International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-10

https://doi.org/10.70917/ijcisim-2026-0084
Article

# Accuracy Improvement of Word Vector-Based Machine Translation Algorithm in the Communication of National Community Awareness

**Mengjia Peng ***

School of Foreign Languages, Wuhan Business University, Wuhan 430056, Hubei, China; mira202418@163.com

**Abstract:** In this paper, word vector alignment is introduced into the machine translation model to initialize the word embedding layer of the neural machine translation model for joint training. A monolingual corpus is used to train the model for denoising and self-encoding to improve its encoding ability, decoding ability and translation accuracy. The People's Daily is chosen as the data source, and the study shows that interactive metadiscourse is more frequently used in the news discourse of People's Daily, with 332,433 words, accounting for 81.61% of the total number of metadiscourse, so as to enhance the effect of the communication of national community consciousness. Through the questionnaire survey, it is found that the national community consciousness enhancement score of college students is 4.69, and through the method of this paper, it can provide intelligent solutions for the dissemination of the national community consciousness, which greatly improves the dissemination effect of the excellent traditional Chinese culture.

**Keywords:** machine translation; word vector; national community; dissemination effect

## 1. Introduction

Machine translation, which uses computer technology to automatically convert between different languages, has greatly facilitated international communication and effectively reduced labor costs [1-2]. Machine translation can translate one language into another or multiple other languages, meeting the communication needs of people from different countries [3]. As the pace of information technology and cultural-economic globalization continues to accelerate, machine translation has become an indispensable component of cultural exchange [4-5]. Compared to traditional rule-based and statistical machine translation methods, neural machine translation incorporating word vectors demonstrates superior translation performance and potential due to its unique advantages [6-7].

Most research on machine translation applications aims to improve the accuracy of model translations and promote cross-cultural and linguistic communication. Reference [8] employs embedded neural network technology to overcome language barriers, reducing human-imposed limitations and errors, and significantly improving the accuracy and quality of cross-cultural communication translation. Reference [9] combines speech-to-text recognition with computer translation systems to effectively enhance the understandability and translation accuracy of different languages in cross-cultural communication, and the method's usability has received high praise from interviewees. Literature [10] proposes a realistic open vocabulary network machine translation model, which improves the BLEU score of existing unknown word translation models by 2.1 to 11.4. The translation system equipped with this method achieves a BLEU score of 20.7, yielding superior translation results. Literature [11] conducted a study on translation accuracy using neural network machine translation technology, with Shaanxi red culture texts as the research object. The results indicate that the use of a neural network translation system improves the accuracy of translating Shaanxi red culture texts, thereby promoting the overseas dissemination of ethnic culture. Literature [12] proposes a semantics-based out-of-vocabulary

(OOV) handling method, which calculates the semantic similarity of word vectors and replaces OOV words with the most similar words. The results show that this method achieves a higher BLEU score than the Seq2seq model, effectively improving translation accuracy.

Machine translation has indeed played a significant role in cross-cultural communication, but there are few translation applications related to the dissemination of ethnic collective consciousness. Therefore, further research is needed on the translation of ethnic collective consciousness. Additionally, machine translation has the following shortcomings regarding ethnic collective consciousness:

First, ethnic-specific terms in external propaganda materials lack direct corresponding terms in the target language during translation, and there are no authoritative reference materials on how to accurately convey the cultural information contained in these terms. Second, the limited storage of terms related to ethnic collective consciousness in machine translation corpora leads to cultural gaps in machine translation results. How to effectively utilize existing machine translation corpora to assist human translation and improve the current translation situation is a topic that requires further research.

In this paper, we use word vector-based machine translation algorithm to analyze the discourse description and expression system, accurately grasp the “preferences” and “habits”, and improve the communication effect of forging the national community consciousness. Starting from the perspective of national community consciousness, we discuss the influence of metadiscourse on the dissemination of national community consciousness with the help of machine translation algorithm based on word vectors, taking the news articles of People's Daily about “Chinese national community” as the data source. Questionnaire survey and one-sample t-test are adopted to analyze the effect of the word vector-based machine translation algorithm on the dissemination of national community consciousness in terms of cognitive enhancement, emotional enhancement, volitional enhancement, behavioral enhancement and consciousness enhancement.

