J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 9285--9302

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

A Linear Regression Model-Based Approach to Assessing Translation Efficiency in Business English Translation Teaching

Feng Dong[1,✉], Yue Cheng[2]

[1] Department of Foreign Languages and Business, Jiaozuo Normal College, Jiaozuo, Henan, 454000, China
[2] Youth League Committee, Kaifeng University, Kaifeng, Henan, 475000, China

ABSTRACT
As a core course of Business English majors, Business English translation plays a crucial role in the cultivation of Business English talents, and how to realize the assessment of translation efficiency in teaching has become a hot topic nowadays. This paper builds up a translation efficiency assessment index system in the teaching of business English translation around five aspects: vocabulary, syntax, context, society, and translator's factors. Random forest and Lasso regression methods were used to select 15 feature variables including sentence order and collocation between words. The multiple regression linear model was chosen to construct a model for assessing translation efficiency in business English translation teaching, and the model was estimated and tested. The least squares method was used for estimation, and all the parameters were significant (Sig<0.05) except for the variables compound sentences, sentence structure and situational intermingling. The distribution of the residuals of the model approximates to the normal distribution, which satisfies the assumption of normality and the assumption of independence, and possesses a good fit and some explanatory power.

Keywords: random forest, Lasso regression, multiple regression linear model, business English translation efficiency

1. Introduction

Nowadays, with the development of economic globalization and the further expansion of China's foreign trade, Business English has become an indispensable means of international trade and communication in China, and is increasingly giving full play to its great potential, and is also receiving more and more attention from enterprises, colleges and universities, as well as teachers and students [1-4]. Business English Translation is a compulsory course for business English based on business

✉ Corresponding author.
E-mail address: garrettstudy@126.com (F. Dong).
Received 15 January 2024; Revised 10 February 2024; Accepted 30 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-508

English, the main purpose and task of which is to systematically teach students the basic knowledge of business English translation, the basic content and characteristics of international business activities, and the basic procedures of business activities [5-8], and through a large number of practices, to deeply understand the characteristics of business English and its translation strategies, and to improve the students' ability of translation and interpretation in the field of business [9-11]. The course aims to cultivate interdisciplinary talents who are proficient in English and engaged in foreign business activities required in the field of commerce and trade, and to help English majors continue to lay a good language foundation while teaching them the basic knowledge of international business [12-15], so as to cultivate students' intercultural communication skills and international business communication skills, so that they can use English to directly engage in and research practical and theoretical work on foreign trade and economic cooperation [16-18].

Translation efficiency assessment in Business English translation teaching refers to evaluating and judging students' translation efficiency under the guidance of Business English translation teaching objectives [19-21]. Translation efficiency assessment has the functions of guidance, judgment, feedback, motivation and regulation for business English translation teaching, which can effectively improve and enhance the quality of translation teaching and promote the improvement of students' translation ability [22-24].

In this paper, we take the five factors of vocabulary, syntax, context, society, and translator as the criterion layers to build up a system of indicators for assessing translation efficiency in business English translation teaching, and there are different indicator layers under each criterion layer, which include 18 assessment indicators, such as translation of culturally specific vocabulary, sentence structure, situational integration, historical factors, and translator's bilingualism, and so on. Lasso regression and random forest are used as the methods of feature selection for translation efficiency assessment. The former reduces the dimensionality of the data and calculates the regression parameters through the residual squares of the regression model and the RSS measure of the distance between the response variable and the model prediction. The latter uses Bootstrap resampling to extract multiple samples from the original data, modeling each sample as a decision tree, combining the predictions of multiple decision trees and finally voting to arrive at the final prediction. The training of a single decision tree utilizes the same information gain criterion as the discrete attribute to select the optimal division point. After using Lasso regression and random forest to complete the screening and selection of the features of translation efficiency assessment, facing the characteristics of the various factors affecting translation efficiency, the multiple regression linear model is introduced to construct a model of translation efficiency assessment in business English translation teaching. The multiple linear regression equations are established, the significance test is carried out with the regression coefficients and the regression equations, the residuals are calculated according to the principle of least squares, and the coefficients of the multiple linear regression equations are solved. Finally, the linear regression model of this paper is estimated and tested to verify the validity of the model.

2. Indicator System for Evaluating Translation Efficiency in Business English Translation Teaching

With the deepening and strengthening of commercial exchanges between countries around the world, it has undoubtedly become an inevitable demand to transform the language or culture of another country into that of one's own. Business English plays an indispensable role in the continuous exchanges and mutual understanding between countries around the world, and the demand for complex foreign language talents is even more important today. The construction of the community of human destiny, especially in the economic field, requires the constructive participation of a large number of excellent composite foreign language talents, which is the responsibility and mission of foreign language education and teaching workers in the process of modernization. In this environment, the importance of business English translation teaching in major universities is also becoming more and more prominent, and the assessment of translation efficiency in business English translation teaching, which is the basis for improvement and development of teaching, is also particularly important. In this chapter, we will discuss and analyze the factors affecting translation efficiency and quality in business English translation teaching, and construct a system of evaluation indexes of translation efficiency in business English translation teaching, so as to provide clear evaluation indexes for the evaluation of translation efficiency.

The assessment index system of translation efficiency in business English translation teaching is shown in Table 1. It can be seen that the assessment index system of this paper mainly centers on five criterion layers: lexical factors, syntactic factors, contextual factors, social factors, and translator's factors, and covers 18 index layers such as translation of proper nouns, collocations between words, increase or decrease in vocabulary, derivation of vocabulary meanings, translation of culturally specific vocabulary, and sentence order.

