2023 IEEE 39th International Conference on Data Engineering (ICDE) | 979-8-3503-2227-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDE55515.2023.00318

Attribute Graph Neural Networks for Strict Cold Start Recommendation (Extended Abstract)

1st Tieyun Qian
School of Comput. Sci., Wuhan Univ.
Wuhan, China
qty@whu.edu.cn

2nd Yile Liang
School of Comput. Sci., Wuhan Univ.
Wuhan, China
liangyile@whu.edu.cn

3rd Qing Li
Dept. of Comput., Hong Kong Poly. Univ.
Hong Kong, China
qing-prof.li@polyu.edu.hk

4th Hui Xiong
AI Thrust, HKUST (Guangzhou)
Hong Kong, China
xionghui@ust.hk

***Abstract*—Recently, deep learning based methods, especially graph neural network (GNN), have made impressive progress on rating prediction problem in recommender systems. However, the performance of existing methods drops quickly in the cold start scenario. More importantly, such methods are unable to learn the preference embedding of a *strict cold start* user/item since there is no interaction for this user/item. In this work, we develop a novel framework *Attribute Graph Neural Networks* (AGNN) by exploiting the attribute graph rather than the commonly used interaction graph. AGNN can produce the preference embedding for a strict cold user/item by learning on the distribution of attributes with an extended variational auto-encoder (eVAE) structure. It also contains a new graph neural network variant (gated-GNN) to effectively aggregate various attributes of different dimensions in a neighborhood. Empirical results demonstrate that AGNN achieves the new state-of-the-art performance.**

***Index Terms*—recommender systems, strict cold start recommendation, rating prediction, graph neural networks.**

I. Introduction

Rating prediction aims at inferring a user's ratings for the items which are not rated yet by the user. Traditional methods first learn the user's and item's latent representation, also known as *preference embedding* of a user or an item, and then use a score function over these embeddings to generate ratings for the missing entries in the user-item matrix.

The *normal cold start users/items* refer to those not appearing in the training data. Conventional approaches to this problem adopt side or external information to generate *feature embedding*. Nevertheless, they will suffer from an inevitable performance drop. In this paper, we are interested in an extreme scenario, i.e., *the strict cold start users/items* that neither appear in the training data nor have any interactions at the test stage. Conventional methods are inappropriate for the strict cold start problem since they all require the users/items have interactions in test. Several recent deep learning models can handle this issue, but they disregard the potential of GNN in absorbing information from neighbors.

To address the strict cold start problem, we develop a novel framework *Attribute Graph Neural Networks* (AGNN) by exploiting the attribute graph rather than the commonly used interaction graph. This leads to the capability of learning embeddings for the strict cold start users/items. Our AGNN can produce the preference embedding for a strict cold user/item by learning on the distribution of attributes with an extended variational auto-encoder (eVAE) structure. Moreover, we propose a new graph neural network variant, i.e., gated-GNN, to effectively aggregate various attributes of different dimensions in a neighborhood. Empirical results prove that our model achieves the state-of-the-art performance in both the strict cold start and warm start scenario.

This work was supported by the National Natural Science Foundation of China (NSFC) projects (62276193, 61572376). It was also supported by the Joint Lab. on Credit Sci. and Tech. of CSCI-Wuhan University.

II. Motivation and Background Information

Overall, our model is based on two insights. Firstly, the attribute is the most important side information since the user's/items' preference is determined by the natural properties reflected in their attributes. Secondly, GNN has the ability of borrowing information from neighbors with same attributes, yet the interaction graph is not suitable for strict cold start problem due to the lack of interactions. With the above motivations, the advantages of our model are as follows.

- We highlight the importance of exploiting the attribute graph rather than the interaction graph in addressing strict cold start problem in neural graph recommender systems.
- We design a novel eVAE structure to effectively infer the users'/items' preference embeddings from their attribute distributions with the empowered approximation ability.
- We address the key challenges in aggregating various attributes in a neighborhood by developing a gated-GNN structure which greatly improves the model capacity.

