Deep-Cross-Attention Recommendation Model for Knowledge Sharing Micro Learning Service

Jiayin Lin[1(✉)], Geng Sun[1(✉)], Jun Shen[1,2(✉)], David Pritchard[2(✉)],
Tingru Cui[3(✉)], Dongming Xu[4(✉)], Li Li[5(✉)], Ghassan Beydoun[6(✉)],
and Shiping Chen[7(✉)]

[1] School of Computing and Information Techonology,
University of Wollongong, Wollongong, Australia
jl461@uowmail.edu.au, {gsun,jshen}@uow.edu.au
[2] Research Lab of Electronics, Massachusetts Institute of Technology,
Cambridge, MA, USA
dpritch@mit.ed
[3] University of Melbourne, Melbourne, Australia
tingru.cui@unimelb.edu.au
[4] UQ Business School, The University of Queensland, Brisbane, Australia
d.xu@business.uq.edu.au
[5] Faculty of Computer and Information Science, Southwest University,
Chongqing, China
lily@swu.edu.cn
[6] School of Information, System and Modelling,
University of Technology Sydney, Sydney, Australia
ghassan.beydoun@uts.edu.au
[7] Data 61, CSIRO, Sydney, NSW, Australia
shiping.chen@csiro.au

Abstract. Aims to provide flexible, effective and personalized online learning service, micro learning has gained wide attention in recent years as more people turn to use fragment time to grasp fragmented knowledge. Widely available online knowledge sharing is one of the most representative approaches to micro learning, and it is well accepted by online learners. However, information overload challenges such personalized online learning services. In this paper, we propose a deep cross attention recommendation model to provide online users with personalized resources based on users' profile and historical online behaviours. This model benefits from the deep neural network, feature crossing, and attention mechanism mutually. The experiment result showed that the proposed model outperformed the state-of-the-art baselines.

Keywords: Recommender system · Neural network · Micro learning · Machine learning · Information retrieval

1   Introduction

As a novel online learning style, micro learning aims to utilize users' fragmented spare time by helping them to carry out effective personalized learning activities [1–3]. Such online learning activities could be formal, informal, and non-formal [4], and online knowledge sharing is one way of non-formal learning. Quora,[1] Zhihu, [2]and Stack-overflow[3] are the most representative and successful online knowledge platforms, where users share knowledge by asking and answering questions. In the meantime, the online platforms continuously recommend questions and topics to the users based on their interests, background, and learning requirements.

As the key to the personalized online learning service, the recommendation strategy determines what information will be finally delivered to the target user [5]. As for a new online learning service in the big data era, conventional recommendation strategies, such as collaborative filtering and content-based filtering [6], are no longer suitable for catering the personalized learning requirements. A recommender system always needs to handle and merge different types and format of information ranging from the user's profile to the resource's profiles. Moreover, higher-order feature interaction is crucial for good performance [7]. How to precisely weight different features is also vital for a recommender system, as different features have various importance levels for a personalized recommendation task [8].

In this paper, we propose a novel model, which combines several advantages from different state-of-the-art recommender systems and offers them in a smooth one-stop manner. The rest of this paper will be organized as follows. Section 2 discusses some prior related work about recommender system used in micro learning. The proposed model is introduced and explained in Sect. 3. The relevant experiment of this study is discussed and analysed in Sect. 4. The conclusions are discussed in Sect. 5.

2   Related Work

The recommendation problem has been investigated for many years in different domains. However, the recommendation task in online education always involves some unique requirements or characteristics [9, 10]. In one prior study [11], the ant colony optimization (ACO) algorithm was proposed to recommend personalized learning paths to users based on the demographic information. The ontology-based method was used to add extra user's profile information and relieve the cold-start problem for micro learning service [12, 13]. Another study [14] investigated the learning path recommendation from micro learning service from an exploitation perspective. So far, there are little efforts on deep learning solutions to this problem.

Feature interaction means features involved in a recommendation task tend to influence each other with various combinations. Factorization machine (FM) [15] uses

---

[1] https://www.quora.com/.