## 2. Paths to Greater Precision in the Dissemination of National Community Awareness

Based on the word vector machine translation algorithm empowers the casting of college students Chinese national community consciousness of the intelligent picture of the smart media era, algorithmic technology empowers the casting of college students Chinese national community consciousness, aims to the users of college students of all races as the algorithm structure of the landing point, to realize the casting of college students Chinese national community consciousness of the ubiquity of the educational environment, the way of the closeness of the process of the dynamics of the casting of the Chinese national community consciousness of the college students as a means to promote the casting of the college students Chinese national community consciousness of the intelligent development. Intelligent development of Chinese national community consciousness.

(1) Cyberspace, as the main practical activity space for college students' daily interaction and recreation, is an important environment for college students to shape values and form value consensus, and is also a key experience environment for forging a strong sense of the Chinese national community. Based on the virtual, decentralized and interactive nature of cyberspace, college students not only gain the initiative to choose information, but also become the publishers and disseminators of information content. To a certain extent, cyberspace gives college students more autonomy and amplifies their individual initiative. Therefore, it is especially urgent for students to realize the ubiquity of the educational environment of forging students' sense of community through the network so that the sense of community of the Chinese nation can really cover and penetrate into the daily life of the students, and the sense of community of the Chinese nation can be forged firmly.

(2) On the one hand, the algorithm makes the discourse dissemination of Chinese national community consciousness from abstract to concrete, from indoctrination to experience, and matches the cognitive schema and meaning production of college students with the logic of intuitive infiltration and micro-narrative life. The algorithm can accurately grasp the “preferences” and “habits” of college students by analyzing and predicting their behaviors and interactions, so that through the construction of discourse descriptions and discourse expression systems, the educational method of forging a strong sense of Chinese national community is constantly close to the best state of college students' demands. Through the construction of discourse description and discourse expression system, the educational method of forging the consciousness of the Chinese national community is constantly close to the optimal state of college students' demands, more in line with the expectations of college students, realizing the real connection between abstract theoretical discourse and the life world of college students, and effectively enhancing the dissemination effect of the content of the education of forging the consciousness of the Chinese national community.

(3) Supporting the dynamization of education process. The algorithm, through the data capture and

data analysis of college students' network action information, is able to explore the “potential relationship” and obtain the “multi-source, multi-constructed, multi-state dynamic data information” so as to grasp the development trend and flux situation of college students' consciousness of the Chinese national community in depth. Based on the formed “data hive”, we can intelligently research and judge the potential risks in the process of forging a strong sense of the Chinese national community among college students, and form a precise grading and dynamic control of risks.

## 3. Word Vector-Based Machine Translation Algorithms

### *3.1. Model Architecture for Machine Translation*

Machine translation refers to the process by which a computer can translate the source language sentences in which the sense of national community is spread to the target language sentences that are semantically equivalent to them, and it is an important research direction in the field of natural language processing. In a narrow sense, machine translation generally refers to the translation between texts in different languages, i.e., translating a text in one natural language into a text in another natural language. In a broader sense, machine translation can refer to both text translation and speech translation, the input and output of text translation system are text, while the input and output of speech translation system are speech. In this thesis, machine translation refers to text translation. The language to be translated is called the source language, and the translated language is called the target language. The whole process of machine translation can be divided into three stages: source text analysis, conversion between source text and translation, and translation generation. The model architecture of machine translation is shown in Fig. 1. Since the birth of machine translation, its research has been transformed twice around the two trends of rationalist approach and empiricist approach:

1) The rationalist approach to translation means that human experts generate algorithms using the conversion laws between different natural languages by way of compiling rules, and the computer translates through such rules. The advantage of this method is that it is theoretically capable of grasping the deep-rooted conversion laws between languages. The disadvantage of this method is that it is extremely demanding for experts, who not only need to know the source language and target language, but also need to have certain knowledge of linguistics and translation, and moreover, need to be skillful in the relevant operation skills of computers. Therefore, the acquisition of translation knowledge and linguistic knowledge becomes the main problem faced by the rational-based machine translation method.

