2020

# Attention-based High-order Feature Interactions to Enhance the Recommender System for Web-based Knowledge-Sharing Servic

Jiayin Lin
*University of Wollongong*, jl461@uowmail.edu.au

Geng Sun
*University Of Wollongong*, gsun@uow.edu.au

Jun Shen
*University of Wollongong*, jshen@uow.edu.au

Tingru Cui
*University of Wollongong*, tingru@uow.edu.au

David Pritchard

*See next page for additional authors*

# Attention-based High-order Feature Interactions to Enhance the Recommender System for Web-based Knowledge-Sharing Servic

## Abstract

Providing personalized online learning services has become a hot research topic. Online knowledge-sharing services represents a popular approach to enable learners to use fragmented spare time. User asks and answers questions in the platform, and the platform also recommends relevant questions to users based on their learning interested and context. However, in the big data era, information overload is a challenge, as both online learners and learning resources are embedded in data rich environment. Offering such web services requires an intelligent recommender system to automatically filter out irrelevant information, mine underling user preference, and distil latent information. Such a recommender system needs to be able to mine complex latent information, distinguish differences between users efficiently. In this study, we refine a recommender system of a prior work for web-based knowledge sharing. The system utilizes attention-based mechanisms and involves high-order feature interactions. Our experimental results show that the system outperforms known benchmarks and has great potential to be used for the web-based learning service.

## Keywords

web-based, servic, system, attention-based, high-order, feature, interactions, enhance, knowledge-sharing, recommender

## Disciplines

Engineering | Science and Technology Studies

## Publication Details

Lin, J., Sun, G., Shen, J., Cui, T., Pritchard, D., Xu, D., Li, L., Wei, W., Beydoun, G. & Chen, S. (2020). Attention-based High-order Feature Interactions to Enhance the Recommender System for Web-based Knowledge-Sharing Servic. International Conference on Web Information Systems Engineering United States: IEEE.

## Authors

Jiayin Lin, Geng Sun, Jun Shen, Tingru Cui, David Pritchard, Dongming Xu, Li Li, Wei Wei, Ghassan Beydoun, and Shiping Chen

# Attention-based High-order Feature Interactions to Enhance the Recommender System for Web-based Knowledge-Sharing Service

Jiayin Lin[1], Geng Sun[1], Jun Shen[1], Tingru Cui[2], David Pritchard[3], Dongming Xu[4], Li Li[5], Wei Wei[6], Ghassan Beydoun[7], Shiping Chen[8]

[1]School of Computing and Information Technology, University of Wollongong, Wollongong, Australia

[2] School of Computing and Information Systems, The University of Melbourne, Melbourne, Australia

[3]Research Lab of Electronics, Massachusetts Institute of Technology, Cambridge, MA

[4]UQ Business School, The University of Queensland, Brisbane, Australia

[5]Faculty of Computer and Information Science, Southwest University, Chongqing, China

[6]School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, China

[7]School of Information, System and Modelling, University of Technology Sydney, Sydney, Australia

[8]Data 61, CSIRO, Sydney, NSW, Australia

jl461@uowmail.edu.au, {gsun, jshen}@uow.edu.au, tingru.cui@unimelb.edu.au, dpritch@mit.ed, d.xu@business.uq.edu.au, lily@swu.edu.cn, weiwei@xaut.edu.cn, ghassan.beydoun@uts.edu.au, shiping.chen@csiro.au

**Abstract.** Providing personalized online learning services has become a hot research topic. Online knowledge-sharing services represents a popular approach to enable learners to use fragmented spare time. User asks and answers questions in the platform, and the platform also recommends relevant questions to users based on their learning interested and context. However, in the big data era, information overload is a challenge, as both online learners and learning resources are embedded in data rich environment. Offering such web services requires an intelligent recommender system to automatically filter out irrelevant information, mine underling user preference, and distil latent information. Such a recommender system needs to be able to mine complex latent information, distinguish differences between users efficiently. In this study, we refine a recommender system of a prior work for web-based knowledge sharing. The system utilizes attention-based mechanisms and involves high-order feature interactions. Our experimental results show that the system outperforms known benchmarks and has great potential to be used for the web-based learning service.

