ZIBELINE INTERNATIONAL™ PUBLISHING

ISSN: 2521-0874 (Print)
ISSN: 2521-0505 (Online)
CODEN: AIMCCO

Acta Informatica Malaysia (AIM)

DOI: http://doi.org/10.26480/aim.02.2024.64.68

CrossMark

RESEARCH ARTICLE

DESIGN OF TRANSLATION AMBIGUITY ELIMINATION METHOD BASED ON RECURRENT NEURAL NETWORKS

Jianzhou Cui*

Wuxi City College of Vocational Technology, Wuxi, Jiangsu, 214153, China
****Corresponding Author Email: 75686867@qq.com***

ARTICLE DETAILS

Article History:

Received 23 May 2024
Revised 08 June 2024
Accepted 16 July 2024
Available online 19 July 2024

ABSTRACT

The ambiguity of language inevitably leads to the ambiguity of translation, and how to deal with translation ambiguity has become a persistent focus of attention for both human translation and machine translation. Traditional machine translation mainly adjusts the threshold of ambiguity similarity to deal with translation ambiguity, but the effect is not ideal. The machine translation model based on recurrent neural networks provides us with a new perspective. In this new perspective, the candidate set calculates the similarity, obtains the source language and target language of the reference translation, and then nested in the neural network to complete the ambiguity elimination in language translation. This translation model based on recurrent neural networks effectively eliminates the gradient imbalance problem generated during the translation ambiguity process. Comparative experimental results also show that with a reasonable setting of the similarity threshold, the advantages of the new method are more evident and can better improve the translation results.

KEYWORDS

Translation Ambiguity; Neural Networks; Machine Translation; Translation Model; Language Ambiguity

1. Introduction

The acceleration of internationalization and globalization trends has raised higher demands on the quality and quantity of language translation. Human translation is no longer enough to meet the market needs, and machine translation has become an important supplementary means, even dominant in certain fields. Ambiguity and precision are objective attributes of human natural language, which are essential characteristics of human language (Wen, 1996). Furthermore, due to the cognitive abilities of humans, the purposes and behaviors of language communicators, and the inherent limitations of language itself, ambiguity in language arises. The ambiguity of language leads to ambiguity in translation. Scholars and experts both domestically and internationally have never ceased to explore methods to deal with the ambiguity of translation.

With the rise of machine translation, handling the ambiguity in machine translation has always been a technical challenge that scholars eagerly seek to conquer. In traditional methods of handling translation ambiguity, the most common approach is to adjust the fuzzy similarity threshold in algorithms, but this method faces a real contradiction: when the fuzzy similarity threshold is low, the performance of the method is poor, whereas with a higher threshold, the translation effect is better, but it may lead to difficulties in retrieval and hamper translation speed. In such a situation, it becomes especially necessary to redesign a method to eliminate translation ambiguity in English language translation. In recent years, neural network-based machine translation has surpassed traditional statistical machine translation methods in numerous tasks, becoming the main method in the field of machine translation. Building upon the advantages of traditional methods of handling translation ambiguity, this study proposes a model based on recurrent neural networks for eliminating translation ambiguity, that is, in English translation, calculating the similarity of candidate sets, obtaining the source language and target language of the reference translation, and then nesting them in a neural network to achieve the elimination of translation ambiguity in language translation.

2. The Ambiguity of Translation and its Handling

Engels pointed out in “Dialectics of Nature” that dialectics not only involves “either this or that”, but also recognizes “both this and that” where appropriate, and makes the opposites mutually interdependent (Engels, 1995). This “both this and that” is referred to as ambiguity. The complexity of the human world determines the complexity of the important means of expression of the human world - language. Based on L.A. Zadeh's paper "Fuzzy Sets" published in 1965, he introduced the concept of fuzzy sets and applied it to language research. In this paper, Zadeh argued that the ambiguity in language is inevitable because language itself is a fuzzy tool, often fraught with ambiguities and vagueness. His research suggests that incorporating fuzzy set theory into language studies can lead to a better understanding of linguistic ambiguity, and that linguistic fuzziness and uncertainty can be addressed using fuzzy set methods (Zadeh, 1965). In 1979, Professor Wu Tieping from the Chinese Department of Beijing Normal University introduced “Fuzzy Language Exploration”, marking the formal introduction of fuzzy linguistics in China. The article emphasizes the intrinsic factors that determine the ambiguity of language: language breaks the boundaries of objective things or concepts and uses the same word to express various sensations, which can greatly save language units (Wu, 1979).