[2] https://www.zhihu.com/.

[3] https://stackoverflow.com/.

embedding techniques to model the latent features in low dimensional space and represents the pair-wise feature interactions by using the inner product. It also shows a satisfactory performance when the dataset is in high sparsity, whereas SVMs fails [15]. However, due to the high computational complexity, in many cases, only 2-order feature interactions are involved in the FM.

Deep learning has demonstrated its powerful strength in modelling non-linear transformation in various AI tasks. Besides using deep neural for a recommendation task in isolation (for example [16]), many researchers argue that combining the advantages of deep neural networks (DNN) with classical methods such as linear model or FM could better learn sophisticated feature interactions [17–19].

3   The Proposed Model

In this study, we aim to effectively combine these functionalities: mining and generating high-order feature interaction, distinguishing the importance difference of both implicit and explicit features, and maintaining the original input information in a single network. To this end, we proposed a new deep cross attention network (DCAN) model for the recommendation task of the online knowledge sharing service. The input of the model contains both user-side and question-side information, and the embedding layer maps such information onto a low dimensional space. The embedding vectors are then passed into the DNN network and crossing network separately for mining latent information and high-order feature interactions. The processed results are combined together, and an attention network is used to distinguish the importance differences of different features. Finally, the output layer is used to make predictions with weighted features.

4   Experiments and Analysis

4.1   Evaluation Metrics and Baselines

Evaluation Metrics. As a binary classification task, the first evaluation metric used is Area Under Curve (AUC), which indicates how much a model is capable of distinguishing the two labels. Another metric used in our experiments is mean squared error (MSE), which directly reflects the prediction error of the involved models. Moreover, we also compared the binary cross entropy of the involved models.

Baselines. We compared our model with several state-of-the-art recommendation models, ranging from DeepFM [17], AutoInt [7], DCN [20], AFM [21], and FM [15]. The characteristics of used baselines are introduced in the previous sections.

Table 1. Experiment results of different models

Model	AUC	MSE	Binary cross entropy
FM	0.6934	0.1243	0.4060
DCN	0.7603	0.1134	0.3690
AFM	0.6881	0.1255	0.4094
AutoInt	0.7613	0.1130	0.3679
DeepFM	0.7404	0.1128	0.3671
Proposed model	**0.7848**	**0.1071**	**0.3442**

4.2   Dataset

The dataset is collected from an online knowledge-sharing platform, which contains around 1.8 million questions and users, and more than 4 million answers for the questions. Nearly 10 million <question, user> pairs are involved in this dataset.

4.3   Experiment Results

Based on the experiment results from Table 1, we can clearly see FM and AFM have lowest AUC values and highest MSE scores. These two models only involve low-order feature interactions. While others involve high-order feature interactions. Hence, high-order (complex) feature interactions are vital in the online learning resource recommendation tasks.

According to Table 1, the AUC scores of our proposed model and AutoInt model are the highest two. These two models refine the results of high-order feature interaction via the attention mechanism [22]. Such performance improvement demonstrates that different features/feature combinations are not equally important for personalized learning service, and attention mechanism can automatically distinguish the importance differences of the latent features or the feature combinations generated by the prior layers of the network.

5   Conclusions

In this study, we proposed a deep cross attention network (DCAN) for recommending personalized online learning resources to online learners. The experiment results clearly demonstrated that our model had potential in handling complex online learning recommendation problem. More specifically, according to the experiment results with authentic online knowledge sharing data, the strengths of DCAN can be concluded into two points: 1.this model can automatically mine and generate high-order feature interactions in both explicit and implicit ways; 2. the proposed model can further distinguish the importance differences of different features.

