Research Article

Optimization of Multimodal Japanese Teaching Model Using Virtual Reality

Xiuhua Ye

School of Foreign Languages, Jiaying University, Guangdong, Meizhou 514015, China

Correspondence should be addressed to Xiuhua Ye; 200401049@jyu.edu.cn

Received 26 April 2022; Revised 23 May 2022; Accepted 13 June 2022; Published 4 July 2022

Academic Editor: Muhammad Babar

Copyright © 2022 Xiuhua Ye. 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.

The potential of virtual reality technology to aid education is well acknowledged. It has already been implemented in at least 20 public schools and institutions, with many more participating in assessment and research initiatives. The proposed methodology is to use virtual reality technology in Japanese teaching in order to create an immersive virtual teaching environment for the learners. They can see and hear Japanese teaching through visual, auditory, tactile, and other senses. They can also interact through virtual reality. The objects simulated by this technology interact with each other, thus creating a feeling of living in a super world. The products of virtual reality technology can be used as a new tool for learning the Japanese language. On the one hand, the advantages of virtual reality technology are assisting Japanese teaching and training due to its immersive, interactive, and systematic characteristics. On the other hand, we can use multimodal data fusion to make the data sources more diversified and optimize the Japanese teaching model through virtual reality technology. In addition, more effective and interesting teaching methods provide references for others. The experimental results show that our multimodal fusion BP (backpropagation) neural network algorithm provides a better result than the single decision tree algorithm and support vector machine algorithm.

1. Introduction

Virtual reality (VR) technology, often called virtual environment, computerized simulation, or artificial environment, is a type of computer-generated environment [1]. Aaron Lacier, a computer scientist from the United States, was the first to introduce the concept of VR in the 1980s. There are several electronic technologies integrated into this virtual world computer simulation system. These technologies include computer science, sensor and imaging technology, computer simulation technology, and various types of other electronic technologies. Recently, new developments have occurred. Key technologies in virtual reality like the integration of audio localization, optical tracking and spatial tracking, model building, and a variety of other techniques have increased the importance of VR. This technology gives users an immersive sensation and allows them to engage with objects in the virtual world by stimulating their touch, visual, and auditory senses simultaneously. Students can immerse themselves in the cognitive experience of language learning, reducing the anxiety caused by language communication in the real world, cultivating the habit and ability of independent learning, and promoting the transfer of virtual scene knowledge to the actual language application environment. It also provides Japanese learners with contextualized, diverse, collaborative, and personalized target language scenes. This is a significant step forward in the development of educational technology because it introduces a new learning method in which learners acquire new knowledge and skills through interaction with the information environment with the help of an autonomous learning environment provided by virtual reality technology.

VR is not a brand-new notion. Since the late 1960s, it has existed in various incarnations in numerous locations. Before VR became widely accepted, it was known by a variety of other names, including synthetic environments, cyber-space, artificial reality, simulator technology, and others. The most recent iteration of VR is desktop VR. Desktop VR is also known by various names, such as Window on World (WoW) or nonimmersive VR [2]. Desktop VR is becoming
increasingly popular, and as a result, developers are creating applications that are not completely immersive. These non-immersive virtual reality applications are far less expensive and technically challenging to produce. They have found their way into industrial training and development programs. VR may finally be approaching the range of possibilities for widespread development and usage, particularly in the field of education. According to some reports [3], computer-based virtual learning environments (VLEs) are packaged as desktop VR. As a result, in the future, it will be easier to include it in instructional programs. Computer-based VLEs are paving the way for new horizons in teaching, learning, and practice in subjects as diverse as the military, space expedition, digital marketing, physical sciences, engineering, and others. Students can attain their learning objectives through the use of a virtual learning environment. As a result, VLE-based apps have gained popularity in mainstream education, particularly in schools and colleges. Here, it has been proved to be an effective tool to supplement traditional teaching techniques. Researchers have discovered that these types of learning settings have a better educational influence on students. A VLE can give three-dimensional insight into the structure and function of any chosen system or system component. Consequently, by engaging with and moving through the environment designed for such systems [4, 5], students may learn the fundamental concepts of such systems in a quick, efficient, and pleasurable manner. A well-known fact of virtual reality is that it can make man-made items appear as lifelike as genuine objects [6].

