Proceedings of the 4th International Conference on Computing and Data Science (CONF-CDS 2022)
DOI: 10.54254/2755-2721/2/20220664

Survey on Click-Through Rate Prediction Based on Deep Learning

Yafei Liu[1]

[1]Capital Normal University High School, Beijing, 100048, China

liuyafei2023@163.com

Abstract. With regards to computer science, deep learning forms an essential research area. Recently, deep learning has made achievements on image processing, language understanding, data analysis, speech recognition, online advertising and so on. At the same time, display advertising has also become the most popular means of publicity with the rapid development in terms of the Internet, as well as the rapid expansion of the market of online advertising. Accurate advertising recommendation is the guarantee of Internet platform profits. The premise of accurate recommendation is accurate advertising click through rate prediction. Since 2015, the deep learning success has made the estimation of CTR results more accurate. Many CTR models have been used widely on a large amount of online platforms. This paper reviews several deep learning-based click-through rate prediction models for online advertising recently; classifies these prediction algorithms in the aspect of basic structure, complexity and main functions; and analyses the differences, advantages and the application. Finally, the survey is summarized and the future prospects of this field are envisaged.

Keywords: Deep learning, click-through prediction, online advertisement, marketing intelligence, demand forecasting

1. Introduction

In recent decades, various applications on the internet have resulted in the increase of data scale. Cloud computing, big data, artificial intelligence and other technologies have developed rapidly. A concept of "big data" has appeared in recent years, resulting in the creation and application with regards of large-scale datasets. This urges enterprises to use the data in decision-making to make improvements[1]. Big data has a great potential. It will not only bring significant development to human society, but also a serious “information overload” problem. At present, a key issue in the development of big data is how to obtain valuable information from large-scale datasets. As an effective method of solving the problem of “information overload”, the recommendation system has been noticed widely. According to user needs and interests, the recommendation system excavates the interested records from the massive datasets through the recommendation algorithm, builds click-through rate models, and recommends the results to users. At present, recommendation systems have been successfully applied in many fields.

At the same time, the development of the internet also brings users too much information. The targeted presentation of information to users will bring unexpected benefits to both producers and users. Salesperson usually need to manually adjust them and get the best possible performance in order to improve the attraction and performances of various advertising elements in real time. At the same time, it is also a problem to launch multiple advertisements on a large scale across platforms. For marketers

DOI: 10.54254/2755-2721/2/20220664

and media experts, turning to AI solutions seems to be a good choice if they want to carry out marketing efficiently. Deep learning helps optimize application advertising on a large scale. Deep learning can illustrate human behavior and patterns on an amazing scale. Deep learning can optimize automatic allocation and improve flexibility. While finding the best performing ideas, the system will put the best matching ideas into the right forum at the right time by making a CTR model.

Different from the applications of other artificial intelligence, deep learning is a standard of algorithms that can imitate intelligence of human with no manual interpretation and input rules. Due to its beneficial effects on the strategic performance of the enterprise, deep learning is going to alter the environment of business wholly in the future[2]. The company started to generally adopt deep learning technology. Marketing is important because of its aim to add value and satisfaction[3] through paying attention to demand, demand, targeting customers, customer attraction, management of relationships, sales, competition and other phenomena. Deep learning has become conducive to the sharing of customized services and advertising content in the housing rental industry [4].

It is noted that although there are several surveys in literature on click-through rate prediction, there is a lack of studies about how these algorithm are different from each other and their features with deep learning. Thus, the motivation of this research is to sort out different Click-Through Rate Estimation algorithm to let readers with little relevant knowledge to learn. Under this research motivation, the aim of this study is classifying these models in the aspect of basic structure, complexity and main functions; analysing the differences and advantages with the main research directions with the application progress.

The study’s structure is stated as follows. ParT 2 gives a general introduction in terms of the CNN model with DNN model. Part 3 focuses on analyzing the LSTM model. Part 4 includes analysis of DeepFM model. Finally, Part 5 provides the future development trend of machine learning-based advertising recommendation system in terms of academic and managerial perspective.

