ISSN: 1813-4890

The Improved XdeepFM Algorithm Based on Attention Mechanism

Yingqiao Wang

Tianjin University of technology and education, TianJin 300222, China.

1098957231@qq.com

Abstract

This study proposes an improved recommendation model called MHSA-XdeepFM, which incorporates multi-head self-attention mechanism and an enhanced residual network into XdeepFM to enhance the performance of click-through rate prediction tasks. Click-through rate prediction is one of the key tasks in recommendation systems, aiming to predict the probability of users clicking on candidate items. XDeepFM is a click-through rate prediction model that combines deep neural networks(DNN) with Compressed Interaction Network (CIN) to capture both low-order and high-order feature interactions, while introducing linear layers to emphasize the importance of first-order features. However, in its final output, XDeepFM simply concatenates the outputs of various sub-models without fully considering the connections between them and the importance of features, which may lead to information redundancy and imbalance in feature fusion and representation, resulting in poor accuracy in click-through rate prediction.To address this issue, the study introduces a multi-head self-attention mechanism that allows the model to adaptively focus on the importance of different sub-models. Additionally, the Adaptive Feature Interaction Modeling AFM is incorporated to adaptively model second-order feature interactions. Through these improvements, the model can better capture the correlations and interaction patterns between features, thus improving the accuracy and effectiveness of click-through rate prediction.Experimental results demonstrate that the model outperforms the traditional XdeepFM on the publicly available Criteo dataset, providing important improvements and guidance for the application of recommendation systems in real-world scenarios. The model shows enhancements in performance metrics such as AUC and LogLoss, proving the practical significance of this research in improving the modeling of feature interactions in click-through rate prediction tasks

Keywords

CTR prediction, Feature Interaction, Multi-Head Self-Attention, Factorization Machines, Neural Networks.

1. Introduction

CTR prediction (Click-through Rate Prediction) is one of the key tasks in recommendation systems [1] and online advertising [2]. With the rapid development of the Internet, personalized recommendations and targeted advertising have become essential requirements for major platforms. The goal of CTR prediction is to forecast the probability of users clicking on candidate items based on user information and historical behavior data. Accurate CTR prediction can provide more relevant and personalized recommendations to users, enhancing user satisfaction and advertising effectiveness. Research on ad CTR prediction has a long history, and traditional machine learning algorithms such as Logistic Regression (LR) [3] and Factorization Machines [4] have achieved good results in dealing with raw features and second-

ISSN: 1813-4890

order feature interactions and have been applied in industries. However, such CTR prediction models typically rely on manually designed feature combinations to capture feature interactions, which are limited by the manually designed feature combinations and struggle to capture complex higher-order interactions. Additionally, traditional models often require extensive feature engineering and domain knowledge, increasing the complexity of model design and maintenance.In recent years, deep learning techniques have made significant progress in CTR prediction tasks. Deep learning models, such as DNN (Deep Neural Networks) [5], have been widely applied and successful in CTR prediction by learning nonlinear representations of features through multiple layers of neural networks. Deep & Cross[6] is an improvement that uses a method of building cross networks in contrast to DNN. In this network, each layer achieves explicit high-order feature interactions by performing dot products between feature vectors.Wide & Deep [7] introduced the idea of combining traditional models with deep learning models to achieve a balance between breadth and depth in feature representations for CTR prediction. The model captures the breadth of features through linear models and learns the depth representations of features through a DNN. Wide & Deep model can simultaneously utilize shallow features and deep features, thus improving CTR prediction performance.Building upon this, DeepFM [8] combines FM and DNN to more effectively extract breadth information. Empirical evidence shows that this combination can effectively capture both low-order and high-order feature interactions, thereby improving CTR prediction accuracy.To further enhance CTR prediction models, DCN (Deep & Cross Network) [9] and xDeepFM [10] were proposed. The DCN model learns high-order feature interactions through cross-network structures, allowing for complex interactions between features. xDeepFM, building on DCN's Cross Network, introduces the CIN (Compressed Interaction Network) to implement vector-wise high-order explicit feature interactions, showing the best performance among network structures that combine deep and shallow models.However, in xDeepFM, each sub-model operates in parallel, and this parallel output scheme isolates the information between sub-modules, limiting their potential correlations. The key to addressing this issue is to introduce a mechanism that promotes information flow and interaction between different sub-modules.In recent years, attention mechanisms have been widely applied in various fields of deep learning, such as machine translation, computer vision, etc. [11-12]. CFM introduced self-attention to pool the output of the embedding layer and used a CNN to learn the relationships between features.In this paper, the multi-head self-attention mechanism [13] and residual network [14] are introduced into the CTR prediction model to enhance its performance and accuracy. The multi-head self-attention mechanism can adaptively focus on the importance of different features, extracting feature correlations and interaction patterns. The residual network can learn feature representations and transformations at a deeper level, enhancing the model's representational and generalization capabilities.

