WSEAS TRANSACTIONS on COMPUTERS
DOI: 10.37394/23205.2023.22.2

A Proposed Artificial Intelligence Algorithm for Development of Higher Education

AMIN AL KA'BI
Australian University, KUWAIT

Abstract: Higher education has delved into a new stage of rapid development focusing on quality improvement, while encountering new challenges and obstacles. In this research work, an artificial intelligence algorithm for education improvement is proposed. Firstly, deep feature abstraction in temporal and special dimensions is performed using Long Short-Term Memory (LSTM) artificial neural network and convolutional networks. Consequently, multiscale attention fusion techniques are used to improve the articulateness of the characteristics and come up with better recommendations with the assistance of multilayer perceptron. Moreover, the proposed model helps in improving the cognitive capability of students and enhances their overall quality of perception. Moreover, it has been proven that the performance of the proposed model provides better recommendation outcomes and better robustness compared to existing models through conducting extensive experiments based on real data.

Keywords: Artificial Intelligiance (AI); Communication Sysems; Higher Education; Convolutional Neural Networks (CNN); Recurrent Neural Networks (RNN).

Received: February 19, 2022. Revised: November 15, 2022. Accepted: December 9, 2022. Published: January 19, 2023.

1. Introduction

With the pervasive application of the new generations of mobile communications, computer networking, Internet of Things (IoT), cloud computing, and artificial intelligence (AI), the world is delving into a new era of intelligent information technology and large scale of data mining applications. Artificial Intelligence, machine learning, and big data processing are increasingly becoming the main motives for digital transformation all over the world and control all aspects of our lives [1], [2]. These revolutionary developments in information technology resulted in new methodologies in teaching and learning processes at higher education institutions. Artificial Intelligence helps in the enrichment of students' theoretical background and enhances their university education in terms of diversifying the development of their educational needs, taking into consideration the huge impact of using the Internet, and the wide opportunities to experiment new things and more complex behavior patterns. Moreover, AI, networking, and big data help in providing positive educational and cultural recommendations to users in real time, which can assist in meeting the expanded development requirements of university students [3]. On the other hand, employment of AI in higher education provides universities and colleges with the ability to create a friendly relationship with students, and helps in improving, to some extent, the identity and self-confidence of university students, besides developing harmonious relationships throughout the campus environment in the new educational system.

The optimization of the relationship between students and teachers and the promotion of friendly collaboration between different departments at the university lead to the creation of healthy atmosphere throughout the whole campus and establishing the basic requirements of ethical and moral education at the university. The establishment of morality lies in encouraging students to apply right moral principles and qualities and to apply distinguished moral reality. Therefore, it is possible for the colleges and universities to help students to be more competent and develop desirable qualities by applying various educational methods, and to be main facilitators, participants, and promoters for the students under the enormous wave of big data which includes wide range of data resources, advanced data processing, and usage of artificial intelligence in this regard. Due to its countless advantages and benefits, artificial intelligence has created new approaches to big data acquisition and data processing, besides enabling the usage of new educational methods to promote innovation of thinking, reconstruction of paradigms, and implementing scientific and logical solutions for daily life problems. The role of artificial intelligence and deep learning in improving higher education resulted in establishing robust tools to fix the weaknesses and shortcomings in educational systems such as cold starts, data sparsity, and deprived interpretability.

The introduction of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) enabled linguistic processing tasks, such that students can practice various deep learning approaches such as Deep Cooperative Neural Networks (DeepCoNN) to excavate user preferences and suggest recommendations from the user's perspective [4], [5]. DeepCoNN is primarily based on CNN model, and it consists of two parallel neural networks which learn the students' behavior and the learning outcomes of the university course, respectively, and creates the necessary connections between them at the top of the network to emphasize the interaction and attention mechanisms, which is used for suggesting recommendation algorithms that improve the learning/teaching process at higher education institutions. Neural Attentional Rating Regression (NARRE) employs attention mechanisms to comprehend textual depictions to better understand users' behaviors, predict users' interests, and create proper interpretations, based on two-channel attention mechanism. On the other hand, the advent of Neural Language Processing (NLP) helped in the development of text in the area recommendation. Unfortunately, NLP is unidirectional which limits the power of interpretation. Hence, a bidirectional pertaining model with a high generalization capability called BERT was proposed in [26], which reads the whole text at once using

Transformer's Encoder, and this procedure allows the algorithm to learn on both sides of the word and hence digest the meaning of the word in the sentence more accurately and provides a good ground for downstream tasks [6-9].

