Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

Classroom management of English distance education based on improved machine learning for classification and feature extraction in artificial intelligence

XiaoYan Peng*

College of Modern Economics & Management, Jiangxi University of Finance and Economics, Email Address: 625016371@qq.com

Abstract:

As globe revolves around Internet for infinite access to materials available in a remote region, today's era is referred to as Internet era. Internet has penetrated education sector to a greater extent; traditional offline classrooms are being replaced with online as well as offline classes to facilitate teaching as well as learning. Multimedia courseware is becoming increasingly significant in classroom instruction as part of the education integration movement. We look at how AI multimedia courseware can be used in the classroom and investigate the classroom teaching mode, with a particular focus on the features of AI multimedia courseware classroom teaching method. This research propose novel technique in classroom management for English distance education based on machine learning data classification and feature extraction. Here the classroom infrastructure data has been analysed initially. Then based on the opinion of students and their grades in English subject, the positioning has been carried out. Then the historical data of classroom monitoring has been collected and processed. The features has been extracted using back propagation neural network then based on the extracted features the classification has been carried out using genetic algorithm with regression model. Based on this classified data the student positioning has been carried out. This study uses comparative experiments to validate the performance of model to verify methods performance. Findings demonstrate that approach developed in this work has some effect on real-time datasets.

Keywords: classroom management, English distance education, machine learning, classification, feature extraction, positioning

I. INTRODUCTION

AI (Artificial intelligence) is advancing at an alarming rate, and it has become ingrained in our daily lives. Global usage of method in education has revolutionised way people teach as well as learn. Teachers experience AI, which is one of the disruptive ways for customising distinct learning groups. In all industries, including education, ML and AI are the primary drivers of growth as well as innovation [1]. In next 3 years, 47 percent of learning management solutions are handled with AI talents, based on E-learning industry. Although AI-driven solutions are in use in EdTech for some time, industry has been sluggish to catch up.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

According to 86 percent of academics, method is at the centre of education. AI has potential to enhance both learning as well as teaching, and education sector can assist in the development of the greatest outcomes for students and teachers. People's learning methods have evolved as a result of AI [2]. In the practise of online college English education, AI technology as well as online media in data methods are used. The entire activity of teaching is documented and examined. "That under the method of College English instructional information," according to the findings of a teaching practise study. The framework of College English powered by AI technology" improves students' learning capacity as well as improves their English learning outcomes dramatically [3]. However, in realm of education, its implementation is fraught with difficulties as well as ethical issues. Goal of this research is to examine AI's educational prospects, benefits, and challenges. Multiple applications of AI in education have made the world of education highly convenient and customizable. The technique of learning has altered now that smart devices and computers are available to all teachers. Because of the computer and Internet connection, pupils no longer need to go to body courses to read. Companies can cut the time it takes to perform complex operations by automating AI administrative activities, while academics can spend more time with students. It's time to talk about how academic AI is changing. The goal of this research is to determine the impact of AI on education [4]. More particularly, it is attempted to determine how education influenced AI in administration of many parts of education, including instruction as well as learning. The term "AI" is frequently connected with computers. However, an examination of many papers, particularly in context of education, as basis for creation of AI methods found that the system's primary attractiveness is hardware and software, or AI equipment.

With the advancement of technology, the demand for information is increasing. In today's world, how to make education as well as training satisfy various demands of various people at various times as well as at various times is a topic that must be examined. The network is rapidly evolving, and field of data technology has seen seismic shifts. As a result, the creation and design of distant education platform began in this environment. Because of its vast volume of material as well as flexible engagement with students, distant education platform has distinct advantages in teaching [5]. Without a set location or time limits, the distant education platform can maximise students' subjective as well as independent learning. However, as a range of good network education apps continue to emerge, distance education has become increasingly popular, posing new challenges. Additional distance education systems can only move classroom education to Internet [6]. At same time, numerous courses are uninteresting and lack teacher-student interaction, reducing the benefits of the distant education platform. The majority of modern distance education is done through the use of websites. Many domestic and international companies are currently researching and developing web-based remote education methods, primarily distance education platforms based on ASP, PHP, JSP, and other methods, such as network colleges as well as virtual universities. These days, remote education team is growing rapidly, and development of an interactive distant education platform is attracting increasing societal attention [7].

Contribution of this paper is as follows:

1. To novel technique in classroom management for English distance education based on machine learning data classification and feature extraction

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

2. To analyse classroom infrastructure data and collect the historical data of classroom monitoring

3. To extract the features using back propagation neural network (BPNN) then based on the extracted features the classification has been carried out using genetic algorithm with regression model (GA-Reg)

II. RELATED WORKS

According to estimation record of questionnaire as well as designed methodology, system can estimate evaluation result. Classroom management, instructional materials, lectures, teacher ethics, professional achievements, and other applicable materials are used to evaluate teachers [8]. The ability of teachers to teach is a significant determinant of teaching quality. The reputation of the school and the development of pupils are directly tied to the level of teaching quality. As a result, the goal of conducting a teaching assessment is to summarise classroom teaching experience, offer targeted difficulties, and correct problems based on instructor inadequacy, in order to improve teaching level and ensure student quality [9]. The work on AI in LMS is mostly aimed at assisting teachers in developing better methods as well as learning approaches for use in these environments. There's a lot of information out there regarding how specific AI approaches are used in user interactions as well as how they learn from each engagement [10]. These methods are dependable as well as help greatly to progress of this research. Based on review, it is said that suggested work stands out from others in terms of combining two methods such as AI as well as data analysis, in a single setting. By combining all academic management into a single method, a virtual assistant is built that handles each student's information and provides automated and individualised monitoring. The assistant has all of the knowledge obtained from the data analysis [11], further to learning from user interaction. Evaluation is not restricted to data contained in LMS, as integration of multiple sources becomes a critical component in determining each student's needs and expectations. Big data is the technology that enables this capability, and amount and type of data that is integrated into analysis allows for decision-making flexibility [12]. Because of this connection, AI can make quick as well as accurate choices about student performance. The authors of [13] discuss an artificial English composition scoring system. This rating system helps colleges and universities enhance their teaching and learning systems. Science and technology advancements have strengthened China's national position. The English Test Management System (ETMS) [14] is the product of the autonomous test paper generation gradually replacing conventional English learning as well as examination with computer technology. The extraction of question paper is a multi-constrained problem, and the generated questions are solved using an improved genetic algorithm. The authors of [15] explore how a redesigned teaching system with extra features can improve English teaching in middle schools. Author of [16] recommends that English network intelligent teaching be implemented by merging two methods, namely AI as well as WBIETS, to create English Network Intelligent System (ENTS). The role of AI in English teaching as well as learning classrooms was investigated by authors of [17]. The old teaching paradigm is centred on teacher, but modern method is centred on learner, which promotes independent learning. Process of foreign language instruction (English) has been extended by the author of [18] using AI-powered technologies such as ICALL (intelligent computer-assisted language learning). Authors discussed successful English teaching as well as learning method with deployment of DL method in their works [19]. The author proposes in [20] that when English

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

learning is done online, a logical structure for annotations is supplied for simple course streaming. The author of [21] created and analysed a machine learning system for evaluating English education.

