Intelligent English Translation and Optimization Based on Big Data Model

Xiao Yan Wang

Jiang Xi Institute of Economic Administrators, Nanchang, Jiangxi 330088, China

Correspondence should be addressed to Xiao Yan Wang; 000418@jiae.edu.cn

Received 17 March 2022; Revised 19 April 2022; Accepted 21 April 2022; Published 20 June 2022

Academic Editor: Yuan Li

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

With the development and popularization of communication technology, intelligent analysis using big data front-end in various fields is also skillful. In the study of English literature and short sentence translation, we should realize a fast and intelligent translation approach. Based on the analysis and research of big data, a set of online learning algorithms for English translation learning algorithms is compiled, which will greatly improve the efficiency and accuracy of translation. On the premise of online processing of big data English information, this paper shortens the processing time without affecting the translation accuracy. The radial function algorithm for big data reduces the complexity of big data, improves the computational efficiency of dealing with problems, and realizes the generalization of translation performance. The effective application of big data in the optimization practice of English translation will provide an effective theoretical way for accurate translation and realize intelligent model algorithm. The experimental results are summarized as follows: (1) with the increase of the number of translated texts, the translation time will also increase and the difficulty will increase. (2) The approximation value in radial basis function is used to evaluate the translation effect, and its error in 100 texts will decrease with the increase of literature, and its approximation state is consistent with the real state. (3) Among various translation techniques, conventional literal translation is still the usual method, and its feedback effect is also the best, which is the basic method of thinking according to data. (4) The evaluation of accuracy in intelligent calculation methods is different, from 100% accuracy under heterogeneous functions to 50% accuracy under unified algorithms, which shows that optimization methods need to be updated and improved in time.

1. Introduction

In the realization of text translation, this difficult work is based on the efficient analysis of the network and the Internet. In the face of the analysis of some advanced English literature, powerful data are often needed to support the experimental conditions, so the modeling of data sets should be carried out in time. Under the experimental exploration of big data model, the analysis speed is greatly improved, and the task can be completed efficiently and accurately. In this paper, incremental learning and scientific intelligent algorithm of function expression are adopted to realize high-performance intelligent translation. This paper probes into the present situation and existing problems of the training of sci-tech translators in colleges and universities and analyzes the objectives of the current training mode of sci-tech translators in colleges and universities [1]. It reveals the potential relationship between the linguistic features of passive structures in English original language and the translators’ choices in translating passive structures and further confirms and expands the research conclusions [2]. Taking “A Brief History of Time”, a scientific and technological reading, and the translation of some lines in “The Theory of Everything” as examples, this paper analyzes and discusses that scientific and technological translation also needs artistic theory [3]. On the basis of investigating the demands of non-English major college students’ English listening and speaking based on WeChat, this paper designs a college students’ English listening and speaking learning model based on WeChat [4]. The analysis of college English extracurricular learning activities under the background of “Internet plus Homework” can give some enlightenment to the reform and exploration of college English listening and speaking curriculum [5]. This paper...
analyzes the informationization ability of college students’ English mobile learning and further studies the strategies to improve the ability of mobile learning [6]. Based on the information flow combing of thematic progression mode, the translation forms are the stability of subject and the order of subject replacement [7]. This paper puts forward that the translation strategy is to make the recipient close to the source language and demonstrates that “alienation” strategy is used in translating culture-loaded words in Zhuang language [8]. Starting from the stages of Western translation competence research, this paper explores the origin of its theoretical research and its symbolic academic achievements [9]. Starting with the analysis of competence concept, we reexamine the composition of translation competence model and use appropriate tools and methods to solve the problems encountered in translation [10]. The practical teaching mode of translation for English majors in physical education colleges is constructed, and the teaching method of “process-oriented” is adopted [11]. This paper discusses how to use linguistic translation theory to guide sports English translation practice, defines the types and functions of sports texts by studying text types, and analyzes cases through literature review [12]. Constructing a perfect translation evaluation system makes it clear that scientific translation concepts should be established in college English translation teaching [13]. This paper discusses the teaching of college English translation and its innovation, to improve students’ English translation ability and provide better services for social and economic development [14]. It shows that the first principle of translation is the purpose of translation, and the goal achieved has a great influence on the whole translation behavior [15].

