
Xin Ma

(School of Foreign Languages, Zhengzhou Normal University, Zhengzhou, 450045, China

https://orcid.org/0000-0002-6004-4326, focusmaxine@163.com)

Abstract: Higher vocational education is a self-contained method of higher education that is aligned with global productivity and economic development. Its goal is to develop talented workers who contribute significantly to the economy and industry. Teaching analysis, teaching strategy, teaching practice, and assessment are all part of the course design process in high vocational education. Teaching assessment is one of the most effective methods for improving the quality of course teaching among teaching processes. This research proposes novel techniques in English teaching based on artificial intelligence for course selection based on students’ feedback. Here, the dataset has been collected based on the students’ feedback on courses for Higher Vocational Education in English teaching. This dataset has been processed to remove invalid data, missing values, and noise. The processed data features have been dimensionality reduction integrated with K-means neural network. And the extracted features have been classified with higher accuracy using recursive elimination-based convolutional neural network. Based on this feedback data classification, recommendation for courses in Higher Vocational Education in English teaching has been suggested. The experimental analysis shows various students' feedback dataset validation and training in terms of accuracy of 96%, precision of 92%, recall of 93%, RMSE of 68%, and computational time of 65%.

Keywords: Higher vocational education, artificial intelligence, course selection, students’ feedback, English teaching

Categories: H.5.1, L.3.2, I.2.7, I.2.8, I.6.4, I.7.0, M.7

DOI: 10.3897/jucs.94160

1 Introduction

Artificial intelligence (AI) and digitalization are transforming the way we work, live, communicate, learn, and play. Individuals are increasingly encountering advanced technology such as AI in their daily lives, whether they are aware of it or not, in interactions as different as applying for a loan as well as reading through social media, of which may have a dramatic impact on their lives. [Farahabadi et al., 2021]. Many jobs are being touched by human creativity as well as innovation, which is making it increasingly possible to generate value from new technology two decades into the twenty-first century. AI is increasingly being incorporated into as well as impacting industrial and agricultural processes, services, value chains, and workplace organizations, with more national policies concentrating on it. AI can enhance people's lives, but it also poses ethical and societal concerns, including employment creation and obsolescence [Reddy et al. 2020]. It is a source of social as well as political friction, and it has the potential to exacerbate disparities within and across countries.
Furthermore, a rising body of data suggests that factors like technique as well as globalization are either ‘polarising’ the workforce into high- and low-skilled occupations or ‘hollowing out’ the demand for intermediate-level skills. Intermediate-skilled workers are particularly vulnerable due to the repetitive nature of the activities they frequently perform and the fact that techniques that potentially replace them, such as AI and robotics, can save companies a lot of money [Sarveniazi, 2014]. However, its vital to note that this only applies to intermediate talents as we know them now. The permanence of intermediate occupations is revealed via analyses conducted by occupation rather than the wage percentiles, indicating “the shifting character of intermediate positions” [Zhang et al. 2017]. Equality, as well as openness of education, are important reform markers for modernizing education reform. Use of AI methods can help reach this reform goal more effectively [Perez and Tah, 2020]. In China, science and technology innovation are the primary drivers of MODE reform [Sadr et al., 2019], [Venkateswaran, 2022]. China's AI speech recognition system has now surpassed the world's best, with a 97 percent accuracy rate. With the arrival of the fourth scientific as well as method revolution, China should make full use of more mature AI methods, innovate TECHNIQUES and implement REFORM to facilitate the inefficient and stupid CONDITION [Chen and Jeong, 2007].

The contribution of this paper is as follows:

1. To propose a novel technique in English teaching based on artificial intelligence in course selection based on students' feedback
2. The dataset has been collected based on the student's feedback on courses for Higher Vocational Education in English teaching
3. To extract features using dimensionality reduction integrated with K-means Neural network.
4. To classify the extracted features using recursive elimination-based convolutional neural network.
5. To classify the feedback data, the recommendation for courses in Higher Vocational Education in English teaching has been suggested.