Due to the ongoing growth and development of the open model of Japan and due to the absorption and introduction of other cultures and the increase in the general quality of talents, it has emerged as a significant issue of discussion. According to relevant statistics, it has been estimated that around 600 schools and universities have developed Japanese majors. However, in the eyes of the general public, the Japanese major is viewed as a major with language requirements. Unless we completely understand and develop the academic content, it will be impossible for us to effectively grasp the academic content due to its growth in the number of excellent applied talents and advance societal progress. The conventional teaching model existed in our educational system before the introduction of the Internet. However, the new online teaching model introduced by the internet posed a significant threat to the old teaching model. The traditional classroom teaching style is frequently restricted in terms of teaching time, location, and the number of students. As compared to VR, it is incapable of achieving more ambitious educational objectives. It is true that in a traditional learning environment, it is easier to respond to an educational and teaching reform. Using the e-learning model in conjunction with a traditional learning model helps to avoid the drawbacks of traditional teaching models while also developing the advantages of the e-learning model. This helps in expanding the learning environment and the teaching platform [7].

The essence of the Japanese teaching model [8] is the application, and the capacity to listen and communicate is the foundation of the language. However, even though reforms in Japanese education [9, 10] are currently underway, it has failed to provide learners with an immersive learning environment. For each individual, virtual reality technology develops an imaginary simulated world that is tailored to their requirements and cognition. By incorporating multimodal data fusion technology, the simulated environment is further enhanced and refined. This method can not only increase language application abilities but also integrate Japanese information. It has a favorable impact on the teaching of Japanese as a foreign language. It is discovered after studying the relevant literature on the promotion of Japanese education by virtual reality technology that virtual reality technology can successfully increase the interest of students in practice as well as aid them during the course of their learning, thus increasing the efficacy of Japanese language instruction.

The rest of the research paper is organized as follows: Section 2 will explain the related work which includes virtual reality, multimodal data, and the Japanese teaching model. Section 3 will explain the research design related to BP neural network architecture and principle and data processing. Section 4 will elaborate the results extracted from the proposed methodology. Finally, the concluding remarks are described in Section 5.

2. Related Work

This section explains the virtual reality details, multimodal data, and the Japanese teaching methodology. This will help elaborate all the related work that has been done in this field. The explanation is as follows.

2.1. Virtual Reality (VR). Morton Heilig, an American cinematographer, was the first person to effectively deploy virtual reality technology in the United States. He allowed the spectators to enjoy an imagined voyage through Manhattan in the US due to the lack of technical assistance, lack of a distribution carrier, lack of hardware processing, and other factors. Until the late 1980s, virtual reality technology was not extensively employed. As suggested by reference [11], the implementation of virtual reality technology in basketball instruction is still considered to be in its infancy. The use of virtual reality technology in basketball education is discovered by an examination of the literature on the subject. In this paper, virtual reality technology in basketball teaching is explored further. According to the findings of the study, virtual reality technology and simulation systems in virtual reality technology have a positive role in the promotion of basketball instruction. The study in [12] suggests the use of virtual reality in music education as a teaching tool. The interactive and intuitive learning environment generated by virtual reality technology is favorable to raising the interest and excitement of learners in studying music. Consequently, it improves the overall quality and effectiveness of teaching and learning in musical classrooms. The study in [12, 13] developed a virtual reality-based Japanese educational environment. It was based on current technologies [14, 15]. Physical
computing, software development, human considerations, and the delivery of virtual reality over high-speed networks are among the fundamental hurdles of virtual reality technology [16, 17].

2.2. Multimodal Data. In comparison to the single-modal data, multimodal data (MMD) refers to how people get information about the external environment. It is referred to in this context as a modality. Obtaining information in a single mode is referred to as unimodal. However, obtaining information in more than two modes is referred to as multimodal [18–20]. Multimodal Learning Analytics (MMLA) is the use of a variety of analysis techniques to synchronously collect, integrate, and analyze the MMD in the learning process to reveal the learning mechanism in a complex learning environment. It is also defined by the International Society for Multimodal Learning [21–23]. Learning analysis technology provides scientifically sound and reliable data support. It has become increasingly popular in educational research in recent years [24–26].