2. OUR PROPOSED MODEL

2.1. Xdeepfm

2.1.1. Sub-section Headings

In Deep&Cross, Cross Layer operates the inner product of the vectors of all domains, which belongs to the feature interaction at the element level, which causes the model to lose the awareness of the domain, which means that the elements under the same domain may have different weights when interacting with features. xDeepFM is an improvement of the Wide&Deep model in terms of structure, and in terms of function, it is an improvement of the Deep&Cross model in terms of the above defects.

ISSN: 1813-4890

2.2. Multi-head-self-Attention

The Multi-Head Self-Attention mechanism is a powerful attention mechanism.Attention mechanism is first proposed in the context of neural machine translation and other fields. Vaswani further proposed multi-head self-attention to model complicated dependencies between words in machine translation.It achieves attention to different aspects and correlations between different features through multiple independent attention heads. In the task of click-through rate prediction, the Multi-Head Self-Attention mechanism can be applied to different components of the model, such as feature cross layers or attention mechanisms, to enhance the modeling capacity of feature interactions and improve the accuracy and effectiveness of click-through rate prediction.

2.3. FM

Factorization Machines (FM) are a class of predictive models that are particularly effective for handling sparse data and capturing the interactions between features in a dataset. FM was introduced as a solution to the limitations of traditional linear models, which often struggle to represent the complex relationships between features in large-scale data.

At the core of FM is the concept of feature interactions. Unlike linear models that assume features are independent of each other, FM recognizes that the effect of one feature can be modified by the presence of other features. This is achieved through a matrix factorization approach, where each feature is associated with a latent vector. The interaction between any two features is then computed as the dot product of their respective latent vectors.

2.4. General Description

In response to the drawbacks of information isolation between modules and insufficient deep feature representation capacity in the xDeepFM model, this section introduces a new model called MHSA-XdeepFM. It consists of three sub-models arranged in parallel layers: the FM layer for second-order interaction, the CIN layer for high-order explicit feature interaction, and the DNN layer. Additionally, a Multi-Head Self-Attention Network layer is incorporated to explore the connections among the sub-models. The following sections will provide a detailed introduction to each component of the model.

References

[1] K. R, W. Z, Y. R, H. Z, Y. Y, and J. W: User response learning for directly optimizing campaign performance in display advertising.in Proceedings of the 25th ACM International on Conference on Information and Knowledge Management,vol.25(2016), 679–688.

[2] H.B.McMahan, G. Holt, D. Sculley, M. Young, D. Ebner, J. Grady,L. Nie, T. Phillips, E. Davydov, D. Golovin et al.Ad click prediction:a view from the trenches,in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining,vol.25(2013), 1222–1230.

[3] D.f.Zou, Z.D.Wang,L.M.Zhang,et al.Deep Field Relation Neural Network for click-through rate prediction.Information Sciences,vol.46(2021),128-139.

[4] Rendle S. Factorization machines.Proceeding of the Tenth IEEE International Conferenceon Data Mining,2010:995-1000.

[5] G. Huang, Z. Liu, L. Van Der Maaten,et al.Densely connected convolutional networks.Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 4700–4708.

[6] R.F.Wang,G.Fu,et al.Deep&cross network for ad click predictions.Proceedings of the ADKDD17,2017:1-7.

[7] H.T.Cheng,K.Levent,H.Jeremiah,et al.Wide&deep learning for recommender system.Proceedings of the 1st Workshop on Deep Learning for recommender Systems,2016:7-10.

[8] H.F.Guo,R.M.Tang,Y.M.Ye,et al.A factorization-machine based neural network for CTR prediction.arXiv preprint,2017,arXiv: 1703.04247.

[9] J.X.Lian,X.H.Zhou,F.Z.Zhang, et al.Combing explicit and implicit feature interactions for recommender systems.Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery&Data Mining,2018:1754-1763.

[10] Devlin Jacob,Chang Ming-wei,Lee Kenton， et al.Bert: pre-training of deep bidirectional transformers for language understanding.arXiv preprint,2018,arXiv: 1810.04805.

[11] Lee Jinhyuk,Yoon Wonjin,Kim Sungdong,et al.BioBERT: a pre-trained biomedical language representation model for biomedical text mining.Bioinformatics,vol36(2020),1234-1240.

[12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, et al.Attention is all you need. In Advances in Neural Information Processing Systems. 2007,6000–6010.

[13] Xiao Jun,Ye Hao,He Xiang-nan,et al.Attentional factorization Machines:learning the weight of feature interations via attention networks[J].arXiv preprint,2017,arXiv:1708.04617.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition.In Proceedings of the IEEE conference on computer vision and pattern recognition.2016:770–778.