2 Related Works

There are various effective ways in NLP that help in educational contexts, such as the function of empirical data, corpora, and other linguistic features that are important and effective in the language learning process. Reference [Sadr et al., 2019] addressed new prospects for improving natural language processing (NLP) and its utility in developing educational tools such as reading and writing materials. The use of linguistics in the classroom can help students manage and deal with reading and writing challenges. This can be accomplished by examining syntactic and morphological parameters. Motivation in language acquisition is a powerful tool that may also improve students' educational practices and academic achievement [Venkateswaran, 2022]. DL research has seen a significant increase in activity in recent years. In [National body, 2021], popular architectural models and training techniques were used to introduce DL in NNs briefly. Because of the rise in data volume and processing power, neural networks with increasingly complicated topologies have got a lot of interest. They have been used in
a variety of sectors [Radic et al., 2020]. A considerable number of research investigations focused on practical applications, yielding many study findings. DL methods have been utilized in image analysis, text analysis, speech recognition, and other fields, providing solutions to various real-world problems [Mosavi et al., 2020]. The basic knowledge of transfer learning, the numerous types of methodologies utilized to accomplish transfer learning, and how transfer learning was being used in many subfields of medical image analysis were all evaluated by the author [Ilic et al., 2020]. The review demonstrates that recent developments in DL, particularly advances in transfer learning, have enabled the identification, classification, and quantification of specific patterns from many medical pictures [Taylor et al., 2021]. The authors [Dhawan and Batra, 2021] introduced the Char-CNNS (Character-level CNN with Shortcuts) method to provide an automated approach for determining whether the material in social media comprises cyberbullying. To assist people in comprehending how DL methods are tailored to meet the challenge of speaker recognition, the work [Xiao and Yi, 2020] presented a new technique to extract speaker features by developing CNN filters linked to the speaker. Researchers have employed DL methods to build efficient processes to calculate students’ learning state as well as behaviour [Bucea-Manea-Tonis et al. 2020] [Bernardo et al., 2021] [Du, 2021]. The success of teaching and learning depends on interaction. Researchers use DL methods to create ways to assist and encourage students’ learning enthusiasm because of their strength in natural language generation as well as processing [Dong and Tsai, 2021].

### 3 System Model

This section proposes novel techniques in English teaching based on AI in course selection based on students’ feedback. Here, the dataset has been collected based on the students’ feedback on courses for Higher Vocational Education in English teaching. This dataset has been processed to remove invalid data, missing values, and noise. The processed data features have been dimensionality reduction integrated with K-means Neural network. And the extracted features have been classified with higher accuracy using a recursive elimination-based convolutional neural network. Based on this feedback data classification, recommendation for courses in Higher Vocational Education in English teaching has been suggested. The overall proposed architecture is shown in Figure 1.
4 Dimensionality Reduction Integrated with K-means Neural Network

Let \( \mathbf{w} = [w_1, \ldots, w_n] \) be the coefficients of the first principal component, which is a linear combination of \( \mathbf{X} \). eq. (1) in matrix form:

\[
\mathbf{U}_1 = \mathbf{w}^T \mathbf{X}
\]

\[
\text{var}(\mathbf{U}_1) = \text{var}(\mathbf{w}^T \mathbf{X}) = \mathbf{w}^T \mathbf{S} \mathbf{w}
\]  

(1)

Where \( \mathbf{S} \) is \( X \)'s \( n \times n \) sample covariance matrix. Clearly, increasing the amount of \( \mathbf{w} \) can make \( \text{var}(U1) \) arbitrarily big. As a result, we select \( \mathbf{w} \) to maximize \( \mathbf{w}^T \mathbf{S} \mathbf{w} \) while requiring \( \mathbf{w} \) to be of unit length in Eq (2).

\[
\max \mathbf{w}^T \mathbf{S} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^T \mathbf{w} = 1
\]  

(2)

In eq. (3), a Lagrange multiplier \( \alpha 1 \) of 1 is introduced to solve this optimization problem:
\[ L(w, \alpha) = w^T Sw - \alpha (w^T w - 1) \]  