The process of multimodal interaction includes the perception of the external environment by humans [27], as well as human communication and contact with others. When it comes to processing sensory information, research in cognitive neuroscience has revealed that multimodal interaction is the rule rather than the exception process [28]. Finding one’s way around the Japanese language is a complicated and diverse task. Multimodal Japanese teaching of virtual reality technology involves the interaction between students and teachers, students and peers, and between students and technical products and tools. Through the dual encoding-decoding of verbal and nonverbal information, the complete meaning of the construction process is achieved. As a result, in the multimodal Japanese teaching of virtual reality technology, there is a huge amount of multimodal data created by the interaction between a human and a computer or an item through many senses like sight, hearing, and touch.

2.3. Japanese Teaching Mode. Literature is a term that refers to the study of literature [29]. Education professionals in Japan should clarify the theoretical underpinnings of the use of the empirical teaching model as well as the essential elements of the teaching process itself. Instructors may effectively integrate contemporary information technology to enhance the atmosphere of experiential learning by using their knowledge and expertise. To create an immersive experience environment, there is a need to make full use of training centers and training base facilities. According to the literature, research on improving the effectiveness of Japanese instruction has significant practical value. Computer-assisted multimedia teaching technology offers a broad range of application value in Japanese education [30]. Accordingly, systematic network resources (referred to as NR) may be maximized through the use of a network database. It can aid students in their Japanese language learning efforts [31].

3. Research Design

In this section, the BP neural network architecture and principle and data processing will be explained. It also explains the forward and backward propagations. The explanation is as follows.

3.1. BP Neural Network Architecture and Principle. BP is an acronym that stands for backpropagation. It is also referred to as backward propagation as illustrated in Figure 1. According to the BP neural network, the neurons between each layer can only interact with the neurons in their adjacent layers. The data propagation is also directional which means that the processing results of the neurons in their previous layer are received, while the processing results of the neurons in their subsequent layer can only be propagated to the neurons in their subsequent layer [32, 33]. The network often has to train from a large number of data samples before it can be used when using a BP neural network to grade prediction. This neural network exhibits the traits of self-adaptation, self-tuning, and other qualities. Based on the learning of the data, the weights and thresholds of the neurons in each layer can be established. It allows the formation of associations and the realization of the prediction of the data that needs to be achieved. In addition to the forward propagation and backpropagation processes, the BP neural network method incorporates two other processes that are involved in the process of performance prediction.

3.1.1. BP Neural Network Architecture. When the proposed BP neural network is activated, it has the activation function \( x(x) \). Here, \( x \) represents a variable. The function employed is the sigmoid distribution function.

\[
\phi(x) = \frac{1}{1 + e^{-x}},
\]

\[
\phi'(x) = \frac{e^{-x}}{(1 + e^{-x})^2} = \phi(x)(1 - \phi(x)).
\]

This means the input vector of the input layer must be normalized to bring it within the range \([0, 1]\).

In addition, the input vector must be standardized, with the goal of improving the regularity of the data. The standardization function utilized in this system is as follows:

\[
X_i = \frac{x_i \max - x_i \min}{x \max - x \min} \times \mu + v, \quad 2\mu + v = 1.
\]

\(x_i\) denotes the input vector of the \(i\)th neuron, \(x \max\) denotes the highest value of the input vector, and \(x \min\) denotes the minimum value of the input vector, among other things. When the previous function is used, the input vector \(X\) is normalized to the given value.

To begin with, insert both the input vector and the goal output vector into your computer. To obtain the final result,
normalize both the input vector and the target output vector. Afterward, the results of the processing are delivered in a sequential manner from neurons in the higher layer to the lower layer. As previously explained, they are further processed. The output vector is obtained by the output layer after which it is compared with the goal output vector. The average squared error of the comparison is calculated. If the average squared error and all squared errors are less than 3 percent, the calculation is completed. The revised weight is obtained, and the performance is predicted further. If a mistake occurs, the backpropagation process is carried out to identify the error gradient. Following this, the feedback is given back into the normalization process. Here, the forward propagation process is carried out until the average square error is less than 3 percent.