Language is a highly complex, highly abstract, and infinitely derivational sound-meaning system. Ambiguity is the essential attribute of language (Clark, 2021). Translation, as a conversion activity of cross-language expression, is the practice of presenting the ideas, styles, and techniques

Quick Response Code	Access this article online	
	Website: www.actainformaticamalaysia.com	**DOI:** 10.26480/aim.02.2024.64.68

Cite The Article: Jianzhou Cui (2024). Design of Translation Ambiguity Elimination Method Based on Recurrent Neural Networks. *Acta Informatica Malaysia, 8(2): 64-68.*

in one language form as another language form. The ambiguity of language then leads to the uncertainty in the process of cross-language conversion, that is, the ambiguity of translation. Therefore, precision in translation can only be relative. Both the experts and scholars from China and abroad have long been concerned about the issue of "the ambiguity of translation". Robinson explores the inherent ambiguity in the act of translation, arguing that it is a fundamental challenge that translators grapple with. He discusses how ambiguity arises from linguistic, cultural, and contextual differences between languages, making it difficult to achieve a perfect translation. Robinson's paper provides insights into the complexities of translation and the strategies translators employ to navigate ambiguity while striving for accuracy and fidelity to the original text. (Robinson, 1979). Zhong proposed that "we should pay attention to the ambiguity in translation." (Zhong, 1991). Wu regarded "ambiguity" as an orientation for literary translation research, calling literary works "carriers of fuzzy information." (Wu, 1996). Zhang analyzed three semantic features of ambiguous language: graduality and both this and that, uncertainty and relativity, and their manifestations in translation: phonetic intonation ambiguity, lexical ambiguity, grammatical ambiguity, and pragmatic ambiguity (Zhang, 2008). Lu and Hou proposed that "in many cases, the ambiguous language in Chinese-English tourism texts is a higher level of precision, which can enhance the beauty of language and the readability of the text." (Lu and Hou, 2018).

In response to the ambiguity of translation, scholars have also continued to pay attention and put forward solutions. Zhao proposed methods such as "retaining vague information equivalently, omitting translation ambiguities, and transforming vague information into precise information." (Zhao, 1999). Wang proposed the translation strategies of "literal translation, free translation, and omission." (Wang, 2006). Wang analyzed the ambiguity issues in legal English using translation strategies such as equivalent translation, adding words translation, omitting words translation, and eliminating ambiguity (Wang, 2021). Tokowicz put forwards that the level of translation-ambiguity in the input and its integration into the cognitive mechanism is likely to join the competition that is already occurring among all the other candidates that are activated by the context (Tokowicz, 2023).

In order to deal with the ambiguity of translation, researchers in machine translation have successively developed models such as the concept tree model of context semantic mapping of the topic word table in machine translation, the fuzzy English machine translation semantic reordering model based on fuzzy theory, the fuzzy semantic selection technology in English semantic machine translation based on concept set context matching, the fuzzy semantic representation method of rare words in neural machine translation, and the English hierarchical machine translation model based on intelligent fuzzy decision tree algorithm. Pilault et al. curated a dataset showcasing various linguistic phenomena leading to ambiguities in inference across four languages and it is noteworthy that interactive-chain prompting, illustrated through eight interactions, consistently outperforms prompt-based methods that have direct access to background information for disambiguation (Pilault et al., 2023).

