Research Article

English Translation Teaching Algorithm in Colleges and Universities Using Data-Driven Deep Learning

Yuan Wang

1Foreign Language Teaching Department, Ningxia Medical University, Yinchuan, 750001 Ningxia, China
2Law School, Ningxia University, Yinchuan, 750001 Ningxia, China

Correspondence should be addressed to Yuan Wang; 20060034@nxmu.edu.cn

Received 9 March 2022; Revised 26 March 2022; Accepted 28 March 2022; Published 21 April 2022

Copyright © 2022 Yuan 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.

The ongoing evolution of the information age presents an excellent opportunity for the advancement of education. With the rapid development of corpus linguistics in the late 1960s, corpora became familiar to researchers as a potential resource in language teaching. Corpora are increasingly engaged in current curriculum requirements. Modern educational principles emphasize learners and require learners to participate in the construction of their own knowledge system. Therefore, with computer-assisted language teaching and learning, data-driven learning seems to be an appropriate approach, mixing interest knowledge and curriculum design that requires reassessment of teachers.

The innovations of this paper are as follows. (1) This paper combines data-driven learning with college English translation teaching and introduces the theory and learning methods of deep learning in detail. (2) In the face of two translation teaching methods, use traditional teaching and DDL teaching to design experiments, respectively. This paper compares the performance of the two methods by evaluating the experimental results and draws the conclusion that DDL is superior to traditional teaching.

2. Related Work

With the rapid development of corpus linguistics in the late 1960s, corpora became familiar to researchers as a potential resource in language teaching. Corpora are increasingly
3. Application Methods in Data-Driven Teaching

3.1. Concept Introduction

3.1.1. Corpus. A corpus is of Latin origin and refers to any text in oral or written form. Today, it is commonly used to represent large collections of text representing different domain-specific information. The text is stored in the computer and presented to readers and researchers in a readable form.

According to the above definition, corpora can reveal real language phenomena and overcome the limitations of intuition and introspection, both of which are poor guides for language learning. Natural and frequently occurring real language is represented in the corpus, at least in terms of collocation, frequency, prosody, and diction. As a result, corpus is capable of overcoming subjectivity and one-sidedness. In computer-generated corpora, however, the analysis results of correlation lines and frequencies are more reliable, scientific, and accurate. The advantage of corpus is that it provides a large database of naturally occurring texts that can be analyzed in terms of real textual structures and patterns. Researchers, teachers, and learners do not rely solely on intuition to find explanations that fit the evidence because it is impossible for people to experience all instances of language use. A corpus in computational linguistics typically consists of millions of words. The benefits of corpus include a large amount of information, quick retrieval speed, and accurate retrieval, which offers a new way to solve the problems of “what to teach” and “how to teach” in English teaching [8].

3.1.2. Data-Driven Learning. Data-driven is recognized by more and more people at the same time as the rapid development of educational big data. Data-driven generally refers to the method of discovering, analyzing, and solving problems through data mining [9] and data analysis. The author used “data-driven” to search for keywords on CNKI, and the time ended in December 2021. The attention index of “data-driven” keywords is shown in Figure 1.

For the past fifteen years, researchers have used large-scale corpora to study the languages that are actually used. These corpora significantly improve the quality of teaching reference materials. A new approach to language learning has emerged where students work with “raw” information taken directly from corpora, and it is called DDL, or data-driven learning. Generalization plays an important role in the learning process of DDL. Students need to actively participate in learning activities and outline language rules. A generalization curriculum will be completely foreign to many Chinese students trained in an educational system that emphasizes memory rather than generalization or theory generation.

DDL goes beyond traditional foreign language instruction by combining corpus linguistics research with modern information technology and applying it to second language instruction. Students actively use the material in the corpus to find meaning and language rules on their own, according to the data-driven learning theory, which helps to improve
comprehensive skills. Learners must seek, identify, and infer language rules from the context in order to discover the grammatical rules of the material being studied. When students are encouraged to follow this model, they examine authentic language material and come to their own conclusions. Students will be able to effectively acquire language in this manner. Students learn rules and definitions in a traditional classroom by deducing them from the teacher’s deductive approach and reference books. As a result, this inductive method complements the deductive method currently in use.

