



Deep Learning for FM-Based Recommendation: A Systematic Study on DeepFm and Its Application

Zhiheng Zhu^(✉)

Shanghai University of Electric Power, Shanghai 201306, China
zhu18018637648@126.com

Abstract. As one of the classic topics in artificial intelligence, the recommender system aims to use historical data to make intelligent data recommendations for users to meet their specific needs. Lately, with the maturity of deep learning-related technology, the introduction of deep learning into recommender systems to improve their performance has gradually become a hot research topic. A representative work is called DeepFm, which can capture both low- and high-order feature interactions as traditional FM models and deep learning networks (DNNs), and thus can improve efficiency and correctness. Although numerous research works around DeepFm have been carried out in recent years, there is a lack of systematic research to allow researchers to grasp the general development of related research. Therefore, we focus on the relevant literature with high citation rates in recent years and provide a comprehensive description of the DeepFm technology, highlighting DeepFm-related research and its improvement. In addition, the modeling principles, methods, advantages, limitations, and suggestions of these studies are discussed in detail from different perspectives. We also discuss the main features of these studies and sort out future development directions.

Keywords: Recommender System · Factorization Machine (FM) · Deep Learning Network (DNN) · Deepfm · Study

1 Introduction

The rapid development of the Internet and modern technology has filled our life with more and more information, which will greatly improve our living standard if it can be properly used. Therefore, we need to convert this information into usable data, process and analyze these data, and finally use them for recommendation or other applications. However, the key point is that these data are usually cumbersome and multi-dimensional, which can lead to a lot of trouble in processing. For this reason, we need to investigate how to process and use this quickly and effectively.

In this paper, we focus on how DeepFm and its improvements handle data and introduce in which fields DeepFm is applied, as well as traditional recommendation and deep learning models and their improvements respectively. During this research, we can find common characteristics in the research materials of recommendation systems, deep

learning models, DeepFm, and their improvements and applications. Firstly, they are high-dimensional in number, which requires extensive reading and browsing. Secondly, they are heterogeneous in scope, which requires targeted research and greatly increases the difficulty of organizing these works. Finally, it requires repeated and focused research so that we can more accurately filter out the important part. At the same time, we also found a problem in this process that materials about the recommendation model and deep learning model usually only study their own fields, and there is rarely research on the conversion process or combination of the two, which causes a certain gap in research. This paper mentioned the relationship between the two models and why they need to be combined. It is like a bridge between the two so that the reader can quickly and precisely conduct relevant research. It can also serve as an extended introduction to the improvement and application fields, providing the reader with certain inspiration.

In this project, we systematically studied relevant research in recent years and chose the ones with high citation rates. In this paper, we investigate the improvement and the application of DeepFm-related research. The former is further divided into the user level and the modeling level. At the user level, content-based collaborative filtering and neural collaborative filtering are mentioned [1, 2]. There is also a deep knowledge-aware network (DKN) [3], which is used in personalized recommendation, which, however, lacks an end-to-end training approach and can only contain text; MKR [4], a knowledge graph-embedded assistance; deep interest network (DIN) [5], which can effectively capture interests but does not consider the possibility of interest failure; deep interest evolution network (DIEN) [6], able to effectively solve interest changes over time, but does not address the stability of long-term behavior; Binn [7], which considers long-term stable behavior and evolution, but needs to study the impact of different types of users' behavior; and long and short-term user representations (LSTUR) are able to deal with user representation of short-term versus long-term behavior [8]. At the modeling level, there are deep and cross networks (DCNs) [9], fuzzy neural networks (FNNs) and self-normalizing neural networks (SNNs) [10], product-based neural networks (PNNs) [11], and AutoInt [12], which address sparse, dense, and high-dimensional datasets; Wide & Deep learning [13], which considers the problem of low-order high-order interactions but requires professional experience in feature engineering; and xDeepFm [14], a combination of a compressed interaction network (CIN) and a DNN, where the former is able to construct automatic fork multiplication of finite higher-order features. Clustering prior to interaction such as WebAPIs and K-means clustering are also proposed to improve the reliability and speed for mining. Finally [15, 16], applications and partial improvements in other domains are also proposed.

