ICIC Express Letters
Volume **14**, Number **4**, April **2020**

ICIC International ©2020 ISSN 1881-803X
pp. **369–377**

MEDICAL QUALITY ASSESSMENT AND PROFESSIONALIZED RECOMMENDATIONS BASED ON DEEP LEARNING

Qiqi Chen, Xingwei Liu, Mingyang Liao, Yi He and Feng Mu

School of Computer and Software Engineering
Xihua University
No. 999, Jinzhou Road, Jinniu District, Chengdu 610039, P. R. China
lxw@mail.xhu.edu.cn

Received September 2019; accepted December 2019

Abstract. *Medical quality is the standard to measure the level of medical staff. Due to the large population and restricted medical resources in China, medical quality has attracted significant attention from both the government and the public in recent years. At present, one of the shortcomings of medical quality assessment is lack of objectivity, and the most important challenge for existing healthcare review websites is the lack of personalized and professionalized guidelines for users to choose medical services. In this paper, we develop a novel medical quality assessment and healthcare recommendation algorithm which is based on deep learning. The proposed algorithm differs from previous work in the following aspects: through big data analysis, objective evaluation of the medical quality is carried out; besides, it can timely obtain a doctor's or a department's specialty and deal with temporal changes of it, and a deep learning approach has been successfully employed in recommender systems to improve accuracy on recommendation. The experimental results show that the proposed algorithm provides a high-performance healthcare recommendation.*

Keywords: Medical quality assessment, Healthcare recommendation, Clustering algorithm, Convolutional neural network, Deep learning

1. **Introduction.** Medical quality is the standard to measure the level of medical staff. Generally, the technical level, medical effect and work quality of medical staff are measured. Due to the large population and restricted medical resources in China, medical quality has attracted significant attention from both the government and the public in recent years. Governments hope to allocate limited resources to hospitals based on medical quality. The public craves highquality healthcare services, and people choose appropriate hospitals on the basis of medical quality information they can collect about target hospitals. At the same time, hospital managers seek to improve medical quality which is the key factor for attracting public or private funding and healthcare service consumers [1]. Therefore, how to assess medical quality objectively and comprehensively so as to achieve a convincing quality ranking of doctors, departments or hospitals has become a hot current research topic in China.

In general, expert judgments are frequently used in practice audits, peer reviews, or practice visits for assessing medical quality. Due to the fact that expert judgments can mainly measure medical quality by experience; besides, most objective indicator assessment methods used in the literature are based on random samples, not all data, they may not be able to objectively reflect the quality of medical care of one hospital as a whole [2].

The way that people choose medical service is changing with health related reviews websites, WeChat and mobile Apps [3]. Through these modes, detailed information about doctors, departments or hospitals can be obtained for choosing doctor with an online appointment. However, several challenges exist to enable personalized and accurate medical

DOI: 10.24507/icicel.14.04.369

services: unfortunately, at present such personalization and professionalization demand cannot be satisfied intelligently according to user and doctor feature. It is a great issue for the medical crowd-sourced reviews that rating accuracy is often interfered by users emotion [4].

Recommender systems are effective tools of information filtering that are prevalent due to increasing access to the Internet, personalization trends, and changing habits of computer users [5]. Although existing recommender systems are successful in producing decent recommendations, they still suffer from challenges such as accuracy, scalability, and cold-start [6-8]. In the last few years, deep learning, the state-of-the-art machine learning technique utilized in many complex tasks, has been employed in recommender systems to improve the quality of recommendations [9-12]. To address these challenges, this paper develops a novel medical quality assessment and recommendation algorithm which is based on deep learning. The proposed algorithm differs from previous work in the following aspects: through big data analysis, objective evaluation of the medical quality is carried out; besides, it can timely obtain a doctor's or a department's specialty and deal with temporal changes of it, and a deep learning approach has been successfully employed in recommender systems to improve accuracy on recommendation.