(3)

Standardization of raw data: In eq. (4), the raw data should have unit variance and zero mean

\[ x'_j = \frac{x_j - \bar{x}}{\sigma_j} \forall j \]  

(4)

In eq. (5), calculate the raw data's covariance matrix as follows:

\[ \Sigma = \frac{1}{m} \sum x_j (x_j)^T, \Sigma \in R^{n \times n} \]  

(5)

Data must first be standardized before the covariance matrix can be calculated. To do so, we use eq. (6) to calculate the primary vector of empirical mean:

\[ U_m = \frac{1}{n} \sum_{i=1}^{n} X_{[m,i]} \]  

(6)

On matrix lines, the empirical mean would be used. Distance matrix with the mean would then be calculated as eq (7)

\[ B = X - uh, \]  

(7)

In each of the entries, \( h \) is a vector with a size of \( 1 \times n \) and a value of 1. Equation (8) would be used to produce a covariance matrix \( \Sigma \) with \( m \times m \) dimensions:

\[ \Sigma = E[B \otimes B] = E[B \cdot B^*] = \frac{1}{n} B \cdot B^* \]  

(8)

Consider the random vector \( X' = [X_1, X_2, \ldots, X_n] \) and assume that it exhibits matrix covariance \( \Sigma \) with exceptional values \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0 \). Take a look at the linear compositions in eq. (9):

\[
\begin{align*}
Y_1 &= l'_1X = l_{11}X_1 + l_{12}X_2 + \cdots + l_{1n}X_n, \\
Y_2 &= l'_2X = l_{21}X_1 + l_{22}X_2 + \cdots + l_{2n}X_n, \\
&\vdots \\
Y_n &= l'_nX = l_{n1}X_1 + l_{n2}X_2 + \cdots + l_{nn}X_n.
\end{align*}
\]  

(9)

\[ \text{var} (Y_i) = l'_i \Sigma i, \text{cov} (Y_i, Y_k) = l'_i \Sigma k, i, k = 1, 2, \ldots, n \]

Calculate the covariance matrix’s eigenvectors and eigenvalues as provided in eq (10)

\[ u^T \Sigma = \lambda u \]

\[ U = \begin{bmatrix} 1 & u_1 & u_2 & \cdots & u_n \\ 1 & u_1 & u_2 & \cdots & u_n \\ \vdots \end{bmatrix} \]  

(10)
It is necessary to project the raw data onto a k-dimensional subspace: Pick up the covariance matrix's top k eigenvectors. These will be the data's new original foundations. The relevant vector's calculation is provided in Eq (11)

\[ x_i^{\text{new}} = \begin{bmatrix} u_1^T x_i \\ u_2^T x_i \\ \vdots \\ u_k^T x_i \end{bmatrix} \in \mathbb{R}^k \]  

For given data \( \{x_i\}_{i=1}^N \) let \( D = [d_{ij}] \) be pairwise Euclidean matrix whose entry \( ij \) and \( d \) shows Euclidean distance between high-dimensional data points \( i \)-x multidimensional scaling finds a linear mapping \( P \) such that enhances the cost function in eq. (12):

\[ \psi(Y) = \sum_{ij} (d_{ij}^2 - \|y_i - y_j\|^2) \]  

The Euclidean distance between the low-dimensional data points \( y_i \) and \( y_j \) is \( \|y_i - y_j\| \) with \( \|v\|^2 = 1 \) for all column vector \( v \) of \( P \), and \( y_j \) is constrained to be \( x_j A \). The eigen decomposition of the Gram matrix \( T = XX^T = A A^T \) is demonstrated to produce the minimum of the cost function \( a(Y) \). Actually, we may get the Gram matrix by computing eq. (13): double-centering pairwise squared Euclidean distance matrix

\[ g_x = -\frac{1}{2} \left( d_{ij}^2 - \frac{1}{n} \sum_i d_{ii}^2 - \frac{1}{n} \sum_j d_{jj}^2 + \frac{1}{n^2} \sum_{i=1}^m \sum_{j=1}^m d_{ij}^2 \right) \]  

