

Report for Multi-Level Deep Cascade Trees for Conversion Rate Prediction in Recommendation System

Xiaobing Yu

Miami University, USA

Keywords: IdcTree, DNN, e-commerce, deep learning.

Abstract: The study conducted by Wen et al. (2019) proposes a multi-Level Deep Cascade Trees Model (IdcTree) for conversion rate prediction in the recommendation system. The report introduces the background of recommendation systems used in the e-commerce industry and the proposed IdcTree for conversion rate prediction. The history and the challenges associated with DNN in previous work are carried out in the previous work of this study. The report also summarized in detail the main techniques and results of this study conducted by Wen et al. (2019) which establishes a basis for future research to investigate incorporating more features and other improvements that can improve efficiency in deep learning.

1. Introduction

Recommendation systems are used in the e-commerce industry to solve problems associated with information overloading. In this case, business owners assess the worth of a recommendation system based on the value it brings to the business. Some of the widely adopted measures in this context include click-through rates (CTR), conversion rates (CVR), sales or revenue, among others. Over the past few decades, deep learning has been employed in many application areas successfully and helped to overcome the obstacles associated with conventional models. However, deep learning is also associated with various deficiencies, such as the numerous hyper-parameters that require tuning, huge amounts of data are involved, and it also the powerful computational facilities are required to facilitate the training. These obstacles have created a need for a more effective and efficient alternative for deep learning.

The proposed IdcTree for conversion rate prediction is aimed at improving the effectiveness and efficiency of recommender systems. Recommender systems play a significant role in the success of e-commerce. By improving the experiences and services offered to shoppers, e-commerce businesses can increase their sales and also establish profitable business relationships with their customers. Businesses can effectively monitor the performance of their recommender systems by monitoring CVR. The CVR is used to measure the rate at which website visitors take any desired action that the e-commerce website owner would like them to take. A good recommender system is associated with a higher CVR. A low CVR is an indication that the recommender system is not optimized enough, as people are not taking the business' desired actions. Eliminating the obstacles associated with deep learning helps to improve the effectiveness associated with the recommender system.

A recent study conducted by Zhou and Feng (2017) proposed gcForest, an alternative to deep neural networks (DNN) that generates a deep forest ensemble and utilizes a cascade structure for representation learning. The gcForest achieves higher performance compared to DNN, and is also associated with a few hyper-parameters. The findings of the study inspired Wen et al. (2019) to come up with their study, where they propose the IdcTree novel model and the EldcTree extension. These are decision tree ensemble methods that utilize the deep cascade structure and a feature representation based on cross-entropy. The innovation will be applicable in all types of recommender systems, including content-based filtering, collaborative filtering, and hybrid systems.

2. Previous Work

Neural networks were first proposed in the 1940s, although the first application of digital neurons occurred in the 1980s, facilitated by the LeNet network utilized for the recognition of handwritten digits. The evolution of DNN models over the years has led to the development of different structures that facilitate various types of applications. These structures include the Multiple-Layer Perceptrons (MLP), Deep Belief Networks (DBN), and Convolutional Neural Networks (CNN).

The study conducted by Zhou and Feng (2017) was the first approach to addressing the problems associated with DNN. The researchers proposed the gcForest as an alternative to DNN. This was the first auto-encoder to be developed based on tree ensemble learning. According to Zhou (2012), ensemble learning is a model that seeks to train multiple learners, and also combines to tackle the problem. Also referred to as forests, the tree ensemble methods are the most effective approaches for supervised learning. Examples of tree ensembles include the Random Forest introduced by Breiman (2001), and the gradient boosting decision trees (GBDT) (Friedman, 2001). According to Liu et al. (2008), tree ensembles for supervised learning exceed expectations through their success in other tasks such as anomaly detection by the isolation forest method. Wen et al. (2019) propose the IdcTree ensemble, which employs the GBDT models. Despite the challenges associated with DNN, there has been widespread adoption and applications of deep learning.