The remainder of this paper is organized as follows. In Section 2, the medical quality assessment algorithm is discussed. In Section 3, the healthcare recommendation algorithm based on deep learning is discussed. In Section 4, the improved recommender algorithm is discussed. Section 5 shows experimental studies for verifying our proposed algorithm. Finally, we conclude this article in Section 6.

2. **Medical Quality Assessment Algorithm.** The workflow of medical quality assessment is shown in Figure 1. Firstly, the clinical cases of different diseases are grouped roughly according to the admission diagnostic code. Then, in order to eliminate the impact of different types of presenting symptoms and clinical history, the clinical cases of the same disease are further subdivided, and subgroups of the clinical cases are obtained.

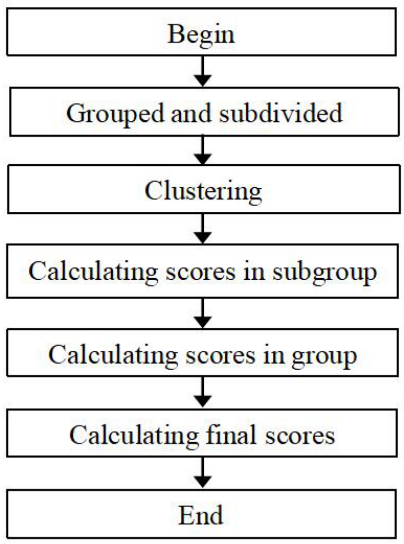

FIGURE 1. Workflow of medical quality assessment

Traditionally, quality indicators include patient demographics, admission date, discharge date, state of illness at admission, admission diagnosis, discharge diagnosis, state of hospital infection, operations, state of illness on discharge, and so on. Because the treatment time, treatment cost and treatment effect not only reflect the "cost" paid by patients for treatment, but also reflect the level of treatment, the three indicators are taken as quality indicators in this paper, and the Sum of Squared Error (SSE) is adopted to optimize the K value of K-means clustering algorithm. Then, each subgroup of the clinical cases is classified by K-means clustering algorithm [13], and the clustering graph is drawn at the same time.

By analyzing the clustering graph, clusters with better treatment effect, shorter treatment time and lower treatment cost are selected. The number of the clinical cases of the selected clusters is divided by the total number of the clinical cases, and the quality score of each evaluation object (i.e., doctor or department) in the subgroup is obtained, which is defined as

$$S = \frac{N_s}{N_t} \times 100\%,$$

where N_s represents the number of the clinical cases of the selected clusters, and N_t is the total number of the clinical cases. Furthermore, through weighted summation of the quality scores of each evaluation object in each subgroup, the quality score of each evaluation object in each group is obtained. Finally, the top-n clinical cases with the highest scores are selected as a doctor's or a department's specialty, and the final score of each evaluation object is obtained by weighted summation of the top-n scores.

3. **The Healthcare Recommendation Algorithm Based on Deep Learning.** At present, the most important challenge for existing healthcare review websites is the lack of personalized and professionalized guidelines for users to choose medical services. In this paper, we develop a novel healthcare recommendation algorithm which is based on deep learning. Specifically, it includes the following modules.

3.1. **A deep neural network model for obtaining patient eigenvector.** The deep neural network model for obtaining patient eigenvector is shown in Figure 2. The first layer of this model is the embedding layer. The input of the embedding layer is patient demographics including patient ID, patient gender and patient age, and the features of patient ID, patient gender and patient age are extracted and obtained by the embedding layer.