2. Online Learning Method for Big Data

2.1. Online Learning Algorithm for Radial Basis Function Construction. The farther away from the center and the distance point, it is a monotone decreasing function, which is defined by Gaussian function theory. Assuming that a function can be approximated by a radial basis function, substituting it into a differential equation and forcing the error of the differential equation to take the minimum under a certain measure on a certain set of data points; thus, determining the coefficient \( a_i \) and even the point \( x_i \), this method has also obtained very satisfactory results in some practical applications.

Gaussian function [16]. Formula expression:

\[
h(x) = \exp \left( -\frac{(x-c)^2}{r^2} \right). \tag{1}\]

Its \( C \) and \( R \) represent the radial center and radius, and the distance decreases monotonously as the distance from the center is farther away.

Multiple quadratic radial basis functions [17]:

\[
h(x) = \sqrt{r^2 + (x-c)^2} \tag{2}\]

Multiquadratic radial basis function is a special function, which only increases monotonously with the distance from the center. The more obvious the meridian center is, the more obvious it is.

Radial basis function is a distance independent variable function between the measured point and the range point, which is formed by superposition linearity. Radial basis function model [18]. The expression is as follows:

\[
f(x) = \sum_{i=1}^{m} \beta_i \phi(||x-x_i||). \tag{3}\]

The interpolation conditions that need to be satisfied are:

\[
f(x_i) = F(x_i) (i = 1, \cdots, m). \tag{4}\]

Equations (3) and (4) are combined:

\[
A \times \beta = F. \tag{5}\]

Among them,

\[
A = \begin{bmatrix} 
\phi(\|x_1 - x_1\|) & \cdots & \phi(\|x_1 - x_m\|) \\
\vdots & \ddots & \vdots \\
\phi(\|x_m - x_1\|) & \cdots & \phi(\|x_m - x_1\|) 
\end{bmatrix}, \tag{6}
\]

\[
F = \begin{bmatrix} 
F(x_1) \\
F(x_m) 
\end{bmatrix}.
\]

Positive definite function, namely,

\[
\hat{\beta} = A^{-1} \times F. \tag{7}\]

Commonly used kernel functions are as follows. Multiquadratic function [19] is:

\[
\phi(r) = (r^2 + c^2)^{1/2}. \tag{8}\]

Inverse multiquadratic function [20], namely,

\[
\phi(r) = (r^2 + c^2)^{-1/2}. \tag{9}\]

Thin plate spline function [21] is expressed as:

\[
\phi(r) = \left(\frac{r}{c}\right)^2 \log \left(\frac{r}{c}\right). \tag{10}\]

The calculation formula of logarithmic path function [22] is:

\[
\phi(r) = \frac{1}{1 + e^{r/c}}. \tag{11}\]

In the above expression, \( c > 0 \).

2.2. Data-Based Unified System Representation Algorithm. Realize structured and semistructured feature quantization, and form a tensor model, which also represents a big data...
fusion model of data eigenvalues. When constructing features based on structured, semistructured, and unstructured data models, the basic fusion of models is obtained according to the corresponding data subtensor models. Under the fusion of the acquisition system of big data analysis platform, the data is converted to ensure the integrity of the structure. Implement data formulas for unstructured data, semistructured data, and structured data. Semistructured data is the data set of nodes, which has a hierarchical structure on each node, and the data model is obtained according to the tag type and object value. The flow chart is as follows in Figure 1.

The unified tensor model is obtained by coding high-dimensional data. Defined as:

\[ T \in R^{l_1 \times l_2 \times \cdots \times l_p}, T = R^{l_1 \times l_2 \times \cdots \times l_p}, T \in R^{l_1 \times l_2 \times \cdots \times l_p}. \]  

(12)

Among them, \( R \) defines the different attributes of data in each stage, the extension of tensor, and the definition of operator.