This paper investigates various techniques used in deep learning such as CNN and LSTM to suggest proper recommendation models to be employed in higher educational institutions, which helps the students in locating their knowledge of interest and real attention. Moreover, it helps in improving the cognitive capability of the students and the overall quality of their mindsets.

2. Methodology

The proposed model of higher education presented in this paper is supported by artificial intelligence and it is based on convolutional neural networks (CNN) and progressive features obtained from the progressive dimension by using bidirectional neural networks. Sec. 2.1 explains the theory of CNN, and the procedure of feature extraction from progressive dimension is introduced in Sec. 2.2. The scale fusion attention is explained in Sec. 2.3. The procedure of calculating the prediction scores and suggesting recommendations is explained in Sec. 2.4.

2.1. Convolutional Neural Networks (CNN) Model

The Convolutional Neural Networks (CNN) Model consists of an input layer, pooling layer, fully connected layer, convolutional layer, and an output layer [10], [11]. In this model, the convolution layer is the outcome of the inner product of the filter matrix and the embedding vector. The pooling layer groups each feature map acquired from the convolution layer. It is found that the better features can be extracted using the maximum pooling approach in comparison with the mean pooling approach. The output features of the pooling layer are taken by the fully connected layer as an input and activated by the activation function to generate a fixed dimension feature vector. The performance of CNN is excellent in dealing with classification problems; however, its recommendation algorithms produce small number of results. This is due to the fact that the recommendation process is a regression process, and they have different objectives [12], [13]. To obtain more precise recommendation results for text data, it is possible to build a hybrid recommendation model by merging a classical recommendation algorithm with CNN. Fig. 1 depicts the hierarchy of the Convolution neural network (CNN) model.

2.1.1. Input Layer.

In CNN model structure, the role of the input layer is to transform the textual data of the learning resource into an embedding matrix, where the dimensionality of the embedded data is very low, with the possibility of converting the discrete data sequences into continuous data vectors.

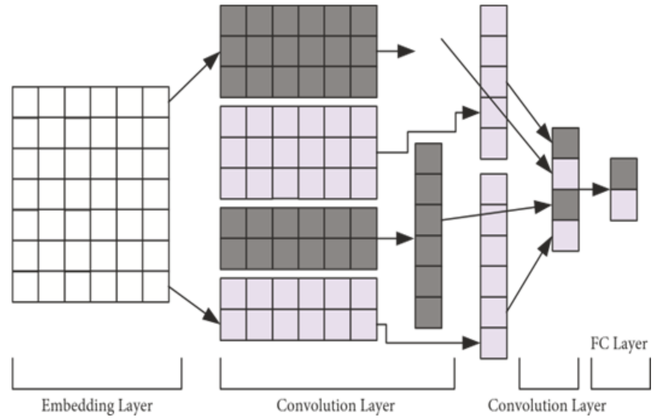


Fig. 1. Convolution neural network (CNN) model.

Equation 1 illustrates the embedding matrix D where each row in the matrix represents a clause element.

$$D = \begin{bmatrix} w_{11} \cdots w_{1i} \cdots w_{1m} \\ w_{21} \cdots w_{2i} \cdots w_{2m} \\ \cdots\cdots\cdots \\ w_{n1} \cdots w_{ni} \cdots w_{nm} \end{bmatrix}, \quad (1)$$

where m denotes the embedding dimensionality, n denotes the number of the words, and $[w_{i,1:\,m}]$ denotes the vector representation of the i-th word.