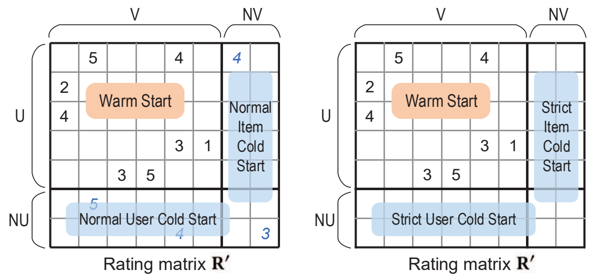
III. Problem Definition

Let $U = \{u_1, u_2, ..., u_M\}$ be a set of users and $V = \{v_1, v_2, ..., v_N\}$ be a set of items, where M and N denote the corresponding cardinalities. Each user or item is associated with a set of attributes from different fields. All attributes are concatenated into a multi-hot attribute encoding $\mathbf{a} \in \mathbb{R}^K$.

$$\mathbf{a}_u = \underbrace{[0,1]}_{gender}\ \underbrace{[1,0,0,...,0]}_{age}\ \underbrace{[0,1,0,...,0]}_{occupation},$$

Let $\mathbf{R} \in \mathbb{R}^{M \times N}$ be the user-item interaction matrix, where each $r_{ij} \in \mathbf{R}$ is either a rating score denoting u_i gives a rating to v_j, or 0 denoting the unknown ratings of items

2375-026X/23/$31.00 ©2023 IEEE
DOI 10.1109/ICDE55515.2023.00318



(a) warm start and normal cold start scenarios (b) warm start and strict cold start scenarios

Fig. 1: Warm start, normal cold start, and strict cold start scenarios in rating prediction task. ($\mathbf{R}' \in \mathbb{R}^{(M+\Delta M)\times(N+\Delta N)}$)

that the users have not interacted yet. The classic *warm start recommendation* predicts the unknown ratings for the users/items that already exist in the original interaction matrix $\mathbf{R}$ and also have interaction history, while *the normal cold start recommendation* predicts the ratings for the users/items unseen during training but having interactions at the test stage, and *the strict cold start user/item recommendation* aims to predict new users' ratings on items or to predict users' ratings on new items, as shown in Figure 1.

IV. Summary of the Technical Ideas

Our AGNN model consists of an input layer, an interaction layer, a gated-GNN layer, and a prediction layer.

Input layer We present an input layer *to construct the user (item) attribute graph* via two types of proximities, i.e., preference proximity and attribute proximity, which measure the historical preference and the attribute similarity between two nodes and are calculated by cosine distance. We then build a *k*-NN attribute graph based on the overall proximity that is summed over two types of proximity.

Interaction layer We get the user and item representation $\mathbf{p}_u$ and $\mathbf{q}_i$ by fusing the historical preferences and the attributes in the interaction layer.

$$\mathbf{p}_u = \mathbf{W}_u[\mathbf{m}_u;\ \mathbf{x}_u] + \mathbf{b}_u, \quad \mathbf{q}_i = \mathbf{W}_i[\mathbf{n}_i;\ \mathbf{y}_i] + \mathbf{b}_i, \tag{1}$$

where $\mathbf{m}_u$ and $\mathbf{n}_i$ encodes the user u's preference and the item i's property, and $\mathbf{x}_u$ and $\mathbf{y}_i$ denotes the attribute embedding for a user u and for an item i, respectively.

Injected eVAE structure in the interaction layer For strict cold start nodes without any interactions, we develop an injected eVAE structure to reconstruct the node's missing preference embedding from its attribute embedding. Our eVAE contains three parts: inference, generation, and approximation. The first two parts are the standard VAE and the third one is our extension. Briefly, we require the reconstructed embedding to be similar with both the preference embedding (by the constraint) and the original attribute distribution (by the standard VAE) in the approximation part.