2) Empirical translation method refers to the data-driven basis, which advocates that computers automatically learn the conversion laws between natural languages from large-scale data. The advantage of this method is that due to the continuous growth of text data on the Internet and the increasing computing power of computers, the data-driven statistical translation method has gradually become the mainstream technology of machine translation. But at the same time, statistical machine translation is also facing problems such as data sparsity, difficult to design features, etc. Deep learning can better alleviate the challenges faced by statistical machine translation, neural machine translation is in the rising period of research, and it is the hot spot in the field of machine translation at present.

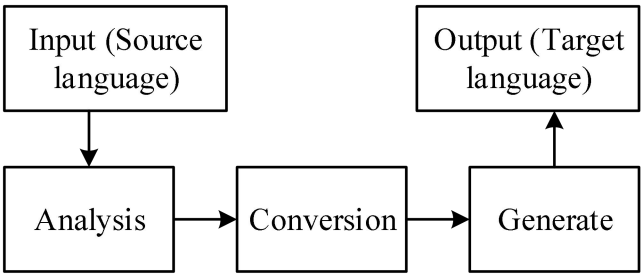


***Figure 1.*** Machine Translation process.

### *3.2. Neural Machine Translation Based on Word Vectors*

In neural machine translation, for a certain language $L$, each word in the model input is represented by a corresponding word vector $w$, and the space where these word vectors are located is the word vector space of the language $L$. Mikolov et al. found that the resultant constructions of the word vector spaces of different languages are similar, even if they vary greatly from one language to another. This situation still holds for Chinese and the target language. It is well known that a good initialization is very important for neural network models, and similarly a better initialization of the word embedding layer is

also conducive to the convergence of neural machine translation models. In this paper, we first use the Chinese monolingual corpus to train their respective word vectors, then use the open-source tool MUSE to align the word vector space, and finally use the aligned word vectors to initialize the word embedding layers in the model respectively [13].

Word vector alignment, also known as word vector space alignment, refers to mapping the word vectors of two or more different languages into an identical potential vector space, so that the word vectors of words with similar meanings in different languages have similar distances in the word vector space. Let $X_e \in \mathbb{R}^{d \times n}$ be the bilingual word vector matrix, and $Y_e \in \mathbb{R}^{d \times n}$ be the Chinese word vector matrix, where $d$ is the dimension of the word vectors, and $n$ is the number of the first $n$ most frequent words.

Then word vector alignment is to find a matrix $W' \in \mathbb{R}^{d \times d}$ so that the structure of the two word vector spaces is aligned as much as possible, as shown in equation (1):

$$W' = \arg\max_{w} \| WX_e - Y_e \|_F \tag{1}$$

where $\| \|_F$ denotes the $F$ paradigm.

Assuming the spatial dimension is two-dimensional, the dashed irregular shape denotes the Chinese word vector space, and the solid irregular shape denotes the word vector space, the shape structures of the word vector space and the Chinese word vector space are roughly aligned in the aligned potential vector space. Circles indicate the position of certain words in the Chinese word vector space, and triangles indicate the position of certain words in the word vector space, then in the aligned word vector space, some words with similar meanings that occur more frequently have similar distances in the vector space. In the aligned word vector space, the Chinese word “University” is the closest.

After aligning the Chinese word vector space, the aligned word vector matrix of Chinese can be obtained, and then these word vectors can be used to initialize the word embedding layer of the proposed neural machine translation model in this paper, so as to carry out the joint training of neural machine translation. The objective function of training is shown in equation (2):

$$L_{mt}(\theta) = \mathbb{E}_{X \sim D_x, Y \sim D_y}[(-\log P_{x \to y}(Y \mid X)) + (-\log P_{y \to x}(X \mid Y))] \tag{2}$$

where $\theta$ denotes the parameters of the model, $D_x$ and $D_y$ denote the set of target utterances and the set of Chinese utterances in the parallel corpus, respectively, $X$ denotes the target utterances in the intertranslated sentence pairs $(X, Y)$ in the parallel corpus, $Y$ denotes the Chinese utterances in the intertranslated sentence pairs, $P_{x \to y}$ and $P_{y \to x}$ denote the translation models, and $E$ denotes the mathematical expectation.