DDL involves setting up situations in which students can answer questions about language on their own by studying corpus data, and setting up language learning situations may vary by purpose. Teachers and students can use the original index tables and observe them, not necessarily knowing what they will find, but exploring rules, patterns, and meanings. Alternatively, teachers can carefully select and edit index lines and possibly create material to reveal linguistic features.

3.1.3. Advantages and Disadvantages of Data-Driven Learning. Data-driven learning enables learners to use computer network technology and related software tools, such as search engines, to improve students’ cognitive processes in modern information teaching. Of course, this is also very important for their future social development. Therefore, we are changing the problems that exist in college English translation classes. We change the current teaching status and adopt advanced teaching methods to meet the high-performance requirements of single translation, so that students can master language learning skills, improve autonomous learning ability, and cultivate system management ability and individual and autonomous learning.

Corpus are considered beneficial for language teaching, mainly because they provide learners and teachers with a variety of real target language inputs and information about the frequency of use of certain language items and the most common word pairs or phrases. The advantages and disadvantages of data-driven learning are shown in Table 1.

3.2. Deep Learning and Shallow Learning. The study of learning based on deep neural network [10–12] is called deep learning [13, 14]. Figure 2 plots two different neural networks.

Deep learning is clearly defined as the goal of cultivating superior thinking, where learners critically construct the basics of the subject based on intrinsic motivation to learn positive emotions, attitudes.

Shallow learning is a type of learning that is simple and mechanical. Trainees passively acquire new knowledge through repeated memorization based on external learning incentives, and their behaviours are passive treatment and lack of knowledge interaction. In order to pass the test and for other purposes, the trainee studies the content of the test and memorizes it as personal and irrelevant facts. Long-term knowledge retention and flexible application are difficult to achieve. This study compared deep learning and shallow learning in terms of goal level, learning motivation, learning goals, learning behaviour, learning style, knowledge system, reflective state, transfer ability, terms of focus, cognitive results, learning environment, evaluation methods, emotional attitudes, and so on by sorting out the research status at home and abroad, as shown in Table 2.

It should be noted that emphasizing deep learning does not deny shallow learning, and learning is a continuous process from shallow learning to deep learning [15]. As shown in Figure 3, shallow learning creates a platform for deep learning. The richer the background knowledge of students, the more connections between knowledge can be constructed, which can only be achieved after multiple thinking steps.
3.3. Learning Methods

3.3.1. Logistic Regression. The simplest regression is linear regression, such as \( y = f(x) \), which represents the relationship between the independent variable \( x \) and the dependent variable \( y \). However, the robustness of linear regression is poor, mainly because the sensitivity of linear regression is consistent in the whole real number domain, and the classification range needs to be in \([0, 1]\). Based on linear regression, logistic regression applies a logistic function, but because of this logistic function, logistic regression has a favorable place in the field of machine learning [16].

The assumptions of the logistic regression model are

\[
h_{\psi}(t) = g(\psi^T t), \tag{1}\]

where \( t \) represents the feature vector and \( g \) represents the logistic function. A commonly used logic function (and also the function mainly used in this paper) is the sigmoid function, and its expression is shown in formula (2):

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

The graph of this function is in Figure 4:

| Table 1: Advantages and disadvantages of data-driven learning. |
|---|---|---|
| DDL | Advantages | Shortcoming |
| 1 | Rich learning resources | Teachers’ knowledge of corpus is not enough |
| 2 | Freedom to study | Not always suitable for learning purposes |
| 3 | Learning is more motivated and fun | May be limited by real language |

| Table 2: Comparison of deep learning and shallow learning. |
|---|---|---|
| Deep learning | Shallow learning |
| Motivation to learn | Intrinsic motivation | Extrinsic motivation |
| Learning method | Active learning that focuses on meaning construction of concepts and principles | Passive machine learning biased towards symbolic memory |
| Learning method | Active learning that focuses on meaning construction of concepts and principles | Passive machine learning biased towards symbolic memory |
| Thinking level | Application, analysis, evaluation, creation, etc. | Thinking memory, understanding, and other low-order thinking |
| Evaluation method | Process evaluation and qualitative evaluation | Summative evaluation and quantitative evaluation |