2 Background and Principle of DeepFm

DeepFm is a factorial decomposer-based neural network, an end-to-end learning model emphasizing low- and high-order interactions [17], that combines decomposer recommendations and deep learning in an algorithm designed to solve the CTR problem [18].

From Fig. 1, it is clear that the model consists of DNN and FM. It learns from embedding layers and uses latent vectors to value its interactions with other features.

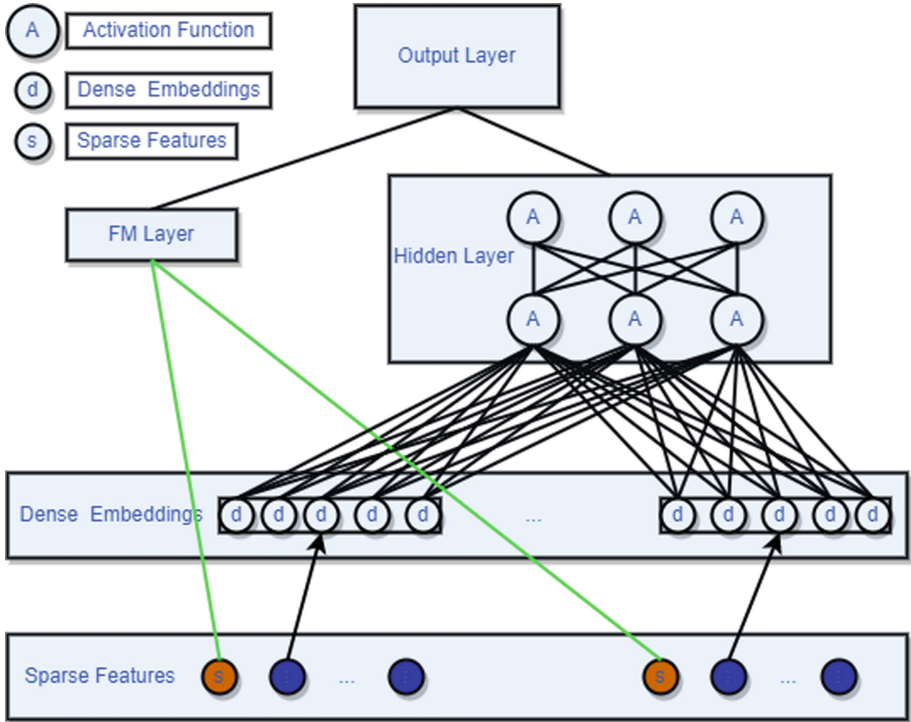


Fig. 1. DeepFm framework

Then it models

$$\hat{y} = \text{sigmoid}(y_{FM} + y_{DNN}) \tag{1}$$

An embedded FM modeling framework is combined with a DNN-based neural network framework in parallel that captures a combination of low- and high-order feature interactions [19]. With a sparse data matrix of large dimensionality, such as when the user has a specific preference or provides great many features, the dense embedding of FM allows the embedding vector to correspond to every nonzero feature, thus enhance the model’s generalization capability. The DNN multilayer perceptron can enable the high-order feature mining of FM giving the model certain ability to generate high-order features before their appearance. Furthermore, DNN and FM can use an identical input vector learnt through the embedding layer, which can avoid repeated learning.

