Improvement Research on Computer Assistant Proverb Translation

Jianyu Zheng

Schoo of TranslationStudies, Jinan University, Zhuhai, Guangdong, China

e.Jianyu@163.com

Keyword: CAT, BP neural network, Translation preference, Foreignizing translation, Domesticating translation.

Abstract. The thesis focuses on the efficiency problem caused by the unordered reference version provided by CAT software when calling the translation memory database. The BP neural network is used to study the connection between the translator's background and his/her preference for the goal of providing translators with ordered reference version. The thesis first acquires the research datum of the translator's background and their preference of translation.And the datum are quantified by options. Then, based on the BP neural network, an analysis model telling the preference between foreignizing and domesticating translation will be set up .Furthermore, the neural network will be trained through the training data set. Finally, with the help of detected set, we can check the grading accuracy between the predicted results and the expected value. Analysis model of the preference between foreignizing translation and domesticating translation of translators based on BP neural network can better recognize the translator's foreignizing and domesticating translation preference.

Introduction

With the increasingly fierce competition in translation industry, CAT, compared with human translation, can largely improve the quality and efficiency of translation. Translation memory and terminology management are two core technologies of CAT[1]. While in the real situation, when finding the matched translation for the source language, if the reference versions provided are unordered and messy, it will without doubt prolong the reading time and make it harder to discriminate, thus decreasing the translation efficiency.

To iron out this flaw, ordering the reference version by preference analysis will be a reasonable try. Wu Xiaoli[2] proved from the subjective angle that the subjective feelings, diction habit, ideology and cultural background of translators will impact on the quality of translation. Cai Suying[3]pointed out that the translation thoughts and attitudes, in most cases, are related to the preference of translators whether to choose foreignizing translation or domesticating translation. In order to find out the connection between the subjective backgrounds of translators and their translation preference, we need a more scientific method. Gao Jun[4] put forward the possibility that artificial neural network can be used to abstract, simplify and simulate human brains, and further to reflect characteristics of human brains, which can be widely used in related areas of analysis and prediction. Wu Jiansheng[5] pointed out that BP neural network is the most widely used neural network learning algorithm. Most neural networks are adopting BP algorithm or the variation of it. In fact, the algorithm shows the best part of neural network.

All the above research results taken into consideration, this thesis will use the BP neural network to build a preference analysis model for translators. We will first quantify the datum from the questionnaire, then input the datum in order to get the model. After that, input the background information of the translators for predicting their translation preference when translating proverbs. Finally, compare the predicted results and the expected value and make comments on the model.

Analysis model of the preference between foreignizing translation and domesticating translation of translators based on BP neural network

Background factors often impact on the translation preference of translators[2], if CAT software can adjust the reference versions by order according to their translation preference, it will without

doubt shorten the time for choosing ideal reference versions, thus improving the efficiency of translation. Therefore, this thesis adopts three layers BP neural network to build an analysis model for telling the preference of translators, and use E-C proverb translation as an example to verify the rationality of the model.

Introduction of BP Neural Network. BP neural network, as the most widely used artificial neural network, possess complete theory system and learning mechanism[6]. The main idea of the algorithm is to divide the learning process into two stages: one is positive signal propagation and the other is error back propagation. During the process of positive signal propagation,the information from input layer transmits to output layer through the hidden layer, and it will become into output signal at the output end. During this process, the network weights remain fixed, the state of each layer of neurons will only affect the next layer of neurons. When the output layer fails to get the expected outputs, it will resort to the process of error back propagation. During the process of error back propagation, error signals unable to meet the accuracy requirements will start from the output end, be backward propagated layer by layer in a certain way, proportion the error to all units on layers, and dynamically adjust the connection weights of the whole unit based on error signals. Through a cycle of positive propagation and back adjustments, the weights between neurons can be continually revised. When the errors of the output signal finally meet the accuracy requirement, the learning process will come to an end[6]. The model is just like Figure 1 shown:

