A Personalized Recommendation System for English Teaching Resources Based on Learning Behavior Detection

Zhang Yuanfei

College of International Studies, Beibu Gulf University, Qinzhou 535011, China

Correspondence should be addressed to Zhang Yuanfei; zhangyuanfei@bbgu.edu.cn

Received 9 July 2022; Revised 12 August 2022; Accepted 16 August 2022; Published 7 September 2022

Academic Editor: Muhammad Zakarya

Copyright © 2022 Zhang Yuanfei. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

When confronted with a plethora of resources, many students struggle to quickly filter out the content that is relevant to them. Because there are many English teaching resources and it is difficult to accurately recommend suitable teaching resources for students. Therefore, in this paper we suggest a personalized recommendation system for English teaching resources, which is founded on learning behavior detection. To begin with, a spatiotemporal convolutional network is introduced to effectively identify students’ online classroom behavior, and a global attention module is added to increase the model’s ability to learn global feature information. Furthermore, the identified characteristics of student behavior are incorporated into the recommendation module. Similarly, the differential evolution (DE) algorithm is implemented to the smoothing factor and kernel function center of a generalized regression neural network (CRNN) for resource recommendation mode, while taking into account the strong dependence of the GRNN training effect on the smoothing factor and the kernel function center. The smoothing factor and offset factor are optimized and solved, and the optimized smoothing factor and offset factor are used to recommend GRNN resources. Experiments show that the approach described in this work first has a high precision (i.e., 90.98%) in behavior recognition, and second, the recommendation performance is superior to both of the comparison algorithms (i.e., 85.23% and 78.33%), resulting in better resource recommendation accuracy. The fundamental goal of this work is to deliver several important guidelines for the informatization and intelligence of the English educational resources and services.

1. Introduction

Alterations will be made to the manner in which education and learning are carried out, as well as the background of the Internet and education, which will be investigated during the transformation of English teaching in schools and institutions of higher education. The proliferation of technology in schools has led to the development of a novel instructional approach known as blended learning. Mobile devices are reshaping both the teaching and learning process, as well as the relationship between instructors and students. Students are able to gain exposure to a variety of cultural perspectives when the audio-visual English instruction is integrated into a blended learning environment. This allows students to rethink the traditional approach to education that centers on the role of the teacher [1, 2]. This is particularly true when it originates to education in the actual world. For the reason that of the diminishing role of teachers and instructors, students are incapable to build their own understanding networks or pick out the most suitable learning approaches and methodologies [3], even though there are massive amounts of resources available to them. This leads to an excessive amount of information. I am unable to independently build upon and actively acquire new knowledge. According to this analysis, in order for teachers to help students integrate learning resources and improve their learning efficiency, they need to explore personalized teaching modes within the context of AI adaptation, understand their students’ current learning status and needs, and rely on intelligent teaching systems [4, 5].

It is possible, particularly, for schools to improve the quality of their teaching if they have a better understanding of how engaged students are in their own educational institutions. When evaluating the educational program at a particular university, the most important metric to look at is the degree to which students participate in the process of
their own education [6]. As an essential component of students’ active participation in their own education, researchers have spent a significant amount of time analyzing the conduct of students in the classroom. The traditional method of evaluating students’ behavior in the classroom is to do so manually, which is a time-consuming process. Because of the rapid progress that has been made in AI in recent years, we are now in a position to turn this disadvantage into a strength by utilizing AI technology [7, 8]. It has become a major issue in the development of education, which will lead to the development of an intelligent, efficient, and comprehensive education analysis system. Recognize how students learn in a classroom setting.

Because of the explosive growth of the Internet, human behavior recognition has found widespread use in a variety of contexts, including video surveillance and video comprehension, to name just two examples [9, 10]. The most important aspect of human action recognition is learning how to extract rich and discriminative features for the purpose of fully describing the spatial and temporal information of human actions. By comparison, the RGB video-based methods of behavior recognition are currently receiving less attention than skeleton-based ones. This is, in fact, due to the latter’s poor adaptability to dynamic environments and certain complex backgrounds. By contrast, the skeleton-based methods of behavior recognition are currently receiving more attention.

