A Graph Attention Network Model for GMV Forecast on Online Shopping Festival

Qianyu Yu, Shuo Yang, Zhiqiang Zhang, Ya-Lin Zhang, Binbin Hu, Ziqi Liu, Kai Huang, Xingyu Zhong, Jun Zhou, and Yanming Fang

Ant Group
{qianyu.yqy, kexi.ys, lingyao.zzq, lyn.zyl, bin.hbb, ziqiliu, kevin.hk, xingyu.zxy, jun.zhoujun, yanming.fym}@antgroup.com

Abstract. In this paper, we present a novel Graph Attention Network based framework for GMV (Gross Merchandise Volume) forecast on online festival, called GAT-GF. Based on the well-designed retailer-customer graph and retailer-retailer graph, we employ a graph neural network based encoder cooperated with multi-head attention and self attention mechanism to comprehensively capture complicated structure between consumers and retailers, followed by a two-way regression decoder for effective predition. Extensive experiments on real promotion datasets demonstrate the superiority of GAT-GF.

Keywords: GMV forecast · GAT · e-commerce sales promotion.

1 Introduction

GMV (Gross Merchandise Volume) forecast is an essential problem for shopping festivals (e.g., Double 11[1] and Double 12) on e-commerce platforms, since accurate GMV estimation can help platforms assess the sales ability of retailers and then provide better services. However, its materialization is non-trivial, with two major challenges based on the analysis on real data in Taobao.com[2]. **1). Abnormal sales in shopping festivals.** Fig. 1 shows the average daily GMV of each month and shopping festivals for all the online retailers. We can find that the GMV distributions are indeed very skew, where the sales of Double 11 (Double 12) is extremely larger than usual days. Due to the lack of trend and seasonal patterns, classical statistical and time series based methods [1, 2] are not suitable. **2). Different contributions derived from neighbors.** Through the intuitive analysis for the consumption behaviours of customers in the shopping festival from the perspective of gender in Fig. 2, we conclude that the consumption behaviours of different consumers vary greatly. Thus it is impressing to develop an ingenious module to capture structural impact derived from related retailers and customers in two main aspects: i) Different relationships between retailers have

[1] https://en.wikipedia.org/wiki/Singles%27_Day
[2] https://en.wikipedia.org/wiki/Taobao

different impacts. ii) Distinct consumption preferences of consumers in shopping festivals need to be carefully considered.

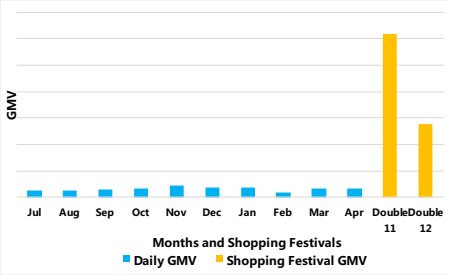


Fig. 1: GMV comparison of usual days and sales promotion.

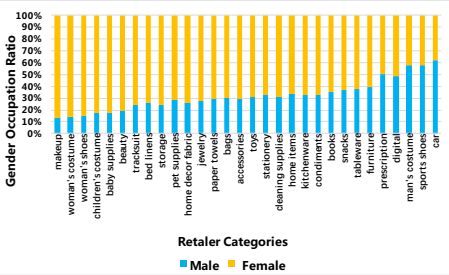


Fig. 2: Customers Gender Analysis of Different Retailer Categories.

In this paper, we design an end-to-end graph neural network based model to tackle the aforementioned challenges. Firstly, an R-C graph (short for **R**etailer-**C**ustomer graph) consisting of the transactions between retailers and their consumers and an R-R graph (short for **R**etailer-**R**etailer graph) consisting of various relationships between retailers (e.g., supply chain [3] and warehouse sharing) are conducted to comprehensively explore complicated structure between consumers and retailers. Following [4], we proposed GAT-GF, a novel Graph ATtention network for GMV Forecast, consisting of a graph neural network based encoder cooperated with multi-head attention and self attention mechanism and an effective two-way regression decoder. Extensive experiments on large-scale real shopping festival datasets prove the effectiveness of our proposal.