**Keywords:** Recommender System, Neural Network, Web-based Learning, Machine Learning, Information Retrieval.

# 1 Introduction

Web-based learning services aim to effectively utilize mobile devices to conduct real-time personalized learning activities [1]. Online knowledge sharing is one representative and popular informal learning style [2]. Online users post and answer questions from various discipline in a knowledge-sharing platform (like Quora[1] and Stackoverflow[2]), and the platform engages users with new questions based on their profile and historical activities. However, the plethora of user interests and backgrounds could easily result massive volumes of options and can induce disengagement, i.e. producing questions that have the opposite effect. Hence, a web-based knowledge-sharing platform has to rely on a sophisticated recommender system to filter out irrelevant information to truly create a personalized learning service.

An effective recommender system needs to handle and merge different types and formats of information from both the users' profiles and historical activities and the resources profiles. Higher-order feature interaction (combination) is also crucial for good performance [3]. Generating high-order feature interaction manually requires strong domain background, and is very time consuming and labour-intensive, which make it impractical for the large-scale online system, in the context of big data. Furthermore, different features have various importance levels for a personalized recommendation task [4]. How to precisely distinguish the importance differences of different features for a specific user is also vital for a personalized web-based learning service. Conventional recommendation strategies such as simple collaborative filtering and content-based filtering [5] are no longer adequate to handle massive, complex, dynamic data due to their drawbacks in scalability and modelling higher order features.

In this paper, we combine high-order feature interactions and the attention mechanisms to refine a recommender system proposed in one previous work [6]. This enables automatic exploration of high-order feature interactions, differentiating the important degrees for different features, and mining latent features from the original input.

The rest of this paper will be organized as follows. The prior work of recommendation strategies used for online learning services will be firstly reviewed in Section 2. Different state-of-the-art mechanisms proposed in recent research of deep learning will also be introduced in Section 2. In Section 3, the architecture of the proposed model and the relevant technical details of each component will be presented and explained. The experimental results will be presented and analysed in Section 4. The summary of this research and the future plan will finally be discussed in Section 5.

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

# 2 Related Work

## 2.1 Recommendation in Online Learning

Recommender systems have been studied for many years based in various application areas. However, due to the pedagogical considerations [7, 8], a recommendation strategy used in other domains cannot be transformed to fulfil the delivery of online learning services. This is particularly true for the delivery of micro learning services. Hence, this knowledge sharing task still lacks a sophisticated solution to enable recommending personalized online resources to target users.

A blended model was proposed in prior study [9], which combined a learning management system, a set of web 2.0 tools and the e-learning recommender system to enhance personalized online learning. However, this study did not provide technical innovations of a recommender system for online learning service. In an early survey [8], several recommendation approaches for e-learning service are listed and analysed. However, they were all too preliminary to be applied to web-based informal learning services. Another study [10] proposed a hybrid recommendation algorithm which combined the collaborative filtering and sequential pattern mining together for a peer-to-peer learning environment. Learning path recommendation is investigated in [11] and [12]. However, the proposed models were constructed mainly based on the demographic information, which does not provide much scope for exploring individual preferences.

## 2.2 Factorization Machine and Deep Network

In many recommendation scenarios, features involved impact each other (called feature interaction). For example, the feature pair (learning interests, knowledge level) determines the difficulty of a specific course for a learner. A classic work Factorization machine (FM) [13] use the inner product to model the feature interaction. This alleviates the data sparsity problem by using *embeddings* to represent the user and the item. However, due to computational and space complexity, only up to second-order feature interaction can be applied to many real-world applications.

Deep learning has been used in many application areas [14-16] and has demonstrated its outstanding ability to model complex problems. Hence, many researchers are investigating combining deep learning technique with conventional recommendation strategies. One representative model proposed by Google is ‘Wide&Deep’. This combines the benefits of memorization and generalization by using a linear model and a deep network [17]. Both low-order and high-order feature interactions are also investigated in [18] through FM and DNN components. As the DNN models high-order interactions in an implicit manner, the learned results can be arbitrary; in another work, an extreme deep factorization machine (xDeepFM) is proposed for generating and modelling feature interactions of a bounded degree [19]. Similarly, model [19] also contains two components, a compressed interaction network (CIN) and a DNN. The CIN and DNN learn the explicit and implicit feature interaction, respectively.