3.1.2. Forward Propagation. It represents the transmission of information from one layer to another. It starts with the input layer and ends with the output layer. This is called forward propagation. It is the process of propagating information from the neurons of the \( k \)th layer to the neurons of the \( k+1 \)th layer of the pyramidal neural network. The connection between the input and output of the two layers is represented as follows:

\[
y_j^k = \phi_j^k \left( \sum W_{ij}^k y_i^k - \theta_j^{k+1} \right).
\]  

The number of neurons in the \( k \)th layer is represented by the letters \( j = 1, 2, 3, \ldots, n_k, n_{k+1} \), \( k = 1, 2, 3, \ldots, M - 1 \); \( n_k \) is the total number of neurons in the \( k \)th layer. The total number of layers in the BP neural network system is denoted by the letter \( M \). The weight transmitted from the \( i \)th neuron in the \( k \)th layer to the \( j \)th neuron in the \( k+1 \)-th layer is denoted by the symbol \( \theta_j^{k+1} \) in the equation. The threshold of the \( k+1 \)-th layer neuron is denoted by the symbol \( \theta \). The output amplitude of the neuron can be restricted to the range of \((0, 1)\) or to the range of \((0, 2)\). The activation function is a nonlinear function with a nonlinear mapping function. It can limit the output amplitude to the range of \((0, 1)\) and \((-1, 1)\). The sigmoid function is the activation function that was employed.

3.1.3. Backpropagation. The vector \( X_p \) is input into the input layer, and the output layer will output the vector \( Y_{p,k+1} \). Once the data propagation and computation have taken place in the input layer, it is inevitable that there will be some degree of inaccuracy between the computed vector and the theoretical output vector. Suppose that the squared error of the \( j \)th neuron in the \( k \)th layer is \( 1/2 e_j^2(n) \) and the total squared error of the output vector of the output layer is \( E(n) = 1/2 \sum e_j^2(n) \).

For instance, if the total number of samples used in the whole computation procedure is \( N \), the associated mean squared error is

\[
E_A = \frac{1}{N} \sum_{n=1}^{N} E(n).
\]  

The mean squared error is a function of modifying the weights and thresholds of each layer of the BP neural network. The ultimate goal of learning is to make the mean squared error as lower as possible. In the process of weight correction, the method used is the gradient correction method. It can be corrected according to the following formula:

\[
\Delta w_{ij}(n) = \alpha w_{ij}(n - 1) + \eta_j \delta(n) y_i(n).
\]  

Here, \( \eta \) is the learning step size in the calculation process. \( \alpha \) represents the momentum factor in the learning process, and \( \alpha \) is in the range of \((0, 1)\). The main function is to control \( \eta \) to not be too large.
For instance, \( \eta \) is the learning step size in the calculating process, \( \alpha \) is the momentum component in the learning process, and \( \alpha \) is in the range of \((0, 1)\); the major role is to regulate \( \eta \) so that it does not become too big.

Using the above computation methods and techniques, it is possible to construct a BP neural network model. By incorporating the learning of previously known data and weight correction, it is possible to further refine the weights of the backpropagation neural network (BP neural network), the connection between the input vector, and the output vector for each layer of neurons. BP neural networks are used to predict multimodal data fusion. To apply this model to multimodal data fusion prediction, also known as sample data, it must be used to complete the learning process. The goal of learning is to reduce the average squared error between the output layer and the output vector of the known data samples. However, this is not completely possible. The learning process can be terminated once the value is less than a predetermined value. In this study, when the weight adjustment is carried out by the BP neural network proposed, the weight is changed once for each sample data set that is input into the network.

3.2. Data Processing. Autoencoders are a type of deep neural network. They are one of the most often used deep neural networks in unsupervised machine learning [34]. An encoder and a decoder are the two building components of a standard autoencoder.