Semitensor product [23]. Defined as:

\[ X \circ Y := \sum_{i=1}^{m} X_i Y_i \in R^n. \]  

(13)

\( X \) belongs to a row of vectors of \( R \), and \( Y \) belongs to a row of vectors of \( R \).

The half vectors of \( X \) and \( Y \), namely,

\[ X \circ Y = (Y^T \circ X^T) \in R^n. \]  

(14)

Tensor expansion multiplication follows the function [24]. Expression:

\[ f : A \circ B \rightarrow C, C \in R^{l_1 \times l_2 \times \cdots \times l_s}. \]  

(15)

AB associative law. Among them,

\[ (A \circ B) \circ C = A \circ (B \circ C). \]  

(16)

Two time operators of tensor extensibility constitute heterogeneous data sets, with time orders \( T1 \) and \( T2 \), and their dimensions are:

\[ I_{t-1} \in \{i_1, i_2\}, \]  

\[ I_{t-2} \in \{i_2, i_3\}. \]  

(17)

This can be expressed as a merging time order.

2.3. Quantitative Model of Multisource Heterogeneous Data. In order to ensure the quantitative model of each kind of data, explain its data under different characteristics and stage characteristics.

Quantization of unstructured data. The formula is as follows:

\[ T_{\text{video}} = R^{l_1 \times l_2 \times l_3}, \]  

(18)

where \( l \) is the stage data, \( R \) is the quantitative expression, and the data eigenvalue is obtained according to the fourth-order tensor model.

The semistructured data formula is as follows:

\[ X \in R^{l_0 \times l_1 \times l_3 \times l_4}, \]  

(19)

where \( er \) is the row that identifies the matrix, \( ec \) is the column of the matrix, and \( en \) is the encoding of the elements.

In the managed database, the main data is stored and exiled, which is often represented by numbers or symbols, and then has matrix expressions.

2.4. Simulation Experiment for Big Data. Discretization algorithm in incremental efficient equation.

\[ x_1 = x_{2s}, \]  

(20)

\[ x_2 = -p_1 x_1 - p_2 x_2 + x_1^3 + q \cos (w t_1). \]  

Discretize it to get:

\[ x_1(t + 1) = x_1(t) + x_2(t) \Delta t, \]  

\[ x_2(t + 1) = x_2(t) + \left(-p_1 x_1(t) - p_2 x_2(t) - x_1^3(t) + q \cos (w t_1) \Delta t\right), \]  

(21)

When \( t = 0.01 \), \( p = 1.1 \), and \( q \) is the sample size, the incremental algorithm will be used to calculate the standard error to achieve the algorithm accuracy in order to avoid accidental errors. Standard error rate [25], that is,

\[ \Delta = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \frac{y(j) - y(j)^2}{y(j)^2}}. \]  

(22)

Through the results of simulation experiments, the average sample error rate of training set and test set can be obtained, and the error can be improved again to a reliable range. With the increase of sample size, the accuracy is improved. After many experiments, the learning standardization is obtained, which is suitable for big data.

3. Analysis Based on Data Structure Model

3.1. Formal Theory for Big Data Model. Big data mainly takes two forms: static big data and dynamic streaming data. Dynamic big data will be used when dealing with limited data, while static data is mainly taught by learning methods. The main form of them is to reduce the training time of the process, but there are disadvantages of single collection and reverse data collection. Therefore, it is necessary to improve the characteristics of the model, so that the data flow has the characteristics of infinity, disorder, real-time, and suddenness. The number of texts translated each time is not fixed, but changes in different phrases and long sentences based on the special meaning of the article. By adding and deleting the training set of the text, no matter the text changes dynamically based on the number of words, the time of each practice training set will not change, which improves the translation efficiency and can also translate strong and difficult sentences by 90.
First, the essence of data analysis is to verify rather than explore to get a conclusion.

Second, the verification in data analysis can only be falsification rather than confirmation in essence. Strictly speaking, all data-based verifications of the reliability of models and assumptions need to pay attention to the p value.