3. Main Results

The researchers conducted online and off-line evaluations and experiments aimed at determining the effectiveness of the proposed approach. The offline evaluation results show that the proposed method IdcTree achieved higher AUC value than DNN, GBDT, GBDT+LR, and other methods. The EldcTree proposed method was found to have better AUC than the IdcTree, which was explained by the notion of ensemble learning. The F-EldcTree was found to have the best results compared to the other competitive methods. The results were explained by the full utilization of weak and strong correlation figures, in combination with the notion of ensemble learning. The online evaluation results revealed the effectiveness associated with level-by-level learning and the effectiveness of the F-EldcTree. In evaluating the effectiveness of level-by-level learning, the researchers implemented the IdcTree and the EldcTree methods with similar features from Naïve GBDT.

The methods were deployed in the recommender system after which increment in CVR was monitored. The IdcTree was found to have a CVR increase of 4%, while the EldcTree gained over 7%. The difference in CVR gain was attributed to the more reliable feature representation capability of the EldcTree. The researchers also employed the F-EldcTree to the online environment. Results revealed that DNN and the gcForest had better results than Naïve GBDT. The proposed method had the best results, with a record 12% CVR increment. These results led to the conclusion that the proposed approach has a more robust feature representation ability, which is attributed to its deep cascade structure and the implementation of strong correlation features and weak correlation features in an adequate manner. The proposed feature learning methods, the IdcTree, and the extension EldcTree, are made up of a deep cascade structure, achieved through sequential stacking of multiple GBDT units. The proposed method was found to have the best performance in both the offline and online experiments.

4. Techniques

The researchers proposed the IdcTree, which employs the deep cascade tree structure; based on inspirations from DNN representation learning characterized by level-by-level feature abstraction. The IdcTree is aimed at addressing the CVR prediction problem in the recommendation system. The structure of the IdcTree was made through sequential stacking of multiple GBDTs. The cross-entropy associated with every leaf node in the preceding GBDT was calculated to determine the input feature representation for the next unit. The study illustrates a two-level IdcTree characterized by three trees

and two trees in each level respectively. The following mathematical notations were used to define the cross-entropy for every leaf node:

$$S_{ijk} = -\frac{1}{I_{ijk}} \sum_{m=1}^{I_{ijk}} [y_m \log(h_{ij}(x^m)) + (1 - y_m) \log(1 - h_{ij}(x^m))] \quad (1)$$

Where, S_{ijk} represents the cross-entropy of the k -th leaf node of the j -th tree at level i ; I_{ijk} represents the number of instances falling in the k -th leaf node of the j -th tree at level i ; L_{ij} is the number of leaf nodes of the j -th tree at level i ; δ_{ijk} represents the split threshold for the k -th node of the j -th tree at level i ; F_{ij} represents the feature value at the j -th of the feature at level i ; N_i represents the number of trees for the GBDT model at level i ; $h_{ij}(x_n)$ is the predicted probability of the n -th instance x_n on the j -th tree at the level i ; and y_n is a representation of the ground truth label of the n -th instance x_n .

The following two equations were also formulated:

$$S_{ij}^+ = \bigcup_{p=1}^{L_{ij}} I(S_{ijp} > \delta_{(i+1)pg}) S_{ijp} \quad (2)$$

$$S_{ij}^- = \bigcup_{p=1}^{L_{ij}} I(S_{ijp} \leq \delta_{(i+1)pg}) S_{ijp} \quad (3)$$

An improved structure for the *ldcTree* based on the notion of ensemble learning was proposed and named, *Ensemble ldcTree (EldcTree)*. The proposed *EldcTree* enhances the diversity associated with the model and also improves the representation ability. The researchers also use the *Statistic Boosting Feature Importance* as a measure of feature importance. This leads to the classification of features into *Strong Correlation Features (SCF)* and *Weak Correlation Features (WCF)*.

In the evaluation settings of the study, Wen et al. (2019), implement both the *Area Under Curve (AUC)* and *F1score* based on precision and recall as the evaluation metrics. Precision and recall were defined through the following equation:

$$\text{Precision} = \frac{|P \cap T|}{|P|} \quad (4)$$

$$\text{Recall} = \frac{|T \cap P|}{|T|} \quad (5)$$

$$F_1 = \frac{(2 * \text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \quad (6)$$

The F_1 score was defined as:

The experiments involved a comparison of the proposed method with other closely related approaches. *Naïve* GBDT is used to refer to a single model of GBDT without the typical level-by-level learning. Another method for comparison was the GBDT + LR. The method's feature representation is based on the GBDT model. The method is utilized in the prediction of CVR by Logistic Regression (LR). A DNN structure with three hidden layers and a prediction layer was designed. After which ReLU was implemented as the activation function for the hidden layers. *gcForest* by Zhou and Feng (2017), was the last comparison method integrated into the study. The researchers replaced the Forests found in the *gcForest* with GBDTs. The researchers also explore the performance differences between the proposed *ldcTree*, *EldcTree*, and *F-EldcTree*.