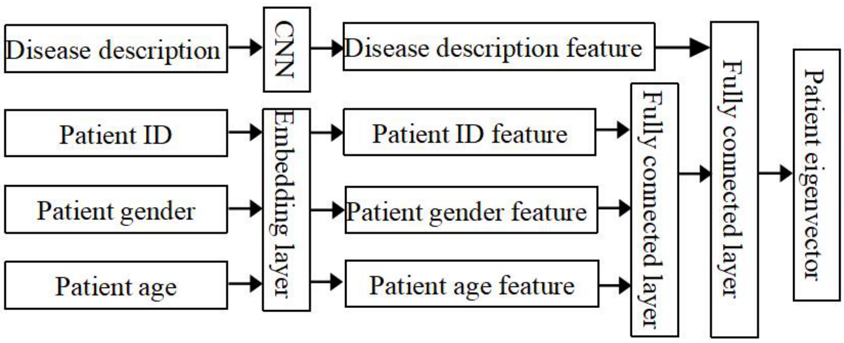

Figure 2. A deep neural network model for obtaining patient eigenvector

A Convolutional Neural Network (CNN) is applied in disease description [14, 15]. This CNN consists of three components that transform the input volume into an output volume, namely, convolutional layers, pooling layers, and fully connected layers. As being the core operation, convolutions aim to extract features from the input which is disease description in this paper. Feature maps are obtained by applying convolution filters with a set of mathematical operations. Pooling reduces the dimensionality of the feature maps to decrease processing time. The output from the convolutional and pooling layers represents highlevel features of the input, and these features can be used within the fully connected layers.

The three features obtained from the embedding layer, i.e., patient ID, patient gender and patient age, are concatenated together as the input data used within a fully connected layer. After the output from the CNN (i.e., disease description feature) and the first fully connected layer are concatenated together as the input data used within the second fully connected layer, the patient eigenvector will be obtained.

3.2. **A deep neural network model for obtaining candidate eigenvector.** The deep neural network model for obtaining candidate (i.e., doctors or departments) eigenvector is shown in Figure 3. The first layer of this model is the embedding layer. The input of the embedding layer is the candidate ID and candidate specialty, and the features of candidate ID and candidate specialty are extracted and obtained by the embedding layer; besides, it is necessary to further deal with the feature of candidate specialty. In this paper, we will sum the feature vectors of candidate specialty together. The two features obtained from the embedding layer, i.e., candidate ID and candidate specialty, are concatenated together as the input data used within a fully connected layer. After the output from the first fully connected layer are concatenated together as the input data used within the second fully connected layer, the candidate eigenvector will be obtained.

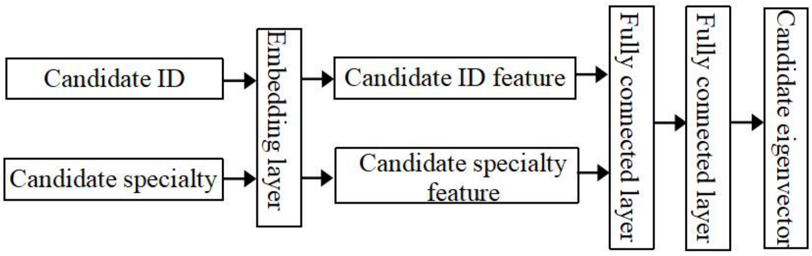

FIGURE 3. A deep neural network model for obtaining candidate eigenvector

3.3. **A deep neural network model for healthcare recommendation.** A deep neural network model for healthcare recommendation is shown in Figure 4. It is to multiply the patient eigenvector and the candidate eigenvector, and we incorporate predicted rating and actual rating into a regression model which is a Multi-Layer Perceptron (MLP) neural network. In this paper, the experimental dataset is divided into training sets and test sets. The cost or loss function is optimized by the Mean Square Error (MSE), and the experimental parameters including the number of epochs, learning rate, and batch size are set. Through training this model, the patient eigenvector matrix and candidate eigenvector matrix will be obtained. The patient eigenvector matrix and candidate eigenvector matrix are used to predict the scores of candidate, and the top-n candidates with the highest scores are taken out to generate a recommended list for patients.

