DEEP MATRIX COMPLETION ON GRAPHS: APPLICATION IN DRUG TARGET INTERACTION PREDICTION

Aanchal Mongia[1], Angshul Majumdar[2]

[1]Dept. of Computer Science and Engineering, IIIT-Delhi, India
[2]Dept. of Electronics and Communications, IIIT-Delhi, India

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

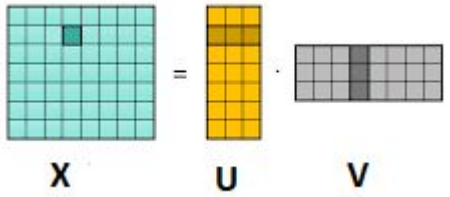
Introduction

- Matrix completion: a crucial statistical problem in the research community used in missing data imputation and prediction problems.
- Conventional algorithms cannot leverage the metadata associated with the row and column entities of a matrix.
  Solution: Graph regularization (prevents overfitting and enables the algorithms to take into account the associated metadata)
- Deep learning: shown to learn meaningful representations
- Motivation: success of two very recent studies on (shallow) matrix completion on graphs and deep matrix factorization (without graphs).

Background

- Two conventional Matrix completion algorithms: Nuclear Norm Minimization (NNM) and Matrix Factorization (MF)

- MF based techniques assume that the matrix to be completed is low-rank and can be decomposed into a product of two low-rank latent factor matrices



- **Graph regularized version** adds a graph laplacian penalty to the cost function. The graph laplacians are derived from the weights between the nodes in row/column graphs and encode the information of row/column entities.

- **Going deeper** with the no of factors is another way of improving over the standard MF

Background

- MF (Matrix factorization)

$$\min_{U,V} ||Y - M(UV)||_F^2$$

- GRMF (Graph regularized matrix factorization)

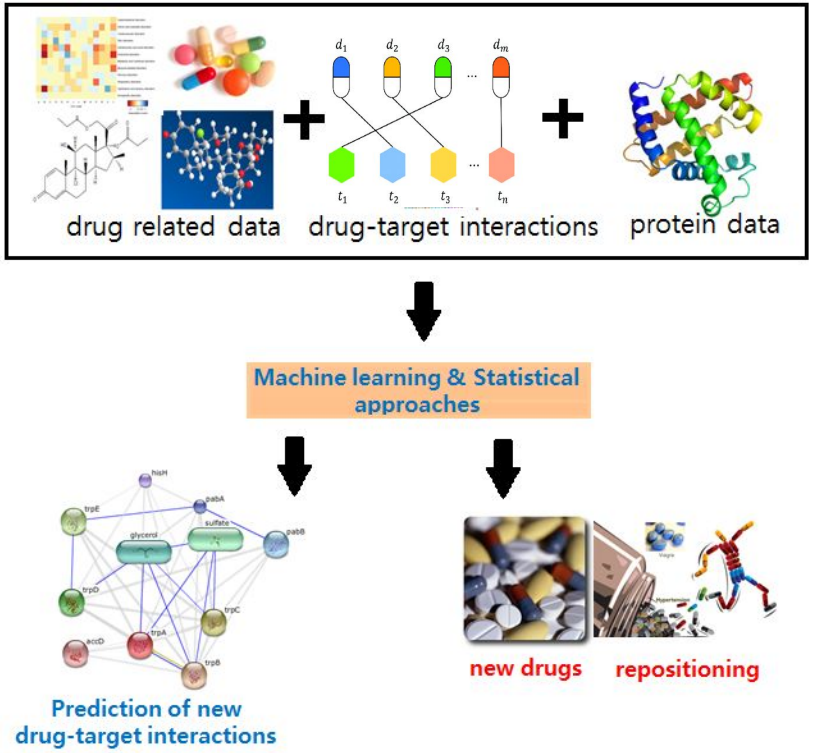
$$\min_{U,V} ||Y - M(UV)||_F^2 + \mu_1 Tr(U^T L_d U) + \mu_2 Tr(V L_t V^T)$$