2 The Proposed Model

Before the elaboration of the the proposed GAT-GF, we briefly describe the input of GAT-GF, which aims at fully exploring complicated structure between consumers and retailers for facilitating GMV forecast. Specifically, it contains an R-C graph, where retailers and consumers are connected with transaction relationship and an R-R graph, where retailers are connected based on some domain knowledge (e.g., house sharing or supply chain). Now, as shown in Fig. 3, we are ready to zoom into each well-designed part of the proposed GAT-GF, i.e., a graph neural network based encoder and a two-regression decoder.

2.1 Graph Neural Network based Encoder

Information Extraction with Multi-head Attention To fully consider the different contributions derived from neighbors, we employ the multi-head attention mechanism [7] to adaptively learn the representation for each retailer, where each head may extract one aspect of customers' or retailers' influence,

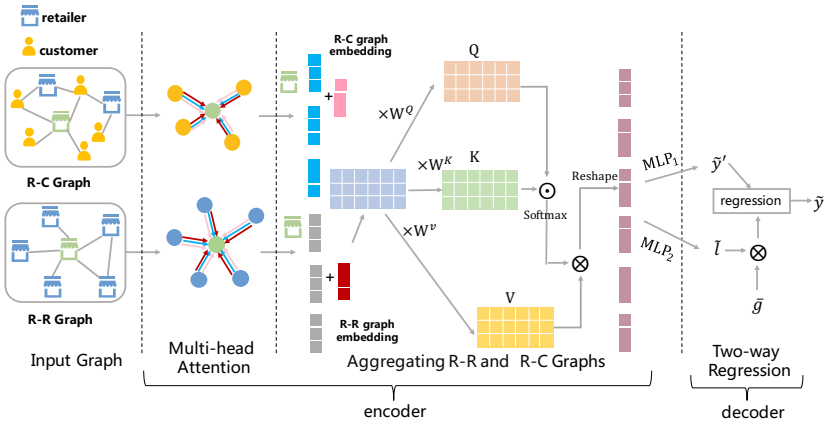


Fig. 3: The overall framework of GAT-GF.

and different heads pay attention to different aspects. Therefore, the customers' and retailers' contribution can be divided without any manual defined strategy.

For each graph $\Phi \in (RR, RC)$, at the t-th aggregation (we suppose neighbour in T hops), we design our graph attention network layer using multi-head attention mechanism in [5], and we get the initial node representation $e_{i,\Phi}^{(t+1)}$:

$$e_{i,\Phi}^{(t+1)} = \Big\|_{k=1}^{K} \sigma\Big(\sum_{j \in \mathcal{N}_{i,\Phi}} \alpha_{ij,\Phi}^{k,(t)} W_{\Phi}^{k,(t)} h_{j,\Phi}^{(t)}\Big), \tag{1}$$

where $\mathcal{N}_{i,\Phi}$ denotes node i's one-hop neighbour set in R-C or R-R graph. $h_{i,\Phi}^{(t)} \in \mathbb{R}^F$ represents node i's intermediate embedding at the t-th aggregation in graph Φ and $h_i^{(0)}$ is the original feature vector of node i, $||$ represents concatenation, K is the number of attention heads, σ represents the activation function, and W is input linear transformation weight matrix. Here we define attention coefficients α_{ij}^k with a learnable parameter H as

$$\alpha_{ij}^k = \frac{\exp(H_{ij}^k(h_i \,||\, h_j))}{\sum_{j' \in \mathcal{N}_j} \exp(H_{ij'}^k(h_i \,||\, h_{j'}))}, \tag{2}$$

Information Aggregation with Self-attention In macroscopic view, we adopt a self-attention mechanism to aggregate representations generated from R-R and R-C graph. In particular, inspired by the idea of positional encoding in [6], we also generate two learnable network type embedding $p_{i,\Phi} \in \mathbb{R}^{K \times F}$ for the R-R and R-C graph and obtain:

$$e_i^{(t)} = (e_{i,RR}^{(t)} + p_{i,RR}^{(t)}) \,||(e_{i,RC}^{(t)} + p_{i,RC}^{(t)}). \tag{3}$$