1.1. Neural network

From 2015, deep learning started to make significant improvement on estimation performance of CTR by transferring the classic architectures or developing the new ones. The applications of deep learning programming libraries with GPU computing stack and the universal approximation property of neural network[5] starts the training of models of deep neural network for the goal to catch the interaction patterns of feature with high-order and accomplish superior performance in the estimation of CTR.

1.2. Convolutional neural network (CNN)

Convolutional neural network (CNN) represents a set of feed-forward neural network imitating the learning process in the human brain. Compared to traditional neural networks, CNN reduces the connection between layers of the network, reduces the risk of over-fitting with the model parameters, thus simplifies the complexity of the model efficiently. Convolutional neural network comprises one or more convolutional layers, a fully-connected layer and one or more pooling layers, as illustrated in Fig. 1. In CNN, a composed instance with various fields is incorporated into the vectors of dense: each field is mapped to a vector of embedding feature. The convolutional layer utilizes one kernel which slides across the matrix of instances constructed by the vectors of embedding feature [6]. The pooling layer utilizes specified algorithms to reduce the dimension of the input. An function to activate is used for the output from the pooling layer. The activation function pooling layer, and convolutional layer can accomplish the learning feature of multiple-order. The features learned are connected by the fully-connected layer and the prediction has accomplished through the function of Softmax. CNN performs efficiently on finding interactions of local characteristics and eliminate the parameters nimber due to the shared weights mechanism design[7].

DOI: 10.54254/2755-2721/2/20220664

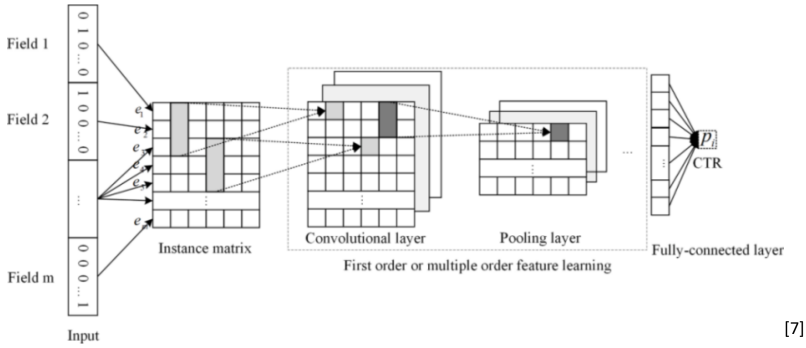

Figure 1. The CNN modeling framework.

Convolutional Neural Networks (CNN) are widely utilized for feature interaction modeling. Convolution click prediction model (CCPM)[8] excerpts local and global key characteristics from input instances of different parts, and repeatedly executes convolution, pooling and nonlinear activation that generate arbitrary order feature interaction. In CCPM, a flexible p-max pool layer can be used for selecting outstanding characteristics, where p can be defined as a function of convolution layers number and the input instances length. The embedding order of feature vectors might largely influence the feature interaction, because the convolution layer and the pool layer capture information in their respective local fields. Different embedded eigenvector sequences do not alter the meaning implied in most practical cases, like online advertising.

However, because CCPM is sensitive to field order, it is able to only study some interactions of characteristics between adjacent characteristics. In order to overcome the shortcomings of CCPM, researchers have designed other CNN models. Feature generating convolutional neural network (FGCNN) improves CCPM by introducing non adjacent features of reconstruction layer modeling. Then, the new features produced by CNN are combined with the original features to make the final forecast. FGCNN then verifies the features made by CNN can expand the space of the former characteristics and decrease the difficulty of optimizing the present deep structure. FGCNN has advantages over CCPM in AUC and logloss.