These models and methods have achieved certain research results in evaluating language semantic features, model construction, and feature combination matching in English translation, but their usual approach to eliminating translation ambiguity is through adjusting the fuzzy similarity threshold in the algorithm, which is faced with the practical contradiction that the fuzzy similarity threshold is directly proportional to the translation effect but inversely proportional to the translation speed. In response to this situation, it is particularly necessary to redesign a method to eliminate translation ambiguity in English language. Building on the advantages of traditional translation ambiguity handling methods, this study proposes a translation ambiguity elimination model based on recurrent neural networks. That is, in English translation, candidate sets are compared for similarity, generating source language and target language translations of the reference translation, and then nested in neural networks to achieve translation ambiguity elimination in language translation.

3. Design of Ambiguity Elimination Method for Machine Translation from the Perspective of Neural Networks

The neural network-based machine translation method directly adopts the neural network framework and builds the translation model in an "end-to-end" manner (Bahdanau et al., 2014). This translation method completes the translation at the sentence level of the source and target languages by using the "encoder-decoder" architecture. This is different from the traditional approach of optimizing statistical machine translation models using deep learning technology. Based on the neural network-based machine translation method, we first use an "encoder" to encode the source language sentence to a dense vector, and then use a "decoder" to decode the target language sentence from the vector mentioned above (Zhu, 2019).

3.1 Translation Fuzzy Retrieval

In the traditional machine translation process, the translation algorithm itself has certain limitations and cannot meet the actual needs of the translation process. In the specific field of English professional language translation, only when the vocabulary in the text reaches a certain repetition rate, can the translation memory be utilized. However, language itself can generate infinitely, so for new translation content, the coverage probability of the translation memory is relatively small. Therefore, this study needs to improve the coverage probability of the translation memory through translation fuzzy retrieval, to enhance the translation performance. To accomplish translation fuzzy retrieval, first, a candidate set needs to be determined based on the reverse indexing of words, followed by similarity calculation. The translation fuzzy retrieval framework designed in this study is as follows, as illustrated in Figure 1:

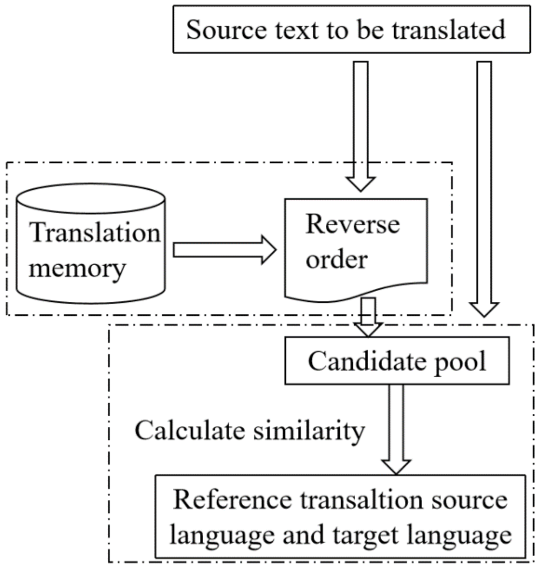

Figure 1: Translation framework for fuzzy retrieval.

In the realm of professional English translation, to ensure the effectiveness and quality of the translation, a sizable translation memory is required. As a result, directly computing the similarity between the target language and the vocabulary in the translation memory is computationally intensive and not very efficient. To address this issue, this paper proposes a reverse indexing structure for the translation memory to generate a candidate set. The reverse indexing structure is a type of data structure that facilitates the establishment of a "sentence-vocabulary" matrix. The structure of this matrix is as follows, detailed in Table 1:

Table 1: Matrix of "Sentences-Vocabulary".