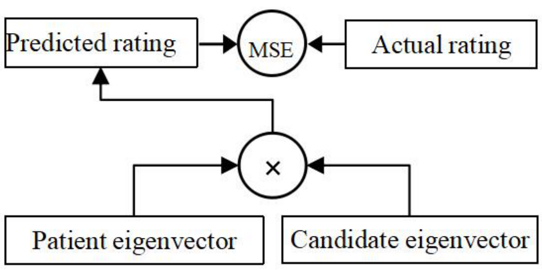

FIGURE 4. A deep neural network model for healthcare recommendation

4. **The Modified Healthcare Recommendation Based on Deep Learning.** In Figure 3, because there are many types of the feature of candidate specialty which is extracted and obtained by the embedding layer, it is necessary to further deal with the feature of candidate specialty. As mentioned earlier, we will sum the feature vectors

of candidate specialty together. However, in the improved method, we will average the feature vectors of candidate specialty. In the model in Figure 4, as mentioned earlier, it is to multiply the patient eigenvector and the candidate eigenvector. However, in the improved method which is shown in Figure 5, the patient eigenvector and the candidate eigenvector are concatenated together as the input data used within a fully connected layer, and predicted rating will be obtained.

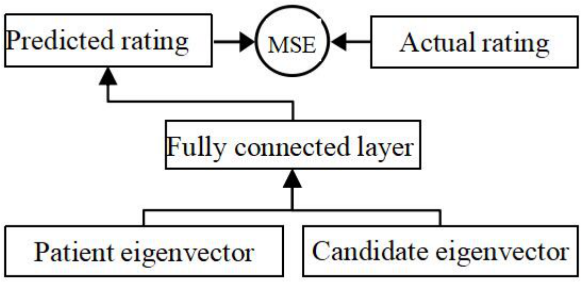

Figure 5. A modified model for healthcare recommendation

Fully connected layers connect every neuron in one layer to every neuron in another layer. It is in principle the same as the traditional MLP. The core operation of MLP is matrix vector product, and it can be said that the predicted rating is the weighted sums of the patient eigenvector and the candidate eigenvector.

Training a deep neural network can be slow. So far we have seen some ways to speed up training (and reach a better solution): using a good activation function, using Batch Normalization. Another huge speed boost comes from using a faster optimizer than the regular Gradient Descent optimizer. In this paper, ReLU activation function and Adam optimization algorithm will be used.

5. Experiments and Results Analysis.

5.1. Experiments and results analysis of medical quality assessment algorithm.

Since the 1990s, the hospital in China has gradually completed the clinical medical informationization including Hospital Information System (HIS), Electronic Medical Records system (EMRs), Laboratory Information System (LIS), PACS, Nursing Information System (NIS), etc. The clinical cases data used to assess medical quality in this study are derived from Electronic Medical Records (EMRs), Electronic Health Records (EHRs) and administrative data of a special hospital in Chengdu, China. Figure 6 shows the K value of K-means clustering algorithm determined by the SSE. For each subgroup, treatment

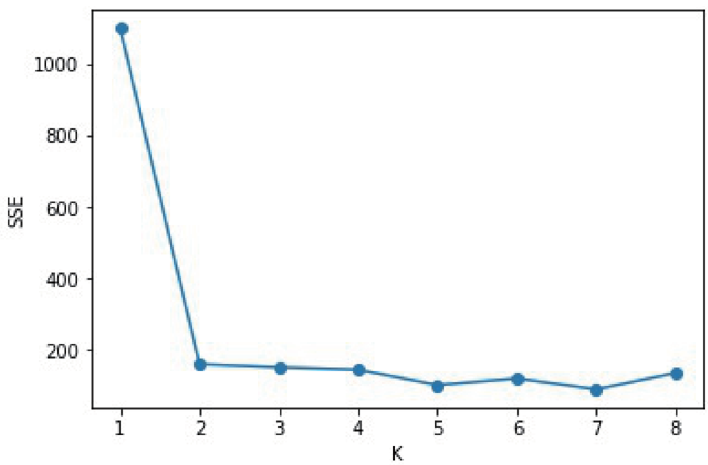

Figure 6. The K value determined by the SSE

time, treatment cost and treatment effect are taken as quality indicators in this paper. Firstly, the three quality indicators are standardized; then, the K value of K-means clustering algorithm is determined by the SSE.