Table 1. The evaluation index system of translation efficiency

Target layer	Symbol	Criterion layer	Symbol	Index layer	Symbol
Evaluation of translation efficiency in business English translation teaching	A	Vocabulary factor	B	Translation of proper nouns	B_1
				Combination of words	B_2
				Decrease of vocabulary	B_3
				Lexical extension	B_4
				Translation of culture-specific words	B_5
		Syntactic factor	C	Sentence word order	C_1
				Sentence components	C_2
				Compound sentences	C_3
				Sentence structure	C_4
		Contextual factor	D	Influence style	D_1
				Scene blending	D_2
		Social factor	E	Historical factors	E_1
				Cultural factors	E_2
				Political factors	E_3
		Translator factor	F	The translator 's bilingual level	F_1
				Translator 's comprehensive literacy level	F_2
				Translator 's time management	F_3
				Translator 's sense of professional responsibility	F_4

3. Translation Efficiency Evaluation Feature Selection Method

In the process of modeling based on actual data, the phenomenon of too many feature variables is often encountered, which causes dimensional disaster and reduces the accuracy of model evaluation. In order to alleviate the dimensionality disaster, this chapter will use the combination of Random Forest and Lasso regression to select the optimal subset of features, and based on the index system of translation efficiency assessment in business English translation teaching constructed above, select some of the factors among the many indexes in the translation efficiency assessment system as the more important features to train the model.

3.1. Feature selection method

3.1.1. LASSO regression. Due to the original data features are more, the data dimension is larger, and there are features that have less impact on translation efficiency, in order to reduce the time of model training and remove some features that have less impact on translation efficiency, thus reducing the model complexity and improving the model generalization ability, the LASSO regression can solve the data covariance problem as well as select the features that have a strong impact on the target variables [25]. In the process of predictive modeling, LASSO regression can be used first to reduce the dimensionality of the data.

The process of Lasso regression analysis method is as follows:

We use the residual sum of squares (RSS) of the regression model to measure the distance between the response variable and the model prediction, this statistic reflects the predictive ability of the model, the coefficients of Lasso are obtained by solving the minimum of the following equation, see equation (1):

$$\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij})^2+\lambda\sum_{j=1}^{p}|\beta_j|=RSS+\lambda\sum_{j=1}^{p}|\beta_j| \tag{1}$$

where n is the number of samples, p is the number of parameters, y_i is the response variable, β_0 is the intercept term, and $\lambda \geq 0$ is a conditioning parameter that will be determined separately. When the conditioning parameter λ is sufficiently large, l_1 the penalty term has the effect of forcing the estimates of some of these coefficients to 0. Lasso's coefficients $\hat{\beta}_\lambda^L$ were obtained by solving the above equation for the minimum value of

Using a statistical approach, Lasso employs l_1 penalty terms to define the l_1 paradigm of the coefficient vector β as equation (2):

$$\|\beta\|_1=\sum|\beta_j| \tag{2}$$

The most basic estimation method for the regression parameter β is the ordinary least squares estimation (OLS).The OLS is obtained by minimizing the residual sum of squares RSS and its expression can be calculated as equation (3):

$$\hat{\beta}^{OLS}=(X^TX)^{-1}X^TY \tag{3}$$

Lasso's coefficient estimates are equivalent to solving the problem of the following equation, see equation (4):

$$\min_\beta\left\{\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2\right\},\sum_{j=1}^{p}|\beta_j|\leq s \tag{4}$$

where $s \geq 0$ is a tuning parameter that controls the strength of the compression.

First we look for the coefficient where the above equation reaches its minimum, in the form of the following equation (5):

$$\hat{\beta}_j^L=\begin{cases} y_j-\dfrac{\lambda}{2}, y_j>\dfrac{\lambda}{2} \\ y_j+\dfrac{\lambda}{2}, y_j<-\dfrac{\lambda}{2} \\ 0,|y_j|\leq\dfrac{\lambda}{2} \end{cases} \tag{5}$$

Lasso's compression in this way is known as soft-thresholding, such that our coefficient vector $\lambda\sum_{j=1}^{p}|\beta_j|$ is completely compressed to zero also explains why Lasso can do variable selection.

3.1.2. Random forests. The basic constituent unit of random forest is decision tree, also known as categorical regression tree (CART) [26]. The basic idea of categorical regression tree is a bifurcated recursive partitioning method, which makes full use of bifurcated trees in the computational process, and splits the current sample set into two sub-sample sets under certain partitioning rules, so that each non-leaf node of the generated decision tree has two branches, and this process is repeated on the sub-sample set until it can not be subdivided into leaf nodes. The randomness of the random forest is reflected in the selection of samples with put-back random selection. It will result in different trees, each learning a part of the characteristics of the overall dataset, and eventually everyone votes (regression problems by taking the average) to get the final prediction. Since a single decision tree tends to be less accurate and prone to overfitting problems, it is necessary to improve prediction accuracy by aggregating multiple models. Random Forest uses the Bagging method to combine decision trees. The basic idea is to use Bootstrap resampling method to extract multiple samples from the original sample, model the decision tree for each sample, and then combine the predictions of multiple decision trees, and arrive at the final prediction result by voting [27].

In the process of random forest training, each time to establish a decision tree, through resampling to get a data training to participate in the decision tree training, which is there will be roughly one-third of the data is not utilized, not involved in the establishment of the decision tree, this part of the data is called out-of-bag data (OOB), this part of the data can be used in the inch on the performance of the decision tree to evaluate, and to calculate the prediction error rate of the model is called out-of-bag data error .

When training a single decision tree, the decision tree is usually selected for optimal attribute partitioning when dealing with discrete attributes using the following method:

1) Information gain: Assuming that discrete attribute a has V possible values $\{a^1, a^2, \cdots, a^V\}$, we use information gain to evaluate the “purity enhancement” of using attribute a as a division node with the following formula:

$$Gain(D,a) = Ent(D) - \sum_{v=1}^{V}\frac{|D^v|}{|D|}Ent(D^V) \tag{6}$$

where $Ent(D)$ is the information entropy of the sample set D. Then the optimal division attribute a_* is as in equation (7):

$$a_* = \arg\max_{a\in A} Gain(D,a) \tag{7}$$

2) Gini index: the purity of data set D is measured by the Gini value, see equation (8):

$$Gini(D) = \sum_{k=1}^{|y|}\sum_{k\neq k} p_k p_k^{'} = 1 - \sum_{k=1}^{|y|} p_k^2 \tag{8}$$

The smaller $Gini(D)$ is, the purer the dataset is.