Scatters are measured by utilizing scatter matrices. Consider \( r \) class \( C_i \) each including \( n_i \) points \( x_i^j \in \mathbb{R}^d \) and set \( X = [\hat{C}_1, \cdots, \hat{C}_r] \in \mathbb{R}^{dn} \), where \( \hat{C}_i = [x_1^i, \cdots, x_{n_i}^i] \) and \( n = \sum_{i=1}^r n_i \). Let \( \bar{X} = \frac{1}{n} \sum_{i=1}^r x_i^j \) and \( \bar{x} = \frac{1}{n} \sum_{i=1}^r x_i^j \)  

Now define 3 scatter matrices:

- Between-class scatter matrix \( S_b = \sum_{i=1}^r n_i (\bar{x} - \bar{X})(\bar{x} - \bar{X})^T \),
- Within-class scatter matrix \( S_w = \sum_{i=1}^r \sum_{j=1}^{n_i} (x_i^j - \bar{x})(x_i^j - \bar{x})^T \),
- Total scatter matrix \( S_t = \sum_{i=1}^r \sum_{j=1}^{n_i} (x_i^j - \bar{x})(x_i^j - \bar{x})^T \). LDA is a strategy for solving Eq. (14)'s optimization problem:

\[ \arg \max_{u \in \mathbb{R}^k} \frac{|u^T S_b u|}{|u^T S_w u|} \]  

Let \( X = [x_1^{n_1}, \cdots, x_n^{n_r}] \) and \( Y = [y_1^{n_1}, \cdots, y_n^{n_r}] \) be two data set of \( n \) points in \( \mathbb{R}^p \) and \( \mathbb{R}^q \), associated with them have two matrices by eq. (15):

\[ A_X = [x_1 - \bar{x}, \cdots, x_n - \bar{x}] \in \mathbb{R}^{p \times n}, A_Y = [y_1 - \bar{y}, \cdots, y_n - \bar{y}] \in \mathbb{R}^{q \times n} \]
where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=-y_i}$ are means of $x_i$ and $y_i$.

CCA is, in fact, a strategy for solving the following optimization problem in Eq. (16):

$$\arg \max_{U_X, U_Y} \frac{\mathbf{u}_X^T A_X A_Y^T \mathbf{u}_Y}{\sqrt{\mathbf{u}_X^T A_X A_X^T \mathbf{u}_X \mathbf{u}_Y^T A_Y A_Y^T \mathbf{u}_Y}}$$

which can be modified as eq. (17)

$$\arg \max_{U_X, U_Y} U_X^T A_X A_Y^T U_Y, U_Y^T A_Y A_X^T U_X = 1, U_Y^T A_Y A_Y^T U_Y = 1$$

Assuming that the solution to the previous optimization issue is the pair of projective directions $(U_X^*, U_Y^*)$, we can identify another pair of projective orders by solving eq. (18)

$$\arg \max_{U_X, U_Y} U_X^T A_X A_Y^T U_Y, U_Y^T A_Y A_X^T U_X = 0, U_Y^T A_Y A_Y^T U_Y = 1$$

We get an $m$-dimensional specification of the linear combination of these vector solutions by repeating the previous technique $m-1$ times. In reality, by solving the paired eigenvalue problem eq. (19), we can obtain this $m$-dimensional space:

$$A_X A_Y^T (A_Y A_X^T)^{-1} U_X = \lambda A_X A_Y^T U_X, A_Y A_X^T (A_X A_Y^T)^{-1} U_Y = \lambda A_Y A_X^T U_Y$$

and eigenvectors $(U_X^{(i)}, U_Y^{(i)})$, $i = 1, \ldots, m$ corresponding to $m$ largest eigenvalues are pairs of projective directions. Hence from eq. (20)

$$\{U_X^{(i)^T} A_X, i = 1, \ldots, m\} \text{ and } \{U_Y^{(i)^T} A_Y, i = 1, \ldots, m\}$$