5. Discussion

In the e-commerce domain, several decision attributes usually influence the preferences of the decision-maker. Under normal conditions, the decision-maker aims at maximizing their utility function.

Apart from improving user experiences, the recommender system gets the web visitors to spend more time on the e-commerce platform and purchase more goods and services. CVR has been used

for a long time as a measure of recommender systems and other online marketing strategies. An improvement of CVR in this case implies that the recommender system is effective and optimized in the right manner. Recommender systems for modern e-commerce websites are required to be innovative in a way that they serve users with diversified content in regards to the recommended items. Persistence in recommendations is also another key feature necessary for recommender systems. This involves re-showing recommendations to users based on their past activities alongside new recommended items. Effective recommender systems also have to consider the privacy of users, by ensuring that the process of building user profiles does not interfere with the privacy of data collected from users. There are other multiple factors such as product labelling, trust, robustness, and user demographics, among others, that recommender systems have to account for, to improve accuracy and effectiveness.

6. Conclusion

The study conducted by Wen et al. (2019) introduces an alternative learning method that is more effective and efficient than DNN. The researchers propose the *ldcTree* and its extension, the *EldcTree*. The proposed method is characterized by a deep cascade structure that is constructed through sequential stacking of several GBDT units. The implementation of a cross-entropy feature representation facilitates clear explanation and the desired distributed feature ability. Wen et al. (2019) found the proposed method *F-EldcTree* to have the best performance compared to the other models, in both the offline and the online experiments. When deployed in a recommender system for an e-commerce platform, the proposed methods achieved better results than other competing methods. A CVR increase of up to 12 percent was recorded. Unlike with other methods, the proposed *F-EldcTree* is associated with minimal training cost and are capable of supporting parallel implementation.

The study conducted by Wen et al. (2019) adds up to the growing literature on deep learning. The study establishes a basis for future research to investigate on incorporating more features and other improvements that can improve efficiency in deep learning. Future researchers can also focus on how improvements made by the proposed *F-EldcTree* can be beneficial for a wide array of applications, particularly recommender systems. Fields such as social media recommendations and film viewing platforms may also be greatly impacted by algorithm changes. Researchers can also focus on exploring the end-to-end training method proposed by Wen et al. (2019), which is utilized in jointly feature learning and classification based on the deep cascade tree structure. The study by Wen et al. is groundbreaking in the field of deep learning and overall application of AI models.

References

- [1] Coates, A., Huval, B., Wang, T., Wu, D., Catanzaro, B., & Andrew, N. (2013, February). Deep learning with COTS HPC systems. In International conference on machine learning (pp. 1337-1345).
- [2] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- [3] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).
- [4] Yang, B., Lei, Y., Liu, J., & Li, W. (2016). Social collaborative filtering by trust. *IEEE transactions on pattern analysis and machine intelligence*, 39(8), 1633-1647.
- [5] Hinton, G., Osindero, S., & Yee-Whye Teh. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554,
- [6] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

- [7] Raina, R., Madhavan, A., & Ng, A. Y. (2009, June). Large-scale deep unsupervised learning using graphics processors. In Proceedings of the 26th annual international conference on machine learning (pp. 873-880).
- [8] Wen, H., Zhang, J., Lin, Q., Yang, K., & Huang, P. (2019, July). Multi-Level Deep Cascade Trees for Conversion Rate Prediction in Recommendation System. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 338-345).
- [9] Zhou, Z. H. (2012). Ensemble methods: foundations and algorithms. CRC press.
- [10] Zhou, Z. H., & Feng, J. (2017). Deep forest: Towards an alternative to deep neural networks. arXiv: 1702.08835 v1.