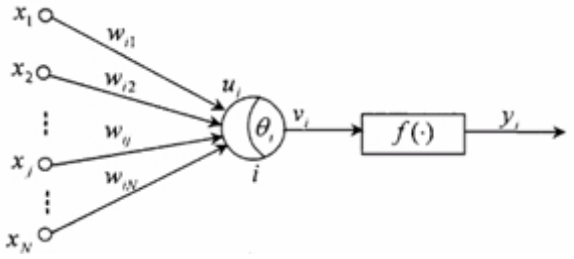

Fig. 1. BP Neuron Model

Let $x_j(j=1,2,\cdots,N)$ represents the input signal of neuron j; w_{ij} represents the connection weight; u_i represents the output of linear combination of the input signals, also the net input of neuron i; θ_i represents the threshold value of the neurons; v_i represents the value adjusted by the threshold value; $f(\cdot)$ represents the excitation function of the neurons, among which Sigmoid function is the most common one, and the expression is as follows:

$$f(x)=(1+e^{-\alpha x})^{-1},\ \alpha \text{ is a constant} \quad (1)$$

The mathematical model of the sensory transfer of the input signal in units:

$$u_i=\sum_{j=1}^{N} w_{ij}x_j \quad (2)$$

$$v_i=u_i+\theta_i \quad (3)$$

$$y_i=f(v_i) \quad (4)$$

Analysis model of the preference between foreignizing translation and domesticating translation of translators. Considered that if we can adjust the order of reference versions according to the translation reference of translators ,so that the time spent on reading reference versions will be shorten and the efficiency of translators will improve a lot. Whereas in real situations, there exist many more factors (such as education background, quantity of translation, and duration of translation.etc.) that will impact on the preference of translation[2]. Therefore, the thesis adopts the BP neural network with a hidden layer in building the model to simulate the real situations. The Network Structure(See Figure 2):

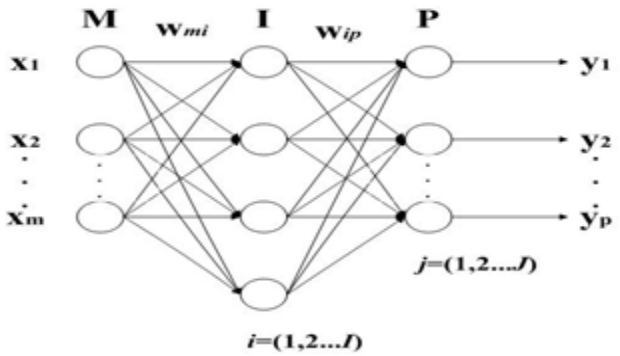

Fig. 2. The BP neural network with a hidden layer

In Fig. 2, we first quantify all the background factors of every translators(such as the education background, the quantity of translation, and the duration of translation.etc.). Let them as M input signals on the input layer, and m represents any input signal; Set there is I neurons in the hidden layer, and i represents any neuron; There are P neurons on the output layer, each representing different option values of the quantified preference of translators, and p represents any neuron; w_{mi} represents the weight from the input layer into the hidden layer, and w_{ip} represents the weight from the hidden layer into the output layer; u represents the input of the neuron, and v represents the driver output. Therefore, let the subscript of u or v as a certain neuron, for example u_i represents the input of the i th neuron on the hidden layer; Let the set of quantified factors of the translators in terms of the foreignizing translation and domesticating translation of proverbs, namely the training sample set as $X=[X_1,X_2,\cdots,X_N]'$;the corresponding factors of each of translators in terms of the foreignizing translation and domesticating translation of proverbs, namely any single training sample as $X_k=[x_{k1},x_{k2},\cdots,x_{kM}]'\ k\in(1,2,\cdots,M)$; the quantified alignment of the translators in terms of the foreignizing translation and domesticating translation of proverbs, namely the actual output as $Y_k=[y_{k1},y_{k2},\cdots,y_{kP}]'$; and the expected output of the network as $D_k=[d_{k1},d_{k2},\cdots,d_{kP}]'$. Besides, let n represent iterative times, and weight and the actual output are considered as function of n[6]. The general steps of the algorithm are as follows:

Step 1: Weight initialization: Randomly assign a set of small nonzero values as $w_{mi}(0)$ and $w_{ip}(0)$;

Step 2: Determine the structure parameters and define the variables in the BP neural network: Set $w_{ip}(0)$ as the input vector, and N as the number of the samples. $Y_k(n)=[y_{k1}(n),y_{k2}(n),\cdots,y_{kP}(n)]$ as the n th iterative network actual output, and $d_k=[d_{k1},d_{k2},\cdots,d_{kP}]$ as the expected output.