In addition, DMCNN is an improved version relative to CCPM. DMCNN introduced dense matrix to solve the disadvantage that CNN only captures local adjacent modes[9]. The matrix is composed of a deep layer, a DMCNN layer, and an embedded layer. The embedding layer transforms characteristics of high-dimensional sparse space to low-dimensional dense vectors. The DMCNN layer converts the embedded vector into a matrix of density, and then utilizes CNN to interactively extract the low-order characteristics. The normalized outputs of the embedded layer and DMCNN layer are deeply connected, and the high-order features are extracted interactively to obtain the predicted value. DMCNN predicts CTR better than CCPM in AUC and logloss.

1.3. Deep neural network (DNN)

DNN or sparse neural network (SNN) is another deep learning model that is widely used in professional fields. Using the vector expression of each sparse (or classification) feature, it is easy to construct the instance dense vector by connecting these vectors and inputting the relationship of these vectors into multiple perceptrons(MLPs) with a sigmoid output[10]. For the representation of the high-order characteristics, deep neural networks have efficient performances[11]. In general, DNN is composed of a feature input layer, two or more hidden layers that process input features, and an output layer that passes prediction probabilities through an activation function. The high-order characteristic expression method can be completed by increasing the number of DNN layers.

However, deep neural network cannot be directly applied to high-dimensional input feature space because it is computational complex to train a large number of parameters. Through the parameter initialization research of DNN which estimated CTR, researchers found that the prediction performance of DNN can be improved through the embedded vector initialized by the advanced training of factorization machine(FNN) or stacking automatic encoder. The framework of FNN modeling is shown in Figure 2.

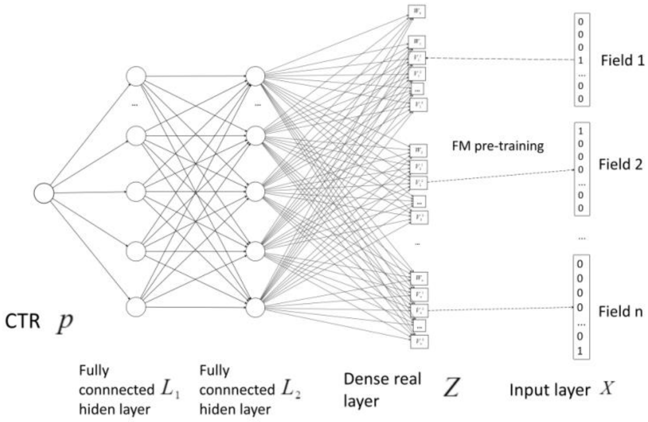

Figure 2. The FNN modeling framework.

As the earliest published deep neural network model for CTR estimation, Wide&Deep neural network[12] adds the logit values of LR (wide part) and DNN (deep part), and then feeding it as inputs into the final sigmoid function, which is shown in the following:

$$f_{\Theta}(x) = \theta_0 + \sum_{i=1}^{n}(x_i\theta_i) + \mathrm{MLP}_{\phi}([v_1, v_1, \ldots, v_n])$$

where $\Theta = (\theta, v, \phi)$. It is worth noting that cross features that were manually designed can also be included into the feature vector. This early deep learning CTR model uses a single MLP to reduce human efforts in feature engineering.It should be noted that other researchers[13] have explored the inverse model of FNN, using DNN for offline learning of high-order non-linearity and FM for online CTR prediction.

FNN can represent high-order feature interaction, learn precious and valuable patterns, and learn from the classification feature interaction. In addition, the computational complexity of the high-dimensional prediction problem is also reduced due to the use of FNN. In addition, some CTR prediction experiments show that FNN is better than some other models[14] in AUC. However, FNN has some disadvantages: (1) the influence of advanced training FM on embedded parameters may be too large; (2) the FNN efficiency is restricted by FM advanced training[15]; (3) FNN, which is also important for CTR prediction, does not include interactions of low-order characteristic[16].