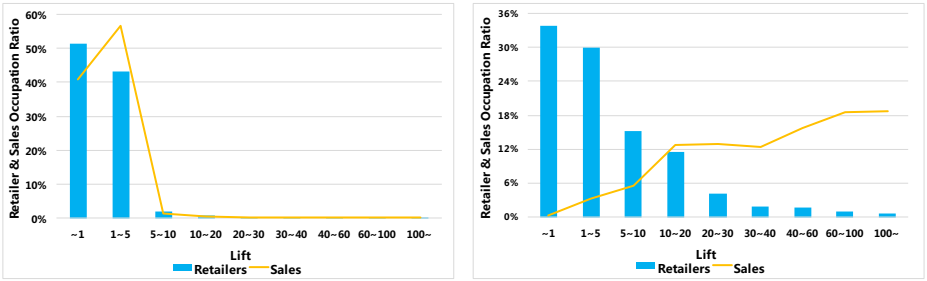
Self-attention mechanism operates on the input encoding $e_i^{(t)} \in \mathbb{R}^{2K \times F}$, and computes the representation $h_i^{(t)}$ as:

$$h_i^{(t)} = \text{softmax}\frac{(e_i^{(t)} W^{Q,(t)})(e_i^{(t)} W^{K,(t)})^{\mathrm{T}}}{\sqrt{d_z}}(e_i^{(t)} W^{V,(t)}). \tag{4}$$

W^Q, W^K and $W^V \in \mathbb{R}^{F \times d_z}$ are parameter matrices to learn. Once we obtain $h_i^{(t)}$, we can take it back to Eq.(1) to compute embeddings of the next hop hierarchically or compute the loss function in the next section.

2.2 Two-way Regression Decoder

Given the embedding of each retailer based on our proposed encoder, we design a two-way decoder for final prediction. Beforehand, with the real data from Taobao.com, we give an intuitive analysis about the abnormal sales in shopping festivals, as shown in Fig. 4a and Fig. 4b. We observe that retailers do not change much between two adjacent months (Fig. 4a), while retailers whose lift in the shopping festival exceeds five times over the usual days occupy the overall retailers about 35%, and sales of these retailers compromise 96% sales volume of shopping festival (Fig. 4b). We take the inspiration that large share of sales in the shopping festival is dominated by a small number of retailers, and propose an effective two-way decoder with the consideration of the GMV lift prediction.



(a) Usual sales. (b) Shopping festival.

Fig. 4: Lift comparison on usual sales and shopping festival sales.

In particular, the estimated GMV lift is multiplied by the average daily GMV, and we get the other prediction through this auxiliary task. Next, we combine the two predictions, and learn the parameters to balance them to get the final prediction result. Thus, final prediction is calculated as:

$$\tilde{y}_i = \underbrace{\alpha \mathbf{v}_G^{(T)} h_i^{(T)}}_{\text{GMV estimation}} + \beta \underbrace{(\mathbf{v}_l^{(T)} h_i^{(T)} + b_1)}_{\text{lift estimation}} \cdot \overline{g}_i + b_2, \tag{5}$$

where $\tilde{y}_i$ denotes the final estimation of the GMV of retailer i, $\overline{g}_i$ is the average daily GMV of retailer i, $\mathbf{v}_G \in \mathbb{R}^{d_z}$ and $\mathbf{v}_l \in \mathbb{R}^{d_z}$ are parameters to transform the retailer embeddings to GMV and lift estimation, α and β are the parameters adjusting the two estimations to the final regression prediction, and b_1 and b_2 are the constant offsets. Then we employ a mean square error loss on the final node prediction.

3 Experiment

3.1 Experimental Setup

Samples. We choose the labeled samples (retailers) of Double 11 and Double 12 on the Taobao.com. In practice, online shopping platform predicts sales volume using data of months ahead of the shopping festival to provide services in advance accordingly. Therefore we employ retailers' attributes of June to Sep., which can be represented as Double 11-June (Double 11 using data of June) to Double 11-Sep.. As a result, our model is evaluated on 8 datasets (4 for each shopping festival respectively). Sales volume of shopping festival in 2018 year is considered as the training set, and 2019 as the testing set.