III. SYSTEM MODEL

This section discuss novel technique in classroom management for English distance education based on machine learning data classification and feature extraction. Here the classroom infrastructure data has been analysed initially. Then based on the opinion of students and their grades in English subject, the positioning has been carried out. Then the historical data of classroom monitoring has been collected and processed. The features has been extracted using back propagation neural network then based on the extracted features the classification has been carried out using genetic algorithm with regression model.

Online Education Method

Advancement of ICT has opened up a plethora of opportunities for educational projects to be carried out, ensuring that all individuals have access to high-quality education regardless of when or where they are. Access options that are placed in people's hands have removed time as well as distance as barriers to teaching as well as learning. It has benefit of being an asynchronous study approach, with hours as well as days of week set aside for instructor engagement. Online education is a result of society's present fast-paced lifestyle. Online education provides a mutual educational goal without constraints of geography or time, whether for employment, family, or geographical location of certain people. The following are few characteristics of online education methods:

- Students can interact with the information, their teachers, and their peers in an interactive model.
- Accessible at any time and in any location with an Internet connection.
- Students can participate in tasks or activities at same time as others in both synchronous and asynchronous modes.
- Online resources provide access to resources without necessity for them to be physically present at any given time.

Teachers use smart classroom's modern teaching platform to track students' learning data, like chapter study, sign-in, contribution in classroom activities, exercise completion after class, and so on, to create a done estimation basis as well as provide reasonable as well as effective estimation, ensuring an open, fair, and just estimation as well as arousing students' learning enthusiasm. To facilitate formation of decision tree method, this research uses computer time after class, degree of understanding of course before learning, typical homework situation, and total score attributes that are closely related to score attributes as basis for forming decision tree method of total score classification as well as generating a new basic data table for student score analysis. If Si is number of samples in method Ci, then eq. (1) gives amount of data required to categorise a given data object:

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

$$I(S_1, S_2, \cdots, S_m) = -\sum_{i=1}^{m} p_i \log_2(p_i), \tag{1}$$

where pi is probability that any data object belongs to type $C_i + P_i = S_i/S$. Let an attribute A take v various values $\{a_1, a_2, \cdots, a_v\}$. Data required to partition current sample set utilizing attribute A can then be determined using formula in eq. (2):

$$E(A) = \sum_{i=1}^{F} \frac{S_{1i}+\cdots+S_{mi}}{S} I(S_{1j}, \cdots, S_{mj}) \tag{2}$$

As a result of dividing appropriate sample set of current branch node by attribute A, data gain received is as eq. (3):

$$\text{Gain}(A) = I(S_1, S_2, \cdots, S_m) - E(A) \tag{3}$$

The sum of squares of cluster centre errors of all objects in data set as well as their associated clusters is usually used to calculate it. The clustering operation can be completed when sum of squares errors is modest enough. Clustering criterion function is expressed as eq. (4):

$$\text{RSS} = \sum_{i=1}^{k} \sum_{p\in c_r} |p - m_i|^2 \tag{4}$$

The data set is divided into k clusters, with ci representing set of i-class data items, p representing data objects in cluster ci, mi representing average value of cluster ci, and k representing that data set is divided into k clusters. Cluster analysis is utilized as a pre-processing step in other mining methods and as a module of data mining. The percentage of X, Y contained in the transaction in D is the support degree s, which is the probability P(X ∩ Y), and its expression is as eq. (5):

$$s(X \Rightarrow Y) = P(X \cap Y). \tag{5}$$

Therefore, formula of on f(X ⇒ Y) is as follows $\text{onf}(X \Rightarrow Y) = \frac{\text{supp}(X\Rightarrow Y)}{\text{supP}(X)}$

To build the formula for support value, compare frequency of a project set with reference frequency in various project sets. Project set's benchmark frequency is considered statistically independent. DF factor is calculated as eq. (6):

$$DF(I) = |P(X) - P(Y)| \times \frac{P(X\cup Y)}{P(X)P(Y)} \tag{6}$$

Students' images are constantly improved, their preferences are methodically as well as continuously studied, tailored learning identifies are realised, development trends are forecasted as well as best learning behaviour techniques are examined through behavioural modelling. Teachers provide various assignments based on every student's learning condition, as well as they push various learning materials as well as data

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

into a personalised push task book. Teacher's homework is completed and returned to teacher so that teacher can have a better understanding of students' understanding of objective issues. Teachers' subjective questions are evaluated, accepted, and recorded in microclass, and a overall summary is prepared based on that dat. To attain communication as well as data feedback with teachers, students are taught to use Internet platform media to view teachers' microlesson video gaining knowledge as well as share their own ideas and comments.

Feature extraction using back propagation neural network:

The backpropagation technique is used to train regular feedforward neural networks. In this case, a specific input is initially propagated across network before output is computed. Forward pass is what it's called. A differentiable loss function is then utilized compare output to a ground truth label. Gradients of loss based on all of specifications in network are evaluated using chain rule in the backward pass. A standard gradient descent update rule is given by: A standard update rule is given by eq. (7): The loss function between network output y and ground truth label y is given by L (y, y), and vector of all weights in network is denoted by w.

$$\mathbf{w} \leftarrow \mathbf{w} - \eta - \nabla_{\mathbf{w}} \mathcal{L}(y, \hat{y}) \tag{7}$$

The learning rate, which governs magnitude of steps made with every weight update, η is shown here. In practise, input samples will be sampled from the training dataset in equal-sized batches, with loss averaged or summed, resulting in less noisy updates. Mini-batch gradient descent is the term for this method. RMSprop and Adam are two further gradient descent flavours that extend on Equation (4), making optimization technique even more robust.