This should be noted that learners will find it increasingly difficult to locate helpful learning resources on the Internet as the amount of data that is stored on the Internet, which is still continues to grow very rapidly. The process of extracting useful data from the network will require users to devote a greater amount of their time to it. As a consequence of this, server-side records, statistics, and calculations are utilized in the process of implementing personalized resource recommendation for users. As a result, users are able to rapidly obtain valuable data from the massive amounts of data [11–13]. In fact, users currently prefer cloud-based online learning, but the difficulty of pushing resources is increased due to the large amount of resources in the online environment, the diversification of resource forms, and the limitation of available platforms for resources. However, cloud-based online learning is currently the preferred option for majority of the users. In order to provide accurate resource recommendations, it is necessary to perform an in-depth analysis on both the user and the resource, as well as education on their respective characteristic attributes. After that, you should look for a resource that corresponds to that user’s smallest distinguishing feature as closely as possible and then, potentially, recommend using that resource [14, 15].

At this very moment, a great number of research projects on intelligent recommendation are being carried out. A number of academic institutions manage large amounts of data and make recommendations for resources using the Hadoop platform. A number of researchers make use of the Spark platform in an effort to improve the effectiveness of the resource recommendation process [13]. They are based on massive resource recommendation research carried out in an environment utilizing cloud computing, and their primary focus is on developing a cloud computing data push platform rather than conducting in-depth research into microresources and the methods used by them. The collaborative filtering algorithm and the multiclass support vector machine algorithm can be used to construct an intelligent recommendation system. Although there has been an increase in the recommendation’s accuracy, it is still unable to accurately reflect the specifics of the micro-resources in question [15, 16].

In this paper, both global attention mechanisms and GCN are discussed with the goal of improving the detection of student actions. Deep learning, also known as DL, is a technology that has been making significant contributions to AI in recent years. Some researchers claim that they are able to more accurately identify and classify resource attributes that contain multiple features by using DNN as a recommendation algorithm for educational resources. As a consequence of this, the paper uses the generalized regression neural network (GRNN) algorithm in DL to make recommendations regarding online educational resources. In order to make the GRNN algorithm better suited for resource recommendation, additional optimizations have been made to increase the accuracy of the resource recommendations it generates. The following points discuss the fundamental contributions of our research.

(i) This paper suggests a personalized recommendation system for English teaching possessions founded on learning behavior detection.
(ii) A spatiotemporal convolutional network is introduced to effectively identify students’ online classroom behavior, and a comprehensive attention component is added in order to increase the model’s ability to acquire global feature knowledge.
(iii) The identified characteristics of student behavior are then incorporated into the recommendation module. Finally, the differential evolution (DE) procedure and set of rules are implemented to the smoothing factor and kernel function center of a generalized regression neural network (CRNN) for resource recommendation model.

The remaining portion of the manuscript is prepared as follows. In Section 2, we make available a comprehensive summary and review of the state-of-the-art literature. The methodology of the suggested research work is deliberated in Section 3. Furthermore, a mathematical model of the suggested recommendation procedure is also discussed in this section. Experiments, simulations, and empirical outcomes are deliberated in Section 4. In the last, Section 5 recapitulates this research and provides some guidelines for the future work.

2. Related Work

Students who are participating in a blended learning scenario can benefit from having access to a learning environment that is interactive and immersive thanks to the
utilization of various modes of presentation, including text, images, audio, and video. The idea of learning that continues throughout one’s life can be supported by participating in a wide range of educational pursuits, such as giving speeches, reading, and writing [1, 2].