### 2.3 Mining Latent Information

With the successes of 'Wide&Deep' and DeepFM [18], multi-component network structure are increasingly popular. Such structure shows outstanding performance in applying different techniques to mining latent information from different perspectives simultaneously. Among them, feature interaction and weighting strategies are mostly benefitting from such multi-component structure network.

**Feature Interaction.** This is a fundamental problem and plays a significant role in a recommendation task. There are also many prior studies [3, 20] mainly focusing on feature interaction strategies. The study [20] proposed a novel cross-network, which explicitly applied feature interaction in each layer; and the cross-network consisted of all the cross-terms of degree up to the highest. In another study [3], key-value attention mechanism was used for determining which feature combinations were meaningful.

**Attention and Residual Mechanism.** Attention mechanism has been widely used in many areas, such as computer vision and natural language processing. This allows the network to pay different degrees of attention to different parts. Attentional factorization machine was proposed in [21], which could distinguish the importance differences of various feature combinations. Instead of simple attention network, multi-head attention mechanism [22] was also used another recommendation strategy [3]. That showed the ability to explore meaningful feature combinations through different non-linear transformations. Squeeze-and-Excitation network (SENET) [23] was used in the study of Feature Importance and Bilinear feature Interaction network (FiBiNET) [4]. SENET was used to make the model pay more attention to the important features and decrease the weight of uninformative features through inner product and Hadamard product.

Demonstrated in [16], the idea of residual shows outstanding performance in stabilizing the optimization process of a deep network. Moreover, the residual function can also improve the model performance by providing sufficient information from previous layers of the network. Hence, for the recommendation task, as the network becomes deeper, many researchers start involving the residual connection (unit) in some components of the network. One prior study [24] used residual units to implicitly perform specific regularizations leading to better stability. Similarly, in the crossing component of the 'Deep&Cross' Network [20], the residual unit was used in each crossing layer to add the current input information back. In [3], standard residual connections were added in both interaction and output layers to achieve a hierarchical operation manner.

# 3 Solution

In this section, we firstly propose three hypotheses which might be significant to web-based knowledge sharing service. The design of the model and technical details of each component is then presented and discussed.

## 3.1 Hypotheses

In this study, the proposed model is designed based on the following hypotheses:

1. High-order feature interaction is vital to further improve the performance of a recommender system which used for a web-based big data application. Low-order feature interaction cannot sufficiently mine and model the underlying complex feature interactions for informal learning service.
2. The features involved in a learning platform have different important degrees. Precisely differentiate the feature importance is vital to a personalized learning service, and it can further improve the recommendation results.
3. The proposed model also holds moderate efficiency. For a web-based learning service in the big data era, efficiency is also an important indicator.

For example, the proposed model manages to recommend the question to the user that might be interested in. The give question is about machine learning, the difficult level of this question is entry-level. And we have two users with the following features:

*User_1: (Interested topic: computer science, Occupation: student, Gender: male, Age:23, Location: Australia)*

*User_2: (Interested topic: computer science, Occupation: research fellow, Gender: male, Age 32, Location: China)*

For the example, the proposed model should effectively and automatically generate meaningful feature combinations such as *(Interested topic, Occupation)* and *(Gender, Age)*, distinguish the importance difference between different features such as for a give question the feature *(Interested topic, Occupation)* might be more important than the feature *(Gender, Age)* and decide that the *User_1* might be more interested in this question.

## 3.2 Model Architecture

Based on the above hypotheses and one previous proposed initial idea [6], the general architecture of our model is shown in Figure 1. Our proposed model contains three significant networks: a cross-network for exploring feature interactions, a deep network for mining latent information, and an attention network for distinguishing the importance of different features. The input of the model is high dimensional vectors which contain both user and item relevant information. The output of the model are decimals range from 0~1 indicate how much a user will be interested in a given question.

### 3.3 Embedding

For a recommendation task, the input contains many highly sparse categorical data, such as the genre describing the discipline of the educational resource and it may be multi-valued (for example, subject of 'machine learning' could belong to both disciplines of mathematics and computer science). In our proposed model, we apply an embedding layer to reduce the dimensionality and sparsity of the raw data. The raw input contains history of interaction between user and item and side information of the user and item. The embedding operation could not only reduce the computational workload, but also boost the model performance. This process can be formulated as follows:

$$X_{embed,i} = W_{embed,i}\, X_i \tag{1}$$

Where $X_{embed,i}$ is the embedding result of the i-th categorical feature, $W_{embed,i}$ is the embedding matrix that maps the i-th original categorical feature into the low dimensional space, and $X_i$ is the i-th feature.