	S1	S2	S3	S4	S5
Vocabulary1	V	-	-	V	-
Vocabulary2	-	V	V	-	-
Vocabulary3	-	-	-	V	-
Vocabulary4	V	-	-	-	V
Vocabulary5	-	V	-	-	-
Vocabulary6	-	-	V	-	-

From a sentence-based perspective, each column represents the vocabulary contained in the sentence. For example, in sentence 1, it includes vocabulary 1 and vocabulary 4, and does not contain any other vocabularies. This vertical dimension index is called a forward index. From a vocabulary-based perspective, each row represents the frequency of a vocabulary appearing in sentences. For example, vocabulary 1 appears in sentences 1 and 4, and this horizontal dimension index is called an inverted index. The generated inverted index is saved in a specified file, and during the query process, relevant information can be obtained by reading the file's storage path. The retrieval process is as follows, as shown in Figure 2:

Cite The Article: Jianzhou Cui (2024). Design of Translation Ambiguity Elimination Method Based on Recurrent Neural Networks. *Acta Informatica Malaysia, 8(2): 64-68.*

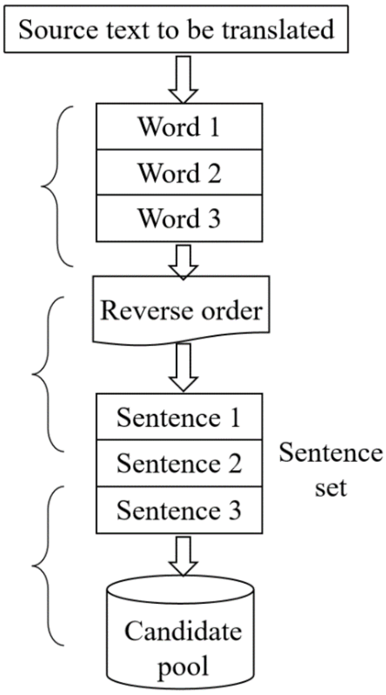

Figure 2: Reverse order file retrieval.

After conducting a reverse order file retrieval, a candidate set will be obtained. In order to determine the utility of a sentence for the intended translated source language, it is necessary to calculate the frequency of occurrence of each word in the sentence. The formula is as follows:

$$hit = \frac{2\times\sum_{w\in W} hit(w|S,S')}{length(S)+length(S')} \tag{1}$$

In the above expression, S represents the source sentence to be translated, S' represents all sentences in the candidate set, W represents all the vocabulary in the set S, w represents a certain word of set S, $length()$ represents the total number of words in the sentence. The main approach of this paper is to select the top 10 sentences with the highest frequency in the sentences set as candidates. Thus, the fuzzy translation retrieval is completed.

3.2 Establishing Recurrent Neural Networks

The core concept behind fuzzy translation lies in the utilization of recurrent neural networks to achieve the transformation of fuzzy translation. It heavily relies on deep learning through recurrent neural networks to extract natural language feature vectors. Neural networks are composed of numerous neurons, and the structure of artificial recurrent neural networks is quite analogous to biological neurons. When a neuron is stimulated by an external input, the artificial recurrent neural network abstracts the response to the stimulation into mathematical language. The structure of a neuron when stimulated is illustrated in Figure 3.

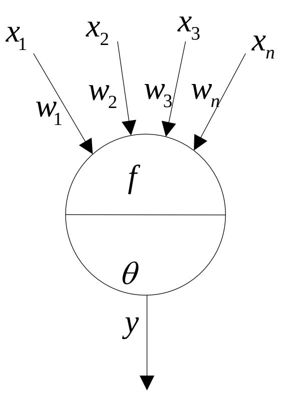

Figure 3: Illustration of neuronal stimulation from the external environment.

In the Figure 3 which is shown above, $X = (x_1, x_2 \ldots, x_n)$ represents the external stimuli received by the neuron; $W = (w_1, w_2 \ldots, w_n)$ denotes the weights corresponding to the external inputs; θ indicating the threshold within the neural unit, a value that can be manually adjusted, and y illustrates the output of the neuron. These relationships can be symbolized as:

$$y = f(\textstyle\sum_{i=1}^{n} w_i x_i - \theta)x_1 \tag{2}$$

In the artificial recurrent neural network structure composed of neurons, there are three main layers: the input layer, the hidden layer, and the output layer. In traditional machine translation processes, sentences are broken down into individual words and phrases for translation, but this method can lead to a loss of semantic coherence in the context. With the establishment of a recurrent neural network, the source sentence can be encoded, processed to obtain a fixed-length sentence vector, and then, through the application of specific algorithms and constraints, decoded with the assistance of the neural network to eliminate ambiguity in translation. The recurrent neural network is then trained as illustrated in Figure 4.

