[Arxiv 2020] Pre-training Graph Neural Networks for Cold-Start Users and Items Representation paper [paper][code]

Node/Graph Tasks:
Node tasks - obtaining node embeddings of high quality and making better recommendations for the cold-start users

Training Type:
Pre-training each model consecutively and joint training all models together

Pretext task data: graph topology (user-item interactions)
The paper proposes a pre-training GNN model to reconstruct the cold-start user/item embeddings in the meta-learning setting. Firstly, the ground-truth embedding for each user/item is learned from the observed interactions (graph topology) by Neural Collaborative Filtering (NCF). Then K neighbors for each user/item are randomly sampled to mimic the cold-start setting where GNN models are further applied to predict the embedding of each user/item by the neighbors' embeddings. The models are trained by maximizing the similarity between the embedding learnt in the cold-start setting and the ground-truth embedding learnt by NCF.

Brief Summary:
Current recommendation systems fail to learn high-quality embeddings for cold-start users/items due to sparse interactions, which can be addressed by 1) incorporating extra information such as content features or knowledge graphs and 2) GNNs which capture the high-order collaborative signal. However, both of these two methods fail to thoroughly solve the cold-start problem. Extra information is not always available and GNNs does not specially deal with the cold-start neighbors. Therefore, this paper proposes a pre-training GNN model as the pretext task to enhance the embeddings of the cold-start users/items.

To simulate the scenario typically encountered by cold-start users/items, we remove neighbors of each user/item and create a new bipartite graph with less interactions between users and items. A GNN model is applied on this new bipartite graph to create the embeddings, the similarity of which to the ground-truth embeddings is further calculated by cosine similarity to train the GNN model. Notably, the ground-truth embeddings are output by NCF utilizing the original graph with full interactions between users/items. Moreover, to address the problems of aggregating inaccurate embeddings of the cold-start neighbors, a meta aggregator and an adaptive neighbor sampler are further incorporated into the above pre-training GNN model.

The embeddings of a cold-start user/item may suffer from the poorly embeddings of its neighborhoods that might also be cold-start items/users during feature aggregation in GNNs. Thus, the paper trains another self-attention encoder g before training the GNN encoder f to learn an additional embedding for each node u only based on its K first-order neighbors that are sampled. The embeddings $\boldsymbol{h}'_u$produced by g is combined with the original embedding $\boldsymbol{h}_u^{l-1}$ and the aggregated neighborhood embeddings $\boldsymbol{h}_{\mathcal{N}(u)}$ at each convolution in f as Eq. (8)-(11):

$$\boldsymbol{h}_u^l = \sigma(\boldsymbol{W}^l \cdot \mathrm{CONCAT}(\boldsymbol{h}_u', \boldsymbol{h}_u^{l-1}, \boldsymbol{h}_{\mathcal{N}(u)}^l)), \tag{8}$$

$$\boldsymbol{h}_{\mathcal{N}(u)}^l = \mathrm{AGGREGATE}(\{\boldsymbol{h}_i^{l-1}, \forall i \in \mathcal{N}(u)\}), \tag{9}$$

$$\{\boldsymbol{h}_1, ..., \boldsymbol{h}_K\} \leftarrow \mathrm{SELF\_ATTENTION}(\{\boldsymbol{h}_1^0, ..., \boldsymbol{h}_K^0\}), \tag{10}$$

$$\boldsymbol{h}_u' = \mathrm{AVERAGE}(\{\boldsymbol{h}_1, ..., \boldsymbol{h}_K\}). \tag{11}$$

Then the randomly sampling strategy of K-neighborhood nodes is replaced by an adaptive sampler which selects nodes probabilistically guided by the feedbacks from the pre-training GNN model. The improved sampler is modeled as a hierarchical Markov Decision Process and sampling t-th node from the l-order neighbors is modeled as a bernoulli experiment where the successful probability is calculated by feeding the state feature s_t^l of the node to be sampled into a MLP. The state feature s_t^l is a high-dimensional vector and each entry measures the cosine similarity between its initial embedding and the target user u's initial embedding, the initial embedding of each formerly selected neighbor by the $l-1$-th subtask and the average embedding of all the formerly selected neighbors respectively. The reward is the similarity gain between the ground-truch embeddings and the predicted embeddings with/without the improved sampler as Eq. **??**.

The paper first pre-train the meta learner g only based on first-order neighbors, then fix parameters of g and train the GNN encoder f incorporating features output by g in each graph convolution step, and next pre-train the neighbor sampler s with feedbacks from the GNN encoder f. Finally all three models g, f, s are jointly trained. To ensure a stable update during joint training, parameters are updated by a linear combination of its old version and the new version.

Bibtex:

@articlehao2020pre, title=Pre-Training Graph Neural Networks for Cold-Start Users and Items Representation, author=Hao, Bowen and Zhang, Jing and Yin, Hongzhi and Li, Cuiping and Chen, Hong, journal=arXiv preprint arXiv:2012.07064, year=2020