Figure 7 shows a clustering graph of a subgroup at $K = 5$. The coordinates of the horizontal axis represent treatment time, treatment cost and treatment effect, respectively. In order to facilitate the application of machine learning algorithm, one-hot coding is utilized to represent treatment effects which range from 1 to 5 (i.e., best to worst). By analyzing the clustering graph in Figure 7, it is found that cluster 1 and 2 are clusters with better treatment effect, shorter treatment time and lower treatment cost. Therefore, we choose cluster 1 and 2 to calculate the quality scores: the number of the clinical cases of the selected clusters (i.e., cluster 1 and 2) is divided by the total number of the clinical cases, and the quality scores of evaluation object (i.e., 7 departments) in Figure 7 is obtained, which are 0.88, 0.79, 0.8, 0.84, 0.93, 0.89 and 0.91, respectively.

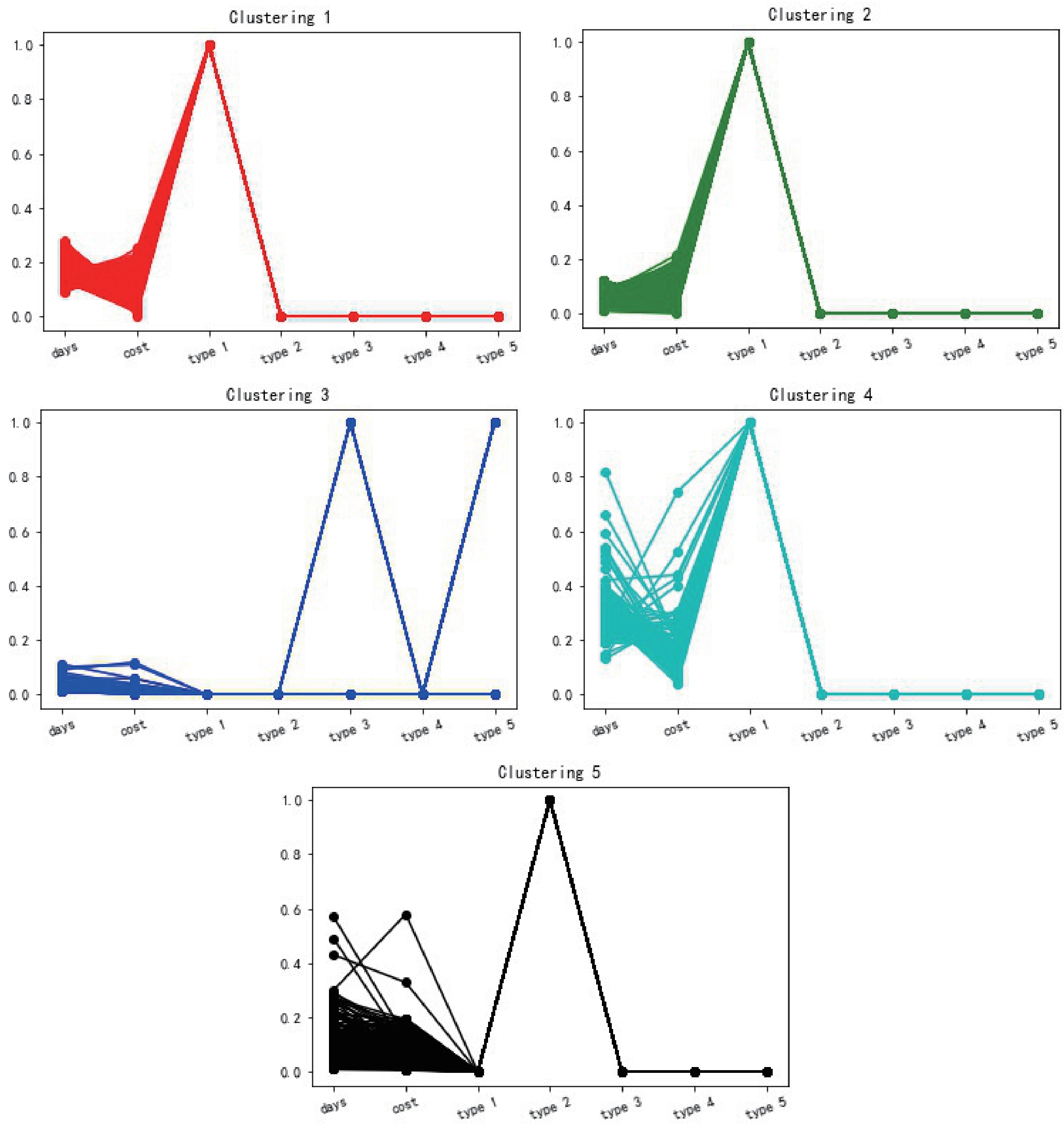

FIGURE 7. A clustering graph of a subgroup at $K = 5$

5.2. **Experiments and results analysis of healthcare recommendation algorithms.** In order to train the model, it is necessary to define a cost or loss function to evaluate the model, and the Mean Square Error (MSE) is used as the cost function here. Furthermore, the experimental dataset is divided into training sets and test sets, and the experimental parameters including the number of epochs, learning