The output activation function ψ in a regression model is usually the identity function; to be more general, we suppose by eq. (8)

$$\psi(a_1, \dots, a_K) = \big(g_1(a_1), \dots, g_K(a_K)\big) \tag{8}$$

where g1, g2,..., gK are R-to-R functions. Let's compute Ri's partial derivatives with respect to output layer's weights as shown in eq. (9)

$$a^{(L+1)}(x) = b^{(L+1)} + W^{(L+1)} h^{(L)}(x)$$

$$\frac{\partial R_i}{\partial W_{k,m}^{(L+1)}} = -2\left(Y_{i,k} - f_k(X_i, \theta)\right) g_k'\left(a_k^{(L+1)}(X_i)\right) h_m^{(L} \tag{9}$$

Differentiating now with respect to weights of previous layer $\frac{\partial R_i}{\partial W_{m,l}^{(L)}} = -2 \sum_{k=1}^{K} \left(Y_{i,k} - f_k(X_i, \theta)\right) g_k'\left(a_k^{(L+1)}(X_i)\right)$ by eq. (10)

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

$$a_k^{(L+1)}(x) = \sum_j W_{k,j}^{(L+1)} h_j^{(L)}(x)$$

$$h_j^{(L)}(x) = \phi\left(b_j^{(L)} + \left\langle W_j^{(L)}, h^{(L-1)}(x)\right\rangle\right) \tag{10}$$

This leads to eq. (11)

$$\frac{\partial a_k^{(L+1)}(x)}{\partial W_{m,l}^{(L)}} = W_{k,m}^{(L+1)} \phi'\left(b_m^{(L)} + \left\langle W_m^{(L)}, h^{(L-1)}(x)\right\rangle\right) h_l^{(L-1)}(x) \tag{11}$$

Let us introduce notations by eq. (12)

$$\delta_{k,i} = -2\left(Y_{i,k} - f_k(X_i,\theta)\right) g_k'\left(a_k^{(L+1)}(X_i)\right)$$

$$s_{m,i} = \phi'\left(a_m^{(L)}(X_i)\right) \sum_{k=1}^{K} W_{k,m}^{(L+1)} \delta_{k,i-} \tag{12}$$

Then by eq. (13)

$$\frac{\partial R_i}{\partial W_{k,m}^{(L+1)}} = \delta_{k,i} h_m^{(L)}(X_i)$$

$$\frac{\partial R_i}{\partial W_{m,l}^{(L)}} = s_{m,i} h_i^{(L-1)}(X_i) \tag{13}$$

The gradient values are used to update the gradient descent algorithm's parameters. We have the following at step r + 1 by eq. (14):

$$W_{k,m}^{(L+1,r+1)} = W_{k,m}^{(L+1,r)} - \varepsilon_r \sum_{i \in B} \frac{\partial R_i}{\partial W_{k,m}^{(L+1,r)}}$$

$$W_{m,l}^{(L,r+1)} = W_{m,l}^{(L,r)} - \varepsilon_r \sum_{i \in B} \frac{\partial R_i}{\partial W_{m,l}^{(L,r)}} \tag{14}$$

where B is a batch and $\varepsilon_r > i$ is learning rate that satisfies $\varepsilon_r \to 0, \sum_r \varepsilon_r = \infty, \sum_r \varepsilon_r^2 < \infty$, for example $\varepsilon_r = 1/r$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

We use Backpropagation equations to evaluate gradient by a two pass method. In forward pass, fix value $\theta^{\cdots}(r) = \left(W^{(1,r)}, b^{(1,r)}, \dots, W^{(L^*_{+1},r)}, b^{(L+1,r)}\right)$ and we evaluate estimated values $f\left(X_i, \theta^{(r)}\right)$ and all intermediate values $\left(a^{(k)}(X_i), h^{(k)}(X_i) = \phi\left(a^{(k)}(X_i)\right)\right)'_{1\le k\le L+1}$ that are stored.

Indeed for sigmoid function by eq. (15)

$$\phi(x) = \frac{1}{1+\exp(-x)},\ \phi'(x) = \phi(x)(1-\phi(x)) \tag{15}$$

For hyperbolic tangent function ("tanh") by eq. (16)

$$\phi(x) = \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)},\ \phi'(x) = 1-\phi^2(x) \tag{16}$$

Let us consider here a K class classification issue. Output is $f(x) = \begin{pmatrix} \mathbb{P}(Y=1/x) \\ \cdot \\ \mathbb{P}(Y=K/x) \end{pmatrix}$. Softmax function is assumed to be the output activation function by eq. (17).

$$\operatorname{softmax}(x_1, \dots, x_K) = \frac{1}{\sum_{k=1}^{K} e^{x_i}}\left(e^{x_1}, \dots, e^{x_K}\right) \tag{17}$$

Let's compute the gradient using some helpful computations.

$$\frac{\partial \operatorname{softmax}(x)_i}{\partial x_j} = \operatorname{softmax}(x)_i(1-\operatorname{softmax}(x)_i) \text{ if } i=j = -\operatorname{softmax}(x)_i \operatorname{softmax}(x)_j \text{ if } i \neq j$$

Then we have $-\log(f(x))_y = -\sum_{k=1}^{K} \mathbf{1}_{y=k} \log(f(x))_k = \ell(f(x), y)$, for loss function l associated to cross-entropy.

we want to evaluate gradients by eq. (18)

$$\text{Output weights } \frac{\partial \ell(f(x),y)}{\partial W_{i,j}^{(L+1)}} \text{ Output biases } \frac{\partial \ell(f(x),y)}{\partial b^{(L+1)}} \text{ Hidden weights } \frac{\partial \ell(f(x),y)}{\partial W_{i,j}^{(h)}} \text{ Hidden biases } \frac{\partial \ell(f(x),y)}{\partial \delta_i^{(h)})} \tag{18}$$

for $1 \le h \le L$. We utilize chain-rule : if $z(x) = \phi\left(a_1(x), \dots, a_J(x)\right)$, then $\frac{\partial z}{\partial x_i} = \sum_j \frac{\partial z}{\partial a_j}\frac{\partial a_j}{\partial x_i} = \left\langle \nabla\phi, \frac{\partial a}{\partial x_i}\right\rangle$.