### 3.4 Cross Network

The cross-network used in this study is based on the method proposed in [20]. The cross network is used to automatically generate the high-order feature interaction. Such network consists of several layers, and each layer is an operation of feature interaction. For each interaction layer, the operation of feature crossing can be simply formulated as follows:

$$X_{l+1} = X_0 X_l^T W_l + b_l + X_l \tag{2}$$

Where $X_l$ is the output of the $l$-th crossing layer, $W_i$ and $b_i$ are weights and bias parameters of each crossing-layer. As demonstrated in [20], such special structure of network can increase the interaction degree as the network goes deep, with the highest $n+1$ polynomial degree of the $n$-th layer. Moreover, regarding the efficiency of the network [20], the time and space complexity are both linear in input dimension.

### 3.5 Deep Network

For the simplicity and generalization, a conventional fully connected neural network is used in the proposed model as the deep component. The deep network implicitly captures the latent information and feature combinations. Each layer of the DNN network can be formulated as follow:

$$h_{l+1} = f(W_l h_l + b_l) \tag{3}$$

Where $h_l$ denotes the output of the $l$-th layer of the deep component, $f(\cdot)$ is the activate function, where ReLU is used in this study. $W_l$ and $b_l$ are parameters of the $l$-th layer of the deep network.

### 3.6 Residual Connection and Attention Network

Before providing transformed information to the attention network, the original input information is continuously added to the output of the deep network and the cross network by using residual. This aims to maintain the original input information, which might suffer information loss after going through several layers of neural network.

An attention network is applied right after the combination layer to interpret the important difference of various features. The attention mechanism can be formulated as follows:

$$a_i' = ReLU(WX_i + b) \tag{4}$$

$$a_i = \frac{\exp(a_i')}{\sum_i \exp(a_i')} \tag{5}$$

where, both $W$ and $b$ are model parameters. The attention score is calculated through Softmax function. The calculated attention scores are projected back to the output of the combination layer. The process can be formulated as follows:

$$X_s = a_i W X \tag{6}$$

Where, $a_i$ is the calculated attention score, $W$ is the weight adopted in the network, and the $X$ is the output of the former combination layer, and $X_s$ is the final value after attention mechanism is applied. The demonstration of the attention operation is shown in Figure 2. Where $\boldsymbol{f_1}$ to $\boldsymbol{f_n}$ are the latent features passed into the attention network, the outputs of the network are attention scores for latent features with Softmax function. Lastly, each score is assigned to the feature by Hadamard product.

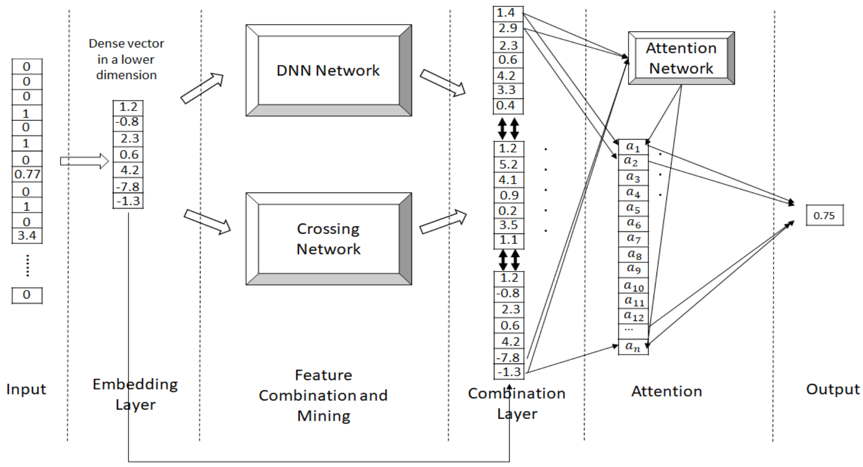


**Figure. 1** The Overall Network Structure of the Proposed Cross Attention Boosted Recommender System

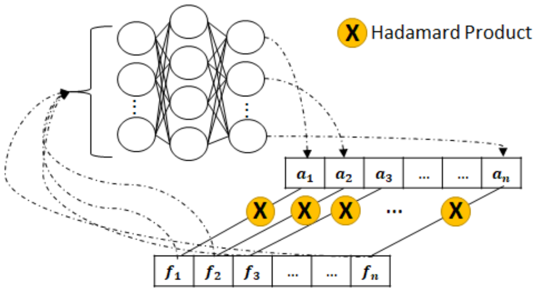


**Figure 2.** Visualization of the Attention Operation

# 4 Experiments and Analysis

In this section, we compare our proposed model with several state-of-the-art recommendation strategies.