Several researchers have investigated blended learning in relation to the recently developed audio-visual instructional strategy for the English language. Several researchers have presented ideas for models of digital education that are based on campus settings [3, 4]. Blended learning can be successfully implemented if the focus is placed on the following six measures: teacher grouping, venue separation, and time dispersion; classification of resources; a wide variety of learning approaches; and time dispersion and dispersion of time, respectively. Some academics believe that students can actively acquire and share educational resources using their familiar mobile devices in BYOD learning, and they provide an example of a student-centered foreign language teaching model based on the WeChat system. BYOD learning refers to the practice of students bringing their own devices to class in order to participate in the learning process. Some academics have attempted to combat the problem of “dumb English,” and as a result, a new classroom model has been developed that includes students in addition to teachers and students. This model was developed as a result of the efforts of these academics. Using news English audio-visual content as an example, a number of researchers have developed a flipped classroom that is based on virtual reality and carried out empirical research [5–9].

On the other hand, rather than focusing on how to better create an immersive learning environment using electronic devices, the majority of attention is centered around the development of platforms and software as well as the integration and development of learning resources. There has not been nearly enough focus placed on conducting in-depth research into the operation of course services. In light of students’ restricted capacity for independent learning, AI-based adaptive education mechanisms are an immediate necessity for the purpose of complementing the work of educators and delivering individualized support to pupils.

On the basis of their level of complexity, human behavior can be broken down into four different categories: posture, individual action, interactive action, and group activity. Gesture can be defined as the movement of the human skeleton [10, 17], which is the most fundamental part of the body. A sequence of well-coordinated movements will eventually result in a single action. Examples of interactions include those between humans, as well as those between humans and objects. This term, when applied to the context of group activities, refers to pursuits that involve a number of people and a number of different things [18, 19]. The actions of students in classroom scenes are not limited to those that are simply related to posture. This category includes not only individuals but also things and activities, such as writing on paper or playing with mobile phones, for example [20].

Visual behavior recognition, in the vast majority of instances, requires both the characterization of behavior and the detection of targets. The data obtained from using pose estimation to determine the position and motion of each joint in the human body can then be used to characterize human behavior. Pose estimation is used to obtain this information. Key is in two dimensions for a number of different players [21]. There are two different kinds of point detection algorithms: top-down and bottom-up. These names refer to the direction in which key points and the human body are detected first. The most traditional bottom-up method available in OpenPose is used to determine the joint points of body parts. This method measures the body parts’ maximum thermal value. After that, it will be possible to construct a human posture skeleton and rapidly connect the different joint points with one another. The OpenPose algorithm can still produce high-quality results and has a high degree of robustness, even if there are more people in an image. This is because the algorithm is able to scale well.

The RNN, CNN, and GCN action recognition methods are the three primary varieties of skeleton-based action recognition methods [11, 14]. The most well-known ST-GCN model encodes skeleton sequences by first constructing a spatiotemporal graph, then stacking a series of spatiotemporal graphs, utilizing convolutional extract features, and finally performing predictive recognition utilizing a model constructed with GCN. This model was developed by Google Brain. Based on GCN, a number of researchers have proposed an attention-enhanced graph convolution LSTM network, which is abbreviated as AGC-LSTM, for the recognition of human skeleton behavior. It is possible to learn high-level semantic spatial and temporal features. Researchers have proposed a structure known as the 2s-AGCN, which learns the structure of the skeleton graph in an adaptive manner. This enables a greater degree of model flexibility and represents a departure from the conventional manual setting. An operator method has been proposed by a number of researchers, and it is capable of extracting multiscale structural features and modeling long-term context dependencies from graph convolutions [5, 7, 9, 18, 19].

When teaching English in the context of information technology, it is imperative to take into account the learning habits of students and the individual differences of those students. This will allow educational resources to be distributed more efficiently and will also allow educational services to become more intelligent. Scholars have conducted research in this field, and some of them have proposed a personalized recommendation model of learning content that is based on user interest, learning preference [12, 13], and knowledge model in order to answer the issue of massive resources and personalization. A number of researchers have developed a framework for the recommendation of personalized learning resources based on the learner model. The framework makes use of a hybrid recommendation system that generates an electronic schoolbag learning database. Researchers have found that taking into account user cognition and the rules of neighboring user groups are the best way to generate an optimal personalized learning path for each individual student.