### *3.3. Training of the Model*

After constructing the completed model and initializing the word embedding layer of the model using aligned word vectors, this paper uses a joint training approach to train the whole model. This is done by first training the denoising self-encoding using a set of Chinese monolingual corpus, then translating using a set of parallel corpus, and finally updating the model using the loss of both in each round of data iteration. In the whole training process, the denoising self-encoding and translation processes are carried out alternately, and the proportion of the loss of denoising self-encoding in the total loss decreases with the increase of the number of training steps. In summary, the total objective function of the model is shown in equation (3):

$$L_{model}(\theta) = \lambda_{ae} L_{ae} + \lambda_{mt} L_{mt} \tag{3}$$

where $\lambda_{ae}$ and $\lambda_{mt}$ denote the weights of the loss of denoised self-coding and translation to the total loss, $L_{ae}$ denotes the loss of denoised self-coding and $L_{mt}$ denotes the loss of translation, respectively.

The loss function curve when the translation task is fine-tuned is shown in Figure 2. The use of aligned word vector initialization to initialize the word embedding layer of the model can not only give the model a better initial value to make the model converge faster, but also make the sentence vectors formed by the encoder of the model after encoding the sentences and the target sentence vectors as close as possible, so that the sentences in the target language can be more accurately decoded during the decoding process, generating a better translation result and improving the translation accuracy.

Enhancement.

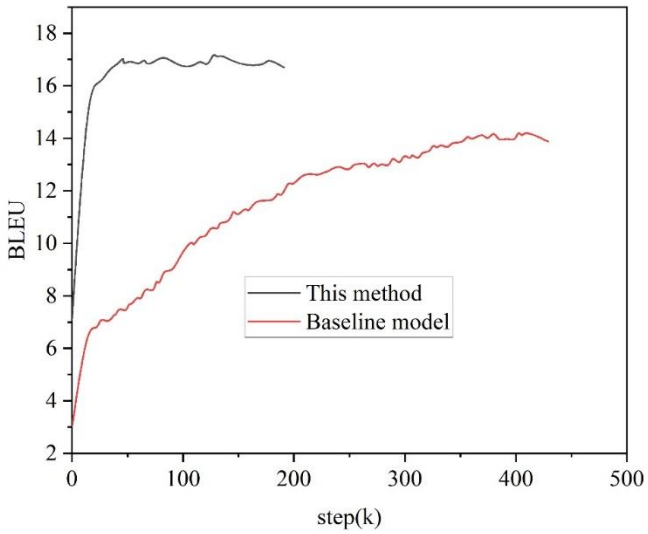


***Figure 2.*** The curve of the bleu value.

## 4. Research on the Application of Machine Translation Algorithms Based on Word Vectors in the Dissemination of National Community Awareness

### *4.1. Experimental Parameterization*

In this study, English and Chinese reports in the People's Daily were chosen as the data source, as news reports on the "Chinese National Community" in the People's Daily reflect the Party and State's goal of building and mainstreaming the understanding of this issue, and have become an authoritative source of information for other media reports. Founded in 1948, the People's Daily has a history of 76 years, and its texts have a chronological span and co-temporal thickness. Its news articles on the "Chinese National Community" reflect the Party and State's important deployments and key nodes of the issue, and have also become an authoritative source of information for other media reports. First, through the "People's Daily Graphic Database", with "Chinese National Community" as the keyword and "title + body" as the search condition, the search time is 2024. From May 28, 2024 to December 31, 2024, a total of 1,020 valid samples were retrieved, including news, newsletters, editorials and other genres. Second, potential errors, inaccuracies, and inconsistencies were identified and resolved with the help of Pandas software for the purpose of data cleaning. Once again, the corpus of "People's Daily's Chinese National Community" reports was initially self-constructed by categorizing them with the information of title, time, disseminator, time, version, and keywords. According to the aforementioned ten metadiscourse markers as the basic parameters of media metadiscourse classification statistics, then the self-constructed corpus is manually identified and labeled. The machine translation algorithm using word vectors is used to analyze the frequency of each type of metadiscourse and count the frequency and number of each type of metadiscourse on national community awareness in the self-constructed corpus. Finally, by analyzing the texts, we will deeply understand the functions and effects of metadiscourse in specific contexts, and discuss the reasons for the differences in the distribution of metadiscourse and its impact on the construction of consensus on the Chinese national community.