- DMF (Deep matrix factorization)

$$\min_{U1,U2,V} ||Y - M \circ (U1U2V)||_F^2 \text{ such that } U1 \geq 0, U2 \geq 0$$

Drug–Target Interaction prediction

- Drugs: chemical compounds
  Targets: Amino-acid sequences/proteins
- Experimental validation traditionally done through wet-lab experiments
- Need for alternatives **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)**



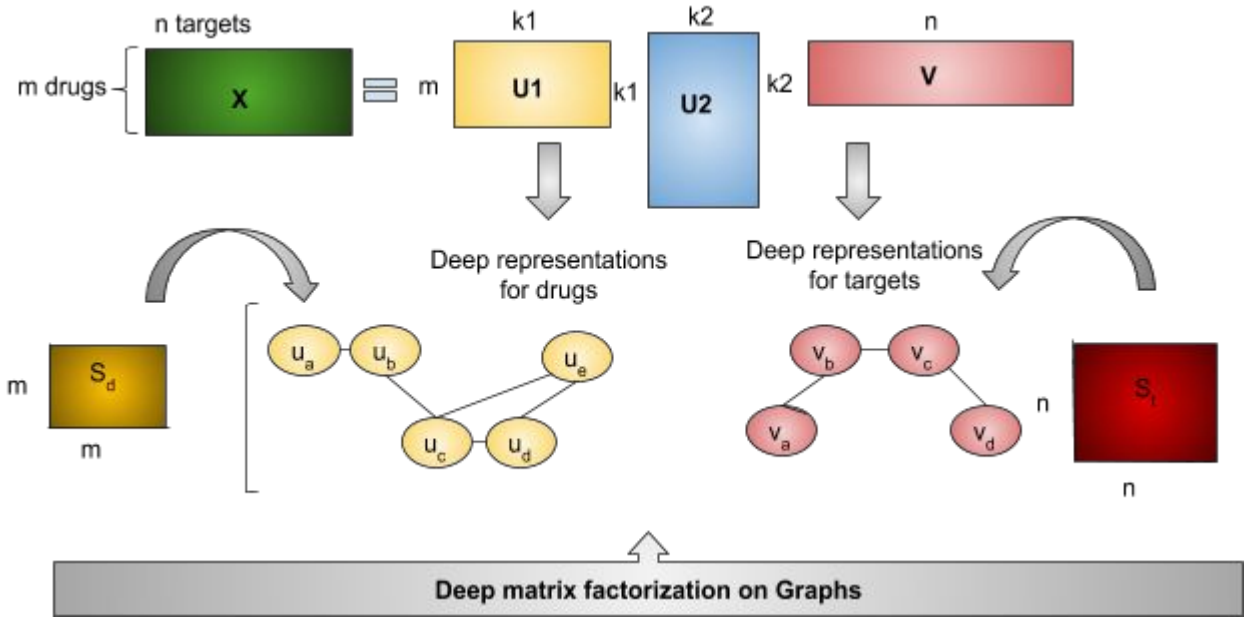
Proposed approach

- DTI prediction modelled as matrix completion
- DMCG: Deep matrix completion on graphs
- Formulation:

$$\min_{U1,U2,V} ||Y - M(U1U2V)||_F^2 + \mu_1 Tr(U1^T L_d U1) + \mu_2 Tr(V L_t V^T)$$

- Finds a low-rank interaction matrix that is structured by the proximities of drugs and targets encoded by graphs using deep matrix completion approach
- Exploits drug and target similarities
- Deep version of matrix completion to learn complex representations

Proposed approach



Algorithm

Algorithm 1 DMCG (Y, M, S_d, S_t)