References

1. Lin, J., et al.: From ideal to reality: segmentation, annotation, and recommendation, the vital trajectory of intelligent micro learning. World Wide Web, 1–21 (2019). http://dx.doi.org/10.1007/s11280-019-00730-9

2. Lin, J., et al.: A survey of segmentation, annotation, and recommendation techniques in micro learning for next generation of OER. In: 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD), pp. 152–157. IEEE (2019)

3. Sun, G., Cui, T., Yong, J., Shen, J., Chen, S.: MLaaS: a cloud-based system for delivering adaptive micro learning in mobile MOOC learning. IEEE Trans. Serv. Comput. **11**(2), 292–305 (2015)

4. Eshach, H.: Bridging in-school and out-of-school learning: formal, non-formal, and informal education. J. Sci. Educ. Technol. **16**(2), 171–190 (2007). https://doi.org/10.1007/s10956-006-9027-1

5. Lin, J., et al.: Towards the readiness of learning analytics data for micro learning. In: Ferreira, J.E., Musaev, A., Zhang, L.-J. (eds.) SCC 2019. LNCS, vol. 11515, pp. 66–76. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23554-3_5

6. Pazzani, M.J.: A framework for collaborative, content-based and demographic filtering. Artif. Intell. Rev. **13**(5-6), 393–408 (1999)

7. Song, W., 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. ACM (2019)

8. Huang, T., Zhang, Z., Zhang, J.: FiBiNET: combining feature importance and bilinear feature interaction for click-through rate prediction. In: Proceedings of the 13th ACM Conference on Recommender Systems, pp. 169–177 (2019)

9. Sikka, R., Dhankhar, A., Rana, C.: A survey paper on e-learning recommender system. Int. J. Comput. Appl. **47**(9), 27–30 (2012). https://doi.org/10.5120/7218-0024

10. Wu, D., Lu, J., Zhang, G.: A fuzzy tree matching-based personalized e-learning recommender system. IEEE Trans. Fuzzy Syst. **23**(6), 2412–2426 (2015). https://doi.org/10.1109/TFUZZ.2015.2426201

11. Zhao, Q., Zhang, Y., Chen, J.: An improved ant colony optimization algorithm for recommendation of micro-learning path. In: 2016 IEEE International Conference on Computer and Information Technology (CIT), pp. 190–196. IEEE (2016)

12. Sun, G., Cui, T., Shen, J., Xu, D., Beydoun, G., Chen, S.: Ontological learner profile identification for cold start problem in micro learning resources delivery. In: 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), pp. 16–20. IEEE (2017)

13. Sun, G., Cui, T., Xu, D., Shen, J., Chen, S.: A heuristic approach for new-item cold start problem in recommendation of micro open education resources. In: Nkambou, R., Azevedo, R., Vassileva, J. (eds.) ITS 2018. LNCS, vol. 10858, pp. 212–222. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91464-0_21

14. Rusak, Z.: Exploitation of micro-learning for generating personalized learning paths. In: DS 87-9 Proceedings of the 21st International Conference on Engineering Design (ICED 17), 21–25 August 2017, vol 9, pp. 129–138. Design Education, Vancouver (2017)
15. Rendle, S.: Factorization machines. In: 2010 IEEE International Conference on Data Mining, pp. 995–1000. IEEE (2010)
16. Zhang, W., Du, T., Wang, J.: Deep learning over multi-field categorical data. In: Ferro, N., et al. (eds.) ECIR 2016. LNCS, vol. 9626, pp. 45–57. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30671-1_4
17. Guo, H., Tang, R., Ye, Y., Li, Z., He, X.: DeepFM: a factorization-machine based neural network for CTR prediction. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, pp. 1725–1731 (2017)
18. Lian, J., Zhou, X., Zhang, F., Chen, Z., Xie, X., Sun, G.: 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. ACM (2018)
19. Cheng, H.-T., et al.: Wide & deep learning for recommender systems. In: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, pp. 7–10. ACM (2016)
20. Wang, R., Fu, B., Fu, G., Wang, M.: Deep & cross network for ad click predictions. In: Proceedings of the ADKDD 2017. pp. 1–7 (2017)
21. Xiao, J., Ye, H., He, X., Zhang, H., Wu, F., Chua, T.-S.: Attentional factorization machines: learning the weight of feature interactions via attention networks. In: International Joint Conference on Artificial Intelligence, pp. 3119–3125 (2017)
22. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)