Based on the Logistic Regression (LR) model, the FM model further considers the relationship between any two features [15]. Citing the second-order interaction, the equation is initially defined as

$$y_{FM} = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n w_{ij} x_i x_j \tag{2}$$

Unlike logistic regression, FM adds the feature combination of the consequent, where the scalar W_i is used to measure the order of the important feature interactions, which

are the weights. If $X_i X_j = 0$, then $W_{ij} = 0$, so if the data is extremely sparse, W_{ij} will eventually become 0. Then, it will have the same form as logistic regression. In order to deal with such problems, the FM model is improved to:

$$y_{FM} = w_0 + \sum_{i=1}^n \sum_{j=i+1}^n v_i v_j x_i x_j \quad (3)$$

where a potential vector V_i can be taken to calculate the interaction effect. An inner product of $V_i V_j$ is used instead of W_{ij} , because if the feature X_i interacts with any other features, the inner product is not zero, that is, the inner product can be used as a weight. Thus, the factorization machine can represent and model the first-order and second-order feature interactions respectively. An additive unit and multiple inner product units together make the FM output.

Another part of DeepFM is the deep neural network (DNN) [17], a simple feedforward neural network with the function of learning higher-order feature interactions.

Instead of using FM to pre-train the latent vector V , DeepFM uses V to initialize the neural network, which is different from other related algorithms [17]. It means that this part of FM does not exist independently, it trains and learns together with the whole model. DeepFM itself does not have pre-trained feature vectors. It relies on shared feature embeddings, which are free from cumbersome feature engineering, and it is also effective for both low- and high-order interaction representations. Therefore, DeepFM is effective in reducing the error rate and improving prediction results. The model is both capable of click-through prediction, and other machine learning tasks, such as various recommendations. Simply, it uses and only uses the combined idea of FNN. DeepFM uses FM for embedding, and the results are widely and thoroughly shared.

3 A Systematic Study on DeepFM-Related Work

In the recommendation process, Li L et al. developed a hybrid algorithm for optimization and simulation of recommendation systems on the basis of content and collaborative filtering [1], which builds a joint model and fuses content and collaborative filtering together [20]. It makes the matrix less sparse thus the collaborative filtering achieves better accuracy. He X et al. also proposed neural collaborative filtering based on DNN [2], which is different from traditional filtering algorithms based on matrix decomposition. It uses the NCF framework to learn user-item interaction through neural networks, which opens up a new research direction for deep learning-based recommendation models. But in the future, it is also possible to extend the NCF framework to the level of developing models for users and building recommendation systems for multimedia items.

Wang H et al. raised a deep knowledge-aware network (DKN) [3], which combines the knowledge graph with news recommendation. As a deep recommendation framework based on content, DKN has multi-channel and word entity aligned knowledge-aware CNN fusing the semantic-level and knowledge-level representations of news. In addition, an attention module is designed to address the different interests of users. However, DKN lacks an end-to-end training approach, and it contains little information other than text. Similarly, for knowledge graph-assisted recommendation, there is MKR, another multi-task learning approach [4]. This end-to-end deep recommendation framework

is capable of assisting in recommendation with knowledge graph embeddings. Items and entities may have identical features. To model these features, MKR contains cross-compression units that explicitly model the higher-order interactions. The item and entity representations can complement each other, thus improving the generalization capability.

Similarly, Wang R et al. proposed the deep crossover network (DCN) [9]. It is capable of handling a large number of sparse or dense features. It is composed of multiple layers, each retaining the interactions of the previous one, so that each layer is capable of generating higher-order interactions. Combined with traditional deep representations, it learns explicit crossover features with bounded degrees, effectively capturing predicted feature interactions without the need to manually construct feature interactions [21]. It has good results in click-through rate prediction.

In solving complex problems of high dimensionality and feature with many categories, Zhang W et al. proposed two DNN-based recommendation models – the FM-supported neural network (FNN) and the sampling-based neural network (SNN) [10]. The former reduces dimensionality by supervised learning embedding and using FM to turn sparse features into dense features. The latter is a fully connected network that uses negative sampling – RBM and DAE. Both neural networks can also be extended to a product-based neural network (PNN) [11], which has a product layer to capture the interaction patterns between domains.

Considering sparse and high-dimensional input features, Song W et al. proposed AutoInt, automatic feature interaction learning based on a self-attentive mechanism neural network [12], which can avoid experts performing a manual selection of feature interactions. It proposes a multi-headed attention neural network with skip-connect that explicitly models feature interactions in a low-dimensional space. It models feature interactions in different orders with multiple layers of the multiheaded attention neural network.