Step 3:Input the training samples: Input in turns the training samples $X=[X_1,X_2,\cdots,X_K,\cdots X_N]$, set $X_k\ (k=1,2\cdots,N)$ as the learning sample.

Step 4: Positive propagation: Based on the concerned formulas, figure out the network outputs and the training error of the samples.

Step 5: Error back propagation: Based on the error signals, update weights and threshold values of all layers; Judge whether K>N, if yes, then shift to step 6; it not, then shift to step 3.

Step 6: Figure out the aggregate error of the network training, if it meets the accuracy requirement, then end the training; if not, then shift to step 3, and a new round of learning starts[7].

Experiments and Result Analysis

Data Collecting and Quantificative Processing. The thesis designed several multiple questions in the questionnaire and collected related data for researching the background factors of translators and their foreinizing translation and domesticating translation preference. The 16 factors involved are as follows: sex, age, vocation, education background, whether having the experience of living

abroad and how long, whether speaking dialects, the starting point of translation, quantity of translation, text type of translation, age groups of the target readers, the status of the target readers, whether having interests in literature writing or just writing, whether highly valuing translation theories, the frequency of using proverbs in daily life, the attitude on the purity of Chinese, whether using certain translation methods in translating. In the meanwhile, the thesis selected 10 English proverbs and listed 3 versions of their Chinese translations, of which the 1st applied the foreignizing translation method, the 3rd applied the domesticating method, the 2nd was the intermediate state and combined the two methods.

The thesis finally collected 124 valid questionnaire data, 114 of which were used in the BP neural network training set, and the rest 10 were used in the testing set for BP neural network preference analysis model.

The thesis first quantified the results. During this process, it's worth mentioning that we set the foreignizing translation option as 1; the intermediate state option as 0; and the domesticating translation option as -1. Then we count the number of the choices of each translation, and consider it as the preference of the translator. When the numbers of the domesticating translation options and he intermediate state options top out and equal to each other, denoted by -0.5; When the numbers of the foreignizing translation options and he intermediate state options top out and equal to each other, denoted by 1.5; After this process, we set the background factors data as the the BP neural network input, the quantified preference value as the the BP neural network output, and then carry out the training. Finally, set the background factors data in the rest 10 questionnaires as the input variable, get the predicted value after the operation of the training, and compare it with the actual expected value in order to test the accuracy of the model.

Model Testing and Results Analyzing. By using the MATLAB software, we adopt the trainrp method in the BP neural network, and set the error value as 0.0001, the learning speed as 0.01and times of training as 10000 times, carry out the experiment according to the steps mentioned above, and get the computational results.(See Chart 1)

Chart 1 The BP neural network computational results on the 10 proverbs translation

computational results	**-0.50**	**1.92 E-10**	**-7.46 E-11**	**-9.92 E-11**	**3.32 E-11**	**-1.03 E-09**	**-1.00**	**-1.77 E-10**	**-1.0 0**	**-3.64 E-10**
Normalization of the results	**-1**	**0**	**0**	**0**	**0**	**0**	**-1**	**0**	**-1**	**0**
Expected values	**-0.5**	**0**	**0**	**0**	**0**	**0**	**-1**	**0**	**-1**	**0**

Chart 1 shows that only the 1st figure of the results after normalization does not assort with the expected values, namely the computational results generally accord with the expected values. Therefore, the model can relatively well predict the foreignizing translation and domesticating translation preference of translators based on their background factors. The thesis owns the above reasonable results to the reasonable setting of questionnaire, the comprehensive factors involved, the relevant datum and the right choice of the algorithm.