The hidden Markov model is a tool that is utilized by some academics in order to model the emotional states of learning that students are experiencing, comprehend the
changes that students are experiencing in their emotional cognition, and adapt teaching strategies to fit the overall cognitive state of learners (HMM) [15, 16, 22–26]. Personal recommendation and learning system adaptability research are currently in the early stages of weak artificial intelligence [27, 28]. This is not ideal for the personalized learning needs of English audio-visual courses because of the limitations of weak artificial intelligence [29, 30].

3. The Proposed Method

3.1. Learning Behavior Recognition. First of all, this paper proposes a method for identifying learning behaviors in school English classrooms. This paper divides learning behaviors into listening, speaking, reading, and writing. After effectively identifying student behaviors, English teaching resources for corresponding behaviors are recommended to students through a personalized recommendation system [31].

This information can be obtained by using pose estimation algorithms like OpenPose in conjunction with depth cameras like the Kinect. A frame’s skeletal information is represented by vectors, and the 2D or 3D coordinates of each human joint are represented by corresponding vectors. Vectors are used to represent skeletal information in frames. The natural human body connections that are shown in Figure 1 of this article are used in order to connect the various joints to one another [32].

The ST-GCN first obtains the skeleton sequence data, which is composed of coordinates, in order to model the structural information that is present between these joints. Joints that exist in one frame are regarded as possessing the same spatial dimension, whereas joints that exist in another frame are regarded as possessing a time dimension that is distinct from the former. A residual mechanism is added in between each ST-GCN, and there are a total of nine ST-GCN units, each of which is composed of the GCN and TCN [33]. These layers make up the TCN module and referred to as the ReLU layer, dropout layer, 1-D convolution layer, and the batch normalization layer [34]. The following is how the spatial graph convolution for each node in the graph representing the human skeleton is calculated using the following equation:

\[ f_o(x_i) = \sum \frac{x_j}{x_j} f_i(x_j) w(I_i(x_j)), \]  
(1)

where \( f_o(x_i) \) and \( f_i(x_i) \) are output and input feature maps, respectively. Furthermore, \( w(I_i(x_j)) \) is the weight function, and \( I_i(x_j) \) is the mapping function.

The calculation method, which is based on the adjacency matrix \( A_i \), is as follows:

\[ f_o = \sum w_i f_i (B_i^{0.5} A_i^{-0.5}). \]  
(2)

This should be noted that the weight of the attention approach is calculated using (3) as follows:

\[ W_{att} = a (g(AP(f_i)) \oplus MP(f_i)), \]  
(3)

where \( AP(f_i) \) is the average pool operation, \( MP(f_i) \) is the max pool operation, \( g \) is convolution operation, and \( a \) is the ReLU function.

The data that have been preprocessed are delivered to the action recognition model by the first branch stream after it has been received. The first thing that this network does is a process the data that have received. After that, it determines which joints are activated via the CAM, and then modifies the joint that corresponds to that activation in the mask matrix. The value of the updated mask matrix is passed through to the second stream, where it is then saved in the variable. The output for the second stream network is created by giving it as input the data that have been preprocessed as well as the mask matrix.

In a similar manner, the input of the third stream network is multiplied by the mask matrix from the second stream, and the results are aggregated. This ensures that the input of the second and third streams is only composed of joints that were not activated by the streams that came before them, which in turn enables the action recognition model to investigate additional feature information that differentiates each joint. The loss function of the network is computed using the following equation:

\[ L = y \ln \tilde{y} - \sum y_j \ln \tilde{y}_j. \]  
(4)

Locating the target detection module at the coordinates of the hand, intercepting the hand in the original image, and inputting the information into the small classification network are the steps that need to be taken in order to extract features from the partial picture of the hand. It is important that the model’s attention be drawn to the position of the hand. The next step is to incorporate information about the student’s posture into the context of the classroom instruction in order to further constrain the behavior that was detected and to improve the accuracy of the four actions (listening, speaking, reading, and writing).

