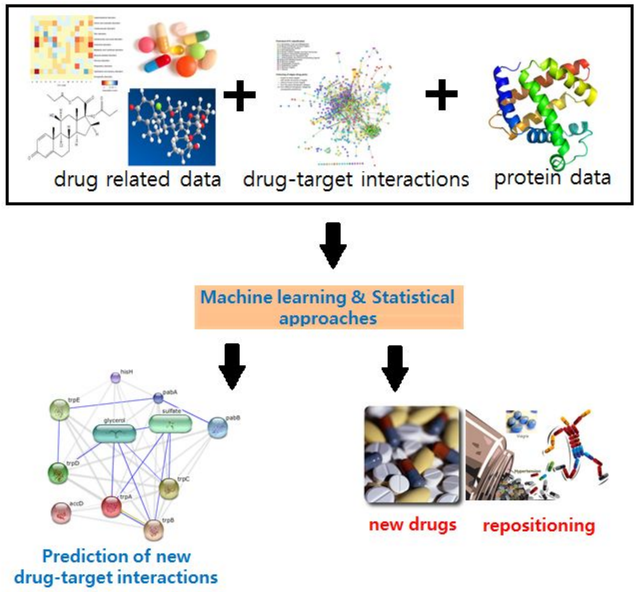
Deep Latent Factor Model for Predicting Drug Target Interactions

Aanchal Mongia[1], Vidit jain[1], Emilie Chouzenoux[2] and Angshul Majumdar[1]
[1] Dept. of CS, IIIT-Delhi , [2] CVN, Inria Sacley, CentraleSupélec

INTRODUCTION



- Drugs: chemical compounds
  Targets: Amino-acid sequences/proteins
- The interactions of some (a subset) drugs on some (a subset) targets are known through wet-lab experiments
- There is a **need to complement wet lab experiments by computational means** to narrow down the search space for experimental verification.
- Interactions can be predicted from: known drug-target interaction network, similarities over drugs and those over targets (optional)

Proposed Approach

MOTIVATION

- In [1], it was experimentally shown that matrix factorization-based techniques (where data matrix is factorized into 2 matrices) yield by far the best results.
- Deep learning has been successful in solving various Machine learning problems.

OBJECTIVE FUNCTION

$$\min_{U_1,U_2,U_3,V} \tfrac{1}{2}\left\|Y - R\cdot(U_1U_2U_3V)\right\|_F^2 \text{ such that } U_1U_2U_3V \geq 0, U_1 \geq 0, U_2 \geq 0, U_3 \geq 0, V \geq 0$$

LAGRANGIAN BASED FORMULATION

$$\min_{U_1,U_2,U_3,V,X} \tfrac{1}{2}\left\|Y - R\cdot(U_1U_2U_3V)\right\|_F^2 + \tfrac{\lambda}{2}\left\|U_1U_2U_3V - X\right\|_F^2 \text{ such that } X \geq 0, U_1 \geq 0, U_2 \geq 0, U_3 \geq 0, V \geq 0$$

Algorithm

Initialize: $X^0, U_1^0, U_2^0, U_3^0, V^0$

For $k = 1, 2, \ldots$

$$X^{k+1} = P_+\left(X^k - \gamma\left(R^T\cdot\left(R\cdot X^k - Y\right) + \lambda\left(X^k - U_1^kU_2^kU_3^kV^k\right)\right)\right)$$

$$U_1^{k+1} = P_+\left(U_1^k - \theta_1\lambda\left(U_1^kU_2^kU_3^kV^k - X^{k+1}\right)\left(U_2^kU_3^kV^k\right)^T\right)$$

$$U_2^{k+1} = P_+\left(U_2^k - \theta_2\lambda\left(U_1^{k+1}\right)^T\left(U_1^{k+1}U_2^kU_3^kV^k - X^{k+1}\right)\left(U_3^kV^k\right)^T\right)$$

$$U_3^{k+1} = P_+\left(U_3^k - \theta_3\lambda\left(U_1^{k+1}U_2^{k+1}\right)^T\left(U_1^{k+1}U_2^{k+1}U_3^kV^k - X^{k+1}\right)\left(V^k\right)^T\right)$$

$$V^{k+1} = prox_{\theta G\left(X^{k+1},U_1^{k+1},U_2^{k+1},U_3^{k+1},\cdot\right)}\left(V^k\right)$$

$$G\left(X,U_1,U_2,U_3,V\right) = \tfrac{1}{2}\left\|Y - R\cdot X\right\|_F^2 + \tfrac{\lambda}{2}\left\|U_1U_2U_3V - X\right\|_F^2$$

$$(\theta_1, \theta_2, \theta_3, \theta) > 0 \text{ and } \gamma \leq \left(\left\|R\right\|_F^2 + \lambda\right)^{-1}$$

RESULTS

TABLE I: DATASET DESCRIPTION

Datasets	NR	GPCR	IC	E
Drugs	54	223	201	445
Targets	26	95	204	664
Interactions	90	635	1476	2926

TABLE II: LATENT FACTORS – INPUTS TO ALGORITHMS

Layers	E	GPCR	IC	NR
1 Layer / SVT / PMF / BMC	10	80	85	25
2 Layer	185-25	40-10	100-15	20-10
3 Layer	180-85-15	65-15-5	100-50-10	20-10-5

- Evaluation metrics: AUC and AUPR
- Train:Test = 70:30% (done 10 times and average reported)
- **On one hand, going deeper improves abstraction of the latent factors but on the other hand, the number of parameters to learn also increases. Owing to the second fact, over-fitting sets in, the algorithm fails to generalize on the unseen data**

TABLE III: TABLE SHOWING AUPR

Dataset	1 Layer	2 Layer	3 Layer	SVT	BMC	PMF	GRMF
E	.639	.714	**.728**	.010	.706	.622	.498
GPCR	.599	.615	**.616**	.036	.604	.556	.442
IC	.792	**.828**	**.828**	.056	.803	.760	.381
NR	.097	.121	**.125**	.092	.107	.091	.097

TABLE IV: TABLE SHOWING AUC

Dataset	1 Layer	2 Layer	3 Layer	SVT	BMC	PMF	GRMF
E	.877	.897	**.899**	.496	.879	.855	.573
GPCR	.870	.881	**.884**	.528	.876	.858	.561
IC	.928	**.942**	.941	.488	.931	.874	.628
NR	.634	**.669**	**.669**	.461	.639	.618	.456

CONCLUSIONS

- *This is the first work that shows how the drug target interaction can be decomposed into more than two factors and estimated via the product of individual factors.*
- *When the associated biological metadata of the drugs/targets is not available, our method yields the best possible results .*
- *The algorithm can be extended to other applications (CF, scRNA-seq imputation, etc) and can be modified to incorporate biological metadata associated with the drugs and targets.*

LITERATURE Reference

1. Ezzat, Ali, Min Wu, Xiao-Li Li, and Chee-Keong Kwoh. "Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey." Briefings in bioinformatics (2018).

CONTACT Information

Angshul Majumdar
Indraprastha Institute of Information Technology, Delhi
India
Contact : angshul@iiitd.ac.in

View full paper at https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8683123