The $i$th row and $j$th columns of a matrix $X = [x_{ij}]$ are given as $x_i$ and $x_j$. Moreover refer to Frobenius norm, 2-norm, 1-norm, and 2,1-norm of a matrix $X$ as

$$||X||_F = \sqrt{\sum_{i,j}\|x_{ij}\|^2}, ||X||_2 = \sqrt{\sum_{i,j}\|x_{ij}\|^2}, ||X||_1 = \sum_{i,j}|x_{ij}|,$$

and $X_{2,1} = \sum_i \|x_i\|_2 = \sum_{i,j} \sqrt{x_{ij}^2}$. Transpose, trace operator and inverse of a matrix $X$ are denoted by $X^T$, Tr ($X$), and $X^{-1}$. As seen in Eq. (21) the least-squares loss function is as follows:

$$\min_{W} \sum_{i=1}^{m} \|X^T w_i - y_i\|^2 = \min_{W} ||X^T W - Y||_F^2$$
PROOF. According to Step 2 in Algorithm 1, by eq. (25) modified as follows:

\[ ||W||_F^2 \]

is a diagonal matrix with \( \frac{1}{2||w_i||} \) as the kth diagonal element. In real-world applications, however, \( XXT \) is not always invertible. To achieve this, the smooth regularisation term is given a conventional objection function, such as a 2-norm.

\[
\min_w \|X^TW - Y\|^2_F + \rho \|W\|^2_F
\]  

(22)

We take the derivative of each row \( w_i (i \leq m) \) and then set it to zero, as shown in Eq. (23):

\[ XXTw_i - Xy^i + \rho_1 D_i w_i + \rho_2 D_i w_i + \rho_3 XLX^T w_i = 0 \]

(23)

where \( Di(1 \leq i \leq m) \) is a diagonal matrix with \( \frac{1}{2||w_i||} \) as the kth diagonal element and \( D \) is a diagonal matrix with \( \frac{1}{2||w_i||} \) as the kth diagonal element. Equation (24) is then modified as follows:

\[
w_i = (XX^T + \rho_1 D_i + \rho_2 D_i + \rho_3 XLX^T)^{-1} Xy^i
\]

(24)

PROOF. According to Step 2 in Algorithm 1b, eq. (25)

\[
W^{(t+1)} = \min_w \text{Tr} ( (X^TW)^T (X^TW - Y) ) + \rho_1 \sum_{i=1}^m |w_{i}^{(t+1)}|^2 D_{i}^{(t)} w_{i}^{(t+1)} + \rho_2 \text{Tr} W^T D T + \rho_3 XLX^T
\]

(25)

\[
\text{Tr} ( (X^TW^{(t)} - Y)^T (X^TW^{(t+1)} - Y) ) + \rho_1 \sum_{i=1}^m |w_{i}^{(t+1)}|^2 D_{i}^{(t)} w_{i}^{(t+1)} + \rho_2 \text{Tr} W^{(t)} D W^{(t)} + \rho_3 XLX^T
\]

\[
\Rightarrow \text{Tr} ( (X^TW^{(t)} - Y)^T (X^TW^{(t+1)} - Y) )
\]

\[
+ \rho_1 \sum_{i=1}^m \sum_{j=1}^n \frac{(w_{ij}^{(t+1)})^2}{2 \|w_{ij}^{(t+1)}\|} + \|w_{ij}^{(t+1)}\|^2
\]

\[
+ \rho_2 \sum_{i=1}^m \sum_{j=1}^n \frac{(w_{ij}^{(t+1)})^2}{2 \|w_{ij}^{(t+1)}\|} + \|w_{ij}^{(t+1)}\|^2
\]

\[
+ \rho_3 XLX^T \leq \text{Tr} ( X^T W^{(t)} - Y ) ( X^T W^{(t+1)} - Y )
\]