The Gini index for attribute a is defined as equation (9):

$$Gini\_index(D,a) = \sum_{v=1}^{v}\frac{|D^v|}{|D|}Gini(D^v) \tag{9}$$

Thus the attribute that minimizes the post-division Gini index is selected as the best dividing attribute in the set of candidate attributes A as in equation (10) below:

$$a_* = \arg\min_{a\in A} Gini\_index(D,a) \tag{10}$$

When dealing with continuous attributes: given a sample set D and a continuous attribute a, assume that a occurs at n different values on D, and sort these values from smallest to largest, denoted $\{a^1, a^2, \cdots, a^n\}$. Based on the division point t, D can be divided into subsets D_i^- and D_t^+, where D_i^- contains samples that do not have a value greater than t on attribute a, and D_t^+ contains samples that have a value greater than t. Next the best division point t can be selected using the same information gain criterion as for the discrete attribute, equation (11) below:

$$Gain(D,a) = \max_{t \in T_a} Gain(D,a,t) = \max_{t \in T_a} Ent(D) - \sum_{\lambda \in \{-,+\}} \frac{|D_t^\lambda|}{|D|} Ent(D_t^\lambda) \tag{11}$$

where $Gain(D,a,t)$ is the information gain of the sample set D based on the bisection of the division point t. Thus, the division point that maximizes $Gain(D,a,t)$ can be selected.

3.2. Characterization of the assessment indicator system

3.2.1. Ranking the importance of feature variables. In this paper, we collect a large amount of text materials related to English translation teaching in colleges and universities, clean the data through automated scripts to ensure the quality of the data, and form the dataset Data1.The Data1 dataset contains many examples of business English translation, which is suitable for the model training of this paper.There are a total of 2,000 pieces of data in Data1, and 1,400 pieces of training data are extracted through stratified sampling, and the remaining 600 pieces of data are used as the test set (below). The remaining 600 data are used as the test set (below), and the division of the data set is randomly processed by python software. Random forest algorithm can output the importance degree of feature variables as well as rank the importance to get the importance metric value of each feature variable, as shown in Figure 1. From the figure, it can be seen that there is a large difference in the degree of importance of different feature variables. The importance measure values of the characteristic variables such as translator's bilingualism, translator's comprehensive literacy level, collocation between words, derivation of the meaning of words, etc. are all greater than 0.04, indicating that they have a greater impact on translation efficiency. Among these variables, translator's bilingualism has a high importance measure value of 0.421, which has the most significant effect on translation efficiency. The importance measure value of translator's professional responsibility, political factors, and translator's time management is small, almost 0, indicating that these characteristic variables have less influence on translation efficiency. According to the importance measure value, all the characteristic variables are divided into the following four categories.

1) Very important factors, including the translator's bilingualism, the translator's general literacy level, collocation between words, derivation of the meaning of words.

2) Important factors, including translation of proper nouns, sentence components, sentence order, increase or decrease in the number of words, translation of culturally specific words, sentence structure, compound sentences.

3) General factors, including affecting style, situational integration, historical factors, cultural factors.

4) Unimportant factors, including translator's professional responsibility, political factors, translator's time management.

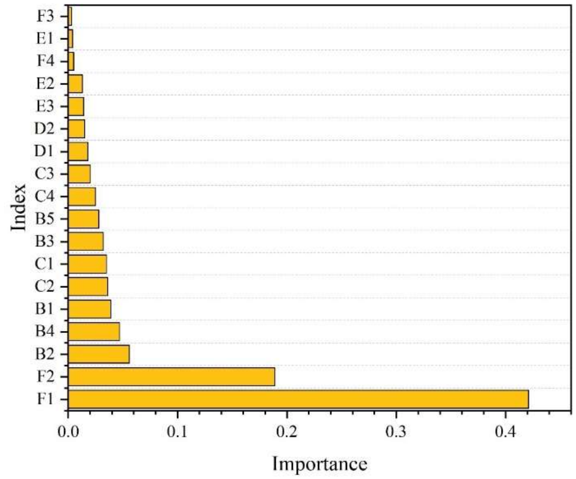


Fig. 1. The importance measure of each characteristic variable

3.2.2. Selection of the subset of features with the highest combined goodness of fit. Not all of the characteristic variables that have a relationship with the explanatory variables should be entered into the random forest regression model. Removing relatively unimportant variables from the model strengthens the level of explanation of the model. The method of screening explanatory variables in this paper is the backward screening method. After many backward screening experiments until the model has the smallest mean square error and the highest goodness of fit, the experimental results are finally obtained as shown in Table 2. According to the table, it can be seen that when the characteristic variable of the translator's time management is excluded, the random forest regression model reaches the lowest mean square error value and the highest goodness of fit of 93.5%. Therefore, the top 17 feature variables of the importance measure value are initially selected as the feature subset.