In order to prove the validity of the models in this paper, the entire corpus was pre-trained using Word2vec with 100, 150 and 200 dimensional word vectors respectively. In the training of the three models dimension size is 100, 150 and 200 respectively, sg is equal to 1, the default value of 5 is selected for the window window, the configured threshold of random sampling is 1e-3, and the number of iterations iter is 2. In order to allow the collection of rare words to be pre-trained to the maximum extent, min_count is set to 3. The corpus is trained to get three models with different dimensions, namely word2vec_100.model, word2vec_150.model and word2vec_200.model, and finally compare the experimental results and trade-off meter. This experiment uses the keras deep learning framework, the back-end environment is TensorFlow, using Python language programming to achieve; the experiment running program for the Anaconda3 software in the Spyder, Win7 system, memory 8GB, etc., the experimental parameter settings as shown in Table 1, in which the input layer of the word vectors are used for the "People's Daily" text language characteristics of the training Word2vec contextual word

embedding model.

***Table 1.*** Experimental parameter Settings.

| Parameter | Value |
|---|---|
| Hidden node number | 100 |
| Loss function | categorical_crossentropy |
| Optimzer | Adam |
| Batch_size | 130 |

Word vectors are vectors used to represent words and are also considered as feature vectors of words, the process of mapping words to real number domain values is called word embedding. Vector space model is a widely used model for information retrieval and has the advantage of using spatial similarity to approximate semantic similarity. The measure of semantic similarity is actually mapped to the measure of vector similarity, that is, for two words that need to calculate semantic similarity can be transformed into the form of values in a multidimensional vector space in order to facilitate the calculation and organization. The semantic scoping experiment loads the above collection of vectors word2vec_100 trained using the monolingual corpus in the petroleum domain for word embedding as the background vectors, and then encodes and transforms the results of the Yodo translation and the standard translation into the input vectors respectively, so that these vectors can better express and compute similarity and analogical relationships between different words. In terms of experimental preprocessing we do some dimensionality reduction on the documents to improve the accuracy of the model. The standard translation codes of several commonly used terms of national community awareness are converted into vector values, and the closest words in the semantic range are extracted in descending order according to their similarity values, and the degree of superposition of the two translation results is measured by judging whether the extracted words cover the results of the machine translation, so as to define the semantic range of the two translation results. The distance value corresponding to the extraction of near-sense words is in the interval of [0, 1], and the closer to 1, the closer the two words are to each other, and the stronger the semantic correlation is. On the contrary, it means that the two words are more distant from each other semantically. Here, the selected typical terms of national community consciousness are organized and listed as shown in Table 2.

***Table 2.*** Key words of the Chinese national community consciousness.

| Synonym | Similarity | Synonym | Similarity |
|---|---|---|---|
| We | 0.7415 | Chinese nation | 0.5849 |
| Everybody | 0.6849 | Compatriots on both sides | 0.5663 |
| China | 0.6425 | Stick to | 0.5627 |
| Big family | 0.6357 | Masses | 0.4846 |
| Family | 0.6394 | Approval | 0.4863 |
| Chinese children | 0.5581 | Stick to | 0.4863 |
| …… | …… | …… | …… |

## *4.2. Chinese National Community Media Metadiscourse Analysis*

The statistical analysis of machine translation based on word vectors found that interactive metadiscourse is more frequently used in news discourse. The results of metadiscourse analysis are shown in Table 3, in 1020 news discourse, all kinds of metadiscourse markers are 407,349 words, among which interactive metadiscourse is 332,433 words, accounting for 81.61% of the total number of metadiscourses, and guided metadiscourse is only 74,916 words. The usage frequency of the intensive marker category metadiscourse in interactive metadiscourse is much higher than that of other interactive metadiscourse, accounting for 48.48% of the total. It can be seen that the Chinese national community media metadiscourse is mainly used in an interactive way to express the certainty and strong confidence

of the communicator on the topic through the intervention or evaluation of the news, and to construct the Chinese national community discourse system. Note: The number of metadiscourse tokens indicates the total number of different categories of metadiscourse appearing in the corpus, the percentage of the total number of metadiscourse tokens indicates the proportion of a certain category of metadiscourse in the total metadiscourse corpus, and the frequency of occurrence shows the frequency of a certain category of metadiscourse appearing in the total corpus.