The encoding process of the autoencoder from the input to the hidden layer is shown in the following equation:

\[
h = g\theta_1(x) = \sigma(W_1x + b_1).
\]

The process of decoding from the hidden layer to the output layer is

\[
\hat{x} = g\theta_2(h) = \sigma(W_2h + b_2).
\]

The goal function for optimization is as follows:

\[
\text{minimise loss} = \text{dist}(X, \hat{X}^R).
\]

The information in this article was obtained from a Japanese training facility in Beijing’s Haidian District. When it comes to Japanese instruction, the data types that are most commonly encountered are video and audio. The autoencoder approach is utilized to produce a data set for the multimodal fusion of video and audio data. Then, the MATLAB platform is utilized to conduct simulation experiments on the data set. First, 80 percent of the data from the fusion data is randomly picked as the training set, and the remaining 20 percent of the data is randomly selected as the test set from the training set.

In this research, the multimodal Japanese teaching mode in virtual reality is predicted using the data presented above and the BP neural network was developed. The following is a detailed description of the procedure.

1. Perform multimodal data fusion.
2. Determine the number of nodes in the input layer, hidden layer, and output layer of the BP neural network, as well as the network parameters. After this, initialize the network.
3. Train network parameters with the preprocessed data to reduce the error between the network output and the previously acquired teaching impact.
4. Perform analysis and evaluation of the output network findings.

4. Results

This paper employs three methods to conduct test experiments to verify the benefits of the method described in this paper in terms of data accuracy in the context of implementing multimodal Japanese teaching in a virtual environment. These methods are the multimodal data fusion BP neural network algorithm (OUR), the support vector machine (SVM), and the decision tree (DT) algorithms. It is possible to compare the convergence speed of the three strategies by examining how quickly the loss function converges in each case. It is illustrated in Figure 2.

The accuracy of the OUR, SVM, and DT was calculated on the test set as seen in Figure 3.

Based on Figures 2 and 3, we can conclude that our approach exhibits good convergence. Our algorithm is faster than DT. The speed of DT algorithm is faster than the speed.
of SVM. In terms of accuracy, the SVM method is the worst. It also proves that our algorithm is better than the DT algorithm in terms of accuracy.

The findings of the BP neural network trained in this research on the test set. It was based on the data in the previous part. It is shown in Table 1.

This research proposes a strategy that can be seen in Table 1. It has high accuracy in the training dataset and can attain an accuracy of 81 percent on the test dataset. At the same time, it has been noticed that the amount of training data is gradually increased; resultanty, the accuracy of the verification may also increase. As a consequence, our suggested BP neural network technique for multimodal data fusion not only successfully fuses video and audio data but also produces more pleasing results.

5. Conclusion

In the virtual reality learning environment, students have direct communication with the scenes provided by the computer. It is similar to personalized tutoring. This provides learners with a positive personal emotional experience and effectively increases their interest in students in learning by increasing their interaction with the scenes. Two unique aspects of virtual reality technology are immersion and interactivity. Learners may actively interact and think in the learning environment about the main points and problems of learning tasks. This enhances their learning motivation and improves their learning autonomy. Vivid scenarios are also used in conjunction with the learning activities. Using virtual reality technology in Japanese education provides learners with simulated situational experience and practice of corresponding scenes. This allows learners to master Japanese language skills and communication logic in the Japanese context in a communication that is closer to reality. It also cultivates their independent thinking and skills. This can be proved helpful in their future success and builds a capacity to respond in an emergency situation.

To summarize, this research offers a BP neural network algorithm model for multimodal data fusion by merging multimodal fusion data with virtual reality technology. This algorithm model is based on a Bayesian neural network algorithm model. It is the practical modal Japanese instruction and has the ability to achieve high accuracy.

In the future, our work will focus on two areas. First, we will apply our method to optimize other language teaching modes like English and Spanish. Second, we will continue to collect multimodal teaching data and comprehensively optimize the teaching mode.

Data Availability

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

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

The work was supported by 2019 Higher Education Teaching Reform Project in Guangdong Province Reform and Exploration of Advanced Japanese Lessons in Local Colleges under the Background of New “National Standard,” under No. 420A011302.

References