Table 2. Feature subset selection

Experimental frequency	Mean square error	Fitting excellence(%)	Remove variables	Number of remaining variables
1	2666010.1	0.93423	None	18
2	2667056.5	0.93486	Translator 's time management	17
3	2636591.2	0.93421	Political factors	16
4	2675355.3	0.93398	Translator 's sense of professional responsibility	15
5	2707761.1	0.93317	Cultural factors	14
6	2717783.4	0.93292	Historical factors	13
7	2723299.1	0.19516	Scene blending	12

3.2.3. Lasso regression for feature selection. Lasso regression has the role of feature selection, which can better deal with the problem of multicollinearity between variables. The more feature variables there are, the greater the complexity of the model, the more likely the model will contain patterns that should not be in the data, and the weaker the generalization ability; conversely, if there are fewer feature variables, the existing explanatory variables are not enough to have an effect on the explanatory variables, and the model's fitting ability is weakened. In this paper, the modeling is performed using Python, and the sklearn library in Python provides the Lasso() method for building a Lasso regression model and for feature selection. The regular term coefficient alpha is the most important parameter of the model. Its initial value is 1. The larger the value, the stronger the penalty for complex models. Therefore, the appropriate alpha should be chosen so that the model contains the right number of features in order to improve the generalization ability and fitting effect of the model training set. Plotting the variation of the regular term coefficients versus the Lasso regression coefficients is shown in Fig. 2. From the figure, it can be seen that as the regular term coefficient alpha increases, more and more feature variables gradually converge to zero, the model complexity decreases, and the goodness-of-fit decreases. The figure shows that when the value of alpha is [0,100], the coefficients of the independent variables tend to be stable, at this time, part of the coefficients of the model are zero, and the other part of the coefficients are non-zero, the model complexity decreases while the goodness-of-fit is not greatly reduced, and the model's ability to fit and the ability to generalize are guaranteed at the same time.

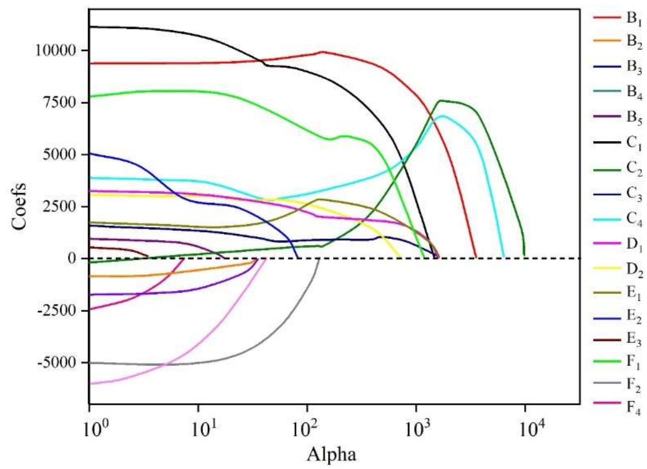


Fig. 2. The change of regularization coefficient and Lasso regression coefficient

Next, the cross validation (CV) method is used to determine the optimal alpha value, and the KFold() function in the model_selection module is utilized to implement the cross validation. The output results are shown in Table 3. After determining alpha=10.0, it can be seen from the output results that among the 17 variable characteristics, the coefficients of translator's sense of professional responsibility and political factors are 0, and these 2 characteristic variables should be discarded. The model finally retains 15 feature variables. The accuracy of the training set and validation set is not much different, indicating that the model has a good fitting ability and generalization ability at this time. The final feature variables identified in this paper that affect translation efficiency are shown in Table 3.

Table 3. The characteristic variables that affect translation efficiency

The system of characteristic variables	Characteristic variables
The system of characteristic variables affecting translation efficiency in business English translation teaching	The translator 's bilingual level, Translator 's comprehensive literacy level, Combination of words, Lexical extension, Translation of proper nouns, Sentence components, Sentence word order, Decrease of vocabulary, Translation of culture-specific words, Sentence structure, Compound sentences, Influence style, Scene blending, Historical factors, Cultural factors

4. A Model for Evaluating Translation Efficiency in Business English Translation Teaching

Linear regression analysis is a predictive modeling technique that refers to a method of statistical analysis that determines interdependent quantitative relationships between two or more variables. The method often uses the basic principles of data statistics to mathematically process a large amount of statistical data and determine the correlation between the dependent variable and certain independent variables, to establish a regression equation with good correlation, and to generalize it for use in predicting future changes in the dependent variable. According to the number of variables involved, it is divided into univariate regression and multiple regression, and the factors affecting translation efficiency in business English translation teaching are obviously diverse, so this study adopts the form of multiple linear regression [28].

4.1. Multiple linear regression models

4.1.1. Modeling Principles. Regression analysis is a statistical method that allows the quantitative relationship between two or more variables to be determined. It is commonly used today and can be subdivided into two different types of regression analysis, multiple and simple, depending on the number of variables involved. Based on the number of dependent variables, there are two different types of regression analysis: multiple and simple, and based on the relationship between the dependent variable and the independent variable, there are two types of regression analysis: linear and non-linear. If the regression analysis involves only a dependent variable and an independent variable, and the relationship between the two can be expressed by a straight line, then this regression analysis is what we call one-way linear regression analysis. If the regression analysis involves two or more independent variables and the relationship between the independent variables is linear, then it is called multiple linear regression analysis.

A univariate linear regression equation is a representation of the linear relationship between an independent variable and a dependent variable when the linear equation $y = a + bx$ has definite values of a and b. Through research and analysis, a large amount of data can be plotted into a right-angled coordinate system to form a scatter plot, although the above points are not in the same straight line, but a straight line can be plotted so that each scatter point to the line has the minimum total vertical distance, that is, the regression straight line, and the linear regression equation is an expression of the equation of the straight line.

4.1.2. Establishment of multiple linear regression equations. One-dimensional linear regression can only solve the problem where dependent variable y is related to one independent variable x. In practice, we often encounter situations where the dependent variable y is related to more than one independent variable $x_1, x_2, ..., x_n$, which presents us with the problem of multiple regression analysis. The simplest of multiple regressions is multiple linear regression. Compared to univariate linear regression analysis, multiple linear regression analysis has the same basic methodology and idea of achieving minimum Q and residual squares. However, because of the relatively large number of variables involved in multiple linear regression analysis, there are more complex problems to be faced.