***Table 3.*** "Chinese national community" media meta-discourse statistics.

| | Categories | Token number (words) | The proportion of the total words | Frequency of emergence (per thousand words) |
|---|---|---|---|---|
| Interactive meta-discourse | Fuzzy markup | 13810 | 3.39% | 4.11 |
| | Enhanced markup | 197482 | 48.48% | 58.42 |
| | Attitude marker | 30014 | 7.37% | 8.90 |
| | Self-mention | 91127 | 22.37% | 27.46 |
| Guided meta-discourse | Interventional marker | 12785 | 3.14% | 3.84 |
| | Guided meta-discourse | 26783 | 6.57% | 7.96 |
| | Transition marker | 17451 | 4.28% | 5.16 |
| | Frame markup | 204 | 0.05% | 0.07 |
| | Internal marker | 13614 | 3.34% | 3.96 |
| | Quote | 4079 | 1.00% | 1.25 |

## *4.3. Questionnaire Study*

The survey took two forms: questionnaire survey and case interview survey, of which online questionnaire survey was the main form and case interview was the auxiliary form. The questionnaire survey was conducted through a specialized data collection and analysis platform, Questionnaire.com. Constructing and enhancing "identity" is the core of the cultivation of the sense of community of the Chinese nation, and the essence is a psychological process, which follows the psychological development law of "knowledge-feeling-meaning-action". Therefore, from the perspective of identity, this paper divides the four dimensions of "knowledge-feeling-meaning-action" to analyze the accuracy and effect improvement of the dissemination of national community consciousness, which is in line with the concept definition and core essence of the Chinese national community consciousness.

The purpose of the questionnaire survey is to analyze the college students' cognitive evaluation of machine translation of word vectors in the dissemination of national community consciousness through the findings of this questionnaire The questions in this questionnaire are designed in a step-by-step manner, with a total of 28 questions, mainly single-choice, multiple-choice, and subjective, with a reasonable number of questions that are highly targeted. This questionnaire survey mainly selects college students of different majors and degrees in more than ten undergraduate and college universities in a province. The questionnaire survey was issued from September to December 2024, which lasted three months, and the survey data was counted on December 20, 2024, with a total of 1,200 questionnaires issued and 1,026 valid questionnaires recovered, with an effective recovery rate of 85.5%, which is in line with the normative standards of questionnaire surveys and can be used as a basis for the analysis of the problems in this thesis. The following will be a statistical analysis of the basic situation of the valid samples returned by the questionnaire, and the results are shown in Table 4.

***Table 4.*** The survey of the object characteristics was descriptive.

| Project | Type | Number | Proportion |
|---|---|---|---|
| Gender | Man | 647 | 63.06% |
| | Female | 379 | 36.94% |
| Period of birth | 2000 | 794 | 77.39% |
| | 1990 | 214 | 20.86% |
| | 1980 | 18 | 1.75% |
| Political appearance | Masses | 497 | 48.44% |
| | League member | 405 | 39.47% |
| | Party members (including preparation) | 120 | 11.70% |
| | Democrats | 4 | 0.39% |
| Educational background | Specialty | 572 | 55.75% |
| | Undergraduate | 302 | 29.43% |
| | Graduate student | 152 | 14.81% |
| Professional background | Literary history | 386 | 37.62% |
| | Technologists | 589 | 57.41% |
| | Art class | 40 | 3.90% |
| | Sports class | 11 | 1.07% |
| Peoples | Han | 874 | 85.19% |
| | Tibetans | 28 | 2.73% |
| | Hui | 53 | 5.17% |
| | Dongxiang | 36 | 3.51% |
| | Uygur | 10 | 0.97% |
| | Mongolian | 15 | 1.46% |
| | Dai peoples | 2 | 0.19% |
| | Li peoples | 1 | 0.10% |
| | Menggu peoples | 1 | 0.10% |
| | Hui peoples | 2 | 0.19% |
| | Zhuang peoples | 2 | 0.19% |
| | Bai peoples | 1 | 0.19% |
| | Yi peoples | 1 | 85.19% |
| | Gelao peoples | 2 | 2.73% |

