Inductive Matrix Completion using Graph Autoencoder

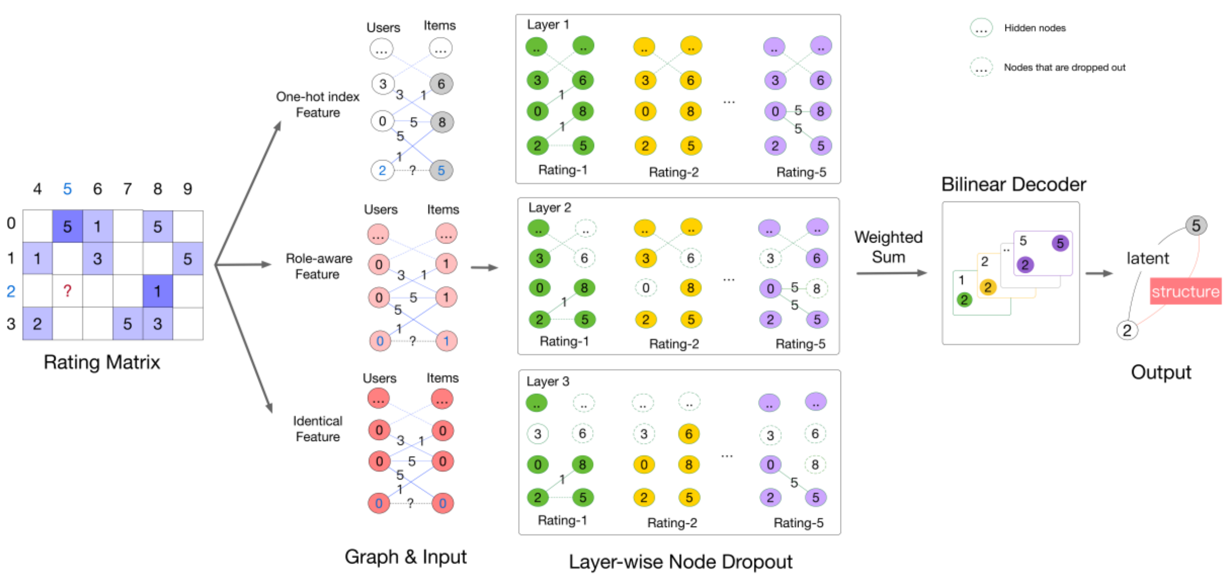

Figure 1: Model Overview. The rating matrix is formulated as a bipartite user-item graph, in which the nodes represent users (or items) and the links represent the corresponding ratings. In addition, the input features of each node in this graph consist of the identical feature, the role-aware feature, and the one-hot index feature. In addition, the encoder of our model has multiple layers (e.g., Layer 1) with multiple rating-subgraph (e.g., Rating 1). As stacking more layers, the node dropout probability increases, which is referred to as layer-wise node dropout. The model aggregated the latent embedding which is learned by one-hot index feature and structure embedding of a node which is learned by role-aware feature and identical feature in all layers by the weighted sum operator. At last, we reconstruct the links by a bilinear decoder. In this way, the output of our model contains the information of both latent link representation and structure representation

In summary, this work makes the following main contributions:

- (Sec. 3.1) To better understand local graph patterns, we conduct a quantitative analysis on five real datasets. Based on this quantitative analysis, we have multiple observations that reveal the properties of local graph patterns in matrix completion. It motivates us to design our model, IMC-GAE.
- (Sec. 3.2) We design two informative features, the identical feature and the role-aware feature, for the model to learn the expressive graph patterns. Moreover, these graph patterns can be easily generalized to unseen graphs.
- (Sec. 3.5) We design a layer-wise node dropout schema that drops out more nodes in the higher layers. With the layer-wise node dropout, link representation in our model contains more node information in a 1-hop local graph around the target link. Accordingly, our model is able to learn local graph patterns associated with the target link, which enhances the capability of the inductive learning of our model.
- (Sec. 5) To illustrate the effectiveness of the proposed IMC-GAE, we conduct empirical studies on five benchmark datasets. Extensive results demonstrate the state-of-the-art performance of IMC-GAE and its effectiveness in learning both local graph patterns and node-specific representations.