Assuming a linear correlation between the p independent variables $x_1, x_2, ..., x_r$ and the random variable p, with a sample size of n, and its i th observation is $x_{i1}, x_{i2}, x_{i3} \ldots, x_{ip}; y_i (i = 1, 2, \ldots, n)$, then its n th observation can be written in the following form:

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2p} + \varepsilon_2 \\ \quad\quad \ldots\ldots \\ y_x = \beta_0 + \beta_1 x_{x1} + \beta_2 x_{x2} + \cdots + \beta_p x_{xp} + \varepsilon_x \end{cases} \tag{12}$$

where $\beta_0, \beta_1, \cdots, \beta_p$ is an unknown parameter, $x_1, x_2, ..., x_p$ is p a general variable that can be controlled and measured accurately, and $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p$ is a random error. The principle is the same as for the one-way linear regression analysis, and we can make the following assumptions: are random variables obeying the same normal distribution $N(0, \sigma)$ and uncorrelated with each other.

If we use a matrix to represent the system of equations (12), we have:

$$Y = x\beta + \varepsilon \tag{13}$$

In the formula:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_x \end{pmatrix} X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{14}$$

The key component of the multiple linear regression analysis is to obtain the valuation b of β to perform the construction of the multiple linear regression equation:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p \tag{15}$$

In turn, the multivariate linear model is described:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_y x_y \tag{16}$$

This paper analyzes the following: estimation of σ^2 and β using the least squares principle, significance testing of the regression coefficients and regression equations, control and forecasting with the help of regression equations, and the use of elimination transformations and Gaussian elimination to solve the system of linear equations in the estimation of β [29].

The construction of multiple linear regression equations is essentially a process of estimation around a multiple linear model (16) to achieve the acquisition of the estimation equation (15). Similar to the one-way linear regression analysis, the basic concept is to solve $b_0, b_1, \ldots, b_p$ in accordance with the principle of least squares in order to make the regression value $\hat{y}_i$ with all observations y_i have the minimum value of Q with the square of the residuals. Since the residual squared vs:

$$Q = \sum_{i=1}^{p} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{p} [y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_y x_{ip})]^2 \tag{17}$$

Is a nonnegative quadratic of $b_0, b_1, \cdots, b_p$, so it must have a minimum value.

In accordance with the principle of extrema, $b_0, b_1, \cdots, b_p$ should be satisfied when Q is an extreme value:

$$\frac{\partial Q}{\partial b_j} = 0 \quad (j = 0,1,2,\cdots,p) \tag{18}$$

is satisfied by equation (17):

$$\begin{cases} \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots b_y x_{ip})] = 0 \\ \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_p x_{ip})] x_{i1} = 0 \\ \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_y x_{ip})] x_{ij} = 0 \\ \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_y x_{ip})] x_{ip} = 0 \end{cases} \tag{19}$$

Eq. (19) is a system of regular equations. It is capable of transformation to the following form:

$$\begin{cases} nb_0 + (\sum_{i=1}^{n} x_{i1}) b_1 + (\sum_{i=1}^{n} x_{i2}) b_2 + \cdots + (\sum_{i=1}^{n} x_{ip}) b_p = \sum_{i=1}^{n} y_i \\ (\sum_{i=1}^{n} x_{i1}) b_0 + (\sum_{i=1}^{n} x_{i1}^2) b_1 + (\sum_{i=1}^{n} x_{i1} x_{i2}) b_2 + \cdots + (\sum_{i=1}^{n} x_{i1} x_{ip}) b_y = \sum_{i=1}^{n} x_{i1} y_i \\ \cdots \\ (\sum_{i=1}^{n} x_{ip}) b_0 + (\sum_{i=1}^{n} x_{ip} x_{i1}) b_1 + (\sum_{i=1}^{n} x_{ip} x_{i2}) b_2 + \cdots + (\sum_{i=1}^{n} x_{ip}^2) = \sum_{i=1}^{n} x_{ip} y_i \end{cases} \tag{20}$$

If the coefficient matrix of the above system of equations is represented using A it can be found that A is a symmetric matrix. Namely:

$$\begin{aligned} A &= \begin{pmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ip} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1} x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1} x_{ip} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \sum_{i=1}^{n} x_{ip} & \sum_{i=1}^{n} x_{ip} x_{i1} & \sum_{i=1}^{n} x_{ip} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ip}^2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{31} & \cdots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{1p} & x_{2p} & x_{3p} & \cdots & x_{np} \end{pmatrix} \\ &= \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ 1 & x_{31} & x_{32} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{z1} & x_{z2} & \cdots & x_{zp} \end{pmatrix} = X'X \end{aligned} \tag{21}$$

In Eq. X is the structure matrix and X' is the transpose matrix of X.
The constant term at the right end of Eq. (20) can also be expressed as the matrix D, i.e:

$$D=\begin{pmatrix}\sum_{i=1}^{n}y_i\\ \sum_{i=1}^{n}x_{ij}y_i\\ \sum_{i=1}^{n}x_{ij}y_i\\ \vdots\\ \sum_{i=1}^{n}x_{ij}y_i\end{pmatrix}=\begin{pmatrix}1 & 1 & 1 & \cdots & 1\\ x_{11} & x_{21} & x_{31} & \cdots & x_{31}\\ x_{12} & x_{22} & x_{22} & \cdots & x_{32}\\ \vdots & \vdots & \vdots & \cdots & \vdots\\ x_{12} & x_{22} & x_{32} & \cdots & x_{32}\end{pmatrix}\begin{pmatrix}y_1\\ y_2\\ y_2\\ \vdots\\ y_2\end{pmatrix}=XY \tag{22}$$

So equation (20) can be:

$$Ab=D \tag{23}$$

Or:

$$(XX)b=XY \tag{24}$$

If it is the case that A is full rank (i.e., the determinant of A is $|A|\neq 0$), then A has inverse matrix A-1, and the least squares of β can be estimated by equations (22) and (23) as:

$$b=A^{-1}D=(X'X)^{-1}X' \tag{25}$$

This is the regression coefficient of the multiple linear regression equation.