Second, the validity test is divided into two aspects: content and structure. In terms of content, the questionnaire was divided into dimensions and set up questions based on mature research, and relevant experts and scholars were asked to review the questionnaire several times to ensure that the content was valid and reliable. Structurally, the scale was analyzed for inter-dimensional correlation, and the results are shown in Figure 3. The correlation coefficients between the dimensions of this scale are between 0.78-0.94, and the correlation coefficients with the total scale are between 0.91-0.93, which indicates that the dimensions are related to each other and independent of each other, i.e., the structure of the scale is appropriate and the validity is good.

To summarize, the reliability and validity of the total scale are relatively good, i.e., the questionnaire of this study is relatively reliable in reflecting the feelings of college students towards the machine

translation of word vectors in the dissemination of ethnic communities.

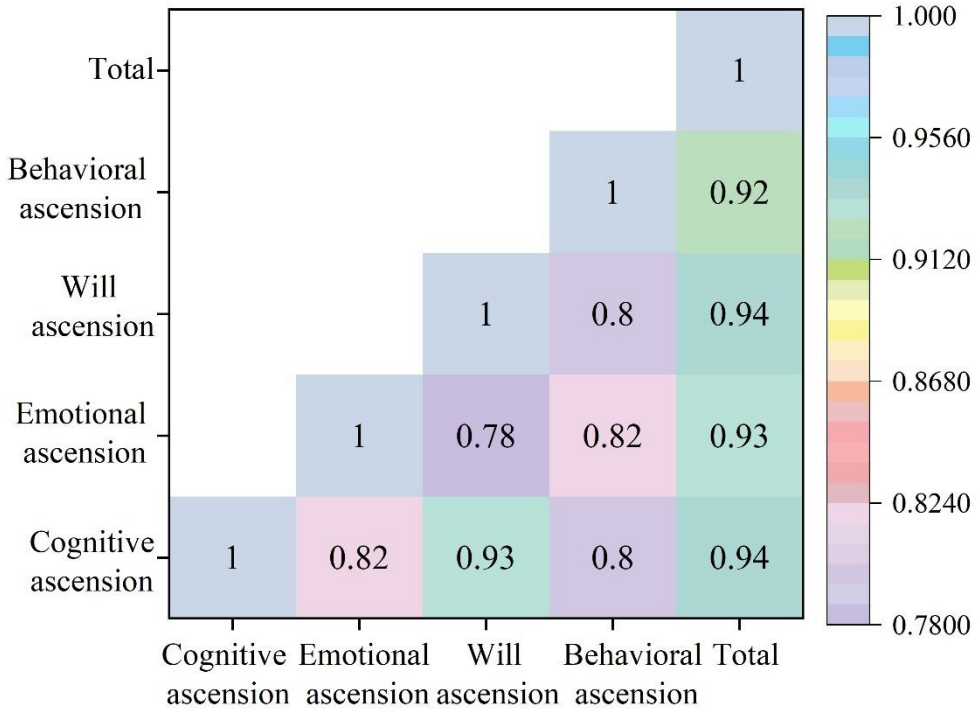


***Figure 3.*** Correlation coefficient matrix.

Based on the inseparable and close connection between consciousness and identity, and the psychological development process of "knowing, feeling, thinking and acting", this study calculates the emotional enhancement of the word vector-based machine translation algorithm in the dissemination of the national community consciousness by using the mean value of the scores of the four dimensions, with the lowest score of 1 and the extreme deviation of 4, where the higher the score means the higher the level of consciousness, and vice versa, the lower the score. . The results are shown in Table 5, the Chinese national community awareness enhancement score is 4.69, with a small standard deviation and non-significant dispersion. Through the one-sample t-test, it is found that the scores of each dimension and the consciousness enhancement score are significantly higher than the median value of 3, which indicates that the machine translation of word vectors by college students is in a relatively good condition in the dissemination of the consciousness of the national community, and that the precision of the dissemination work is effective, and that the next step is to optimize the cultivation work in order to achieve better results.