(a) Simple neural network with only one hidden layer

(b) Deep neural networks with more than two hidden layers

Figure 2: Neural network.
Therefore, the assumptions of the logistic regression model are as shown in formula (3):

\[ h_\varphi(t) = \frac{1}{1 + e^{-\varphi^T x}}. \] (3)

For the LR model, the classification is based on the following: for a given input variable \( t \), the probability of the output variable being 1 is calculated according to the selected parameters, namely,

\[ h_\varphi(t) = P(y = 1|t; \varphi). \] (4)

The ReLu (Rectified Linear units) function is a popular activation function that has gradually replaced the sigmoid function [17]. The function graph of ReLu is shown in Figure 5, it does not tend to saturate with the gradual increase of the input \( x \).

3.3.2. Support Vector Machine. Support Vector Model (SVM) is a learning model developed by many researchers based on statistical machine learning theory. In recent years, extensive research on vector support machines has made support vector machines one of the most important advances in the field of text classification [18]. Support vector machines (SVMs) are developed from optimal grading surfaces with linear separability. The basic idea is to find a hyperplane ordering that satisfies the requirements and keep the points in the training set as far away from the ordering surface as possible, and the distance between them is called the classification interval. The optimal classification surface requires that the classification line not only correctly separates the two classes (with a training error rate of 0) but also maximizes the class separation between the two classes, as shown in Figure 6.

Assuming that the training data \( \{x_a, y_a\} (a = 1, \cdots, j, x_a \in \mathbb{R}^d; y_a \in \{-1, 1\} \) can be separated from an error-free hyperplane \( (\lambda \cdot x) - b = 0 \). The classification hyperplane that is farthest from the sample points of the two classes should obtain the best generalization ability. The optimal hyperplane will be determined by the few sample points (called support vectors) closest to it. The classification hyperplane with sample interval \( \theta \) is

\[ (\lambda \cdot x) - b = 0, \quad \|\lambda\| = 1, \]

\[ y = 1, \text{ if } \lambda \cdot x - b \geq \theta, \]

\[ y = -1, \text{ if } \lambda \cdot x - b \leq -\theta. \] (5)

The classification superlevel for SVM optimization problems should be normalized as follows, at least \( \theta = 1 \), \( \lambda \) and \( b \) can be scaled. The hyperplane is represented as

\[ (\lambda \cdot x) - b \geq 1, \text{ if } y = 1, \]

\[ (\lambda \cdot x) - b \geq -1, \text{ if } y = -1. \] (6)

The distance to the support vector on the hyperplane is \( 1/\|\lambda\| \); therefore, the optimization problem is expressed as

\[ \min \frac{1}{2} \|\lambda\|^2 \]

\[ s.t. y_a (x_a \cdot \lambda + b) - 1 \geq 0 (a = 1, 2, \cdots, j). \] (7)

According to the square programming method in the optimization theory, the problem is transformed into a double
Wolfe problem to solve, constructing the Lagrange function:

\[
L(\lambda, c, b) = \frac{1}{2}\|\lambda\|^2 - \sum_{a=1}^{j} c_a y_a (x_a \cdot \lambda + b)
+ \sum_{a=1}^{j} c_a (c_a \geq 0; a = 1, 2, \cdots, j).
\] (8)

In the formula, \(c_a\) is the Lagrange multiplier. According to the optimization principle,

\[
\frac{\partial}{\partial \lambda} L(\lambda, c, b) = 0,
\]

\[
\frac{\partial}{\partial b} L(\lambda, c, b) = 0.
\] (9)

That is

\[
\lambda = \sum_{a=1}^{j} c_a y_a x_a,
\]

\[
\sum_{a=1}^{j} c_a y_a = 0.
\] (10)

The Wolfe dual problem of the original optimization problem is obtained by operation:

\[
\max W(c) = \sum_{a=1}^{j} c_a - \frac{1}{2} \sum_{a=1}^{j} c_a c_a y_a y_a x_a x_a^T
\]

\[
s.t. \sum_{a=1}^{j} c_a y_a = 0 (c_a \geq 0; a = 1, 2, \cdots, j).
\] (11)