## 4.1 Evaluation Metrics and Baselines

**Evaluation Metrics.** In the experiment, we used the Area Under Curve (AUC) as the main criteria to evaluate the performance of each model. The proposed is a binary classifier which predicts whether a user will be interested in a given question, and the AUC can measure the capability of a model in distinguishing two labels. The calculation of the AUC used in this study is calculated as follow:

$$\mathrm{AUC} = \frac{\sum_{i\epsilon positiveClass} rank_i - \frac{M(1+M)}{2}}{M \times N} \tag{7}$$

Where M and N are the number of positive and negative samples respectively, $rank_i$ is the location of the i-th sample. We also used mean square error (MSE) and binary cross entropy to reflect the errors that made by each model. The binary cross entropy used in this study is formulate as:

$$\mathrm{H} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log\big(p(y_i)\big) + (1 - y_i)\log(1 - p(y_i)) \tag{8}$$

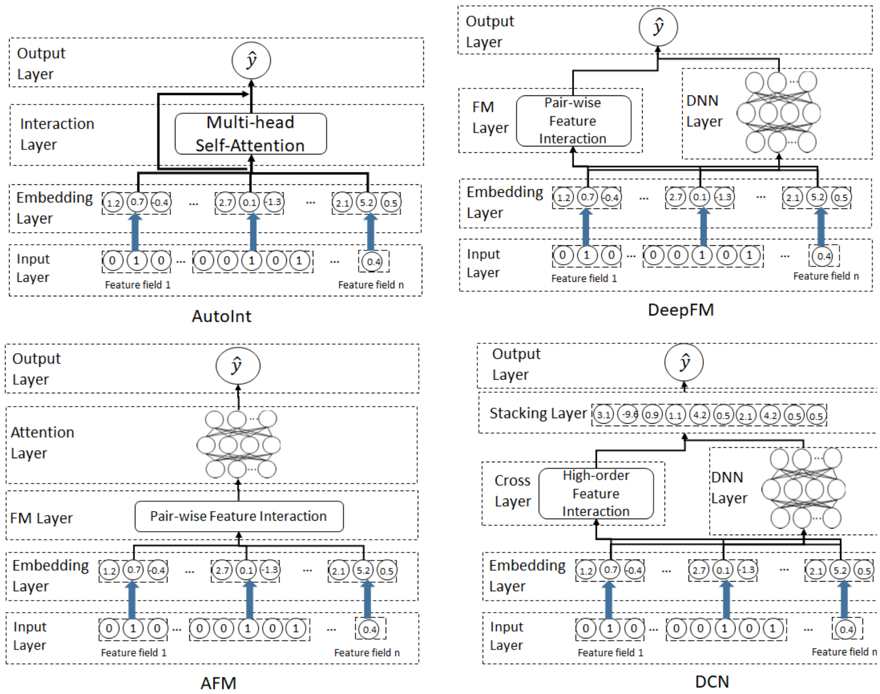


**Figure 3.** Overall Model structure of AutoInt, DeepFM, AFM, and DCN

**Baselines.** AutoInt [3], FM, DeepFM, Deep&Cross Network (DCN), Attention Factorization Machine (AFM) are used as baselines in the experiment. The overall architecture comparison of AutoInt, DeepFM, AFM, and DCN is shown in Figure 3. Each of these models contains an embedding layer, a feature interaction layer, and uses Softmax function to make a prediction. The main difference among these models is the choice of techniques for feature interaction, where multi-head self-attention is used in AutoInt, the combination of the FM and the DNN is used in DeepFM, the combination of FM and simple forward attention network is used in AFM, and combination of the cross-network and the FM is used in DCN.