rate, and batch size are set. The training process is to continuously adjust the model parameters and minimize the MSE. Finally, the patient eigenvector matrix and candidate eigenvector matrix will be obtained. The experimental parameters of the model are set as shown in Table 1.

TABLE 1. The experimental parameters

	Value
Dimensions of embedded matrix	32
CNN sliding window size	2, 3, 4, 5
Number of convolution kernels	8
Number of epochs	5
Learning rate	0.0001
Batch size	256

Thereinafter, we will discuss the impact of the parameters on the performance of our model. Figure 8 shows the training cost curve and test cost curve of the model at different learning rates. As can be seen from the figure, with the increasing of learning rate, the MSE of training rapidly decreases, meanwhile, the MSE of test is smaller and the test cost curve is smoother.

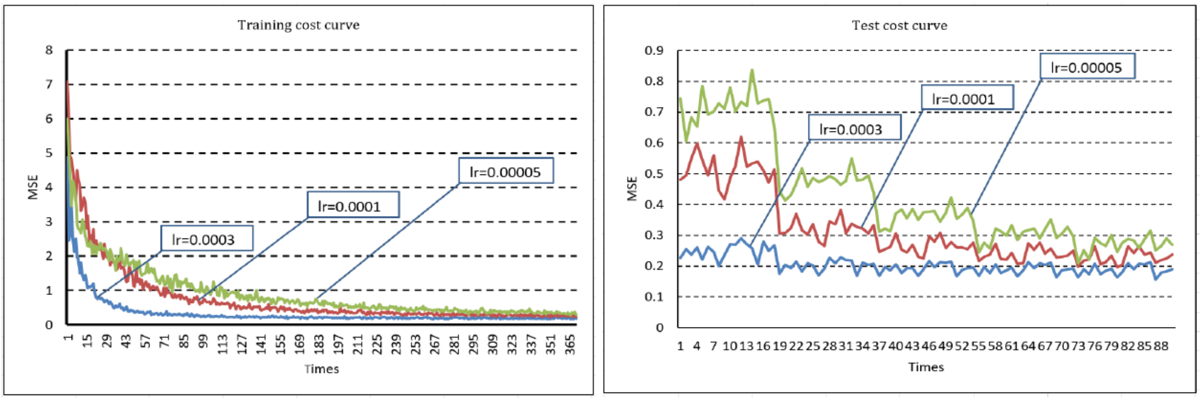

FIGURE 8. The impact of different learning rates on the performance of the model

Figure 9 shows the impact of different processing methods on the performance of the model shown in Figure 3. These processing methods include summation and average operation, which are named sum and mean in Figure 9, respectively. As can be seen from Figure 9, when the average operation is performed, the MSEs of training and test are smaller, and the training and test cost curves are smoother.