2.1.2. Convolution Layer.

The convolution operations are performed on the embedding matrix by using multiple convolution kernels with different dimensions. To cover the entire word embedding vector discussed in this paper, the size of convolution is determined by product of the vector dimension and the number of words. Fig. 2 illustrates the flow of the convolution operation.

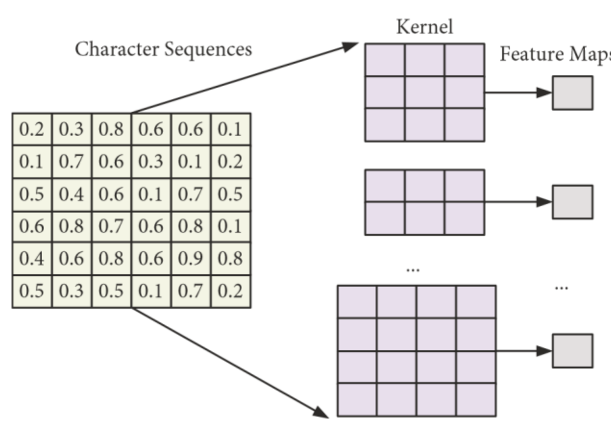


Fig. 2. Flow of the convolution operation.

2.1.3. Pooling Layer.

The main purpose of this layer is to decrease the dimension of the feature map and to decrease the number of parameters used in the network. The main pooling operations include maximum pooling and average pooling. In the pooling operation, little deviations in the feature maps can be discarded and hence the accuracy can be improved while averting the phenomenon of overfitting. If the feature map acquired in the *t-th* convolutional layer is $M_t = \{m_1, m_2, m_3, \dots, m_s\}$, then the maximum value of M_t can be obtained by using the maximum pooling strategy as illustrated in Eq. (2).

$$p_i = max(M_{t)} = max\{m_1, m_2, m_3, \dots, m_s\} \qquad (2)$$

where p_i expresses the pooling outcome of the t_i-th convolutional layer.

2.1.4. Fully Connected Layer. The function of the fully connected layer is to implement the feature vector and the extracted layer values. Assume that the fully connected layer has *m* neurons, and a text feature vector F_i is created after initiating the ReLu activation function such as:

$$F_i = \sigma(w_i p_i + b_i), \qquad (3)$$

where σ denotes the ReLu activation function, p_i is the output of the learning resource text information on the pooling layer, b_i is the bias, and w_i denotes the weight.

2.2. Long- and Short-time Bidirectional Recurrent Neural Network

The neurons of the Long Short-Term Memory (LSTM) artificial neural network accept only the data from the neurons close to the layers, meanwhile the words before and after affect the semantic connections. However, Bidirectional Long Short-Term Memory Recurrent Neural Network (BiLSTM-RNN) consists of two groups of long and short term recurrent neural networks with contradictory learning trends, which enables better understanding of the related semantics in comparison with LSTM. The LSTM is composed mainly from four elements input gate i_t, forgetting gate f_t, memory unit c_t, and output gate o_t. The input gate controls the flow of the data into the memory unit, the forgetting gate identifies the holding of the previous state data in the memory unit, which determines the memory state based on the current input data; and then the output gate identifies the output value of the memory unit for the next state. The relevant computation procedure is illustrated in Eq. 4 to Eq. 9.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \qquad (4)$$
$$f_t = \sigma\big(W_f[h_{t-1}, x_t] + b_f\big), \qquad (5)$$
$$\widetilde{c_t} = \tan h\,(W_c[h_{t-1}, x_t] + b_c), \qquad (6)$$
$$c_t = f_t * c_{t-1} + i_t * \widetilde{c_t}, \qquad (7)$$
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \qquad (8)$$
$$h_t = o_t * \tanh(c_t), \qquad (9)$$

where h_t represents the memory cell state, x_t is the data input, b is the function bias term, W denotes the matrix multiplication operation, $*$ is the dot product operation, and $\sigma(.)$ is the sigmoid function.