3.2. Recommendation Module. The main organization of the GRNN model is presented in Figure 2. The model comprises an input layer and an output layer. In addition, there is a
patterns and a summation layers that help to optimize the learning procedure.

The input of the model is identified by the $X$ matrix as is given in the following equation:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

(5)

The output is defined in the following equation:

$$Y = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix}$$

(6)

The output of the Pattern Layer $p_i$ is given by the following equation:

$$p_i = \exp\left(\frac{-\left(x_i - x_j\right)^T \left(x_i - x_j\right)}{2\beta}\right),$$

(7)

where $\beta$ is the plain language factor.

In the next module, we introduce the offset factor $\chi$, and then, we have definition as given in the following equation:

$$p_j = \exp\left(\frac{-\left(x_i - \chi x_j\right)^T \left(x_i - \chi x_j\right)}{2\beta}\right).$$

(8)

The output of the Summation Layer is given by the following equation:

$$S_i = \sum w_{ij} p_j.$$  

(9)

Lastly, we introduce the DE mechanism to optimize $\beta$ and $\chi$.

The collaborative recommendations of English audiovisual resources can be of assistance in this process as the English educational resource service moves closer to informatization and intelligence. The first thing that needs to be done is to construct a template for the generation of recommendations. According to the findings of a recent study, the intelligent recommendation system has the potential to assist users in broadening their cognitive horizons by analyzing the behavior of user groups. The students’ scores and cognitive abilities are analyzed using mobile devices, which also collect the students’ scores. Matrix decomposition technology is used to create a feature vector, which is then used to generate an automatic recommendation list for each individual learner as well as resource. The correlation between these two variables is used to make predictions regarding the scores. Through the application of the recommendation model to the process of dynamically personalizing educational content for students, a learning effect of sufficient quality has been accomplished.

By integrating information technology and relying on an intelligent teaching system to design an appropriate learning plan for each learner’s level of proficiency, collaborative recommendation mechanisms can be used to help teachers optimize their teaching. This can be accomplished by using collaborative recommendation mechanisms. We develop the students’ capacity for self-directed learning. The specific process and flow of the information of the recommended module are shown in Figure 3.

4. Results and Discussion

The data collection for the online classroom focuses on the four behaviors of listening, speaking, reading, and writing that students frequently demonstrate in English classes. These influencing factors, which include the location of computers and students’ sitting postures, are included in the data gathering in order to create and mark the research on students’ online classroom behavior recognition more realistic. This was done in order to progress the correctness and precision of the results. We use the accuracy and the recall ratio to gauge the precision of the prediction process as given by (10), and (11), respectively.

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN},$$

(10)

$$\text{Rec} = \frac{TP}{TP + FN},$$

(11)

while TN stands for true negative and TP stands for the true false. False positive and false negative are both characterized by FP and FN, respectively. In order to quantify the outcomes of predictions, these indicators are frequently used in artificial intelligence and machine learning research. In addition to these metrics, researchers have employed the RMSE (root mean square error) and MAPE (mean absolute percentage error) indicators to show the value and precision of the prediction outcomes.

Researchers can use AI to not only increase the number of ways they can collect data on student behavior but also the efficiency with which they can do so. Using the recorded classroom videos that were intercepted, the OpenPose human body pose estimation algorithm is applied to obtain each student’s human body. These videos were taken inside the classroom. The major landmarks on the skeleton are identified and analyzed in this process. Abscissa and
ordinate are sorted, key points in the human body of one student are analyzed, and the maximum and minimum abscissa and ordinate values of the ordinate are calculated. It is expanded so that the body area of a single student is proportional to the total scene image's human body area and can be intercepted so that a single student can be found, located, and studied within the video frame. This allows for the detection, location, and study of a single student within the video frame.

