell Line		Per Drug		Micro	
Train/Val Data	Test Data	Model	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
GDSC	GDSC	DeepDSC	58.2	31.6	55.6	28.5	59.8	31.9
		DeepGRMF	**64.6**	**36.6**	**61.4**	**33.8**	**66.9**	**38.9**
GDSC	CCLE	DeepDSC	49.1	49.4	58.5	38.1	55.1	32.2
		DeepGRMF	**56.1**	**55.5**	**69.1**	**49.4**	**61.0**	**44.6**



Kaplan-Meier curves of responder and non-responder group of lung cancer patients that took Cisplatin, Pemetrexed, Paclitaxel, and/or Vinorelbine for adjuvant therapy which these four drugs treated as existing drugs (A) or as new drugs (B). Patients in the responders group survived significantly longer than the non-responders group.