BiLSTM-RNN model consists of two groups of forward and backward LSTM models attached to a learning fearure denoted by $h_t^{forward}$ and $h_t^{backward}$ respectively. Eq. 10 expresses the time-dimensional feature T_i which determines final representation of the LSTM model structure.

$$T_i = h_t^{bilstm} = h_t^{forward} \oplus h_t^{backward} \qquad (10)$$

where $\oplus$ denotes the concatenation operator. This operation enables the BiLSTM model to fully process the input words of the contextual data.

2.3 Multiscale Feature Fusion

The spatial dimensional features *F* and the time-dimensional feature *T* can be merged as depicted in Fig. 3, where multiscale feature synthesis attention process is used to achieve this goal.

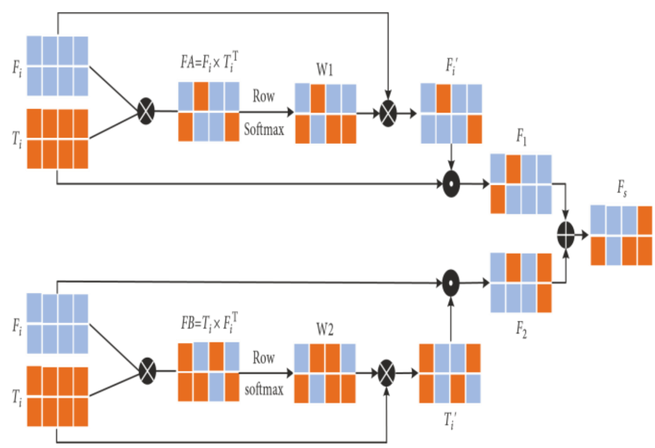


Fig. 3. Multiscale feature synthesis attention process.

Firstly, the matrix *FA* and the matrix *FB* which represent the matching between the dimensional features denoted by F_i and the attribute features denoted by T_i are determined as expressed in Eq. 11.

$$FA = F_i \times T_i^T \quad \& \quad FB = T_i \times F_i^T \qquad (11)$$

Secondly, the function SoftMax is used to find the attention distribution weights w_1 and w_2 of the matching matrices. Then, the attention representation matrices F'_i and T'_I are calculated by multiplying the weights w_1 and w_2 with the individual scale features as expressed in Eq. 12.

$$F_i' = F_i \times w_1 \quad \& \quad T_i' = T_i \times w_2 \qquad (12)$$

where (×) denotes matrix fork multiplication.

At last, the inter-scale mutual attention matrices F_1 and F_2 are calculated by using a multiplicative gating process to multiply the attentional representation with another single-scale feature for the relevant elements as expressed in Eq. 13.

$$F_1 = T_i \cdot F_i' \quad \& \quad F_2 = F_i \cdot T_i' \tag{13}$$

where $(\cdot)$ denotes matrix dot product

To determine the final multi-scale synthesis features, the functions F_1 and F_2 are operated as expressed in Eq. 14.

$$F_S = F_1 \oplus F_2 \tag{14}$$

where $\oplus$ is the Cat operation.

2.4. Scores Prediction and Recommendations Suggestion

The outcome of Eq. 14 is used as an input to the multilayer perceptron the anticipate the scores. A cross-entropy loss function is used to obtain the end-to-end optimization of the model, and to find and to anticipate the weights for each layer. Eventually, anticipation outcomes are determined by relating the activation function to the range [0, 1]. Equations 15-17 show the corresponding calculations.

$$X_t = wh_t + b, \tag{15}$$

where X_t is the fully connected outcome and h_t is the output hidden vector of the decoder.

$$P(y|x) = \frac{e^{h(x,y_i)}}{\sum_{j=1}^{n} e^{h(x,y_i)}} \tag{16}$$

where y is the true description, x is the fully connected, and P is the SoftMax function.

$$L(\theta) = -\sum_{t=1}^{T} logp(y^t | y_{1:t-1}^*) \tag{17}$$

where θ denotes the parameter of cross-entropy balance loss of the model.

3. Experimental Results

Experiments were carried out using PyTorch deep learning framework, NVIDIA CUDA 11.3, and cuDNN deep learning acceleration library. Stochastic gradient descent algorithm SGD was used to train the network with a momentum factor of 0.8, learning decay rate of 0.001, and an initialized learning rate of 0.005. Fig. 4 depicts the accuracy curve and the model training loss curve.