Figure 10 shows the performance comparison between the basic model (shown in Figure 4) and the improved model (shown in Figure 5). As can be seen from Figure 10, the training and test MSEs of the improved model are smaller, and the training and test cost curves of the modified model are smoother. In addition, the improved model converges faster.

6. **Conclusions.** This paper presents a medical quality assessment and healthcare recommendation algorithm based on deep learning. Through big data analysis, objective evaluation of the medical quality was carried out; besides, it can timely obtain candidate

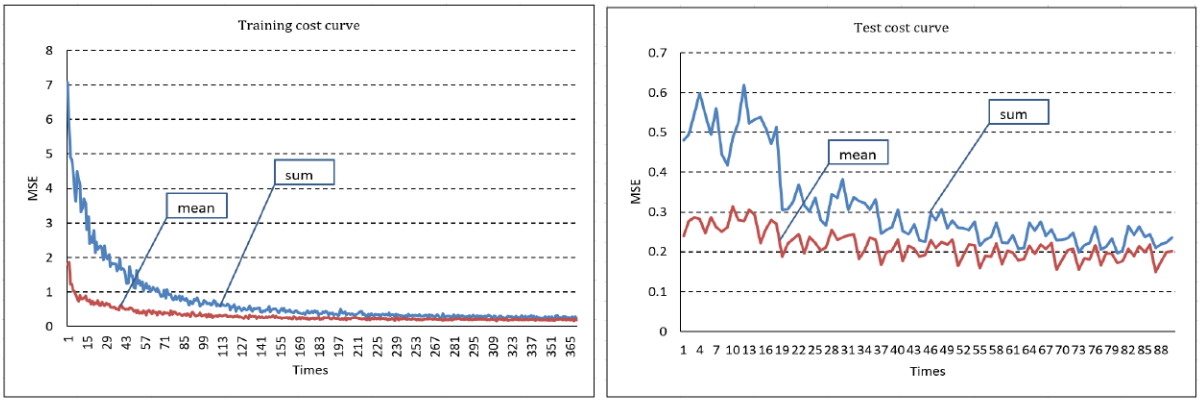

FIGURE 9. The impact of different processing methods on the performance of the model

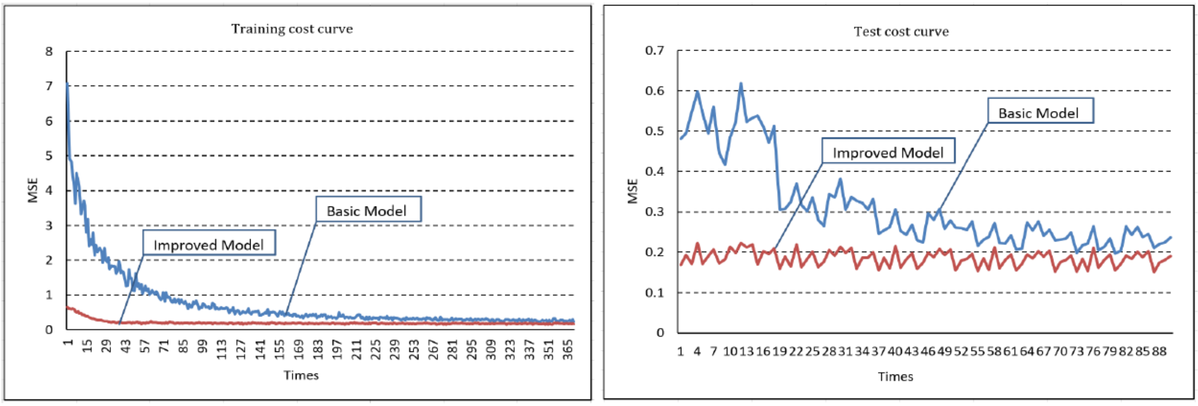

FIGURE 10. The performance comparison between the basic model and the improved model

specialty and deal with temporal changes of it. Finally, deep learning has been successfully employed in recommender systems to improve the quality of recommendations.