Hence we have by eq. (19)

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

$$\frac{\partial \ell(f(x),y)}{\partial (a^{(L+1)}(x))_i} = \sum_j \frac{\partial \ell(f(x),y)}{\partial f(x)_j} \frac{\partial f(x)_j}{\partial (a^{(L+1)}(x))_i}$$

$$\frac{\partial \ell(f(x),y)}{\partial f(x)_j} = \frac{-1_{y=j}}{(f(x))_y}.$$

$$\frac{\partial \ell(f(x),y)}{\partial (a^{(L+1)}(x))_i} = -\sum_j \frac{1_{y=j}}{(f(x))_y} \frac{\partial \mathrm{softmax}\left(a^{(L+1)}(x)\right)_j}{\partial (a^{(L+1)}(x))_i}$$

$$= -\frac{1}{(f(x))_y} \frac{\partial \mathrm{softmax}\left(a^{(L+1)}(x)\right)_y}{\partial \left(a^{(L+1)}(x)\right)_i} \tag{19}$$

$$= -\frac{1}{(f(x))_y} \mathrm{softmax}\left(a^{(L+1)}(x)\right)_y \left(1 - \mathrm{softmax}\left(a^{(L+1)}(x)\right)_y\right) 1y = i$$

$$+ \frac{1}{(f(x))_y} \mathrm{softmax}\left(a^{(L+1)}(x)\right)_i \mathrm{softmax}\left(a^{(L+1)}(x)\right)_y \mathbf{1}_{y<i}$$

$$\frac{\partial \ell(f(x),y)}{\partial (a^{(L+1)}(x))_i} = \left(-1 + f(x)_y\right)\mathbf{1}y = i + f(x)_i \mathbf{1}y \neq i.$$

Hence we obtain, $\nabla_{a^{(L+1)}(x)} \ell(f(x),y) = f(x) - e(y)$,

The partial derivative of loss function based on output bias can now be easily obtained. Since by eq. (20)

$$\left.\frac{\partial \left(\left(a^{(L+1)}(x)\right)\right)_j}{\partial (b(L+1)}\right)_i = \mathbf{1}_{i=j}$$

$$\nabla b(L+1)\ell(f(x),y) = f(x) - e(y) \tag{20}$$

Let's calculate the partial derivative of loss function in terms of output weights now by eq. (21).

$$\frac{\partial \ell(f(x),y)}{\partial W_{i,j}^{(L+1)}} = \sum_k \frac{\partial \ell(f(x),y)}{\partial (a^{(L+1)}(x))_k} \frac{\partial \left(a^{(L+1)}(x)\right)_k}{\partial W_{i,j}^{(L+1)}}$$

$$\frac{\partial \left(a^{(L+1)}(x)\right)_k}{\partial W_{i,j}^{(L+1)}} = \left.a^{(L)}(x)\right)_j \mathbf{1}_{i=k-} \tag{21}$$

$$\nabla w(L+1)\ell(f(x),y) = (f(x) - e(y))\left(a^{(L)}(x)\right)'$$

Now we'll compute the loss function's gradient at hidden layers. The chain rule is applied by eq. (22).

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

$$\frac{\partial \ell(f(x),y)}{\partial\left(h^{(k)}(x)\right)_j} = \sum_i \frac{\partial \ell(f(x),y)}{\partial\left(a^{(k+1)}(x)\right)_i} \frac{\partial\left(a^{(k+1)}(x)\right)_i}{\partial\left(h^{(k)}(x)\right)_j} \tag{22}$$

We recall that by eq. (23)

$$\left(a^{(k+1)}(x)\right)_i = b_i^{(k+1)} + \sum_j W_{ij}^{(k+1)}\left(h^{(k)}(x)\right)_{j+} \tag{23}$$

Hence by eq. (24)

$$\begin{gathered}\frac{\partial \ell(f(x),y)}{\partial h(k)(x)_j} = \sum_i \frac{\partial \ell(f(x),y)}{\partial a^{(k+1)}(x)_i} W_{i,j}^{(k+1)} \\ \nabla_{h(k)}(x)\ell(f(x),y) = \left(W^{(k+1)}\right)^t \nabla_a^{(k+1)}(x)\ell(f(x),y)\end{gathered} \tag{24}$$

Recalling that $h^{(k)}(x)_j = \phi\left(a^{(k)}(x)_j\right)_0$ by eq. (25).

$$\frac{\partial \ell(f(x),y)}{\partial a^{(k)}(x)_j} = \frac{\partial \ell(f(x),y)}{\partial h^{(k)}(x)_j} \phi'\left(a^{(k)}(x)_j\right) \tag{25}$$

Hence by eq. (26).

$$\nabla_{a^{(k)}(x)}\ell(f(x),y) = \nabla_{h^{(k)}(x)}\ell(f(x),y) \odot \left(\phi'\left(a^{(k)}(x)_1\right), \ldots, \phi'\left(a^{(k)}(x)_j\right), \ldots\right)' \tag{26}$$

where $\odot$ represents element-wise product by eq. (27)

$$\frac{\partial \ell(f(x),y)}{\partial W_{i,j}^{(k)}} = \frac{\partial \ell(f(x),y)}{\partial a^{(k)}(x)_i} \frac{\partial a^{(k)}(x)_i}{\partial W_{i,j}^{(k)}} = \frac{\partial \ell(f(x),y)}{\partial a^{(k)}(x)_i} h^{(k-1)}(x) \tag{27}$$

Finally, the loss function's gradient with respect to hidden weights is given by eq. (29)

$$\nabla w^{(k)}\ell(f(x),y) = \nabla \ell(f(x),y) h^{(k-1)}(x)' \tag{29}$$

The gradient must be computed with regard to the hidden biases as the final step. All we have to do now is by eq. (30)

$$\frac{\partial \ell(f(x),y)}{\partial b_i^{(k)}} = \frac{\partial \ell(f(x),y)}{\partial a^{(k)}(x)_i}$$

$$\nabla_{b(k)}\ell(f(x),y) = \nabla_{a^{(k)}(x)}\ell(f(x),y). \tag{30}$$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

Stochastic gradient based back propagation algorithm:

Forward pass: Value of current weights $\theta^{(r)} = \left(W^{(1,r)}, b^{(1,r)}, \dots, W^{(L+1,r)}, b^{(L+1,r)}\right)$, and evaluate predicted values $f\left(X_i, \theta^{(r)}\right)$ and all intermediate values $\left(a^{(k)}(X_i), h^{(k)}(X_i) = \phi\left(a^{(k)}(X_i)\right)\right)_{1 \le k \le L+1}$ that are stored.