1.4. Long short-term memory(LSTM) model

The standard LSTM recursive network unit comprises an input gate i_t, forget gate f_t and output gate o_t. It is worth noticing that the LSTM uses a linear self-loop cell memory consisting of three different gates instead of a simple RNN cell to process the weighted sum of input sequences. The three gates which control the amount of information via the unit memory have the sigmoid activation function: when the gate value is 1, all information flows; when the gate value is 0, there is no information flow. Figure 3 shows the basic learning process of LSTM network.

DOI: 10.54254/2755-2721/2/20220664

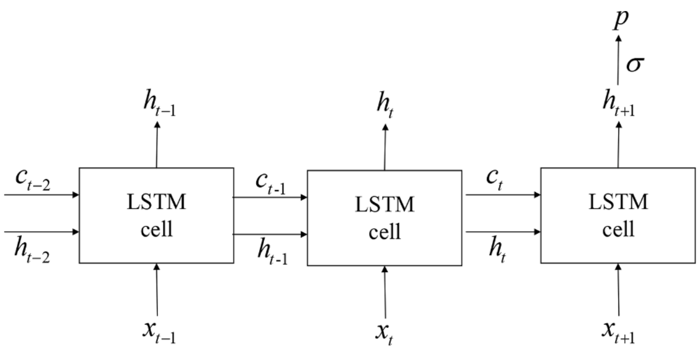

Figure 3. The LSTM modeling framework.

The LSTM network that was put forward by Hochreiter and Schmidhub in 1997[17], is an improvement of RNN, because LSTM can overcome the shortcomings of RNN in the case of long-term data dependence. The user's response to the advertisement displayed to them is regarded as a time event. LSTM and RNN are used as the foundamental framework to predict whether the user wants to click on the advertisement. The CTR prediction model based on LSTM and RNN is better than the linear and nonlinear models in AUC[18]. Because LSTM is trained in sequence, increading time and memory are required in the training steps. Some researchers have proposed an LSTM-RNN architecture to predict the clicking probability of a given advertisement at various locations[19]. At the same time, considering that the advertisements that have a higher ranking, the networks performs better than DNN under certain circumstances. Other researchers have proposed a CTR prediction model called deep interest evolution network (DIEN)[20] to capture the potential interest behind specific advertising behaviors (such as clicking). Other advertisements displayed on the web page together with the advertisement may affect the click times of the advertisement, which is called advertisement externality[21]. DIEN makes good use of this feature.

2. Deep factorization machines(DeepFM)

DeemFM is a neural network based on factorization-machine for prediction of CTR. It combines the DNN structure and FM structure. In DeepFM model, sparse input features are transformed into dense features shared by FM and DNN through an embedding layer, as shown in Figure 4. In DeepFM, FM and DNN are combined to synchronize high-order and low-order characteristics. Compared with Wide&deep model, DeepFM has a shared input for its "wide" and "deep" parts. Except for the original features, there is no need for feature engineering.

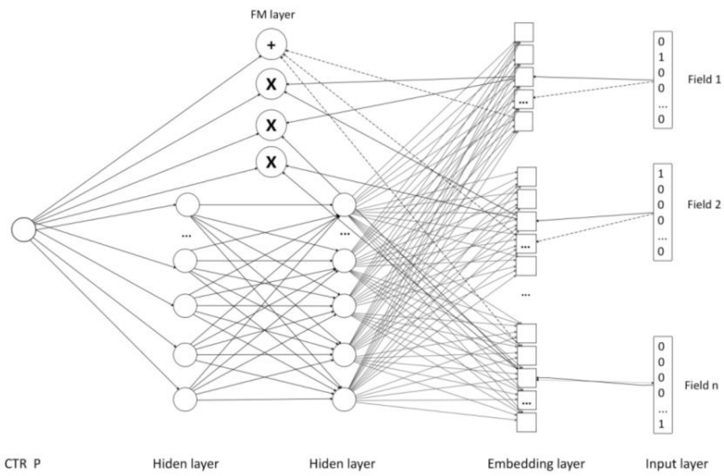

Figure 4. The DeepFM modeling framework.