Third, a model, the biggest assumption is the model itself. Fourth, a model has not been falsified by data, which does not mean that the model is right, and other models are wrong. It is more likely that a model has not been falsified by data, indicating that the model is OK, but it does not rule out that there are other models that are more suitable.

Fifth, the model is not as complex as possible, but as simple as possible on the premise that it can explain the problem.

Sixth, finding valuable variables depends on domain knowledge and mastery of DGP (data generation process).

3.2. The Central Idea of Big Data Model in Computing Convergence. In some specific translation fields, the model cannot make use of all big data. Compared with the traditional model structure, it cannot represent high-dimensional data variables. Therefore, according to the connection of data eigenvalues, the extended format of high dimension and dimension between data is constructed. There are two basic steps in the formation of big data model:

(1) Data collection

Structured data under different circumstances, such as data in database, text data, audio data, and other data with complete structure, need to be collected. Classify the collected data and send them to the big data platform for the next detailed calculation. When collecting, it is necessary to ensure the authenticity of the data without changing the original format of the data.

(2) Quantitative phase of data

According to the structural data submitted in the early stage, the main data feature is to deal with the quantitative situation. Its main situation is to transform the data uniformly, ensure the structure and feature value unchanged in the process, and encode and merge the missing data in time if there are features missing, so as to reduce the small-scale quality weakening of the data.

When considering the internal relations in phased data fusion, first, it is necessary to ensure the new features of data, and second, it is necessary to ensure the stability of the internal structure of data. Tensioned data structures have the same attributes, and multisource heterogeneous represents unified variables with lower order.

3.3. Analyze Big Data Management Technology. Object-relational database management and system domain model are

Figure 1: Flow chart of big data fusion model.
all technical means accompanied by Internet of Things, cloud computing, and spatial data collection. Small data has evolved into the characteristics of big data, and the technical means also present the digital era, dealing with the concept of big data strategically and professionally. Structured data is simply a database, which realizes logical expression and real data by the structure of two-dimensional table. The off-line errors of sample data sets are compared and judged, and the average sample error rate is obtained by dividing the standard error by the total sample data one by one. Institutional data integration must gain insight into and make decisions on huge information assets, so as to realize a new data processing mode. The data set of big data cannot complete all the collection, storage, and management in a certain period of time. The concept of big data management is defined as:

(1) At present, the popularity of software technology and the efficient operation of software tools, in the realization of management, deal with the later improvement of the scale of the operating conditions

(2) After the decision information is quantified, it is modeled, and a data set expressed by grammar is formed for the next summary

(3) It is impossible to implement collection, storage, and management in an all-round way at the same time, so internal management is extremely important and the most solid foundation, to realize comprehensive control for summary work

(i) Understand and support the information needs of enterprises

(ii) Capture, store, protect, and ensure the integrity of data assets

(iii) Ensure data quality

(iv) Ensure data security

(v) Ensure efficient use of data

Turn big data into a basic resource, which is simple and practical, get data management mode from storage, management, and analysis stage, and transform it into reprint of target data planning. The ETL data transformation process is shown in the following Figure 2. Split the process of data flow into E, T, and L. The ETL framework mainly uses the intermediate processes of data extraction, data cleaning, data conversion, and data transfer to complete the data set of goal planning. In the ETL architecture, the data flow is from the source data flow to the ETL tool, which is a separate data processing engine. Generally, all the data transformation work is realized on a separate hardware server, and then, the data is loaded into the target data warehouse. The performance in the framework is easy for people to understand and make the copywriting ideas clearer. Data extraction needs to select useful data. Nowadays, the main means are automatic extraction and full data extraction, which shorten the time and improve the extraction efficiency compared with traditional manual extraction. The data cleaning needs to filter the wrong data, missing data, and repeated data to avoid affecting the quality of the data and ensure the quality of the data in the application process. The filter data for manual screening and machine quality inspection to achieve the secondary use of the standard greatly increase the data versatility. The task of the transformation process is to synchronize some data with inconsistent progress through transformation, and some particles and regular dynamics will be the transformation of coordinate system. Loading is to load the processed data into the target database and provide the data to the system for use.