Backpropagation Method:

- **Evaluate output gradient $\nabla_{a^{(L+1)}(x)} \ell(f(x), y) = f(x) - e(y)$.**
- **For $k = L + 1$ to 1**
- **Evaluate gradient at hidden layer k**

$$\nabla_{W^{(k)}} \ell(f(x), y) = \nabla_{a^{(k)}}(x) \ell(f(x), y) h^{(k-1)}(x)'$$
$$\nabla_{b^{(k)}} \ell(f(x), y) = \nabla_{a^{(k)}(x)} \ell(f(x), y)$$

Evaluate gradient at previous layer

$$\nabla_{h(k-1)(x)^{\ell}} \ell(f(x), y) = \left(W^{(k)}\right)' \nabla_{a(k)(x)} \ell(f(x), y)$$

and

$\nabla_{a(k-1)}(x) \ell(f(x), y) = \nabla_{h(k-1)}(x) \ell(f(x), y)$
$\odot \left(\dots, \phi'\left(a^{(k-1)}(x)_j\right), \dots\right)'$

Fix specifications ε : learning rate, m : batch size, nb: number of epochs.

For $l = 1$ to nb epochs

For $l = 1$ to n/m,

Consider random batch of size m without replacement in leaming sample : $(X_i, Y_i)_{i \in B_i}$

Compute the gradients with the backpropagation algorithm

$$\tilde{\nabla}_\theta = \frac{1}{m} \sum_{i \in B_i} \nabla_\theta \ell(f(X_i, \theta), Y_i)$$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

Update the parameters

$$\boldsymbol{\theta}^{\mathrm{new}} = \boldsymbol{\theta}^{\mathrm{old}} - \boldsymbol{\varepsilon}\bar{\boldsymbol{\nabla}}_{\boldsymbol{\theta}}.$$

The gradient update is frequently terminated after traversing a defined number of time steps to scale backpropagation through time for utilize with extended sequences. Truncated backpropagation across time is the name for this method. We limit the frequency of gradient updates in addition to preventing them from backpropagating all way to commencement of sequence. Following is how abbreviated BPTT works for a particular training sequence. When k1 tokens are processed in so-called forward pass as well as hidden state has been updated k1 times truncated BPTT is launched by backpropagating gradients for k2 time steps. By analogy, the number of time steps between completing a reduced BPTT as well as length of BPTT is given by k1. These parameters will be used throughout the study. Figure 1 demonstrates how the gradients are backpropagated for three (k2) time steps per two (k1) time steps, which is a visual instructive example of truncated BPPT. Note that k1 should ideally be less than or equal to k2 in order to maintain maximum data efficiency, as otherwise some data points might be omitted during training.

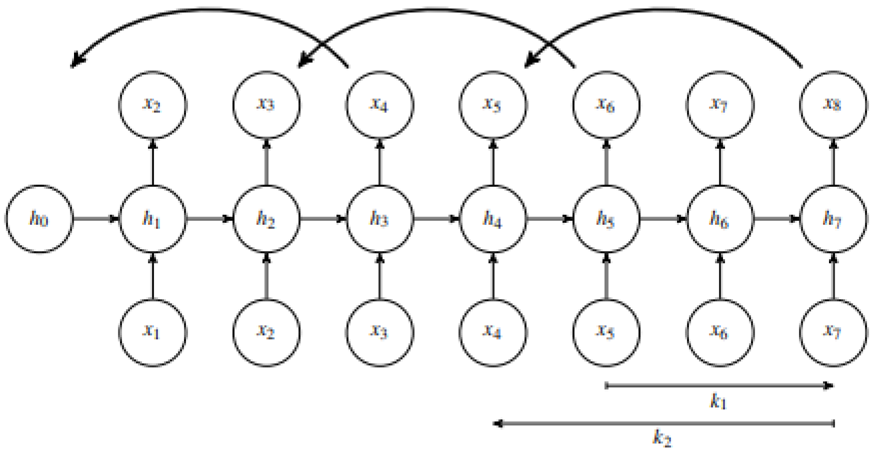


Figure. 1: For k1 = 2 and k2 = 3, an example of truncated backpropagation across time is shown. One backpropagation over time update is indicated by the thick arrows.

Genetic algorithm with regression model in classification:

Piecewise linearity is a characteristic of a quasi-linear function. It also has excellent structure and evaluation performance. Quasi-linear function with degree 2 of freedom is as eq. (31):

$$f(x) = \begin{cases} c_0 + \frac{(x-a_0)(c_1-c_0)}{a_1-a_0}, & a_0 \le x \le a_1, \\ c_1 + \frac{(x-a_1)(c_2-c_1)}{a_2-a_1}, & a_1 \le x \le a_2. \end{cases} \tag{31}$$

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

Given sample data $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, and consider $x_1 \leq x_2 \leq \cdots x_n$. We can construct QRM with freedom degree n, if we utilise quasi-linear function on [a,b] as regression function by eq. (32).

$$y = Q_L\big((x_0, y_0), (x_1, y_1), \cdots, (x_n, y_n)\big) + \varepsilon \tag{32}$$

If $[a, b] \supseteq [\min(x_i), \max(x_i)]$, the final regression curve's results will be unaffected. When range of [a,b] is included in$[\min(x_i), \max(x_i)]$, object of research is subsample rather than the entire sample. As a result, final regression results may be altered. And degree of influence is determined by proportion of [a,b] in [min(x i),max(x i)]. Select $[a, b] = [x_1, x_n]$ first. Second, denote three points on quasi-linear curve from left to right $A(x_A, y_A)$, $C(x_C, y_C)$ and $B(x_B, y_B)$. The regression function is then calculated using eq (33),

$$\mu(x) = \begin{cases} \frac{y_A - y_B}{x_A - x_B} x + \frac{x_A y_B - x_B y_A}{x_A - x_B}, & x \in [x_A, x_B], \\ \frac{y_B - y_C}{x_B - x_C} x + \frac{x_B y_C - x_C y_B}{x_B - x_C}, & x \in [x_B, x_C]. \end{cases} \tag{33}$$

eq. (34) utilizing Least Squares technique to find evaluation value $\hat{y}_A, \hat{x}_B, \hat{y}_B, \hat{y}_C$,

$$\begin{cases} \min \sum_{i=1}^{n} e_i^2 \\ \text{s.t. } x_A \leq x_B \leq x_C, \\ y_A, y_B, y_C \in (-\infty, +\infty). \end{cases} \tag{34}$$

Due to the large number of variables and the function's complexity, solving the function using the Least Square Method is extremely challenging. And as the degree of freedom increases, the number of unknown variables increases in pairs. It becomes more difficult as a result. And we're almost certain we won't be able to acquire the answers using the traditional numerical method. The solution is obtained using the genetic algorithm (GA) in the following section. An evolutionary algorithm is a GA. It's a handy tool for solving optimization problems.