## 4.2 Dataset

The dataset used in the experiment contains 10 million (question, user) pair, which was collected from Zhihu[3]. The user here stands for an online learner who is or is not a participant of a certain question. This dataset also contains other side-information about the users and the questions, such as the answers to the question, categorical information (such as gender) about the user, and the user's learning interests. Moreover, the label of the dataset is unbalanced (the ratio of negative label to positive label

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

is around 4), which reflects the real and typical online application scenario, where for most users, only a small amount of questions they would like to answer due to various reasons, such as ‘pedagogical lurking’. A negative sample stands for recommending a question to a user, but the user does not participate any learning activities; while a positive sample stands for recommending a question to a user and the user participate a certain of learning activity which could be answering the question or commenting the answers given by other users. As discussed in many pedagogical studies [25-27], for the online learning service, it is very difficult to enable learners interacting with each other like offline learning, even if the online learners have great and similar interests in the current learning session.

### 4.3 Experimental Setup

All the models involved in the experiments are implemented using PyTorch [28]. Each categorical feature was represented as an embedding vector with six dimensions. All the non-linear transformations were activated by ReLU except the output layer, which was activated by softmax function. All baselines were implemented strictly follow the suggestions and guideline of their original research. The early-stop mechanism is applied to all models involved in this experiment in order to prevent overfitting. Ten-fold-cross-validation is applied in the experiment.

### 4.4 Experiment Results and Analysis

The comparative experiment result is shown in Figure 4 and Figure 5, which illustrates the overall performance of each model based on three different criteria, AUC,

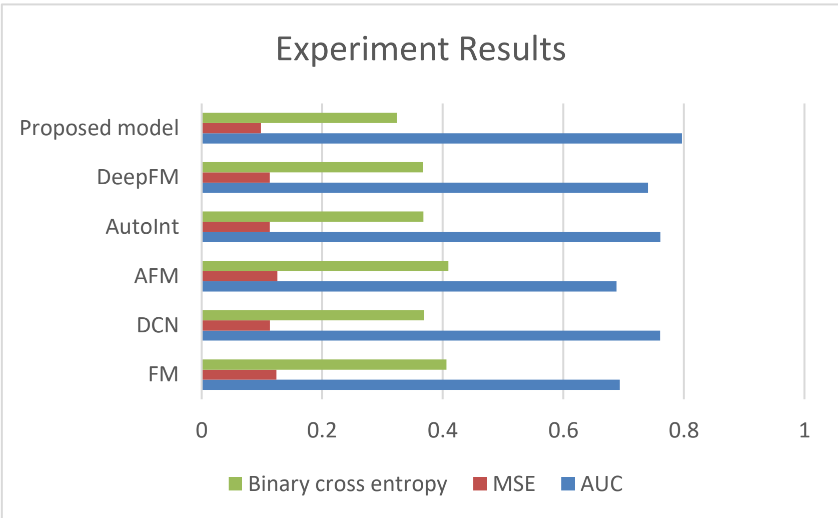


**Figure 4.** The Experiment Result of Different Recommendation Models

MSE, and binary cross entropy. According to the results, we can easily get following 3 conclusions:

**The importance of the high-order feature interaction.** According to the results in Figure 4, we can clearly see that the AUC scores of FM and AFM are the lowest ones, and the MSE and binary cross entropy values of these two models are the highest. These two models are the only two of which involves up to second-order feature interactions. Other baseline methods and our proposed model all involved high-order feature interactions, even though the ways of feature interaction are different. Hence, we argue that the high-order feature interaction (complex feature combination) do reflect how online learners make their decisions and involving the high-order feature interaction is useful and necessary to the large-scale web-based learning recommendation task. This finding proves the first hypothesis that made in Section 3.

**The significance of attention mechanism.** Another conclusion we can get from Figure 4 is the models (AFM and out proposed model) involved the attention mechanism have higher AUC scores than the models which do not. One possible explanation is AFM and our proposed model refine the results of high-order feature interaction via the attention mechanism. Such result approves the second hypothesis that we made in Section 3. The main difference between AutoInt and the proposed model is that they use different techniques to explore the feature interactions. According to the experiment result, we can see that our proposed model outperforms AutoInt when handling the recommendation problem in the online knowledge sharing scenario.