A random removal of some neurons is necessary to overcome the overfitting problem. To achieve this goal a Dropout of 0.5 is used. Fig. 4 shows that when the iteration number approaches 40, both loss curves of test and training get smoother, and the amount of loss decreases below 0.06 which is an indication of the model convergence.

3.1. Experimental Results

3.1.1. Proof of model validity. Experiments are carried out to prove the model validity using identical environmental and evaluation indices. the Gram-based modifieded recommendation algorithm Gram-CF, the traditional collaborative filtering recommendation algorithm User-CF based on user play records, and the collaborative filtering recommendation algorithm FCNN-CF based on user preference statistics are chosen as reference models. The evaluation metrics used are the $F1$ value, precision rate (PR), accuracy rate (AR), and recall rate (RR).

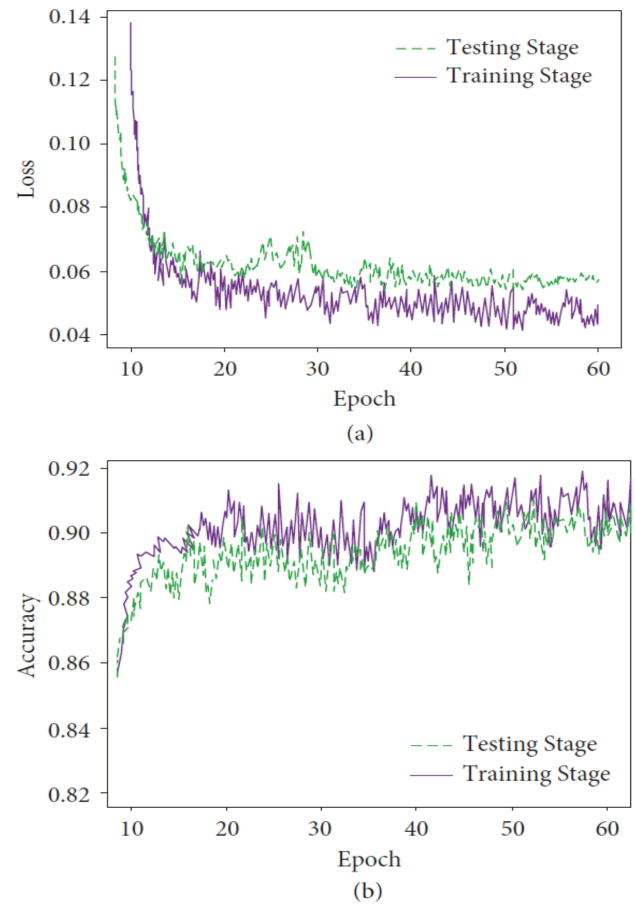


Fig. 4. Loss and accuracy in testing and training and phases. (a) Loss, (b) Accuracy.

Fig. 5 exhibits a significant improvement of the proposed model in this paper by 1.78%, 1.22%, and 3.05% in terms of accuracy in comparison with the reference models FCNN-CF, User-CF, and Gram-CF recommendation models, respectively.