\[
+ \rho_1 \sum_{i=1}^m \sum_{j=1}^n \frac{(w_{ij}^{(t+1)})^2}{2 \|w_{ij}^{(t+1)}\|} + \|w_{ij}^{(t+1)}\|^2
\]
$$+ \rho_2 \sum_{k=1}^{d} \left( \| (w^{(t)})^k \|_2^2 + \frac{\| (w^{(t)})^k \|_2^2}{2\| (w^{(t)})^k \|_2} - \| (w^{(t)})^k \|_2 \right) + \rho_3 X^T X$$

$$\Rightarrow \text{Tr} \left( (X^T W^{(t+1)} - Y)^T (X^T W^{(t+1)} - Y) \right) + \rho_2 \sum_{k=1}^{d} \sum_{j=1}^{m} \| w^{(t+1)} \|_2$$

$$+ \rho_3 X^T X$$

$$\leq \text{Tr} \left( (X^T W^{(t)} - Y)^T (X^T W^{(t)} - Y) \right) + \rho_1 \sum_{l=1}^{d} \sum_{j=1}^{m} \| w^{(t)} \|_2$$

$$+ \rho_2 \sum_{k=1}^{d} \left( \| (w^{(t)})^k \|_2^2 + \rho_3 X^T X \right)$$

---

**Algorithm for Dim_Red_KNN:**

<table>
<thead>
<tr>
<th>Input X,Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output: switch task do</td>
</tr>
<tr>
<td>Case 1</td>
</tr>
<tr>
<td>Case 2</td>
</tr>
<tr>
<td>Forecast value;</td>
</tr>
<tr>
<td>Case 3</td>
</tr>
<tr>
<td>Imputation value</td>
</tr>
<tr>
<td>Endsw</td>
</tr>
<tr>
<td>Enhancing equation (6) to get optimal solution W</td>
</tr>
<tr>
<td>Normalizing X and Y</td>
</tr>
<tr>
<td>Finding optimal k value for test data based on W</td>
</tr>
<tr>
<td>Case 1</td>
</tr>
<tr>
<td>Switch task do</td>
</tr>
<tr>
<td>Finding class labels via majority rule</td>
</tr>
<tr>
<td>Case 2</td>
</tr>
<tr>
<td>Endsw</td>
</tr>
<tr>
<td>Finding prediction value via equation (9);</td>
</tr>
<tr>
<td>Case 3</td>
</tr>
<tr>
<td>Endsw</td>
</tr>
<tr>
<td>Finding imputation value via equation (9)’</td>
</tr>
<tr>
<td>Endsw</td>
</tr>
</tbody>
</table>

---

Recursive elimination-based convolutional neural network (RE_CNN) based classification:

For creating a strong joining method for sentiment analysis, the suggested model integrates both CNN as well as RNN. It consists of 4 layers: embedding, convolutional, recursive, and classification, as illustrated in figure 2.
The Euclidean distances between a point $x_i$ and a point $x_j$ in a vector $v$ of $N$ high-dimensional points $x_1, x_2, x_n$ are translated into a conditional probability $P_{j|i}$, which gives the similarity between point $x_i$ and point $x_j$. As a result, the conditional probability $P_{j|i}$ increases for data points that are close together, whereas $P_{j|i}$ decreases for data points that are far apart. The conditional probability $P_{j|i}$ can be expressed mathematically as eq. (26).

$$p_{j|i} = \frac{\exp\left(-\|x_i-x_j\|_2^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i-x_k\|_2^2 / 2\sigma_i^2\right)}$$  \hspace{1cm} (26)$$

where $\sigma_i$ is the Gaussian variance centred at the location $x_i$. Because the method is focused on modeling pairwise similarities. Similarly, the conditional probability representing the similarities of the map points $y_i$ to $y_j$ and indicates by $q_{j|i}$ is given by

$$q_{j|i} = \frac{\exp\left(-\|y_i-y_j\|_2^2 / 2\sigma_j^2\right)}{\sum_{k \neq i} \exp\left(-\|y_i-y_k\|_2^2 / 2\sigma_j^2\right)}$$

in low-dimensional equivalents $y_i$ and $y_j$ of higher dimensionality $x_i$ and $x_j$.