**The efficiency of the proposed model.** To investigate the third hypothesis that we made in Section 3, we also evaluate the computation efficiency of our proposed mod-

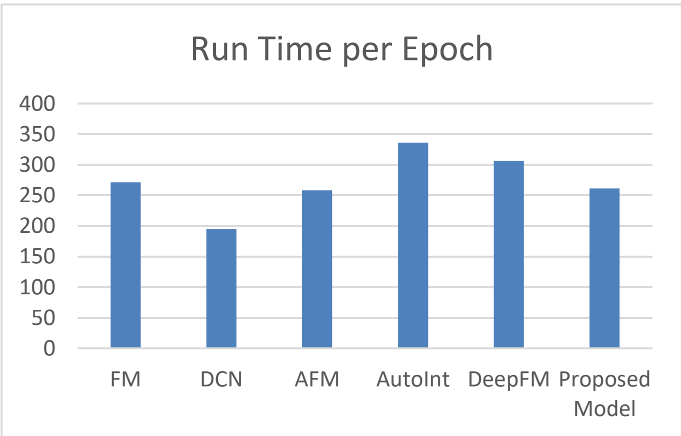


**Figure 5.** Efficiency Comparison of Different Models in Terms of Rum Time (s/epoch)

el and various state-of-the-art recommendation models. The result is also shown in Figure 5. The proposed model is in the third place, outperforming the FM, AutoInt and DeepFM, and very close to the second one (AFM). However, the AFM does not involve high-order feature interactions. The most efficient model is DCN, which only takes around 195 seconds under our experiment setting, while our proposed model takes 261 seconds on average for each training epoch. As the network architecture of our proposed model is extended from DCN and much more complex than other baselines, considering the improvements in recommendation performance and other models have been verified their efficiency on real-world applications, the slight increase of the training time is reasonable and acceptable. The possible reason for this is many operations involved in the proposed model run simultaneously (such as the crossing network and deep network). Hence, we trust under proper configurations our model can reach efficiency requirements.

## 5 Summary and the Future Work

In this study, we refine an existing recommender system [6] using attention mechanism together with high-order feature interaction methods to boost the performance of a web-based knowledge sharing service. By comparing with the-state-of-the-art recommender system, we confirmed three hypotheses about the proposed model: 1. The involved high-order interaction is meaningful and can help further boosting recommendation performance. 2. Features used in the recommender system have different degrees of importance. The attention mechanism can better distinguish such difference comparing to the conventional weighting method. 3. Even though the structure of our model is more complex than the baselines, it still shows acceptable running efficiency.

For future directions, we would like to further explore the recommendation strategy for online learning service. We will continue to investigate how to precisely represent, model and integrate chronological or temporal factor in the recommendation task. As highlighted in [29], with the changing external environment and the internal cognition, a user's interest might evolve over time. Especially for the online informal and non-formal educational activities, many factors are dynamic, such as learning interest drifting and the changes in knowledge level.

**ACKNOWLEDGMENTS.** This research has been carried out with the support of the Australian Research Council Discovery Project, DP180101051, and Natural Science Foundation of China, no. 61877051, and UGPN RCF 2018-2019 project between University of Wollongong and University of Surrey.

## References

1. J. Lin, G. Sun, T. Cui, J. Shen, D. Xu, G. Beydoun, P. Yu, D. Pritchard, L. Li and S. Chen, From ideal to reality: segmentation, annotation, and recommendation, the vital

trajectory of intelligent micro learning. World Wide Web. 1-21(2019), DOI: http://dx.doi.org/10.1007/s11280-019-00730-9.