To furthermore test the accuracy of the model, the thesis takes the first proverb for example test. It reads "absence makes the heart grow fonder", and the three translation versions into Chinese are as follows:

Foreignizing translation: "ren bu zai geng jia shi de liang xin xiang yue"

Intermediate state: : "di yuan xin yu jin"

Domesticating translation: "xiao bie sheng xin hun "

Set the options of every translators on the first proverb as the outputs, the running steps are mentioned above, and the computational results:(See Chart 2)

Chart 2 The BP neural network computational results on the 1st proverb

computational results	-1.89 E-04	-2.70 E-05	-9.99 E-01	-1.0005	-0.9999	-0.9996	-1.00	-1.0002	-1.0007	-0.9999
Normalization of the results	0	0	-1	-1	-1	-1	-1	-1	-1	-1
Expected values	0	0	-1	-1	-1	-1	-1	-1	-1	-1

Chart 2 shows that the results are completely after normalization consistent with the expected values, which further proves the rationality of the model.

Conclusion

For the present, CAT software are not providing preference-based ordered reference version for translators and to some extent influencing translation efficiency . To make up for the flaw, the thesis builds an analysis model of the preference between foreignizing translation and domesticating translation of translators by collecting datum about translators' background factors and reference and running the BP neural network algorithm. And by testing the prediction sets, finally determine that the model can well predict the preference of a certain translators according to his/her background factors.

The thesis determines the translation preference of each translator by quantifying their choices on the ten proverbs. After the learning process of 144 datum, and then input the datum from the rest 10 translators' background information. The computational results are generally consistent with expected value. To furthermore test the accuracy, the thesis conducts the same process on the 1st proverb, and finally determines the high accuracy of the model.

Although the BP neural network can relatively well predict the foreignizing translation and domesticating translation preference of translators based on their background factors, we must acknowledge that it's a progressive process from the foreignizing translation to the domesticating translation[8], and that there still remains several intermediate states that need to be further discussed. If we get a more detailed division, the model constructed will much better predict the preference of translators and also provide more accurately ordered reference versions. Besides, in the thesis, we only discuss the proverb text, our further direction can be other text types in order to expand the application areas of the model.

Acknowledgements

The authors acknowledge the financial support of this research by 2014 Jinan University Undergraduate Training Programs for Innovation and Entrepreneurship

Reference

[1] G.X. Zhang, Y.L. Jiang. Discussing Computer Aided Translation Software. China Modern Educational Equipment, 2009, (15):50-52. DOI:10.3969/j.issn.1672-1438.2009.15.019.

[2] X.L. Wu. The Expression of Translator's Subjectivity in the Process of Translation. Journal of Gansu Radio & TV University, 2009, 19(1):28-31. DOI:10.3969/j.issn.1008-4630.2009.01.009.

[3] S.Y. Cai. Comparative Study on "Foreignizing Translation"and"Domesticating Translation". Scientific and Technical Information, 2008, (29).

[4] J. Gao. Artificial Neural Network Principle and Simulation Examples. Beijing:China Machine Press,2003.7.

[5] J.S. Wu. The Back-propagation Neural Network Meteorological Forecast Model Based on Genetic Algorithms. Guangxi Normal University, 2004.

[6] T.S. Liu. The Research and Application on BP Neural Network Improvement. Northeast Agricultural University, 2011. DOI:10.7666/d.y1973563.

[7] J.Q. Lu., F. Ye. Risk Evaluation and Application in Customs Enterprises Based on BP Neutral Network . Journal of Computer Applications, 2005, 25:333-335.

[8] Z.Q. Zhang. An Exploration of the Conception of Foreignizing Translation. Hunan Normal University, 2013.

[9] J.W. Zhuo. MATLAB Software in Mathematical Modeling. Beijing: Beijing University of Aeronautics and Astronautics Press,2011.4.