Holistic Neural Network for CTR Prediction

Huifeng Guo[1]* , Ruiming Tang[2], Yunming Ye[1]† , Xiuqiang He[2]
[1]Shenzhen Graduate School, Harbin Institute of Technology [2]Noah's Ark Lab, Huawei
[1]huifengguo@yeah.net yeyunming@hit.edu.cn
[2]{tangruiming,hexiuqiang}@huawei.com

ABSTRACT

This paper proposes *HNN*, a holistic neural network structure for click-through rate (CTR) prediction in recommender systems. Empirically, equipped with HNN, the performance of deep neural networks for CTR prediction are improved on Criteo and Huawei App Store datasets.

1. INTRODUCTION

Click-through rate (CTR) prediction is a crucial task in recommender systems, because CTR is an important factor deciding the ranking list that is returned to the user.

Many machine learning algorithms have been proposed to work on CTR prediction problem. As a simple yet effective and efficient approach, generalized linear models (such as FTRL [2]) have shown great benefits in industrial applications. However, these generalized linear models lack the ability of learning complicate feature patterns automatically [1]. Factorization Machines (FM) [4] are proposed to learn (2-order) feature interactions via inner product between embedding vectors of any two features. However, it is very complicated to learn high-order feature interactions using FM.

Since Deep Neural Network (DNN) is able to automatically learn high-order feature interactions, it is promising to adapt it in CTR prediction. [5] proposes FNN, a fully-connected neural network with FM-initialized embedding layer. Unlike FNN which needs a pre-training step, a cross-product layer of pairwise features is introduced in PNN [3]. There are some other works for CTR prediction using neural network, such as RNN-style [6], wide & deep [1].

Among deep learning frameworks for predicting CTR, FNN and PNN are claimed to be the most competitive ones [3]. Therefore, in this paper, we extend them with a holistic neural network structure (*HNN*). The extensions are named as *HFNN* and *HPNN*, respectively. We argue that HFNN and HPNN are able to consider both high- and low-order feature interactions, hence outperform existing FNN and PNN. We demonstrate empirically the effectiveness of our proposed model on Criteo and Huawei App Store datasets.

*This work is done when Huifeng Guo is an intern in Noah's Ark Lab, Huawei.

†Corresponding author

WWW 2017 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054208

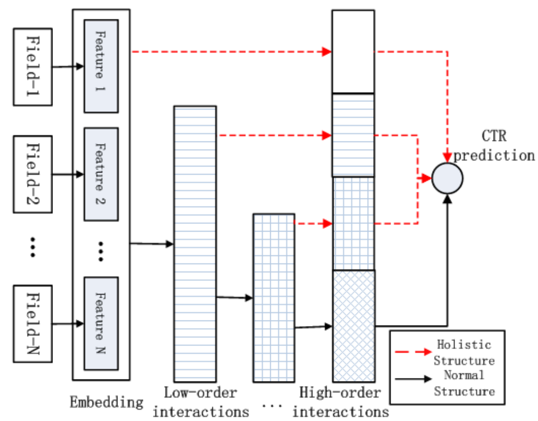

Figure 1: The Holistic Neural Network

2. PROPOSED MODEL

Based on fully-connected neural network, FNN optimizes the embedding layer by introducing FM initialization, while PNN optimizes the feature interaction by introducing a cross-product layer. Stated in [1], to make a good CTR prediction, learning high-order feature interactions alone is not good enough, low-order feature interactions are also needed to "memorize" the historical frequent patterns. Based on this observation, there is one shortcoming of FNN and PNN: the low-order feature interactions cannot make reasonable contributions in the CTR prediction, because only the high-order feature interactions are connected to the output layer. To overcome this shortcoming, we propose HFNN and HPNN, which extend FNN and PNN respectively with holistic structure.

For the ease of presentation, we present Holistic Neural Network structure in Figure 1. The black solid arrows show the forward procedure of FNN and PNN in CTR prediction. A data instance is fed into the neural network through embedding layer, low-order interaction layer(s), and high-order interaction layer(s) to generate a CTR prediction. During the network training procedure, the loss between predicted CTR and the ground-truth is backward propagated to train the parameters in the network.

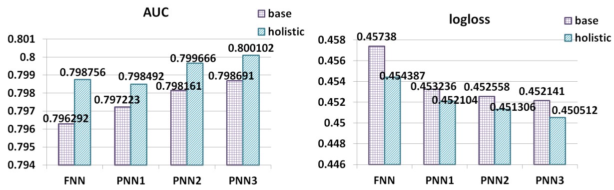

Figure 2: AUC and logloss comparison of different models

In Figure 1, the red dashed arrows present the forward flow of HFNN and HPNN. The holistic structure feeds both high- and low-order feature interactions to the final output layer in order to improve the representation power of the model. More specifically, each layer is connected forwardly to two layers, namely the next hidden layer (**black solid line**) and the final output layer (**red dashed line**). Therefore, the prediction of HNN is influenced by both high- and low-order feature interactions. As we can see, HNN structure considers both "memorization" (captured by low-order feature interactions) and "generalization" (expressed by high-order interactions) to improve model performance. We demonstrate the holistic structure on FNN and PNN, but it is possible to apply it to any neural networks.