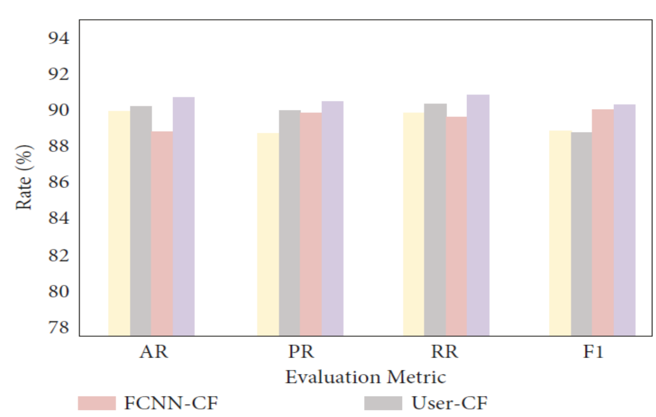
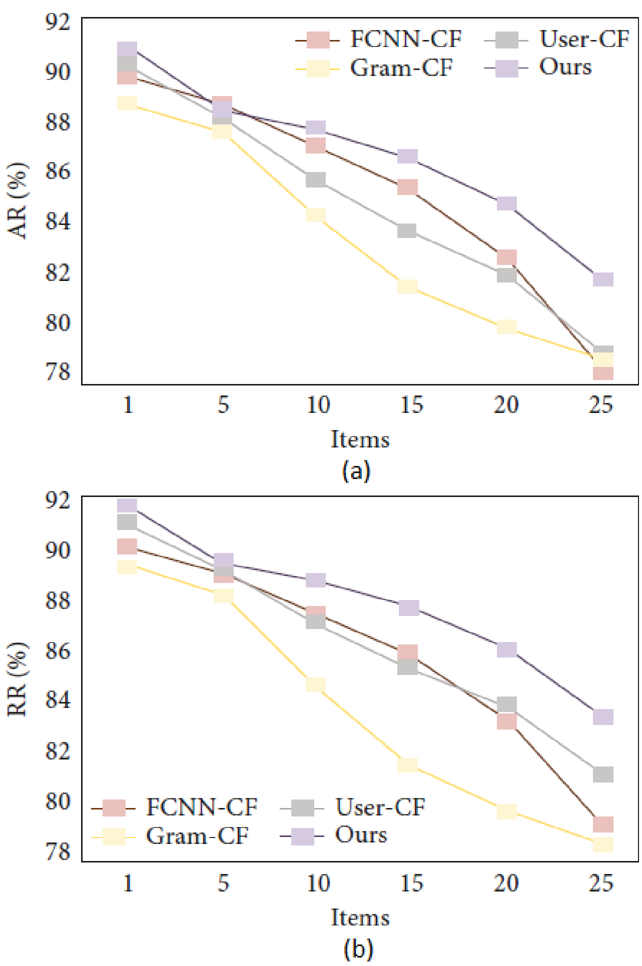


Fig. 5. Comparison of the recommended performance of different models.

However, the proposed model shows and improvement in terms of precision by 2.60%, 0.78% and 1.79%, respectively compared to the reference models, and it shows an improvement in terms of recall, by 1.67%, 1.10%, 91.6%, and 2.58%, respectively. Meanwhile, the improvement in terms of $F1$ is 1.58, 2.38%, and 0.22%, respectively. Consequently, these experimental results prove that the proposed model shows better performance at all aspects compared to the reference models: FCNN-CF, User-CF, and Gram-CF. This improvement is due to two-dimensional feature extraction, so that temporal and spatial are merged to reinforce the features expressiveness, which in turn removes anomalous data, and improves the learning capability of the model.

3.1.2. Robustness Verification. The robustness of the proposed model is verified by performing experiments under similar input data and similar experimental platform. The results of the experiments are depicted in Fig. 6. It can be concluded that the robustness of the proposed model is higher than 81% for AR, PR, RR, and $F1$ metrics compared to the reference models as the number of recommendations approaches 25, which verifies the robustness of the proposed model.