Comparison methods. We take Gradient Boosting Decision Tree (**GBDT**), Neural Network (**NN**) and Graph Attention Network (GAT)-based models [5] (i.e., **GAT-RC** and **GAT-RR** with R-C and R-R graph, respectively) as baseline models. For fair comparison, we train all models with batch size of 64 and utilize Adam optimizer with learning rate of 1e-4 and regularization term of 2e-4. Moreover, we set the number of layer as 2 for NN and GAT based model and use 200 trees for the training of GBDT. We repeat the experiments for 3 times and the averaged results are reported.

3.2 Result Analysis

In order to reduce the prediction error, we take the logarithmic transformation of origin GMV as the regression goal. Besides, all the results reported in this section is on the testing set and the metric we used to measure the performance is the mean squared error (MSE).

Table 1 is the comparison result of the above methods (left) and ablation experiment (right). We can find that GAT-GF outperforms other models, which demonstrates the superiority of the proposed model. GAT-RR and GAT-RC outperforms GBDT and NN. It proves the effectiveness of R-R relations and R-C relations. However, the performance of GAT-RR is better than GAT-RC. It demonstrates the relation between retailers works better, and one reason is that leveraging related retailers' information is more direct to our goal. The performance of GAT-GF is better than GAT-RR and GAT-RC. It shows that fusing the R-R graph and R-C graph is positive. As the days approach the shopping festival, result is more accurate. It is because the feature and the graph are closer

Data	GBDT	NN	GAT-RR	GAT-RC	GAT-(RR,RC)	-SA	-TR	-MH	-NT	**GAT-GF**
Double 11-June	9.01	7.51	7.34	7.35	7.27	7.49	7.32	7.22	7.35	**7.19**
Double 11-July	6.50	6.47	6.33	6.44	6.31	6.33	6.29	6.32	6.26	**6.24**
Double 11-Aug.	5.75	6.18	5.65	5.61	5.57	5.57	5.58	5.53	5.57	**5.51**
Double 11-Sep.	4.68	4.70	4.60	4.61	4.57	4.61	4.54	4.60	4.57	**4.52**
Double 12-June	9.21	7.91	7.77	7.86	7.74	7.83	7.64	7.71	7.64	**7.56**
Double 12-July	7.19	7.27	7.05	7.06	6.99	6.98	6.99	7.05	7.04	**6.94**
Double 12-Aug.	6.63	6.59	6.43	6.51	6.41	6.42	6.44	6.39	6.44	**6.37**
Double 12-Sep.	5.82	5.80	5.68	5.76	5.66	5.73	5.74	5.68	5.64	**5.61**

Table 1: Performance comparison on 8 datasets. "-SA", "-TR", "-MH", "-NT" is short for GAT-GF by removing the self-attention, two-way regression, multi-head and network type embedding module, respectively.

to the day of shopping festival, and the same reason is that the results of the same month in Double 11 are better than that in Double 12. In the ablation experiment, we remove the modules we design and get four models – directly concatenates results from R-R and R-C graphs without self-attention, predicts GMV without auxiliary lift prediction, outputs embedding without multi-head mechanism, and removes network type embedding for input. The result shows that GAT-GF performs better than other models, which demonstrates the effectiveness of each module.

4 Conclusion

In this paper, we study the GMV forecast problem of the online shopping festival, which to the best of our knowledge is the first work. Experiments on the Taobao.com in the Double 11 and Double 12 show the validation our model and the ablation experiments show the effectiveness of each module we design.

References

1. Alon, I., Qi, M., Sadowski, R. J.: Forecasting aggregate retail sales:a comparison of artificial neural networks and traditional methods. Journal of retailing and consumer services **8**(3), 147–156 (2001)
2. Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time series analysis: forecasting and control. John Wiley & Sons (2015)
3. Yang, S., Zhang, Z., Zhou, J., Wang, Y., Sun, W., Zhong, X., et al.:Financial risk analysis for SMEs with graph-based supply chain mining. In: IJCAI. pp. 4661–4667 (2020)
4. Hamilton, W. L., Ying, R., Leskovec, J.: Representation learning on graphs: Methods and applications. arXiv preprint arXiv:1709.05584 (2017)
5. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
6. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al.: Attention is all you need. In: NIPS. pp. 5998–6008 (2017)
7. Li, J., Tu, Z., Yang, B., Lyu, M. R., Zhang, T.: Multi-head attention with disagreement regularization. arXiv preprint arXiv:1810.10183 (2018)