For ease of computation, instead of taking $(X'X)^{-1}$ and then b, b is usually obtained by solving the system of linear equations (20). The first equation of (20) can be reduced to:

$$b_0=\bar{y}-b_1\bar{x}_1-b_2\bar{x}_2-\cdots-b_y\bar{x}_y \tag{26}$$

Inside the style:

$$\begin{cases}\bar{x}_j=\dfrac{1}{n}\sum_{i=1}^{n}x_{ij} \quad j=1,2,\cdots,p\\ \bar{y}=\dfrac{1}{n}\sum_{i=1}^{n}y_i\end{cases} \tag{27}$$

Substituting Eq. (25) in the other equations contained in Eq. (20) yields:

$$\begin{cases}L_{11}b_1+L_{12}b_2+\cdots+L_{1y}b_y=L_{1y}\\ L_{21}b_1+L_{22}b_2+\cdots+L_{2y}b_y=L_{2y}\\ \cdots\cdots\\ L_{y1}b_1+L_{y2}b_2+\cdots+L_{yy}b_y=L_{yy}\end{cases} \tag{28}$$

Among them:

$$\begin{cases}L_{jk}=\sum_{i=1}^{n}(x_{ji}-\bar{x}_j)(x_k-\bar{x}_k)=\sum_{i=1}^{n}x_{ji}x_{ki}-\dfrac{1}{n}(\sum_{i=1}^{n}x_{ji})(\sum_{i=1}^{n}x_{ki})\\ L_{ji}=\sum_{i=1}^{n}(x_{ji}-\bar{x}_j)(y_i-\bar{y})=\sum_{i=1}^{n}x_{ji}y_i-\dfrac{1}{n}(\sum_{i=1}^{n}x_{ij})(\sum_{i=1}^{n}y_i)\end{cases} \tag{29}$$

Using a matrix to represent the system of equations (26) formulas, it can be obtained:

$$Lb=F \tag{30}$$

Among them:

$$L=\begin{pmatrix} L_{11} & L_{12} & \cdots & L_{1P} \\ L_{21} & L_{22} & \cdots & L_{2P} \\ \vdots & \vdots & \cdots & \vdots \\ L_{P1} & L_{P2} & \cdots & L_{PP} \end{pmatrix} b=\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_y \end{pmatrix} F=\begin{pmatrix} L_{1y} \\ L_{2y} \\ \vdots \\ L_{3y} \end{pmatrix} \tag{31}$$

So:

$$b = L - lF \tag{32}$$

Therefore, the coefficients of the multiple linear regression equation can be solved for by first solving L through equation (19) and then substituting L into equation (20). To solve for b, Gaussian transformation or Clem's law can be used. If b is directly substituted into equation (21), the inverse matrix of L needs to be solved, so it is relatively more complicated.

4.1.3. Modeling. Now the multiple linear regression model for the assessment of translation efficiency in business English translation teaching will be constructed, in which the explanatory variables are the various influencing factors affecting the quality of translation, and the explanatory variable is translation efficiency.

$$\begin{aligned} Translation\ efficiency = {} & b_0 + b_1 * B_1 + b_2 * B_2 + b_3 * B_3 + b_4 * B_4 + b_5 * B_5 + b_6 * C_1 + \\ & b_7 * C_2 + b_8 * C_3 + b_9 * C_4 + b_{10} * D_1 + b_{11} * D_2 + \\ & b_{12} * E_1 + b_{13} * E_2 + b_{14} * F_1 + b_{15} * F_2 \end{aligned} \tag{33}$$

4.2. Multiple linear regression model estimation and testing

The model was estimated using the least squares method. Regression analysis was selected in SPSS software, and the variable entry was selected as input. The results of parameter estimation are shown in Table 4. The table shows that most of the parameters are significant, except for the variables compound sentences, sentence structure and situational integration.

Table 4. Estimation of the model

Model	Unnormalized coefficient		T	Significance
	B	Standard error		
Constants	51020.78	835.975	60.979	0.000
B_1	33.311	28.569	2.095	0.112
B_2	136.669	25.344	8.318	0.000
B_3	-9.826	10.509	-3.662	0.000
B_4	1283.122	105.943	12.938	0.000
B_5	-116.202	14.507	-6.272	0.000
C_1	-48.802	15.382	-4.522	0.000
C_2	-1004.03	148.214	-7.829	0.000
C_3	-13.785	74.297	-0.205	0.828
C_4	15.635	75.069	0.876	0.854
D_1	727.318	44.126	13.302	0.000
D_2	60.206	61.674	0.992	0.235
E_1	6400.695	241.583	27.271	0.000
E_2	-125.598	35.812	-4.164	0.000
F_1	1030.005	49.522	19.417	0.000
F_2	-1428.04	26.435	-44.252	0.000

4.2.1. Significance test. The results of the significance test of the model are shown in Table 5, in which R^2=0.536 and adjusted R^2=0.535, indicating that the linear relationship between the independent variables and the dependent variable is strong, and the percentage of the linear form of the characteristic translation efficiency model that can explain the model of the translation efficiency assessment is 53.5%, and the model is generally well-fitted and possesses a general explanatory ability.

Table 5. The test results of significance

R	R^2	adjusted R^2	Standard estimation error
0.735	0.536	0.535	7288.322

4.2.2. Analysis of variance. The results of ANOVA are shown in Table 6, the significance test value of the model is 0.000, which proves that the equation is very significant and rejects the original hypothesis that all the coefficients are zero, and the linear relationship between the independent variables and the dependent variable of the final selected model is established.