Although above approaches can deal with sparse and high-dimensional input features, there are still troubles in embedding and feeding the multilayer perceptron, which cannot effectively capture the different interests of users from so many historical behaviors. So, Zhou G, Zhu X, Song C, et al. proposed the deep interest network (DIN) [5]. Given a certain candidate advertisement, it can learn from its historical behaviors to adaptively compute a representation vector of user interests.

Similarly, An M et al. proposed a neural news recommendation method with both long-term and short-term user representations (LSTUR) [8], containing a news encoder and a user encoder. It first applies a news encoder to learn the representation of news pieces from the headline and topics. The user encoder is then used, allowing for a long-term user representation (LTUR) module and a short-term user representation (STUR) module.

Cheng H T and Koc L et al. proposed Wide & Deep Learning to solve the irrelevant recommendation problem in a sparse environment, which jointly trains a width linear model and a deep neural network to balance the memory capacity and generalization ability of Wide and Deep models to uncover the relevance of rare features that are sparse or even never present to the final label [13].

Lian J et al. proposed a new compressed interaction network (CIN). It explicitly brings about higher-order feature interactions with the benefits of finite higher order,

automatic cross multiplication, and parameter sharing. It further combines a CIN and a classic DNN into an extreme deep factorization machine (xDeepFm) without requiring manual feature engineering [14], which frees data scientists from the tiresome task of feature search.

Meanwhile, in practical applications, low-level and high-level feature interaction does not have the same proportion of information. As a result, the model that they had proposed at first was improved by assigning different weights to linear output [22], second-order combined output, and deep output, respectively.

The description documents of Web API are clustered using Doc2Vec to obtain functional clusters [15]. The multidimensional attributes of the services are extracted with an FM model to mine the higher-order combination and interactions among them. A novel recommendation – Web APIs is proposed by integrating the feature clusters and quality of service of the services. Also, the focus needs to be on multi-feature extraction so that the performance of the model can be improved.

There is another risk evaluation method for imports and exports that consists of fuzzy reasoning and DeepFm [19]. On the basis of pre-processing, numerous historical declarations and inspection data of import and export cargoes are selected. Based on experts' experience, cargo risk evaluation firstly takes the key domain information as its feature index. Then, a fuzzy inference model is constructed using fuzzy theory. In order to carry out the fuzzy normalization of samples, commodity risks could be measured and categorized by DeepFm to build an intelligent comprehensive evaluation system for import and export cargo with fuzzy normalization of commodity risks.

For data dimensionality reduction, the K-means clustering algorithm can cluster the raw log data in consideration of similarity [16]. Later, the relationship of the low-order and high-order feature combination can be obtained by the raw log data by DeepFm parameter sharing strategy, and it will generate a click-through rate prediction model. At last, users will get product recommendation in sequence and give feedback, which comes from the predicted click-through rate. However, more elements such as a spatio-temporal sequence of user behavior, preferences, and surrounding influence need to be modeled.

4 Multi-perspective Discussion

Both the efficiency and the correctness of the model have been improved step by step. This paper considers two levels. Firstly, on the user level, it goes step by step from content and collaborative filtering and neural collaborative filtering to effective interest capture, to the time-to-time interest change, to long-term stable behavior and evolution and the long- and short-term user representation, to diversified interest, and to personalized recommendation and recommendation using recursive diffusion with simulated global social networks. Then, considering the modeling level, it goes from solving the sparse, dense, and high-dimensional dataset, to low-order interaction, high-order interaction, and low- and high-order interaction, to the complex relationship between different features, in order to carry out different interactions, and to the clustering before the interaction. These improve the reliability and speed for mining. In terms of applications, DeepFm and the above improvements can also be applied to product recommendations, technology