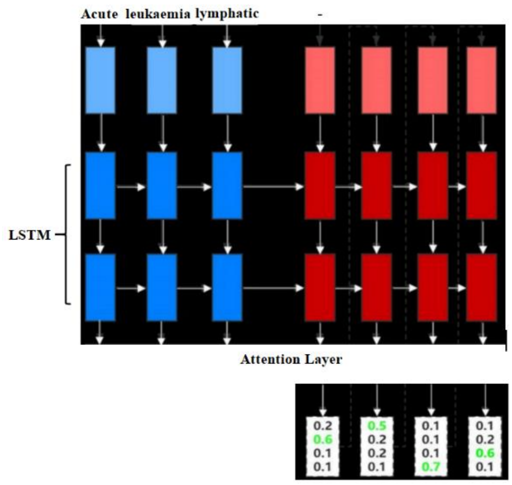

Figure 4: Illustrative diagram of neural network-based translation model training.

In the process of translation model training as shown in the diagram above, it is possible to effectively reduce the engineering design involved in translation work and encode and decode through the encoder and decoder in the recurrent neural network. The encoder can process word vectors and hidden units in the upper layer to obtain hidden units at a certain moment through the following formula:

$$h_t = f(x_t, h_{t-1}) \tag{3}$$

In the above equation, h_t represents the hidden units, f represents the non-linear activation function, t denotes a moment in time, and x_t symbolizes the input sequence vector of t at that moment. With the increase in sentence length and number of iterations, the constructed artificial recurrent neural network can address the deficiencies present in the original methods, thereby eliminating the issue of “gradient imbalance” during the translation process. This marks the completion of the study on eliminating translation ambiguity in the field of English language expertise.

4. Test

In the experimental performance testing experiments of this paper, the experimental dataset selected is the IWSLT2018 machine translation evaluation dataset. This dataset mainly includes multiple machine translation evaluation tasks, with the chosen ones being specialized oral tasks and dialogue tasks. In the specialized oral tasks, BTEC has a total of 15,426 sentence pairs, while the dialogue task DIALOG has 18,645 sentence pairs. In the method proposed to eliminate translation ambiguity in English specialized language translation in this paper, it is mainly divided into translation memory fuzzy retrieval and fuzzy elimination improvement statistics. In the dataset selected in this chapter, there are high-resource parallel corpora and low-resource parallel corpora, with differences in the way these two types of corpora are input into neural network machine translation, requiring the setting of the quantity of these two types of corpora. The quantity setting should satisfy the following formula:

$$ratio = \frac{h_{len}}{l_{len}} \tag{4}$$

In the above equation, h_{len} represents the length of input for high-resource parallel corpora, while l_{len} represents the length of input for low-resource parallel corpora. By analyzing the ratio of processing lengths for these two types of corpora, the appropriate amount of input data can be determined to ensure both corpora are trained simultaneously in the neural network. The experimental environment and parameters for this study are outlined below, as detailed in Table 2.

Table 2: Experimental environment parameters.

No.	Name	Parameter
1	Encoder	Bidirectional LSTM
2	Operating System	Ubuntu 16.04.6
3	Decoder	Unidirectional LSTM
4	Number of Neurons	451
5	GPU	Tesla V 100*4
6	Batch Length of Corpus	128
7	Python	Version 3.6
8	Maximum Sentence Length	40
9	Pytorch	Version 1.46
10	Django	Version 1.11.5
11	Word Embedding Dimensionality	512

This text sets the initial learning rate of the fuzzy neural network to 1, based on the open-source deep learning framework Tensorflow 1.4.0. By correctly setting the experimental parameters, one can effectively ensure that the neural network performs well. In the fuzzy retrieval module, the neural network retrieves from the reverse-ordered index, selecting the top ten translations with the highest hit count as reference candidates in the translation process. In the translation fuzzy elimination method designed in this text, the preprocessing stage mainly uses the preprocessing tools mentioned earlier, with the length limit set between 3 and 100 during length filtering. A length ratio of 0.4 is achieved by using GIZA++ for word alignment, and the results are symmetricized bidirectionally using grow-diag-final-and.