2. H. Eshach, Bridging in-school and out-of-school learning: Formal, non-formal, and informal education. Journal of science education and technology. 16(2). 171-190(2007), DOI: http://dx.doi.org/10.1007/s10956-006-9027-1.
3. W. Song, C. Shi, Z. Xiao, Z. Duan, Y. Xu, M. Zhang and J. Tang. Autoint: Automatic feature interaction learning via self-attentive neural networks. in Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019, pp. 1161-1170, ACM.
4. T. Huang, Z. Zhang and J. Zhang. FiBiNET: Combining Feature Importance and Bilinear feature Interaction for Click-Through Rate Prediction. in Proceedings of the 13th ACM Conference on Recommender Systems. 2019, pp. 169-177.
5. M.J. Pazzani, A framework for collaborative, content-based and demographic filtering. Artificial intelligence review. 13(5-6). 393-408(1999).
6. J. Lin, G. Sun, J. Shen, D. Pritchard, T. Cui, D. Xu, L. Li, G. Beydoun and S. Chen. Deep-Cross-Attention Recommendation Model for Knowledge Sharing Micro Learning Service. in International Conference on Artificial Intelligence in Education. 2020, pp. 168-173, Springer.
7. D. Wu, J. Lu and G. Zhang, A fuzzy tree matching-based personalized e-learning recommender system. IEEE Transactions on Fuzzy Systems. 23(6). 2412-2426(2015), DOI: http://dx.doi.org/10.1109/TFUZZ.2015.2426201.
8. R. Sikka, A. Dhankhar and C. Rana, A survey paper on e-learning recommender system. International Journal of Computer Applications. 47(9). 27-30(2012), DOI: http://dx.doi.org/10.5120/7218-0024.
9. N. Hoic-Bozic, M.H. Dlab and V. Mornar, Recommender system and web 2.0 tools to enhance a blended learning model. IEEE Transactions on education. 59(1). 39-44(2015).
10. W. Chen, Z. Niu, X. Zhao and Y. Li, A hybrid recommendation algorithm adapted in e-learning environments. World Wide Web. 17(2). 271-284(2014).
11. Z. Rusak. Exploitation of micro-learning for generating personalized learning paths. in DS 87-9 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 9: Design Education, Vancouver, Canada, 21-25.08. 2017. 2017, pp. 129-138.
12. Q. Zhao, Y. Zhang and J. Chen. An improved ant colony optimization algorithm for recommendation of micro-learning path. in 2016 IEEE International Conference on Computer and Information Technology (CIT). 2016, pp. 190-196, IEEE.
13. S. Rendle. Factorization machines. in 2010 IEEE International Conference on Data Mining. 2010, pp. 995-1000, IEEE.
14. T. Fischer and C. Krauss, Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research. 270(2). 654-669(2018).
15. J. Liu, W.-C. Chang, Y. Wu and Y. Yang. Deep learning for extreme multi-label text classification. in Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2017, pp. 115-124.
16. K. He, X. Zhang, S. Ren and J. Sun. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, pp. 770-778.
17. H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai and M. Ispir. Wide & deep learning for recommender systems. in

Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. 2016, pp. 7-10, ACM.
18. H. Guo, R. Tang, Y. Ye, Z. Li and X. He. DeepFM: a factorization-machine based neural network for CTR prediction. in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. 2017, pp. 1725-1731.
19. J. Lian, X. Zhou, F. Zhang, Z. Chen, X. Xie and G. Sun. xdeepfm: Combining explicit and implicit feature interactions for recommender systems. in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018, pp. 1754-1763, ACM.
20. R. Wang, B. Fu, G. Fu and M. Wang. Deep & cross network for ad click predictions. in Proceedings of the ADKDD'17. 2017, pp. 12, ACM.
21. J. Xiao, H. Ye, X. He, H. Zhang, F. Wu and T.-S. Chua. Attentional factorization machines: Learning the weight of feature interactions via attention networks. in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. 2017, pp. 3119-3125.
22. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser and I. Polosukhin. Attention is all you need. in Advances in neural information processing systems. 2017, pp. 5998-6008.
23. J. Hu, L. Shen and G. Sun. Squeeze-and-excitation networks. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, pp. 7132-7141.
24. Y. Shan, T.R. Hoens, J. Jiao, H. Wang, D. Yu and J. Mao. 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. 2016, pp. 255-262, ACM.
25. E. Dobozy. University lecturer views on pedagogic lurking. in 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT). 2017, pp. 1-2, IEEE.
26. M.F. Beaudoin, Learning or lurking?: Tracking the "invisible" online student. The internet and higher education. 5(2). 147-155(2002).
27. V.P. Dennen, Pedagogical lurking: Student engagement in non-posting discussion behavior. Computers in Human Behavior. 24(4). 1624-1633(2008).
28. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein and L. Antiga. PyTorch: An imperative style, high-performance deep learning library. in Advances in Neural Information Processing Systems. 2019, pp. 8024-8035.
29. G. Zhou, N. Mou, Y. Fan, Q. Pi, W. Bian, C. Zhou, X. Zhu and K. Gai. Deep interest evolution network for click-through rate prediction. in Proceedings of the AAAI Conference on Artificial Intelligence. 2019, pp. 5941-5948.