Table 1. Deep-Fm-related research

Work	Contribution	Technology	Limitations
A recommendation system based on content and collaborative filtering [1]	Combine the strengths of content and collaboration	Based on content filtering and collaborative filtering	–
Neural collaborative filtering [2]	Learn interactions of users and items with the neural network, giving a new research direction	NCF	Extend NFC to develop models for groups, which can also be applied to recommendations for multimedia projects
DKN [3]	Personalized recommendations	KCNN and attention mechanism	DKN lacks an end-to-end training approach and contains little information other than text
MKR [4]	Knowledge graph-embedded	Cross compression units	How to use this framework to research other neural networks
DCN [9]	Deal with sparse or dense data	Multiple layers, each retaining the interaction with the previous layer	–
FNN and SNN [10]	Solve high-dimensional discrete features	DNN added	FNN has no high-order feature interaction
PNN [10]	Solve high-dimensional sparse features	An embedded layer and a product layer: IPNN based on the inner product and OPNN based on the outer product	The inner product is complex and the outer product is unstable. The performance comparison of network depth and the product layer needs to be enhanced to make it interpretable and visualized

(continued)

Table 1. (continued)

Work	Contribution	Technology	Limitations
Autolnt [12]	Solve high-dimensional and sparse data	Multi-headed attention neural network	Unable to capture different interests. Context can be incorporated and applied to general machine learning
DIN [5]	Effectively capture interests	Local activation unit	Interest dysfunction
DIEN [6]	Interest changes from time to time	Extract potential temporal interest and model the interest evolution process	Need to be more personalized
Binn [7]	Stabilization and evolution of long-term behavior	Neural Item Embedding and Distinctive Behavior Learning	Need to study the impact of the behavior of different types of users
LSTUR [8]	Long- and short-term user representation	News encoder and user encoder	–
Wide&Deeplearning [13]	Low-order and high-order interactions	The Wide model and the Deep model	Wide needs professional feature engineering experience
xDeepFm [14]	Generation of high-order feature interactions	The combination of CIN and a classic DNN	–
The combination of DeepFm and IoMT	Improve DeepFm	Assigned different weights to the linear output, the second-order combined output, and the deep output, respectively	–
Web APIs [15]	Clustering and mining interactions	Doc2Vec and DeepFm models	Focus on multi-feature extraction
DCNN [22]	Multimodal feature extraction	Multimodal feature extractor and rumor detector	Able to predict

(continued)

Table 1. (continued)

Work	Contribution	Technology	Limitations
Personalized product recommendation [20]	Improve the efficiency of personalized product recommendation	K-means	Model more elements

resource recommendations, news recommendations, advertising recommendations, Web services, and many other recommendations (see Table 1). They also have great results and applications in rumor detection, intrusion detection, predicting Chinese Internet users' interest in social software, and risk evaluation of import and export cargoes. In the future, DeepFm can also add expert opinions, after all, some specific fields need professionals to provide a reference. In addition, we found that DeepFm is also applicable to Android app store scoring. But at the same time, manual feature engineering can also be added in some necessary environments. If expert opinion is added, the scoring and accuracy will be improved to a large extent.

5 Conclusion

In this paper, we provide an extensive discussion of existing traditional recommendation models, deep learning models, DeepFm, and their improvements and applications. We explore the following perspectives, including solving data sparsity and high dimensionality, personalized recommendation, capturing interests and changes in interests, short-term and long-term representation of interests, recursive diffusion, diversified interests, and DeepFm and its improvements and applications. We also summarize the advantages and limitations of these models. In addition, we have also investigated methods including clustering and multimodal feature extraction, which simplify data complexity before being used in models to enhance efficiency and precision. Finally, this study also presents research on the applications of DeepFm in other fields and its future improvement. We hope that this study will enable more readers to learn about the methods and efficacy in this field and draw inspiration of other researchers.