Its solution is the optimal solution to the initial optimization problem. An optimization algorithm can be used to solve it; the parameter \(b\) can be calculated according to the Kansh-Kuhn-Tucker condition:

\[
b = y_a - \lambda^T x_a \quad (c_a(0, D)).
\] (12)

So the optimal hyperplane is

\[
f(x) = \text{sgn} \left\{(\lambda \cdot x) + b\right\} = \text{sgn} \left\{\sum_{a=1}^{j} c_a y_a (x_a \cdot x) + b\right\}.
\] (13)

For linear inseparable classification problems, the \(x\) input can be assigned to a high-dimensional feature space \(\phi(x)\) by a nonlinear function, and linear classification is performed in this space, namely,

\[
f(x) = \text{sgn} \left\{\sum_{a=1}^{j} c_a y_a K(x_a \cdot x) + b\right\}.
\] (14)

4. Experiment and Analysis of Translation Instructional Design Based on Data-Driven Learning

4.1. Questionnaire Survey Results. This research conducted a questionnaire survey on 284 non-English majors, including 168 boys and 116 girls, who came from different majors, including civil engineering, computer, and fashion design. The purpose is to make the data universal and authentic.

According to Figure 7(a), the question “Are you good at English translation?” is raised in the questionnaire, in order to investigate the English ability of the students. According to the data, among them, 28 people chose to be good at English translation, accounting for 9.86%, and 56 people chose to be relatively good at English translation, accounting for 19.72%. There are 90 people who choose to be average, accounting for 31.69%, and 110 people who choose not to be good at it, accounting for 38.73%. From this, it can be seen that among non-English majors, fewer students are able to translate English smoothly, and students generally have little interest in English learning.

According to Figure 7(b), it can be seen that in response to the question “What aspects do you think exist in the current college English translation class (multiple choices)?”, 84 people chose the class as boring, accounting for 29.58%. There are 206 people chose not to feel that students are the center of the classroom, accounting for 72.54%, and 42 people chose the content is difficult, accounting for 14.79%. There are 142 people who choose the lecture mode that lacks novelty, accounting for 50.00%, and 184 people who choose to lack basic skills in English and Chinese, accounting for 64.79%.

It can be seen from Figure 7(c) that 156 people chose to have insufficient basic skills in English and Chinese, accounting for 54.93% of the question “What aspects do you have in English translation (multiple choices)” in the questionnaire. There are 192 people chose not to understand some typical differences between English and Chinese, accounting for 54.93%, and 178 people chose to lack basic translation strategies, accounting for 62.68%.
Figure 7: Analysis of questionnaire data about college English translation class.
It can be seen that the college English translation classroom needs to be transformed, and the classroom based on data-driven learning, the construction of a parallel corpus for translation teaching, which can better serve students, better promote students’ learning, and provide learners with authentic English.

4.2. Instructional Design. Most of the traditional translation teaching adopts the teacher-centered evaluation method of translation skills. Under the guidance of teachers, students learn, and students master passive translation knowledge, but cannot actively cultivate translation skills. In addition, they have their own professional courses, and it is impossible to spend a lot of time on corpus screening like English majors. According to the survey, DDL has been divided into three steps: asking questions, solving questions, and summarizing [19, 20]. These are carried out under the leadership of teachers, in order to be more in line with modern learning concepts, and for students to truly master knowledge. We have made relative adjustments and changes on the basis of the original three steps, trying to create a student-centered, bottom-up teaching mode of discovery and translation. This process is summarized as (teacher) corpus presentation-(student) observation and discussion-(student) reporting results-(teacher) guiding comment-(student) practice consolidation-(student/teacher) comparative evaluation. The design process is specifically shown in Figure 8.

This instructional design transforms traditional teaching into a progressive teaching process. It uses an introduction to real expectations and predefined problems, followed by collaborative learning, observation, discussion, analysis, and student reporting. Then, under the guidance and in-depth explanation of the teacher, students will effectively combine translation theory and translation practice at the same time. By analyzing the gap between the actual translation text and the core text, students and teachers finally understand translation knowledge and students master translation learning skills independently. In the whole process of teaching, it is not only necessary to ensure the continuous input of language but also to allow students to exert greater language ability [21].