Gated-GNN layer Intuitively, different neighbors have different relations to a node. Furthermore, one neighbor usually has multiple attributes. Since all these attributes along with the preferences are now encoded in the node's embedding, it is necessary to pay different attentions to different dimensions of the neighbor node's embedding. However, existing GCN [1] or GAT [2] structures cannot do this because they are at the coarse granularity. To solve this problem, we design a gated-GNN structure with an *aggregate gate* $\mathbf{a}_{gate}$ and a *filter gate* $\mathbf{f}_{gate}$ to aggregate the fine-grained neighbor information. The $\mathbf{a}_{gate}$ controls what information should be aggregated from neighbors to the target node, while $\mathbf{f}_{gate}$ controls what information in the target node should be filtered out if it is not consistent with that in the neighbors. We transform the original embedding $\mathbf{p}_u$ and $\mathbf{q}_i$ into the final embedding $\widetilde{\mathbf{p}}_u$ and $\widetilde{\mathbf{q}}_i$ for the user and item node after the gated-GNN layer.

Prediction layer Given a user u's and an item i's representation $\widetilde{\mathbf{p}}_u$ and $\widetilde{\mathbf{q}}_i$, we model the predicted rating of the user u to the item i as:

$$\hat{R}_{u,i} = MLP([\widetilde{\mathbf{p}}_u;\ \widetilde{\mathbf{q}}_i]) + \widetilde{\mathbf{p}}_u\widetilde{\mathbf{q}}_i^T + b_u + b_i + \mu \tag{2}$$

The training objective is to optimize the joint loss of the reconstruction loss in eVAE and the rating prediction one.

TABLE I: Performance comparison. ICS/UCS and WS are the abbreviations for the strict cold start item/user and warm start, respectively. * and † denote the statistical significance between our AGNN and the best baseline at $p < 0.01$ and $p < 0.05$.

RMSE	ML-1M			Yelp		
	ICS	UCS	WS	ICS	UCS	WS
NFM	1.0403	0.9885	0.9130	1.1231	1.1045	1.0620
DiffNet	1.0363	0.9809	0.8622	1.1072	1.1267	1.0444
DANSER	1.1246	0.9808	0.9797	1.1302	1.0927	1.0525
sRMGCNN	1.2978	1.2118	1.1770	-	-	-
GC-MC	1.0526	0.9922	0.8656	1.1229	1.1020	1.0254
STAR-GCN	1.0456	0.9878	0.8573	1.1173	1.0988	1.0232
MetaHIN	1.1162	1.0036	0.9870	1.1184	1.1031	1.0252
IGMC	1.1353	1.0453	0.8883	1.0965	1.0994	1.0512
DropoutNet	1.1008	1.0396	0.9254	1.1891	1.1724	1.1524
LLAE	3.3169	3.3223	3.3384	3.8057	3.8416	3.8008
HERS	1.1219	0.9823	0.9137	1.1977	1.1596	1.0240
MetaEmb	1.0290	0.9863	0.8648	1.0869	1.0928	1.0265
AGNN	**1.0091***	**0.9743***	**0.8533**†	**1.0749***	**1.0657***	**1.0106***
Improvement	2.62%	0.67%	0.47%	1.10%	2.47%	1.23%

V. Results

The comparison results of our AGNN and baselines are reported in Table I. It is clear that the improvements of AGNN over the strongest baselines are statistically significant in both the strict cold start and warm start scenarios on two datasets. The results verify the superiority of our proposed architecture by exploring the attribute graph for recommendation. In the complete version of this paper [3], we conduct rating prediction experiments on a total of three datasets. We also examine the impacts of the ratio of strict cold start nodes, and we conduct ablation and replacement experiments, as well as the parameter and complexity analysis.

References

[1] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *ICLR*, 2017.

[2] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *ICLR*, 2018.

[3] T. Qian, Y. Liang, Q. Li, and H. Xiong, "Attribute Graph Neural Networks for Strict Cold Start Recommendation," TKDE, 34(8): 3597-3610, 2022.