3.4. Technical Analysis Based on Big Data Application. In strict compliance with the correct execution of the database, the complete data storage needs to meet the most basic criteria: consistency, isolation, atomicity, and persistence, that is, a
new structure data storage system with high scalability using NoSQL, which will divide and index data. The structure diagram of its old and new transformation is as follows in Figure 3.

Using NoSQL technology to store data will effectively improve query efficiency, reduce the proportion of memory, and clearly identify the location data that is not good at it. This can not only meet the demand of large amount of data but also ensure the comprehensive and persistent functions of the whole system. Under the advantages of high performance and concurrent function of cluster technology, data can be shared to achieve comprehensive utilization of data, and the overload of the system can also be reduced, and the fault and fault-tolerant points of the database will be upgraded and updated.

3.5. Research on the Operation of Data Model. The model data needs to be analyzed and read metadata, and the corresponding directed graph is constructed and aggregated on the reused model through ring measurement and node matching. The big data analysis work based on the rule engine will be divided into componentization and module reuse requirements. The module diagram divided by function is shown in Figure 4.

In the face of data module interpretation frame diagram, the main metadata analysis, and restoration into a framework analysis diagram, for hierarchical division of the model framework file, the parsing, detection, and pairing under metadata parsing need to be restored to acyclic graphs. The analysis of the model is mainly the directory operation of file copying, weight calculation, and file query. The planning module needs to build rules and build network security to finally realize the matching of facts, workflow generation module is responsible for checking the flow direction of nodes and description files.

4. Strategies of English Intelligent Translation and Optimization

4.1. Intelligent English Translation Method. In today’s education stage, English teaching tasks have become the key training direction, and the basic preparation stage of Chinese-English
translation is supported by big data network platform. Under the condition of intelligent and efficient completion progress in the translation process, it will be a brand-new breakthrough stage for English in actual translation. Compared with the translated texts of data sets, the accuracy of translation is particularly important, especially for some overseas documents. That is to say, intelligent optimization is adopted to improve the rapidity and accuracy of English translation. Compared with English translation in different stages under intelligent method, the usage time is shown in Figure 5.

Before translation, it is necessary to have an early understanding of the types, features, concepts, and translation requirements of all texts. In view of the increase of the number of translated texts, this paper makes a detailed time analysis of the whole translation work. As can be seen from the figure, with the increase of the number of translated texts, the working hours will also increase obviously. It takes the most time in the translation process, which is also the most difficult work, and the later processing and comparison work will be relatively simple.
4.2. Online Translation Effect Based on Radial Basis Function. The effectiveness of the model is verified by online learning method under simulation experiment, and the next algorithm is optimized according to the simulation experimental results. 100 samples of text are selected to realize the translation simulation experiment, which analyzes the accuracy of the algorithm according to the real and predicted results. The experimental results are shown in Figure 6.

The accuracy of radial function is calculated in the form of approximation value, which is displayed in the form of actual state, approximation state, and error. The experimental results show that the calculated values of the approximate state and the actual state are close to coincide in the online translation of 100 samples, which also shows that the online intelligent translation has high calculation accuracy and can accurately translate the text. When the text translation approaches 70, the calculation error will also drop to 0, which is the effect of improving approximation ability and prediction accuracy.

4.3. Analysis of Data Unified Algorithm in English Translation. In order to achieve a comprehensive and detailed translation of English practice results, the data set will be unified and standardized before the work is carried out. On the basis of unified data algorithm, we can realize accurate translation of English contents, which mainly focuses on the investigation and practice of tasks in evaluation. Its various translation accuracy pairs under big data are shown in Figure 7.

As shown in the experimental results in Figure 7, each intelligent computing method can achieve 100% accuracy in word analysis, but with the increase of difficulty, the translation accuracy of text type is only 50%. When words are expressed in multiple meanings, there will be relatively large differences, which will lead to different semantics of text meanings. Text translation based on radial function cannot meet the experimental requirements, which leads to an increase in translation error rate.