Because this method only involves modelling pairwise similarities, the conditional probability $q_{j|i}$ is similarly set to zero ($q_{j|i} = 0$). In t-SNE, this is done repeatedly for a given cost function $C$ using the gradient descent method, so that eq. (27)

$$C = \sum_{i} KL(P_i \parallel Q_i) = \sum_{i} \sum_{j} p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}$$

$KL(P_i \parallel Q_i)$ is Kullback-Leibler divergence function of $P_i \parallel Q_i$ [45]. Kullback-Leibler divergence $KL(P_i \parallel Q_i)$ between them is given as eq. (28)
\[ KL(P_i \parallel Q_i) = -\sum_{x \in X} P_i(x) \log \left( \frac{Q_i(x)}{P_i(x)} \right) \]

Expectation of the logarithmic difference between probabilities \( P_i \) and \( Q_i \) is determined by Equation (29) above. In \( P_i \) and \( Q_i \), it can produce any continuous, random variable \( x \) as

\[ KL(P_i \parallel Q_i) = - \sum_{x \in X} P_i(x) \log \left( \frac{P_i(x)}{Q_i(x)} \right) \]

Subject to: \( y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \)

Where \( y_i \) is corresponding to target, \( y_i = \{ \pm 1 \}, i = 1, ..., m \).

Optimization issue is solved in a dual issue by eq. (30):

\[ W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) \]

Subject to: 1) \( 0 \leq \alpha_i \leq C, i = 1, ..., m \)

2) \( \sum_{i=1}^{m} \alpha_i y_i = 0 \)

where \( \alpha_i \) are the Lagrange coefficients.

The probability densities of \( P_i \) and \( Q_i \) are \( p_{ij} \) and \( q_{ij} \). When \( P_i \) and \( Q_i \) are calculated over continuous sets \( X \) and \( P \), Kullback–Leibler divergence function can be rewritten as eq (31)

\[ KL(P_i \parallel Q_i) = \int_X \log \left( \frac{dP_i}{dQ_i} \right) dP_i \]

where \( \frac{dP_i}{dQ_i} \) in (7) is known as Radon–Nikodym derivative of \( P_i \) with respect to \( Q_i \).

Utilizing the chain rule,

\[ KL(P_i \parallel Q_i) = \int_X \log \left( \frac{dP_i}{dQ_i} \right) \frac{dP_i}{dQ_i} dQ_i \]

Subject to:

\[ 0 \leq \alpha_i \leq C, i = 1, ..., m \]

\[ \sum_{i=1}^{m} \alpha_i y_i = 0 \]
Algorithm for RE\_CNN:

\( F: \) set of ranked features = \{\emptyset\}
\( S: \) score of a criterion function for a certain
\( R: \) set of remaining features = \{1,2,3,\ldots,n\}

Number of remaining features, \( k: = n\),

Step 1: Train a SVM, evaluate score

\[ S_1 = S((x_i(R),y_i)) \]

Step 2: Evaluate \( w_i \) and rank features based on values of \( w_i \), \( S = \{f_1, f_2, \ldots, f_k\} \).

Step 3: Set \( k = k - 1 \) and \( j = 1 \).

Step 4: create a new feature set \( R_i \) by eliminating feature \( f_j \) from \( R \).

Step 5: Train a SVM on samples with remaining features and evaluate score \( S_2 = S((x_i(R_i),y_i)) \).

Step 6: If \( S_1 < S_2 \),

\[ F = F \cup f_j, \quad R = R - f_j \]

go to step 7
else if \( j = k \),

\[ F = F \cup f_j, \quad R = R - f_j \]

go to step 7
else

\[ j = j + 1 \]

go to step 4.