References

1. Li, L., Zhang, Z., Zhang, S.: Hybrid algorithm based on content and collaborative filtering in recommendation system optimization and simulation. *Sci. Program.* **2021**, 1–11 (2021)
2. He, X., Liao, L., Zhang, H., et al.: Neural collaborative filtering. In: *Proceedings of the 26th International Conference on World Wide Web*, pp. 173–182 (2017)
3. Wang, H., Zhang, F., Xie, X., et al.: DKN: deep knowledge-aware network for news recommendation. In: *Proceedings of the 2018 World Wide Web Conference*, pp. 1835–1844 (2018)
4. Wang, H., Zhang, F., Zhao, M., et al.: Multi-task feature learning for knowledge graph enhanced recommendation. In: *The World Wide Web Conference*, pp. 2000–2010 (2019)

5. Zhou, G., Zhu, X., Song, C., et al.: Deep interest network for click-through rate prediction. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1059–1068 (2018)
6. Zhou, G., Mou, N., Fan, Y., et al.: Deep interest evolution network for click-through rate prediction. Proc. AAAI Conf. Artif. Intell. **33**(01), 5941–5948 (2019)
7. Li, Z., Zhao, H., Liu, Q., et al.: Learning from history and present: next-item recommendation via discriminatively exploiting user behaviors. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1734–1743 (2018)
8. An, M., Wu, F., Wu, C., et al.: Neural news recommendation with long-and short-term user representations. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 336–345 (2019)
9. Wang, R., Fu, B., Fu, G., et al.: Deep & cross network for ad click predictions. In: Proceedings of the ADKDD 2017, pp. 1–7 (2017)
10. Zhang, W., Du, T., Wang, J.: Deep learning over multi-field categorical data. In: Ferro, N. et al. ECIR 2016. LNCS, vol. 9626, pp. 45–57. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30671-1_4
11. Qu, Y., Cai, H., Ren, K., et al.: Product-based neural networks for user response prediction. In: 2016 IEEE 16th International Conference on Data Mining (ICDM), pp. 1149–1154. IEEE (2016)
12. Song, W., Shi, C., Xiao, Z., et al.: AutoInt: automatic feature interaction learning via self-attentive neural networks. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 1161–1170 (2019)
13. Cheng, H.T., Koc, L., Harmsen, J., et al.: Wide & deep learning for recommender systems. In: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, pp. 7–10 (2016)
14. Lian, J., Zhou, X., Zhang, F., et al.: xDeepFm: combining explicit and implicit feature interactions for recommender systems. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1754–1763 (2018)
15. Zhang, X., Liu, J., Cao, B., et al.: Web service recommendation via combining Doc2Vec-based functionality clustering and DeepFm-based score prediction. In: 2018 IEEE International Conference on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCLOUD/SocialCom/SustainCom). IEEE (2018)
16. Xu, J., Hu, Z., Zou, J.: Personalized product recommendation method for analyzing user behavior using DeepFm. J. Inf. Process. Syst. **17**(2), 369–384 (2021)
17. Guo, H., Tang, R., Ye, Y., et al.: DeepFm: a factorization-machine based neural network for CTR prediction. arXiv preprint [arXiv:1703.04247](https://arxiv.org/abs/1703.04247) (2017)
18. Bo, L., Xiao, Z., Kewei, L., et al.: Analysis of the click through rate of Chinese Netizens through social software based on DeepFm algorithm. J. Phys.: Conf. Ser. **1284**(1), 012025 (2019). IOP Publishing
19. Xu, Y., Fang, H., Luo, J., et al.: Risk evaluation method of import and export goods based on fuzzy reasoning and DeepFm. Math. Probl. Eng. **2021**, 1–8 (2021)
20. Adomavicius, A., Tuzhilin, A.: Towards the next generation of recommender systems. IEEE Trans. Knowl. Data Eng. (2010)
21. Shan, Y., Hoens, T.R., Jiao, J., et al.: Deep crossing: web-scale modeling without manually crafted combinatorial features. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 255–262 (2016)
22. Yu, Z., Amin, S.U., Alhussain, M., et al.: Research on disease prediction based on improved DeepFm and IoMT. IEEE Access **9**, 39043–39054 (2021)