1: **Sparsify**: Compute N_{ij}, $S_d = N_{ij}. * S_d$, $S_t = N_{ij}. * S_d$
2: **Initialize**: $p, \mu_1, \mu_2, k_1, k_2, X = Y, U2, V$(SVD initialization), $D_d = \Sigma_j (S_d)^{ij}$, $D_t = \Sigma_j (S_t)^{ij}$, $L_d = (D_d)^{-1/2}(D_d - S_d)(D_d)^{-1/2}$, $L_t = (D_t)^{-1/2}(D_t - S_t)(D_t)^{-1/2}$
3: **For loop** , iterate (k)
4: $B_k = (X)_{k-1} + \frac{1}{\alpha} M^T (Y - M(X)_{k-1})$
5: $U1 \leftarrow \text{sylvester}(\mu_1 L_d, U2V(U2V)', B(U2V)')$
6: $U2 \leftarrow U1^{\dagger} X V^{\dagger}$
7: $V \leftarrow \text{sylvester}((U1U2)'U1U2, \mu_2 L_t, (U1U2)'B)$
8: $X = U1U2V$
9: **End loop 1**
10: **Return**: $\hat{Y} = X$

Evaluation

- Datasets used: NR, IC , GPCR, E
- 5 runs of 10-fold Cross validation settings:
  - CVS1/Pair prediction
  - CVS2/Drug prediction
  - CVS3/Target prediction
- **Metrics:** AUPR (Area under the precision-recall curve) averaged across 5 repeats of 10-fold cross-validation.
- **Benchmark techniques:**
  - Matrix completion on graphs
  - Deep Matrix completion (DMC) using Deep matrix factorization

Results

Dataset	DMCG	DMC	MCG
E	0.8906	0.7589	0.7621
IC	0.9301	0.8402	0.8346
GPCR	0.7303	0.5718	0.5956
NR	0.6382	0.3560	0.4558

Table 2. Table showing mean AUPR values for DTI prediction.

Dataset	DMCG	DMC	MCG
E	6.54	1.32	11.12
IC	1.48	0.99	2.12
GPCR	1.20	0.91	1.54
NR	0.92	0.92	1.08

Table 3. Table showing running times (in seconds)

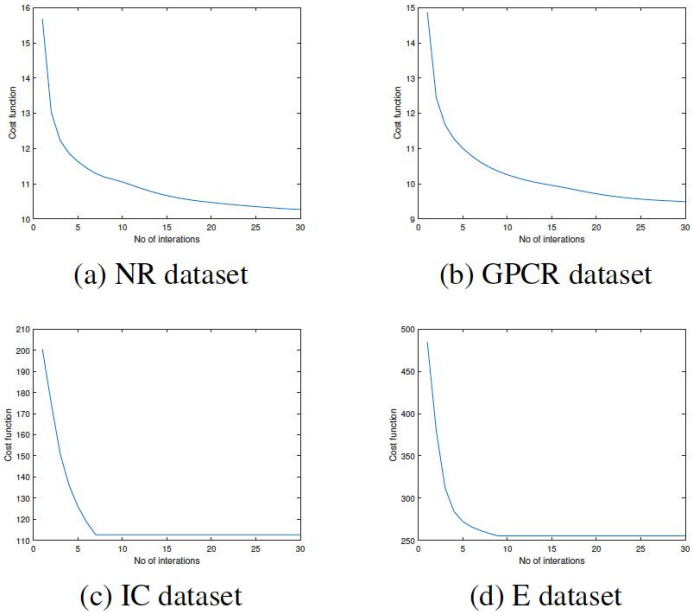


Figure 1: Convergence plots for DMCG

Conclusion

- We proposed a novel deep matrix completion algorithm that learns interpretable features while incorporating the similarity information/metadata associated with row and column entities.

- Algorithm proposed is generic and can not only be used in other bioinformatics problems like protein-protein interaction, RNA-RNA interaction, etc but also other research fields like collaborative filtering

- Future Work: Incorporating different types of similarities in one framework may result in further improvement in results.

Thank You!