Step 7: Repeat steps 1 – 6 until \( R = \emptyset \)

start \( i = 1 \)

\( \sigma: \) variance of Gaussian

For every pair of points \( x_i \) and \( x_j \) in \( V \) do

If \( x_i = x_j \) then

\[ P_{in} = 0 \]

For every counterpart pair of points \( y_j \) and \( y_j \) in low-dimension do

If \( y_i = y_j \) then

Compute \( q_{R_i} = \frac{\exp(-\|y_i-y_j\|_2)}{\sum \exp(-\|y_i-y_j\|_2)} \)

End if

End for

Write /SONd file \( \leftarrow \) class, image_name, \( y_i \), 3/

return (JSON file)

If corresponding \( y_i \) and \( y_j \) are NOT in \([Q_1-k(Q_3-Q_1), Q_3+k(Q_3-Q_1)]\) then

Oulliers[] += image_name

Else

Continue

Return (Oulliers[])]

Call train_vgBl60

Call t-SNE0

Input data is passed through several levels of processing, such as normalization, feature deletion, and classification, as shown in Figure 3.
Performance analysis:
Models of Dim_Red_KNN_RE_CNN were tested on the testing dataset after training using three metrics to help researchers choose the best model. We first generated responses using the proposed technique using 2770 postings in the testing dataset. The postings in the testing set served as model "prompts," meaning the Dim_Red_KNN_RE_CNN models generated responses in response to those prompts.

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Techniques</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>RMSE</th>
<th>Computational time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student feedback classification</td>
<td>Char-CNNS</td>
<td>88</td>
<td>85</td>
<td>81</td>
<td>88</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>91</td>
<td>89</td>
<td>85</td>
<td>82</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Dim_Red_KNN_RE_CNN</td>
<td>96</td>
<td>90</td>
<td>88</td>
<td>78</td>
<td>69</td>
</tr>
</tbody>
</table>
Table 1: Evaluation analysis between proposed and existing techniques

The comparative analysis has been shown in Table 1 based on the evaluation of student feedback classification and course recommendation analysis in terms of accuracy, precision, recall, RMSE, and computational time. Here the proposed techniques compared Char-CNNS and DL with the proposed technique in higher vocational education analysis based on course recommendation and student feedback.

<table>
<thead>
<tr>
<th>Course Recommendation</th>
<th>Char-CNNS</th>
<th>DL</th>
<th>Dim_Red_KNN_RE_CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>88</td>
<td>88</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>83</td>
<td>90</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>82</td>
<td>89</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>77</td>
<td>78</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>69</td>
<td></td>
<td>65</td>
</tr>
</tbody>
</table>

(a) Accuracy
(b) Precision
(c) Recall
(d) RMSE
Figure 4: Evaluation analysis based on Student feedback classification in terms of (a) accuracy, (b) precision, (c) recall, (d) RMSE, (e) Computational time.

(a) Accuracy
(b) Precision
(c) Recall
(d) RMSE
The above Figures 4 and 5 show the proposed technique evaluation in Higher Vocational Education in English teaching. The evaluation has been carried out based on course selection, recommendation, and students' feedback in terms of accuracy, precision, recall, RMSE, and computational time. The selection recommendation for the course has been evaluated based on the students' feedback analysis. For student feedback-based evaluation, the proposed technique obtained an accuracy of 96%, precision of 90%, recall of 88%, RMSE of 78%, and computational time of 69%. From this student's feedback evaluation, course recommendation has been given and its evaluation obtained by the proposed technique in terms of accuracy as 96%, precision of 92%, recall of 93%, RMSE of 68%, and computational time of 65%. The proposed technique obtained optimal results in Higher Vocational Education based on English teaching in course selection from this evaluation.

5 Conclusion

This research proposes novel English teaching techniques based on artificial intelligence in course selection based on students' feedback. The dataset has been collected based on the students' feedback on courses for Higher Vocational Education in English teaching. This dataset has been processed to remove invalid data, missing values, and noise. The processed data features have been dimensionality reduction integrated with K-means Neural network. And the extracted features have been classified with higher accuracy using recursive elimination-based convolutional neural network. Based on this feedback data classification, recommendations for courses in Higher Vocational Education in English teaching were suggested. Experimental
analysis shows various student feedback dataset validations and trainings with 96% accuracy, 92% precision, 93% recall, 68% RMSE, and 65% computational time.

References


Ma X.: English Teaching in Artificial Intelligence-based Higher ... 915