The numerical method is usually based on iterative operations. The general iterative approach is prone to falling into the local minimum trap, resulting in a "endless loop." As a global optimization method, the genetic algorithm avoids this flaw. GA have good convergence, strong currency, and other advantages over standard optimization methods. It has overcome the drawbacks of present analytical techniques. It's also utilised a lot in things like combinatorial optimization, signal processing and ML. It is fundamental technology in the optimization of intelligent and complex systems. The genetic algorithm's basic operation is divided into three parts: selection, crossover, and mutation.

Fitness function. Just as method (4), we can choose $G(x) = [1 + \sum_{i=1}^{n} e_i^2]^{-1}$as fitness function.

• Selection method. Proportional choise operator is chosen.

• Crossover method. Arithmetic crossover operator is chosen bit by bit.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

• Mutation operation. To avoid infeasible solutions, we apply the mutation method as follows: Individual(y_A, x_B, y_B, y_C)as eq. (35):

$$x_B' = \begin{cases} x_B + r(b - x_B), & 0 \le r \le 1 \\ x_B - r(x_B - a), & -1 \le r < 0" \end{cases} \tag{35}$$

$$y_i' = y_i + r_i, i = A, B, C$$

r i is a random number from the normal distribution $N(0, \sigma^2)$ and r is a random number from the $[-1,1], r_i$range.

The mutation operator is employed to guarantee population variety. It improves a string by modifying it locally (parameter solution). After selection and crossover, the values of some genes on all chromosomes in population will never change. This reduces the likelihood of some new chromosomes entering this population, trapping GA-procedure in a local optima trap. To circumvent this, a mutation with an extremely low chance, say 0.01 percent, is necessary. The mutation operator in real-coded GA performs same function as the real parameter crossover operator. The advantage of polynomial mutation operator is that probability distribution does not vary with generations, evading local optima, and hence it is chosen in this work. Eq. (36) changes the ith variable xi (1, t+1) to yi (1, t+1):

$$\mathrm{y}_i^{(1,t+1)} - \mathrm{x}_i^{(1,t+1)} + \left\{\mathrm{x}_i^{(\mathrm{U})} - \mathrm{x}_i^{(\mathrm{L})}\right\}\bar{\delta}_i \tag{36}$$

– δi denotes a polynomial probability distribution. GA has an advantage over other optimization methods in that it uses a population of solutions rather of a single solution. Unlike other traditional methods that use predefined transition rules to move from one answer to another, GA guides the search using probabilistic rules and an initial random population. Another advantage of GA is that it significantly reduces overall computational time. GAs can be handled essentially independently in terms of exploitation and exploration. This gives you a lot of freedom when it comes to building a GA.

IV. PERFORMANCE ANALYSIS

The speaking material consisted of 3,388 words of shadowing samples taken from everyday conversations. The resources were divided into five categories based on the level of English competence. The TOEIC listening as well as reading exam score was utilised to determine the difficulty of items in the current study. Speakers could utilise native reference speech samples as a reference before making necessary speech sample because the materials provided native reference speech samples. This made it easier for speakers to construct complex phrases and eliminated the need for dictionaries. Japanese students (45.53 percent), native Japanese English teachers (11.24 percent), and native English teachers (11.24 percent) were among the speakers who took part in the data gathering (43.23 percent).

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

Table-1 Comparative analysis of Word Perplexity

Number of Epochs	AI_LMS	AI_WBIETS	BPNN_GA-Reg
100	77	79	81
200	79	82	85
300	81	85	89
400	83	89	91
500	85	91	92

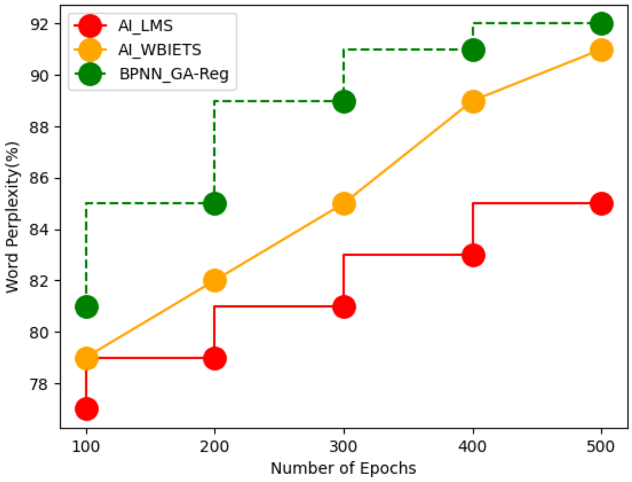
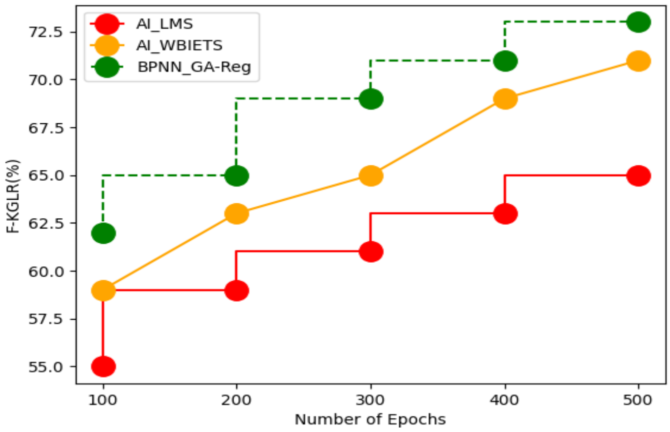


Figure-2 Comparison of Word Perplexity

Table-2 Comparative analysis of Flesch-Kincaid (F-K) Grade Level for Readability

Number of Epochs	AI_LMS	AI_WBIETS	BPNN_GA-Reg
100	55	59	62
200	59	63	65
300	61	65	69
400	63	69	71
500	65	71	73



Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

Figure-3 Comparison of Flesch-Kincaid (F-K) Grade Level for Readability

Table-3 Comparative analysis of Cosine Similarity for Semantic Coherence

Number of Epochs	AI_LMS	AI_WBIETS	BPNN_GA-Reg
100	45	49	55
200	49	51	59
300	51	55	61
400	53	59	63
500	55	61	65

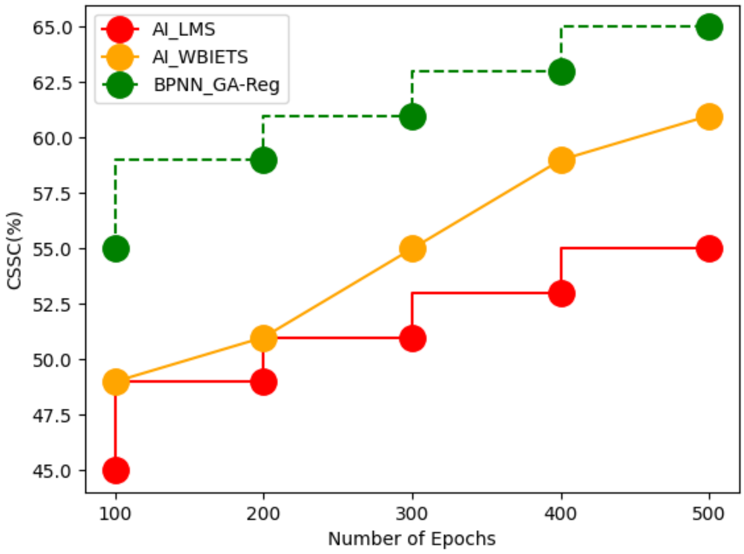


Figure-4 Comparison of Cosine Similarity for Semantic Coherence

Table- 4 Comparison of Gradient change of NN

Number of Epochs	AI_LMS	AI_WBIETS	BPNN_GA-Reg
100	51	55	59
200	52	59	63
300	55	63	65
400	59	65	69
500	61	69	72

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

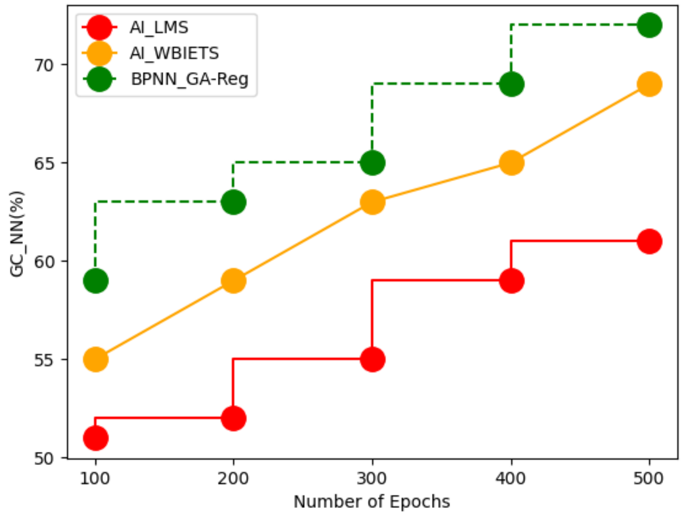


Figure-5 Comparison of Gradient change of NN

Table-5 Comparison of Validation accuracy

Number of Epochs	AI_LMS	AI_WBIETS	BPNN_GA-Reg
100	81	85	87
200	83	89	91
300	85	92	92
400	89	93	93
500	91	95	96

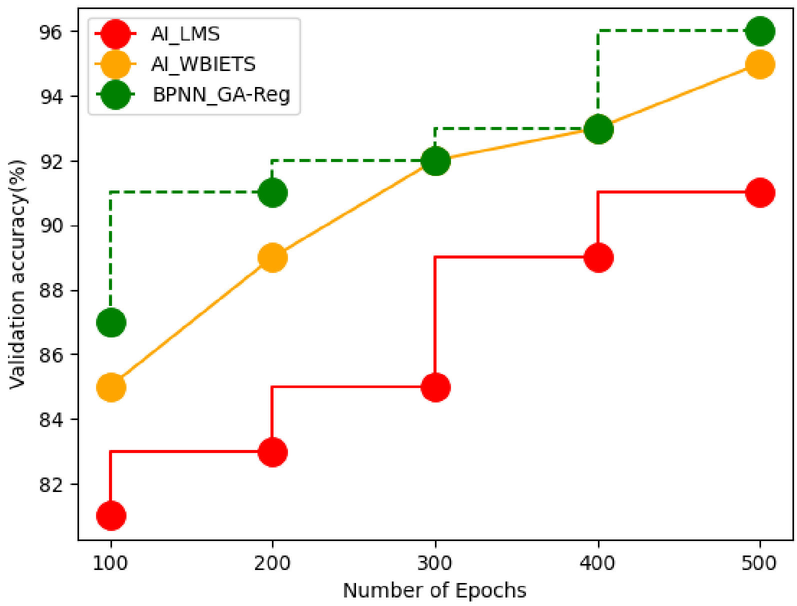


Figure-6 Comparison of Validation accuracy

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

Table- 6 Comparative analysis of Training accuracy

Number of Epochs	AI_LMS	AI_WBIETS	BPNN_GA-Reg
100	75	79	85
200	81	83	89
300	83	85	91
400	85	89	95
500	89	92	97