In practice, DeepFM is much more effective than Wide&Deep, both offline and online, so many iterations of different scenarios and different characteristics have been derived around DeepFM, which is also the main sorting model at present. The reason is that FM crosses two features, calculates the weight coefficients of feature crossing by implicit vectors, and learns the weight coefficients of feature combinations that do not occur before. At the same time, training a DeepFM can also yield a variety of Embedding benefits.

DeepFM has several advantages: (a) it is shareing the original feature input between FM and DNN components; (b) the low-order feature interaction can be realized explicitly, and the high-order feature interaction can be carried out implicitly in a unified framework; (c) it can realize end-to-end learning without advanced-training FMs and manual feature engineering; (d) the Deep component of DeepFM can be replaced by other kinds of deep network architecture, such as PNN[22].

3. Conclusion

CTR models based on deep learning have their own advantages and disadvantages. CNN plays a very powerful role in discovering local feature interactions and reducing parameter quantities. However, useful functional interactions may be lost on CNN. FNN can reduce the problems of computational complexity of high-dimensional input feature only through correct advanced training FM. LSTM can effectively complete tasks and solve the problems of gradient vanishing and exploding. When the data have long-term correlation, it may requires more time and more memory in the training steps. DeepFM has a high level of computational complexity and does not require advanced training for FMs and manual feature engineering, so end-to-end learning can be achieved.

Although the deep learning algorithms in CTR prediction has developed rapidly and achieved great success, the author still note that there are still some major challenges to be solved in this field. Generally speaking, for online advertising, since few users give positive feedback on the advertising displayed to them, the number of clicks is far less than that of non-clicks. As a result, the distribution of sample labels is unbalanced. Since users who are interested in advertisements may still not click, treating non clicking as a negative sample may cause serious noise problems. This requires more intelligent model's guidance for sample advanced processing and feature engineering, filtering noise, and identifying effective negative samples. In addition, the CTR prediction model on the basis of deep learning needs to be improved in the aspect of representational learning (or advanced training) of multi-field categorical data. Various multimedia elements or browsing environments are available in modern information systems. Therefore, it is hopeful to design more CTR estimation models to deal with the feature interactions on multi-modal data.

In this article, ther author gives a literature review of advertisements CTR prediction based on deep learning, discuss the pros and cons of the most advanced CTR prediction models, and evaluate their performance. At the same time, the author summarizes the trends for current research, and major challenges and directions that deserve further exploration in this field. However, there are still some limitations, such as the lack of specific application of CTR models. This review is expected to provide basic knowledge for those who want to engage in this field or are at the entry point. At the same time, this review may provide a theoretical basis for the modeling framework of CTR prediction and promote the development of new models.

References

[1] Bag, S., Gupta, S., & Wood, L. (2020). Big data analytics in sustainable humanitarian supply chain: Barriers and their interactions. Annals of Operations Research, 1–40.
[2] Reis, C., Ruivo, P., Oliveira, T., & Faroleiro, P. (2020). Assessing the drivers of machine learning business value. Journal of Business Research, 117, 232–243.
[3] Sule Birim1·Ipek Kazancoglu2·Sachin Kumar Mangla3·Aysun Kahraman4·Yigit Kazancoglu5 (2021) The derived demand for advertising expenses and implications on sustainability: a comparative study using deep learning and traditional machine learning methods, 3.