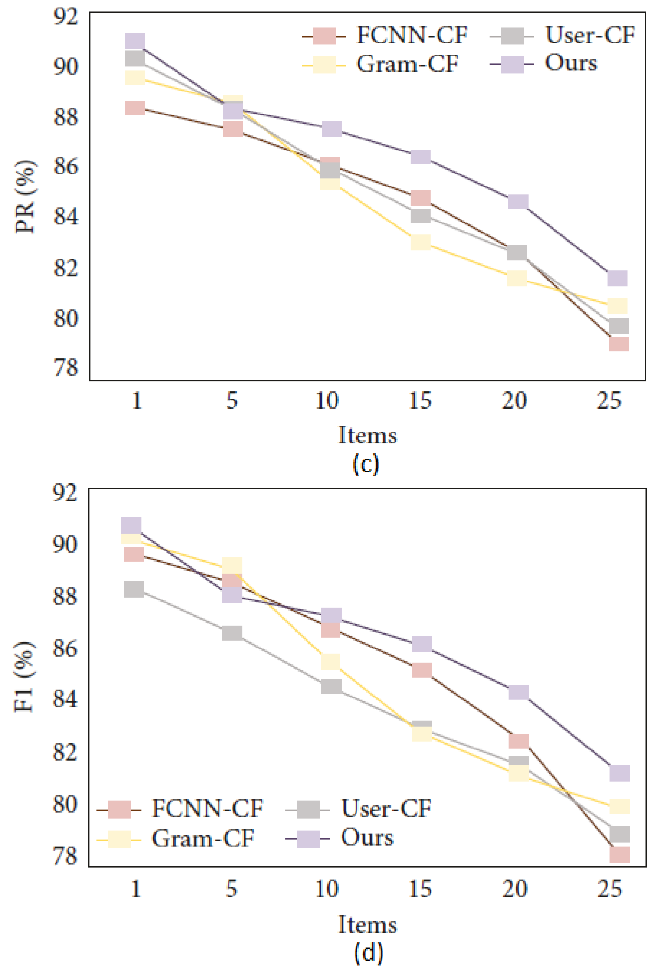


Fig. 6. Recommendation performance versus the number of recommendation items.

3.1.3. Real-Time Verification of Recommendation algorithm. The test results of the real-time performance of the recommendation algorithm of the proposed model are illustrated in Fig. 7, where tests are performed in the same environments with the same data.

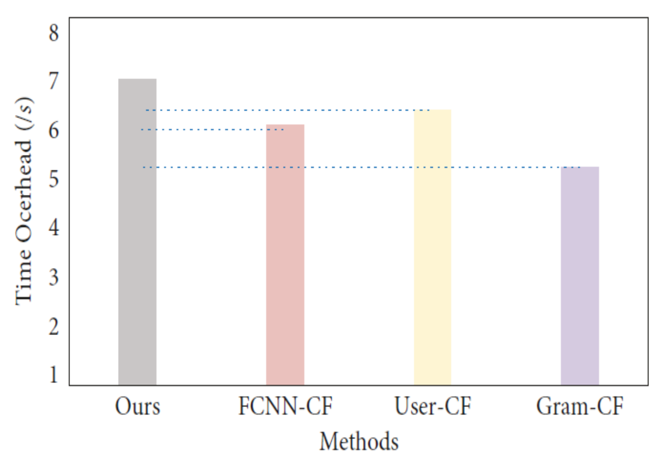


Fig. 7: Success rates of various recommendation reference models.

Fig. 7. Shows that the proposed model has the highest recommendation rate where it has a recommendation rate of 6.9 video/s, FCNN-CF model has a recommendation rate of 6.1 video/s, the User-CF model has a recommendation rate of 6.4 video/s, and the Gram CF model has a recommendation rate of 5.3video/s. The best performance of the proposed model is due to its consideration of the personalized service of users and to merging the attributes of learners and learning resources to create personalized recommendations.

3.1.4. Analysis of a Real-Life Example.

The six sets of real data used in this paper are arranged in the confusion matrix depicted in Fig. 8. The columns of the matrix represent the human models generated by the proposed model, while the rows of the matrix represent the real labels. It can be seen that the accuracy of producing human structural models for the six testers of the six sets of experiments was 92.51%, 93.32%, 93.34%, and 93.62%, respectively. These results prove that the proposed model provides stable performance on multiple data sets besides providing better real-time performance and hence it provides better robustness.