The participants in this study consisted of one hundred different college students. An online classroom simulation was used to collect data from one hundred students who participated in the four activities that were presented earlier in this paragraph. One behavior must be completed by each student in two separate groups for there to be sufficient assurance that at least one group of data has been gathered taking into account the influence factors such as sitting posture. A total of two hundred video files are produced as a result of recording each behavior in the form of a series of video files. The video data file is cropped so that there is always only one student visible in each video. This is done so that the recognition effect of each behavior can be maintained. During the experiment, the skeleton data from the online classroom behavior were scrambled. As a result, only about 80 percent of the students were chosen to be trained on, and only about 20 percent were used for testing.

First, we compared the recognition accuracy of our method and ST-GCN, GCN, and CNN under different iterations. Comparison of accuracy of OUR, ST-GCN, GCN, and CNN is shown in Figure 4.

The comparison of loss rates is made known in Figure 5. This could be easily perceived that the accuracy of the approach suggested in this paper is significantly better than the other three methods, and can achieve higher accuracy and lower loss with fewer iterations.

In order to prove and authenticate the recommendation effect of the approach, as suggested in this paper, TF-IDF and CNN-IDF are selected as comparison algorithms, and
Recall

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF-IDF</td>
<td>0.75</td>
</tr>
<tr>
<td>CNN-IDF</td>
<td>0.80</td>
</tr>
<tr>
<td>OUR</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Figure 7: Comparison of Recall of TF-IDF, CNN-IDF, and OUR.

Acc and Recall are selected as evaluation indicators. On the resource dataset collected in this paper and multiple public datasets, we compare the recommendation effects of different methods, as shown in Figure 6 and 7.

As can be seen from Figure 6 and 7, for multiple sample sets, the resource recommendation correctness of the approach, as suggested in this paper, is the highest, with an average value higher than 0.92, followed by CNN-IDF, Acc, and Recall both converge to above 0.85, and TF-IDF performs poorly.

5. Conclusions and Future Work

The fundamental goal of this paper is to propose a personalized recommendation system for English teaching resources that is based on the detection of learning behaviors. To begin, a spatiotemporal convolutional network is presented, and then, a global attention module is incorporated into the model in order to enhance its capability of acquiring information regarding global features. In addition to this, the recommendation module takes into consideration the observed patterns of behavior exhibited by the students. It is shown in this paper that the differential evolution (DE) algorithm that is built into a generalized regression neural network (CRNN) for resource recommendation mode has an effect on the smoothing factor and kernel function center of the GRNN. The recommendations that GRNN makes regarding resources are arrived at by combining an optimized smoothing factor with an optimized offset factor. Experiments show that the approach described in this work has superior correctness and precision for recognizing behaviors, as well as superior performance when it comes to making recommendations, in comparison with the algorithm that was used in the comparison. The determination of the research work conducted in this study is to shed some light on the informatization and intelligence of English educational resource services. The focus of this study is on English educational resource services.

In the future, we will attempt to implement and put forward new deep learning techniques. Besides this, we will further investigate how different activation functions, as well as the number of layers and hidden layers of the network model, will affect the outcomes of our study. As evidence by the literature review in the earlier sections, graph convolutional networks (GCNs) along with the attention mechanism have the capability to accurately predict the images, and we will implement the GCN model and compare it with our approach. Another interesting direction for the future work is to consider a diversity of the learning mechanism for the recommendation system, that is, different datasets, to study and generalize the outcomes. This work is limited to only a single dataset, and in the future, we will investigate other datasets. Different datasets will have more number of items, and their classification should be investigated. Finally, we will continue to study the execution time, in terms of training and prediction durations, of the proposed model.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

References