The clinical cases data used to assess medical quality in this study are derived from EMRs, EHRs and administrative data of a special hospital in Chengdu, China. In the future research, we will verify and improve our algorithm based on more clinical cases data from more hospitals.

Acknowledgment. This work is partially supported by the Science and Technology Program of Chengdu City under Grant No. 2015-HM01-00527-SF. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] J. Ritchey, E. G. Gay, B. A. Spencer, D. C. Miller, L. P. Wallner, A. K. Stewart, R. L. Dunn, M. S. Litwin and J. T. Wei, Assessment of the quality of medical care among patients with early stage prostate cancer undergoing expectant management in the United States, *The Journal of Urology*, vol.188, no.3, pp.769-774, 2012.

[2] G. L. Kong, D. L. Xu, J. B. Yang and X. M. Ma, Combined medical quality assessment using the evidential reasoning approach, *Expert Systems with Applications*, vol.42, no.13, pp.5522-5530, 2015.

[3] X. T. Zhang, D. Wen, J. Liang and J. B. Lei, How the public uses social media WeChat to obtain health information in China: A survey study, *BMC Medical Informatics and Decision Making*, vol.17, no.2 Supplement, 2017.

[4] Y. Zhang, M. Chen, D. J. Huang, D. Wu and Y. Li, iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization, *Future Generation Computer Systems*, vol.66, pp.30-35, 2017.

[5] C. C. Aggarwal, An introduction to recommender systems, in *Recommender Systems*, Springer, Cham, 2016.

[6] Z. Batmaz, A. Yurekli, A. Bilge and C. Kaleli, A review on deep learning for recommender systems: Challenges and remedies, *Artificial Intelligence Review*, pp.1-37, 2018.

[7] B. T. Betru, C. A. Onana and B. Batchakui, Deep learning methods on recommender system: A survey of state-of-the-art, *International Journal of Computer Applications*, vol.162, no.10, pp.17-22, 2017.

[8] J. Wei, J. H. He, K. Chen, Y. Zhou and Z. Y. Tang, Collaborative filtering and deep learning based recommendation system for cold start items, *Expert Systems with Applications*, no.69, pp.29-39, 2017.

[9] W. Zhang, Y. H. Du, Y. Yang and T. Yoshida, DeRec: A data-driven approach to accurate recommendation with deep learning and weighted loss function, *Electronic Commerce Research and Applications*, no.31, pp.12-23, 2018.

[10] L. B. Zhang, T. J. Luo, F. Zhang and Y. J. Wu, A recommendation model based on deep neural network, *IEEE Access*, no.6, pp.9454-9463, 2018.

[11] H. T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, R. Anil, Z. Haque, L. Hong, V. Jain, X. B. Liu and H. Shan, H_∞ wide & deep learning for recommender systems, *Proc. of the 1st Workshop on Deep Learning for Recommender Systems (DLRS 2016)*, Boston, MA, USA, pp.7-10, 2016.

[12] J. T. Liu and C. H. Wu, H_∞ deep learning based recommendation: A survey, *Information Science and Applications 2017 (ICISA 2017)*, Singapore, pp.451-458, 2017.

[13] A. K. Jain, M. N. Murty and P. J. Flynn, Data clustering: A review, *ACM Computing Surveys*, vol.31, no.3, pp.264-323, 1999.

[14] W. B. Liu, Z. D. Wang, X. H. Liu, N. Zhang, Y. Liu and F. E. Alsaadi, A survey of deep neural network architectures and their applications, *Neurocomputing*, no.234, pp.11-26, 2017.

[15] S. Y. Seo, J. Huang, H. Yang and Y. Liu, H_∞ interpretable convolutional neural networks with dual local and global attention for review rating prediction, *Proc. of the 11th ACM Conference on Recommender Systems (RecSys'17)*, Como, Italy, pp.297-305, 2017.