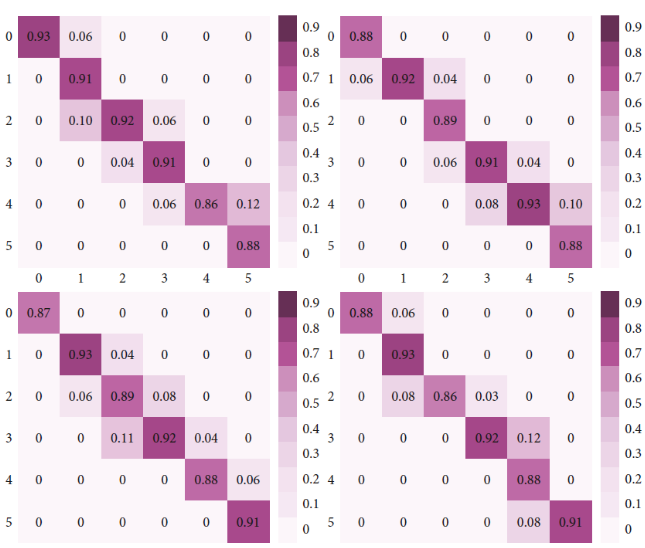


Fig. 8. Confusion matrix.

4. Conclusions

In this research work, a proposed Artificial Intelligence algorithm is presented. The purpose of the algorithm is to improve the learning capability of higher education students by facilitating the quick reading, locating, and comprehending a text section in big data and in real time attention. While improving the cognitive capability of students, it also emphasizes the quality and self-esteem of students. In this paper, it is shown that the proposed algorithm exhibits better real-time performance compared to other standard reference models such as FCNN-CF, User-CF, and Gram-CF recommendation models, besides outperforming these models in many aspects.

References

[1] Canhoto, A.I.; Clear, F. "Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential". *Bus. Horiz.*, *63*, 183–193, 2020.

[2] Dorça, F.A.; Lima, L.V.; Fernandes, M.A.; Lopes, C.R. "Comparing strategies for modeling students learning styles through reinforcement learning in adaptive and intelligent educational systems: An experimental analysis", Expert Syst. Appl., 40, 2092–2101, 2013.

[3] Loftus, M.; Madden, M.G. "A pedagogy of data and Artificial Intelligence for student subjectification", Teach. High. Educ.,25,456–475, 2020.

[4] Alkhatlan A, Kalita J, "Intelligent tutoring systems: a comprehensive historical survey with recent developments", Int J Comput Appl 975:8887, 2018.

[5] Xu, J.; Moon, K.H.; van der Schaar, M. "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs". *IEEE J. Sel. Top. Signal Process.*, *11*, 742–753, 2017.

[6] Khare K, Stewart B, Khare A, "Artificial intelligence and the student experience: an institutional perspective", IAFOR J Educ 6(3):63–78, 2018.

[7] S. Makridakis, "The forthcoming artificial intelligence (AI) revolution: its impact on society and firms," Futures, vol. 90, no. jun., pp. 46–60, 2017.

[8] Abdi, S., Khosravi, H., & Sadiq, S. "Modelling learners in crowdsourcing educational systems", in: International Conference on ArtificialIntelligence in Education (pp. 3–9). Springer, 2020.

[9] A. I. Review, "About the authors," Artificial Intelligence Review, vol. 15, no. 6, pp. 1–6, 2016.

[10] Chounta, I. A., Bardone, E., Raudsep, A., & Pedaste, M. "Exploring teachers' perceptions of artificial intelligence as a tool to support their practice in Estonian k- 12 education", International Journal of Artificial Intelligence in Education, 1–31, 2021.

[11] Yigitcanlar T., Mehmood R., and Corchado J. M., "Green Artificial Intelligence: Towards an Efficient, Sustainable and Equitable Technology for Smart Cities and Futures," *Sustainability*, vol.13, no.16, pp:1-14, August 2021.

[12] J. P. Davis and W. A. Price, "Deep learning for teaching university physics to computers," American Journal of Physics, vol. 85, no. 4, pp. 311-312, 2017.

[13] Belgaum M. R., Alansari Z., Musa S., Alam M. M., and Mazliham M. S., "Role of artificial intelligence in cloud computing, IoT and SDN: Reliability and scalability issues," International Journal of Electrical and Computer Engineering, vol.11, no.5, pp:4458-4470, October 2021.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
The author contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself
No funding was received for conducting this study.

Conflict of Interest
The author has no conflict of interest to declare that is relevant to the content of this article.