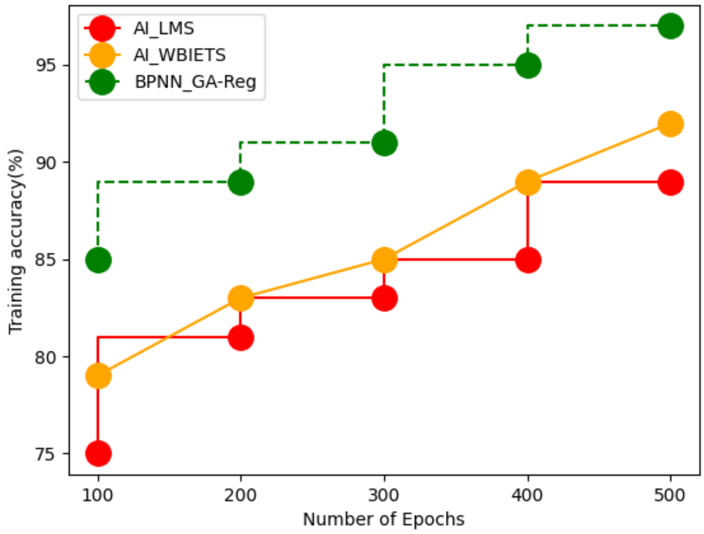


Figure-7 Comparative analysis of Training accuracy

The above table-1 shows comparative analysis of word Perplexity based on existing and proposed technique. Here the comparative analysis has been carried out based on number of epochs. The existing technique compared are AI_LMS and AI_WBIETS with proposed BPNN_GA-Reg. analysis has been carried out upto 500 epochs whereas the proposed BPNN_GA-Reg attained word perplexity of 92% as shown in figure-2, Flesch-Kincaid (F-K) Grade Level for Readability of 73% for 500 epochs as shown in figure-3 and table-2, from table- 3 Cosine Similarity for Semantic Coherence of 65% as shown in figure-4, from table-4 the Gradient change of NN obtained by propose technique is 72% as shown in figure- 5, as shown in table- 5 comparison of validation accuracy of 96% by figure-6; from table- 6 the comparative analysis of training accuracy is shown by 97% as shown in figure-7. From the above analysis the proposed technique attained optimal results in classroom management.

V. CONCLUSION

This research proposed novel technique in classroom management for English distance education based on machine learning data classification and feature extraction. Teachers cannot manage students in real time as well as students have a significant degree of independence, resulting in students' inability to focus on lectures as well as teachers' inability to grasp students' learning status in a timely manner. Based on this, this study develops an intelligent online English education classroom management method by using

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

artificial intelligence and ML methods to enhance methods. Here the aim is to analyse classroom infrastructure data and collect the historical data of classroom monitoring. The monitored data has been extracted the features using back propagation neural network (BPNN) then based on the extracted features the classification has been carried out using genetic algorithm with regression model (GA-Reg). the experimental results shows comparative experiments to validate performance of model to verify system's performance in terms of Word Perplexity as 92%, Flesch-Kincaid (F-K) Grade Level for Readability as 73%, Cosine Similarity for Semantic Coherence of 65%, gradient change of NN of 72%, validation accuracy of 96%, training accuracy of 97%.

REFERENCE

[1] X. Yuan, "Design of college English teaching information platform based on artificial intelligence technology," Journal of Physics: Conference Series, vol. 1852, no. 2, p. 022031, 2021.

[2] Z. Sun, M. Anbarasan, and D. Praveen Kumar, "Design of online intelligent English teaching platform based on artificial intelligence techniques," Computational Intelligence, vol. 37, no. 3, pp. 1166–1180, 2021.

[3] Y. Kang, "English teaching method using flipped learning in the artificial intelligence era," Robotics & AI Ethics, vol. 4, no. 2, pp. 14–21, 2019.

[4] M. Sun and Y. Li, "Eco-environment construction of English teaching using artificial intelligence under big data environment," IEEE Access, vol. 8, pp. 193955–193965, 2020.

[5] Y. Zhang and D. Wang, "Integration model of English teaching resources based on artificial intelligence," International Journal of Continuing Engineering Education and Life-Long Learning, vol. 30, no. 4, p. 398, 2020.

[6] L. Zheng, Y. Zhu, and H. Yu, "Ideological and political theory teaching model based on artificial intelligence and improved machine learning algorithms," Journal of Intelligent & Fuzzy Systems, pp. 1–10, 2021.

[7] N. Ni,"Research on the application of flipped classroom teaching mode in higher vocational English teaching," Advances in Higher Education, vol. 4, no. 9, 2020. M. Tan, "Research on English teaching system based on artificial intelligence and WBIETS wireless network system," EURASIP Journal on Wireless Communications and Networking, vol. 2020, no. 1, 2020.

[8] H. Du, "An English network teaching method supported by artificial intelligence technology and WBIETS system," Scientific Programming, vol. 2021, Article ID 8783899, 9 pages, 2021.

[9] X. Zhang and L. Chen, "College English smart classroom teaching model based on artificial intelligence technology in mobile information systems," Mobile Information Systems, vol. 2021, Article ID 5644604, 12 pages, 2021.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 2564-2583
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

[10] S. Wu and F. Wang, "Artificial intelligence-based simulation research on the flipped classroom mode of listening and speaking teaching for English majors," Mobile Information Systems, vol. 2021, Article ID 4344244, 14 pages, 2021.

[11] Y. Du, "Improvement of English teaching model based on computer network," Journal of Sensors, vol. 2021, Article ID 4756277, 10 pages, 2021.

[12] Z. Wang, L. Cai, Y. Chen, H. Li, and H. Jia, "The teaching design methods under educational psychology based on deep learning and artificial intelligence," Frontiers in Psychology, vol. 12, 2021.

[13] X. G. Xiaobin Guo, "Research on logical structure annotation in English streaming document based on deep learning," 電 腦學刊, vol. 32, no. 4, pp. 109–122, 2021.

[14] L. Zhang, "A new machine learning framework for effective evaluation of English education," International Journal of Emerging Technologies in Learning (iJET), vol. 16, no. 12, p. 142, 2021.

[15] P. Tiwari, J. Qian, and Q. Li, "Implementation of machine learning in the education sector," in Applied Machine Learning for Smart Data Analysis, pp. 155–168, CRC Press, 2019.

[16] H. Jingchao and H. Zhang, "Recognition of classroom student state features based on deep learning algorithms and machine learning," Journal of Intelligent & Fuzzy Systems, vol. 40, no. 2, pp. 2361–2372, 2021.

[17] Wang, Y. (2021). An improved machine learning and artificial intelligence algorithm for classroom management of English distance education. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 3477-3488.

[18] Zhang, E., & Yang, Y. (2021). Music dance distance teaching system based on Ologit model and machine learning. *Journal of Ambient Intelligence and Humanized Computing*, 1-17.

[19] Peng, C., Zhou, X., & Liu, S. (2022). An Introduction to Artificial Intelligence and Machine Learning for Online Education. *Mobile Networks and Applications*, 1-4.

[20] Lu, W., Vivekananda, G. N., & Shanthini, A. (2022). Supervision system of english online teaching based on machine learning. *Progress in artificial intelligence*, 1-12.

[21] Zhou, M., Chen, G. H., Ferreira, P., & Smith, M. D. (2021). Consumer Behavior in the Online Classroom: Using Video Analytics and Machine Learning to Understand the Consumption of Video Courseware. *Journal of Marketing Research*, *58*(6), 1079-1100.