Table 6. Results of variance analysis

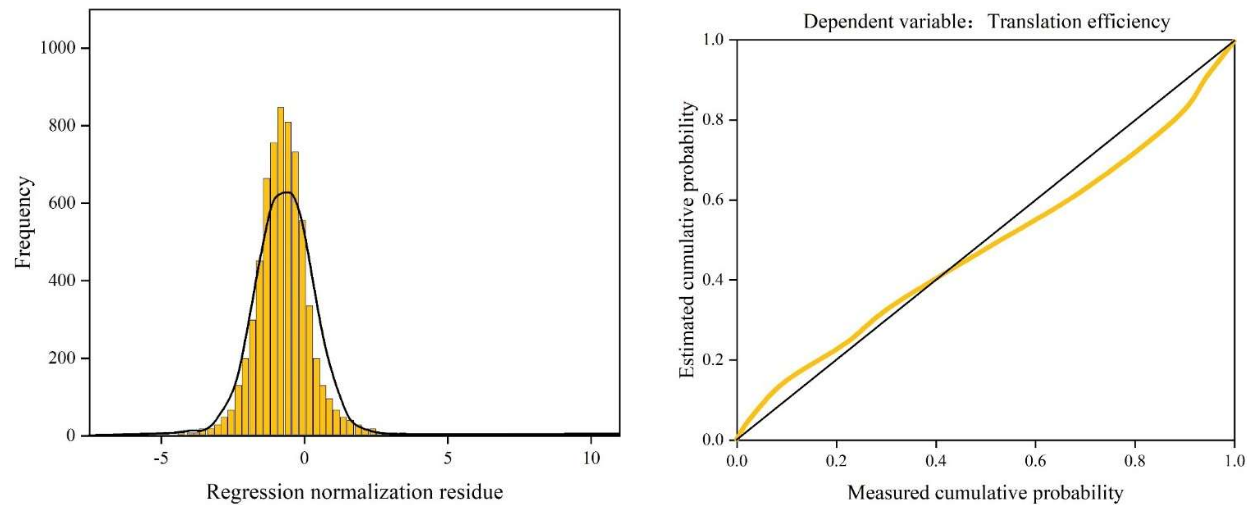
-	Sum of squares	Freedom	Mean square	F	Significance
Regression	417008945103.822	20	27800596360	520.886	0.000
Residual error	357500609873.261	6720	53278797.12	-	-
Total	774509554977.083	6745	-	-	-

4.2.3. Covariance diagnosis. The covariance diagnostic results of the multiple regression analysis of the model are shown in Table 7. The variance inflation factor (VIF) of all variables is much less than 10, with the smallest value of 1.057 and the largest of 3.13. Therefore, the hypothesis of covariance between variables can be rejected by the multiple linear regression model of this study, and it is considered that there is no significant covariance problem between the variables.

Table 7. Results of collinearity diagnosis

Variable	Tolerance	VIF
B_1	0.437	1.942
B_2	0.402	3.045
B_3	0.8338	1.405
B_4	0.749	1.35
B_5	0.305	2.832
C_1	0.851	1.057
C_2	0.886	1.141
C_3	0.373	2.181
C_4	0.477	1.972
D_1	0.534	2.164
D_2	0.345	2.564
E_1	0.88	1.1
E_2	0.472	1.836
F_1	0.832	1.237
F_2	0.24	3.13

4.2.4. Normality test for residuals. The normality test results of the residuals are shown in Fig. 3, Fig. (a) shows the histogram of the residuals, Fig. (b) shows the cumulative probability of the residuals, and the distribution of the residuals is approximate to the normal distribution. From the test results, it can be seen that the model for assessing translation efficiency in business English translation teaching in the form of a linear function satisfies the assumption of normality and the assumption of independence, and the model has a good fit and a certain degree of explanatory power, which is statistically significant.



(a)Histogram of residuals (b)Cumulative probability of residuals

Fig. 3. Normality test of residuals

5. Conclusion

In this paper, the index system for assessing translation efficiency in business English translation teaching is constructed, and random forest and Lasso regression methods are utilized to screen the features of translation efficiency assessment, and 15 feature variables such as translation of proper nouns and collocation between words are screened out. Based on the multiple linear regression model, a model for assessing translation efficiency in business English translation teaching is constructed, and the model is estimated and tested.

In the analysis of the selection of the characteristic variables of translation efficiency, the influencing factors are firstly classified into very important factors, important factors, general factors and unimportant factors based on the importance measure, and the method of screening explanatory variables by backward screening is adopted, which concludes that when the characteristic variable of the translator's time management is excluded, the mean squared error value reaches the lowest and the goodness of fit is the highest and reaches 93.5%, so the variable is firstly excluded. Plotting the variation of the regular term coefficient and Lasso regression coefficient, and determining the optimal alpha value of 10 by the cross-validation (CV) method, it is found that the coefficients of translators' sense of professional responsibility and political factors are 0 when alpha=10, which should be discarded. After eliminating the three characteristic variables of translator's time management, translator's sense of professional responsibility and political factors, the remaining 15 characteristic variables were finally determined to be retained.

With the 15 retained variables as the explanatory variables and translation efficiency as the explanatory variable, the model of translation efficiency assessment is constructed, and the model is estimated and tested. The least squares method was used for model estimation, and it was found that all the parameters were significant (Sig<0.05) except for the variables compound sentences, sentence structure and situational integration. The adjusted R^2=0.535 in the significance test indicates that there is a strong linear relationship between the independent variables and the dependent variable, and the linear form of the characteristic translation efficiency model can explain 53.5% of the translation efficiency assessment model. The significance test value of this model in ANOVA is 0.000, which indicates that the linear relationship between the 15 independent variables and the dependent variable is established in the final retention of this paper. The variance inflation factor of all variables is much less than 10 (VIF<10), and the maximum VIF value is only 3.13, which can be regarded as there is no obvious covariance problem among the variables. In the final normality test of the residuals, the distribution of residuals of this paper's model is nearly normal. Obviously, the model in this paper has a good fit and certain explanatory ability, with statistical significance and utility.

Funding

1. This paper is a phased achievement of a Research and Practice Project on Undergraduate Teaching Reform at Henan University in China, the name of the project is "Course Design and Teaching Practice Research for General Education Course on Introduction to Electronic Sports" (the Number of the project is HDXJJG2023-076).