***Table 5.*** The overall consciousness of national symplete.

| Variable | N | Mean | Standard deviation |
|---|---|---|---|
| Cognitive ascension | 1026 | 4.78 | 0.63 |
| Emotional ascension | 1026 | 4.69 | 0.60 |
| Will ascension | 1026 | 4.79 | 0.58 |
| Behavioral ascension | 1026 | 4.74 | 0.54 |
| Consciousness ascension | 1026 | 4.70 | 0.57 |

## 5. Conclusion

In this paper, from the degree of spreading of national community consciousness, based on the self-constructed corpus of People's Daily's news articles about "Chinese national community", and with

the help of machine translation algorithm technology based on word vectors, we analyze the experiments and results, and the interactive metadiscourse of the national community in the People's Daily is 332,433 words, and the guiding metadiscourse is only 74,916 words. The interactive metadiscourse related to national community in People's Daily is 332,433 words, and the guiding metadiscourse is only 74916 words. Therefore, the national community awareness spreads mainly in an interactive way.

College students of different majors and degrees in more than ten undergraduate and junior colleges and universities in a province are selected as research subjects to reflect the college students' feelings about machine translation of word vectors in the dissemination of national community. It can be found that the scores of all dimensions and awareness enhancement are more than 3 points, indicating that the method of this paper is in a better condition in the communication of national community awareness, and the accuracy of the communication work has been effective.

**Funding**

This research was supported by the Department of Education of Hubei Province with Project: "Research on the Development Path of Translation Technology and the Cultivation of Translators' Competence" (Project No. 22G086).

**References**

1. Peris, Á., Domingo, M., & Casacuberta, F. (2017). Interactive neural machine translation. Computer Speech & Language, 45, 201-220.
2. Garg, A., & Agarwal, M. (2018). Machine translation: a literature review. arXiv preprint arXiv:1901.01122.
3. Clouet, R., & Massot, N. (2024). Designing and Implementing an Online Seminar in Telecollaboration. Machine Translation and Intercultural Communication (MTIC). SKASE Journal of Translation and Interpretation, 17(2), 29-50.
4. Khan, M. I. (2023). Translation: A Tool of Communication and Cultural Exchange. Strength for Today and Bright Hope for Tomorrow Volume 23: 11 November 2023 ISSN 1930-2940, 48.
5. Bal, D. (2025). An Analysis of Neural Machine Translation on Cross-Cultural Communication: A Case Study of Airbnb's Translation Function. Söylem Filoloji Dergisi, 10(1), 608-614.
6. Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
7. Stahlberg, F. (2020). Neural machine translation: A review. Journal of Artificial Intelligence Research, 69, 343-418.
8. Han, Q. (2024). Research on the Accuracy of Machine Translation in Cross-Cultural Communication Based on Embedded Neural Networks. International Journal of High Speed Electronics and Systems, 2540125.
9. Shadiev, R., Sun, A., & Huang, Y. M. (2019). A study of the facilitation of cross-cultural understanding and intercultural sensitivity using speech-enabled language translation technology. British Journal of Educational Technology, 50(3), 1415-1433.
10. Luong, M. T., & Manning, C. D. (2016). Achieving open vocabulary neural machine translation with hybrid word-character models. arXiv preprint arXiv:1604.00788.
11. Wang, J. (2022). Research on cultural translation based on neural network. Mathematical Problems in Engineering, 2022(1), 6330814.
12. Duan, J., Ma, H., & Wang, J. (2023). Automatic Chinese-English Translation Algorithm based on Out-of-vocabulary Words in the Context of Cross-cultural Communication. IEIE Transactions on Smart Processing & Computing, 12(6), 466-471.
13. Jingsong Huang. (2024). Youth Volunteer Services for Red Culture: Mechanisms and Pathways to Strengthen the Sense of Community for the Chinese Nation. Academic Journal of Business & Management,6(12).