4.3. Teaching Applications. According to the courses in the first semester of the sophomore year of the college English final exam, students from two classes of the second year of the same level of non-English majors in a university were selected as the subjects of this experiment. There were 50 students in each class, and a total of 100 students participated in the experiment, and they came from various non-English majors. The 50 people who participated in the posttest and the experimental group participated in the questionnaire survey. The tests and questionnaires of these students were collected, and the questionnaires were all valid questionnaires. Firstly, the data obtained from the posttest is described and analyzed by SPSS, and then, the results of the questionnaire are summarized and discussed. The British National Corpus (BNC) was selected for this experiment. This is a collection of 100 million words of written and spoken language samples from a wide range of sources.

The entire class was treated as a group of participants to reduce between-group effects. Two groups of participants were exposed to two different teaching methods in the experiment. As the control group, class A received traditional translation instruction, while class B received college English translation instruction based on data-driven learning as the experimental group. The experiment was conducted with the same teacher in both groups to reduce the possibility of the experimental results being influenced by differences between the two groups. The target words were chosen based on the results of the pretest, and the students were given 120 words to write down the definition and part of speech of each word. Based on the pretest results, 30 target words were chosen based on the criteria that none of the students responded correctly to them.

The corpus-based materials were distributed three days before the experiment to the experimental class, and they were allowed to discuss but not to refer to books. On the lab day, students will be asked to share their findings with their classmates, and the teacher will offer suggestions or opinions based on their findings. In the control group, the teacher presented the words to the students and explained their meanings, parts of speech, and associations with other words using the traditional method. The participants were then asked to complete sentence-building exercises. In addition, related learning introduces strategies to assist them in learning. Both methods were used under the same conditions throughout the experiments, removing the possibility of external inferences.

Immediately after the experiment, a posttest was performed to test the different effects. In addition, all students in the experimental class were asked to complete a questionnaire. In this experiment, the independent variable was
different vocabulary teaching methods, and the dependent variable was the fractional posttest. Therefore, the results of the test are used to indicate the difference between the two learning methods by means of a paired sample \( t \)-test. For accuracy and efficiency, the posttest results were input into the computer and analyzed with SPSS. The data analysis results are shown in Table 3.

5. Discussion

First of all, through the study of relevant knowledge points of literature works, this paper initially masters the relevant basic knowledge and analyzes how to conduct research on college English translation teaching based on data-driven learning. The depth concept and related technical algorithms are expounded, the learning methods in logistic regression and support vector machines are explored, and the applicability of data-driven learning in translation teaching is analyzed through experiments.

Data-driven learning is a way to increase input and focus on learners. It provides real data that facilitates conscious learning and is the best activity for increasing self-awareness. At the same time, this learning method can not only stimulate students’ autonomous learning but also promote students’ learning ability and improve learning effect [22].

Through experimental analysis, this paper shows that the DDL method is better than the traditional method (Mean = 5.02 > 0). Paired samples \( t \)-tests indicated a significant difference between participants’ performance and DDL-treated words and traditionally treated words \( (p = 0.001 < 0.01) \). So, the trend of significance shows that DDL does play a role in vocabulary learning. Therefore, DDL works better overall than traditional methods.

6. Conclusions

As a new learning method, data-driven learning has been widely used in college English translation teaching. Data-driven learning provides a variety of learning resources for translation teaching and solves the problems of outdated content and insufficient knowledge in college English translation teaching. This way of learning not only helps students master the study skills of English translation but also helps to cultivate students’ autonomous learning ability and meet students’ individual learning needs. Properly used in students’ foreign language learning, DDL is bound to see a broad prospect and have a profound impact on language acquisition. Therefore, it is necessary to study the correlation with other factors, whether it can be generalized, and what form it needs to be further explored. For today’s talent society, the comprehensive quality of students has also been improved to a certain extent, and it is also spread among modern students. In these respects, vigorously promoting data-based learning also requires front-line teachers to understand the concept of classroom teaching and must combine teaching methods with modern information technology to optimize the overall environment for English translation teaching.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author does not have any possible conflicts of interest.

References