4.4. Translation Skills at the Lexical Level. The actual meaning of Chinese in English is much more profound. English sentences often present spherical structure, while Chinese sentences have strong branch relevance, and mainly analyze the feedback effect of text translation in these methods, and the test results are as follows in Figure 8.

Therefore, in the process of translation, many techniques are used to achieve practical results, such as part-of-speech conversion, simplification, information compensation, equivalent translation, order-changing translation, and ellipsis translation.
Summary and Analysis of English Translation Optimization. Starting from the context and getting rid of the surface meaning, we will deal with the translation accurately and flexibly to achieve smooth sentences and clearer regulations. In order to correctly understand the intention of the original text, we should straighten out the internal logical relationship and find reasonable sentence patterns to express it. Compared with the traditional big data processing technology, it can no longer be used in large-scale and highly concurrent network storage structure, and it is in the bottleneck of inability to upgrade and innovate in flexibility and convenience. Therefore, in order to deeply study the development and accuracy of English translation methods, it is the road of change that should be reformed now. With the introduction of the new model, people’s interest in reading is greatly enhanced, and the difficulty of translation is reduced. With the continuous evolution of the index model, the analysis dimension of the business field corresponding to the model is gradually clear, and with the accumulation of historical data, it is natural for machine learning to do big data analysis based on these indicators. Its expanded model analysis checks the consistency of statements as shown in Figures 9 and 10.

In the statistical chart of the comparative experiment between the old and new models, based on the analysis of evaluation indicators, after updating the new big data model, the translation accuracy of English has been greatly improved. As far as the accuracy rate is concerned, it has increased from 98% to 100%, and the proportion of each average index has also increased relatively, giving full play to the advantages of the new model. The recall rate of text has a downward trend, that is, it proves that users can play their good effects and make good feedback after using it.

5. Conclusion

With the development of big data, more and more people will apply it to practice, so according to the significance and value of education and assessment industry, it is extremely valued. Therefore, in the assessment and analysis of English talents training, English translation is an important transitional process. In the face of a large number of text translation work, intelligent calculation is needed to realize the efficient and accurate implementation of data processing. In this paper, based on the intelligent optimization of English translation, the experimental analysis will be carried out, and the incremental learning method and the radial function big data model will be used to deeply study the optimization means.

The development of big data has promoted the all-round development of education, economy, and system and has also achieved new value significance in English education and training in the education industry. In the era of data development, most educators have shown great interest in English learning. Based on the deepening of foreign exchanges, foreign education is extremely important, which will effectively cultivate talents and move towards deeper learning. In the training of professionals to get comprehensive development, to accept the comprehensive and coordinated development of foreign economy, education, and culture, the results of a large number of text experiments are aimed at improving the translation accuracy, that is, confirming the proposed optimization methods for each working stage. The research results of this paper are as follows: (1) the optimization of English translation methods is the main form in the era of big data, and simple analysis and processing of this set of data will be useful numerical variables for variable realization. (2) By improving the basis function and adopting incremental learning algorithm, the overall performance of designing the big data model will verify the effectiveness of its translation. (3) The simulation experiment is to verify the accuracy of the model, realize the processing of flow-like big data, and build an online English translation model in time. (4) It solves the problem of large amount of text, strengthens the applicability of the model to many kinds of word meanings, achieves effective translation of all contents, and studies in batches in time, which is the main method to effectively reduce time.

Prospects and challenges of experimental content: (1) the complexity and fast updating speed of English literature lead to the reduction of the practicability of relevant data, which is also an important challenge. (2) The extensiveness and spatio-temporal characteristics of big data model may lead to the lack of guarantee of data quality and the increase of translation errors. (3) There will be dynamic system fission in modular analysis, so the process data of sustainable development should be used when creating value. (4) There
is a difference in difficulty between simple vocabulary and high-level vocabulary translation, and it is still necessary to solve the difficult problems in mode up to now.

Data Availability

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

Conflicts of Interest

The author declared that there are no conflicts of interest regarding this work.

References