[4] Sengupta, P., Biswas, B., Kumar, A., Shankar, R., & Gupta, S. (2021). Examining the predictors of successful Airbnb bookings with Hurdle models: Evidence from Europe, Australia, USA and Asia-Pacific cities. Journal of Business Research, 137, 538–554.
[5] George Cybenko. (1989). Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems, 2(4):303–314, 1989.
[6] Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An introductory review of deep learning for prediction models with big data. Frontiers in Artificial Intelligence, 3(4), 1–23.
[7] Yanwu Yang , Panyu Zhai (2022). Click-through rate prediction in online advertising: A literature review, 14.
[8] Bin Liu, Ruiming Tang, Yingzhi Chen, Jinkai Yu, Huifeng Guo, and Yuzhou Zhang. (2019). Feature generation by convolutional neural network for click-through rate prediction. In WWW, pages 1119–1129, 2019.
[9] Niu, T., & Hou, Y. (2020). Density matrix based convolutional neural network for click-thr ough rate prediction. In 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD). 46–50, DOI:10.1109/ICAIBD49809.2020.9137 448.org/10.1145/295 9100.2959134.
[10] Weinan Zhang, Tianming Du, and Jun Wang. (2016). Deep learning over multi-field categorical data: A case study on user response prediction. ECIR, 2016.
[11] Ling, X., Deng, W., Gu, C., Zhou, H., Li, C., & Sun, F. (2017). Model ensemble for click prediction in bing search ads. In Proceedings of the 26th International Conference on W orld Wide Web Companion (WWW '17 Companion). International World Wide Web C onferences Steering Committee, Republic and Canton of Geneva, CHE, 689–698. htt ps://doi.org/10.1145/3041021.3054192.
[12] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir,et al. (2016). Wide & deep learning for recommender systems. In 1st DLRS workshop, pages 7–10, 2016.
[13] Huang, Z., Pan, Z., Liu, Q., Long, B., Ma, H., & Chen, E. (2017). An Ad CTR Prediction Method Based on Feature Learning of Deep and Shallow Layers. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. Association for Computing Machinery, New York, NY, USA, 2119–2122. https://doi. org/10.1145/31 32847.3133072.
[14] Zhang, W., Du, T., & Wang, J. (2016). Deep learning over multi-field categorical data. Lecture Notes in Computer Science Advances in Information Retrieval. Springer, 45–57, https://doi.org/10.1007/978-3-319-30671-1_4
[15] Guo, H., Tang, R., Ye, Y., Li, Z., & He, X. (2017). DeepFM: a factorization-machine based neural network for CTR prediction. In Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI '17). AAAI Press, 1725–1731.
[16] Zhang, W., Qin, J., Guo, W., Tang, R., & He, X. (2021a). Deep learning for click-through rate estimation. arXiv:2104.10584.
[17] Hochreiter, S., & Schmidhuber, J.. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
[18] Gharibshah, Z., Zhu, X., Hainline, A., & Conway, M. (2020). Deep learning for user interest and response prediction in online display advertising. Data Science and Engineering, 5(1), 12–26.
[19] Deng, W., Ling, X., Qi, Y., Tan, T., Manavoglu, E., & Zhang, Q. (2018). Ad click prediction in sequence with long short-term memory networks: an externality-aware model. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 1065–1068. https://doi.org/10.1145/3209978.3210071.
[20] Zhou, G., Mou, N., Fan, Y., Pi, Q., Bian, W., Zhou, C., Zhu, X., & Gai, K. (2019). Deep interest evolution network for click-through rate prediction. In Proceedings of the AAAI

DOI: 10.54254/2755-2721/2/20220664

conference on artificial intelligence (AAAI-19). 33(01), 5941–5948. https://doi.org/10.1609/aaai.v33i01.33015941.

[21] Xiong, C., Wang, T., Ding, W., Shen, Y., & Liu, T.Y. (2012). Relational click prediction for sponsored search. In Proceedings of the fifth ACM international conference on Web search and data mining (WSDM '12). Association for Computing Machinery, New York, NY, USA, 493–502. https://doi.org/10.1145/2124295.2124355.

[22] Guo, H., Tang, R., Ye, Y., Li, Z., He, X., & Dong, Z. (2018). DeepFM: An end-to-end wide & deep learning framework for CTR prediction. ArXiv:1804.04950.