Structure

Embedding layer: For each node in each T rating subgraphs (e.g., T=5), there is an initial learnable embedding ($l_t[i]$). Nodes have two additional embeddings which are identical ($u_t[i]$) and role-aware ($r_t[i]$); These two embeddings are shared between rating subgraphs, therefore there are T+2 embeddings per node in general. The output of embedding layer is as follows:

$$x_t^0[i] = Concat(u_t[i], r_t[i], l_t[i]),$$

Message passing layer:

$$x_t^{l+1}[i] = \sum_{j \in \mathcal{N}_t(i)} \frac{1}{\sqrt{|\mathcal{N}_t(i)| \cdot |\mathcal{N}_t(j)|}} x_t^l[j]$$

$|N_t(i)|$ represents the number of neighbors of node i in the t-th rating subgraph. ReLU activation is used between layers.

Layer-wise node dropout is used with increasing rate in higher layers, to promote 1-hop interactions.

Accumulation layer:

$$h_t[i] = \sum_{0 \le l \le L} \frac{1}{l+1} x_t^1[i] \qquad h[i] = \sum_{t \in T} h_t[i] \qquad n[i] = tanh(Wh[i])$$

Bilinear decoder:

$$e_t[i,j] = n[i]^T W_t n[j], \tag{6}$$

where W_t is a learnable parameter matrix. Thus, we can estimate the final rating score as,

$$r[i,j] = \sum_{t \in T} t S_t(\mathbf{e}(\mathbf{i},\mathbf{j})), \tag{7}$$

where $\mathbf{e}(\mathbf{i},\mathbf{j})$ is the vector that concentrate the final link representations of user i and item j on all T rating subgraph, and the S_t is the softmax probability on t-th dimension of $\mathbf{e}(\mathbf{i},\mathbf{j})$ vector.

Node representation regularization:

$$\mathcal{L}_{NRR} = -\sum_{0 \le t < T} \sum_{0 \le i \le N} Cos(h_t[i], h_{t+1}[i]),$$

Inductive Learning

Inductive learning is the ability to perform well for users and items that are unseen during training. Most of recommender systems use side information for unseen nodes, and might work well even without rating data. In this model, side information for users and items is not used, and instead, inductive learning is handled using the links and ratings (cold-start is not handled). To calculate the latent representation of an unseen node, the representations of its seen neighbors is averaged.

$$l_t[i] = \sum_{j \in \mathcal{I}} \frac{1}{|\mathcal{I}|} l_t[j],$$

Experiments

Table 3: RMSE of different algorithms on Flixster, Douban and YahooMusic.

Model	Flixster	Douban	YahooMusic
IGC-MC	0.999	0.990	21.3
F-EAE	0.908	0.738	20.0
PinSage	0.954	0.739	22.9
IGMC	**0.872**	**0.721**	19.1
GRALS	1.245	0.883	38.0
sRGCNN	0.926	0.801	22.4
GC-MC	0.917	0.734	20.5
IMC-GAE (ours)	0.884	**0.721**	**18.7**

Table 4: RMSE test results on MovieLens-100K (left) and MovieLens-1M (right).

Model	ML-100K	Model	ML-1M
F-EAE	0.920	F-EAE	0.860
PinSage	0.951	PinSage	0.906
IGMC	0.905	IGMC	0.857
MC	0.973	PMF	0.883
IMC	1.653	I-RBM	0.854
GMC	0.996	NNMF	0.843
GRALS	0.945	I-AutoRec	0.831
sRGCNN	0.929	CF-NADE	**0.829**
GC-MC	0.905	GC-MC	0.832
NMTR	0.911	NMTR	0.834
IMC-GAE (ours)	**0.897**	IMC-GAE(ours)	**0.829**