2. This paper is a phased achievement of 2024-2025 intangible cultural heritage research project in Henan province in China, the name of the project is "Kites from Song: A practical study of intangible cultural heritage's promotion of cultural tourism, cultural and creative integration".

References

[1] Martins, H. F. (2017). Perspectives on business English as a lingua franca in business communication. *Teacher Education and Curriculum Studies*, 2(5), 61-67.

[2] Tratnik, A., Urh, M., & Jereb, E. (2019). Student satisfaction with an online and a face-to-face Business English course in a higher education context. *Innovations in education and teaching international*, 56(1), 36-45.

[3] Roshid, M. M., & Chowdhury, R. (2024). Power dynamics in business English as a lingua franca discourse. *Business and Professional Communication Quarterly*, 87(3), 432-461.

[4] Barus, I. R. G., & Simanjuntak, M. B. (2020). Evieta-Based Learning Material in English Business Class: Students' Perceptions. *Seltics Journal: Scope of English Language Teaching Literature and Linguistics*, 73-82.

[5] Wu, L., & Wu, L. (2021). Research on business English translation framework based on speech recognition and wireless communication. *Mobile Information Systems*, 2021(1), 5575541.

[6] Zhang, L. (2016). Business English translation strategies in the perspective of Skopos: In the case of advertisement translation. *Open Journal of Modern Linguistics*, 6(03), 243.

[7] Meng, Q., Liu, W., & Lv, Y. (2023). Research on the Teaching Reform of Translation Online Course for Business English Majors in Application-oriented Universities--A Case Study of S University. *In SHS Web of Conferences* (Vol. 159, p. 01005). EDP Sciences.

[8] Li, S. (2017, February). Business English Translation Study under Perspective of Functional Equivalence. *In 2017 International Conference on Humanities Science, Management and Education Technology* (HSMET 2017) (pp. 259-263). Atlantis Press.

[9] Bowker, L. (2020). Machine translation literacy instruction for international business students and business English instructors. *Journal of Business & Finance Librarianship*, 25(1-2), 25-43.

[10] Fan, H. (2017). Strategies of Business English Translation from the Perspective of Functional Equivalence. *Advances in Literary Study*, 6(1), 19-29.

[11] Al-Sowaidi, B. (2021). Use of project-based training in teaching business translation. *European Journal of Education and Pedagogy*, 2(4), 11-16.

[12] Gao, L. (2018). A study on the application of functional equivalence to business English EC translation. *Theory and Practice in Language Studies*, 8(7), 759-765.

[13] Lin, X. (2024). Importance of Basic Translation Course in Business English Teaching. *Advances in Educational Technology and Psychology*, 8(2), 45-50.

[14] Xu, Y. (2021). Research on business English translation architecture based on artificial intelligence speech recognition and edge computing. *Wireless Communications and Mobile Computing*, 2021(1), 5518868.

[15] Lyu, J. (2020). The Modularized Construction on Translation Competence for Business English Majors in China. *English Language Teaching*, 13(7), 124-129.

[16] Yapeng, H., & Lili, J. (2022). Research on the Multimodal Teaching Model of Business English Translation. *Humanities and Social Sciences*, 10(3).

[17] Huang, J. (2022). Main Features of Business English Translation and Teaching Model Optimization Based on the Logistic Model. *Computational Intelligence and Neuroscience*, 2022(1), 2896349.

[18] Zheng, J. (2017, May). Teaching business translation-A project-based approach. *In 3rd Annual International Conference on Management, Economics and Social Development* (ICMESD 17) (pp. 178-183). Atlantis Press.

[19] Deng, L. (2018). The project-based flipped learning model in Business English Translation course: Learning, teaching and assessment. *English Language Teaching*, 11(9), 118-128.

[20] Wu, X., & Zhang, S. (2018). Evaluation of Computer Assisted Translation and its Teaching Application in English Translation. *Educational sciences: theory & practice*, 18(5).

[21] Manipuspika, Y. S. (2021). The effectiveness of translation quality assessment (TQA) models in the translation of Indonesian students. *International Journal of Linguistics, Literature and Translation*, 4(5), 287-297.

[22] Liu, S., Chen, Y., Xu, K., & Lin, J. (2022). Emotional analysis of evaluation discourse in business English translation based on language big data mining of public health environment. *Frontiers in Public Health*, 10, 981182.

[23] Huang, Y. (2024). Developing a fuzzy comprehensive assessment model for English translation for college studentso. *Journal of Computational Methods in Sciences and Engineering*, 24(3), 1709-1725.

[24] Wang, A. (2021). A study on the problems and skills of business English translation. *Learning & Education*, 10(3), 186-187.

[25] Hang Nguyen,Michael Sherris,Andrés M. Villegas & Jonathan Ziveyi. (2024). Scenario selection with LASSO regression for the valuation of variable annuity portfolios. Insurance Mathematics and Economics27-43.

[26] Sami Tabib & Denis Larocque. (2024). Comparison of random forest methods for conditional average treatment effect estimation with a continuous treatment. Statistical methods in medical research(11-12),9622802241275401.

[27] Björn Koneswarakantha,Yves Barmaz,Timothé Ménard & Donato Rolo. (2020). Follow-up on the Use of Advanced Analytics for Clinical Quality Assurance: Bootstrap Resampling to Enhance Detection of Adverse Event Under-Reporting. Drug safety(1),1-3.

[28] Janet Del Rocio Gordillo Cortaza,Efigenia Gonzabay Bravo,Katherine Camacho Vargas,Miguel Castro Mattos,Yuliana Gomez Rutti,Walter Gonzalez Garcia... & Giomar Viteri Gomez. (2024). Variables Related to the Behavior of Body Mass Index Through a Multiple Linear Regression Model. Current Developments in Nutrition(S2),102453-102453.

[29] Zhen Wang,Ang Li & Tianhao Yang. (2024). Research on upper bound limit analysis of slope stability in open-pit mine based on least square principle. Heliyon(17),e37005-e37005.