3. EXPERIMENTS

3.1 Experiment Setup & Evaluation Metrics

We evaluate the effectiveness of our proposed holistic structure on two datasets.

1) Criteo Dataset: Criteo dataset [1] includes 45 million users' click records. We split the dataset randomly into two parts: 90% is for training, while the rest 10% is for testing.

2) Huawei App Store Dataset: We collect 7 consecutive days of users' click records from the gamecenter of Huawei App Store for training, and the next 1 day for testing. There are around 1 billion records in the whole collected dataset.

We analyze two evaluation metrics in our experiments: AUC (Area Under ROC) and *Logloss* (cross entropy).

3.2 Performance on Criteo Dataset

On Criteo dataset, we evaluate 10 models: LR, FM, FNN, HFNN, PNN{1,2,3}, HPNN{1,2,3}[2]. The comparison between normal models and holistic models are based on the setting given in [3], which is the best for FNN and PNN. Note that the activation function used in the output layer of all the models is *sigmoid*.

Since the performance of LR and FM is significantly worse than the deep models, we do not include them in Figure 2. The AUC of LR and FM are 0.7689 and 0.7895, while the logloss are 0.4773 and 0.4605. Two observations are be concluded from the performance of these models in Figure 2: (1) deep models works significantly better than LR and FM; (2) extended with our proposed holistic structure, the performance of FNN and PNN models are both improved.

We also perform hyper-parameters study on HFNN, as shown in Figure 3: (1) we vary the dropout rate as 1, 0.8, 0.6, 0.5, 0.4; (2) we set the number of hidden layers to be 1, 3, 5, 7. We find that HFNN works best when the dropout rate is set to 0.8 and the number of hidden layers is 5.

[1] http://labs.criteo.com/downloads/2014-kaggle-display-advertising-challenge-dataset/

[2] HPNN is based on 3 different PNN variants.

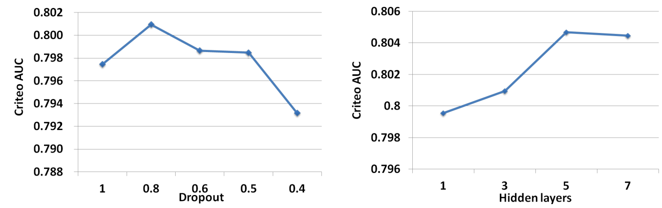

Figure 3: Performance Comparison of Dropout & Network Depth

Table 1: AUC relative performance of different models

	basic features		more features	
	mode 1	mode 2	mode 1	mode 2
LR	0%	0%	0%	0%
FM	+0.36%	-1.06%	-0.79%	-0.72%
FNN	**+0.37%**	+0.76%	+0.10%	+0.33%
HFNN	+0.34%	**+0.95%**	**+0.31%**	**+1.23%**

3.3 Performance on Huawei App Store Dataset

In Huawei App Store dataset, we evaluate four models: LR, FM, FNN and HFNN (we do not present the performance of PNN and HPNN since their performance is relatively worse). Table 1 presents AUC relative performance of the four tested models, while LR is set to be the baseline.

more features refers to the case that we add extra features (such as *app's income info, users' searching info, and etc*) besides *basic features* to the model. *mode 1* refers to the case that we are using the same features in the test set as in the training set. While, in *mode 2*, we remove some features (such as *display position of an app in the list*) that cannot be retrieved during the online serving, from the test set.

From the evaluation results, we can get the following observations: (1) HFNN achieves the best performance in almost all the cases, except for *basic features (mode 1)*, where HFNN has comparable performance to FM and FNN; (2) HFNN and FNN work better than LR and FM; (3) improvement of HFNN over the other models is more significant when using *more features*, than using *basic features*.

4. CONCLUSION

In this paper, we proposed a holistic neural network structure (HNN) to extend FNN and PNN, capturing both high- and low-order feature interactions. Experiments on Criteo and Huawei App Store datasets validated the effectiveness of our proposed holistic neural network structure.

Acknowledgement. This research was supported in part by NSFC under Grant No.61572158, National Key Technology R&D Program of MOST China under Grant No.2014BAL05B06, Shenzhen Science and Technology Program under Grant No.JCYJ20160330163900579 and JSGG20150512145714247.

5. REFERENCES

[1] H. Cheng et al. Wide & deep learning for recommender systems. *CoRR*, abs/1606.07792, 2016.
[2] H. B. McMahan et al. Ad click prediction: a view from the trenches. In *KDD 2013*.
[3] Y. Qu et al. Product-based neural networks for user response prediction. *CoRR*, abs/1611.00144, 2016.
[4] S. Rendle. Factorization machines. In *ICDM 2010*.
[5] W. Zhang et al. Deep learning over multi-field categorical data. In *ECIR 2016*.
[6] Y. Zhang et al. Sequential click prediction for sponsored search with recurrent neural networks. In *AAAI 2014*.