The performance of traditional translation fuzzy elimination methods varies based on similarity thresholds. The method designed in this text is based on neural networks and involves deep learning and development. Therefore, it is necessary to select the final translation memory units from the neural network and the fuzzy translation memory bank as references. The translation results are shown below, as detailed in Figure 5:

```
INPUT : 这些 代码 说明 了 什么?
SMT : Thsse cedes what that means?
FUZZY TM : What do the code say?
Ref Source : 这些 数据 说明 了 什么?
Ref Target : What do the numbers say?

INPUT : 把 我 当 一个 正常人 来 看待，
SMT : Give me when a person.
FUZZY TM : Think of me as a person.
Ref Source : 把 她 当 一个 正常人 来 看待
Ref Target : THink of her as a person.

INPUT : 我们 正在 寻找 烟灰缸，这 需要 时间.
SMT : We’ re looking for an ashtray, it needs
time.
FUZZY TM : We are looking for an ashtrayr, and
it takes time.
Ref Source : 我们 正在 立案，这 需要 钱.
Ref Target : We are baailding a case and It
takes money,
```

Figure 5: Example of translated results.

In light of the translation results provided above, this article incorporates various levels of fuzzy interval similarity thresholds in the experimental design, conducting performance comparisons accordingly. This approach aims to validate the efficacy of the proposed method in enhancing the performance of translation systems.

5. Discussion

This study compared the performance of two methods by setting different similarity thresholds for fuzzy interval ranges. The experimental results are presented below, as detailed in Table 3.

Table 3: Performance comparison under different thresholds of fuzzy interval similarity.

Fuzzy Interval Similarity Threshold	Traditional Method	Proposed Method
(0.0,0.3)	4.85	28.64
[0.3,0.4)	7.56	34.45
[0.4,0.5)	14.54	40.65
[0.5,0.6)	21.58	44.39
[0.6,0.7)	38.564	50.34
[0.7,0.8)	51.33	61.78
[0.8,0.9)	62.65	70.56
[0.9,1.0)	68.35	75.69
(0.0,1.0)	28.31	46.33

The experimental results from the Table 3 indicate that, whether using our proposed method or traditional approaches, as the similarity threshold in the fuzzy interval increases, the similarity between the test set data and training set data also increases, approaching the correct answer. In the translation fuzzy elimination method designed in this paper, during the initial phase of translation fuzzy retrieval, the output translation units only have a significant impact on the final translation result when the similarity threshold between the pending translation dialogues and reference translation segments is the closest to the threshold of similarity.

Analysis of the results from the table reveals that in the retrieval process, when the output units are only identical to the pending sentence for translation, traditional methods include translation units in the machine's translation process but do not output them as the final translation result, hence resulting in slightly lower performance in various fuzzy intervals. The data in the table shows that when the fuzzy interval is below 0.5, the performance of the translation fuzzy elimination method proposed in this paper is significantly better than the traditional methods. As the similarity threshold increases, the performance gap between the two methods gradually decreases. However, it is important to note that a high similarity threshold in practical applications may lead to retrieval difficulties, affecting the progress of translation.

6. Conclusion

In conclusion, with the continuous advancement of artificial neural network deep learning, machine translation has become widely popular. The quantity of translation datasets has seen a significant increase, from general translation to specialized translation in the English language field. Translation ambiguity elimination during the translation process is a commonly used technique in machine translation. Traditional methods perform poorly at low similarity thresholds. However, setting the similarity threshold too high in practical applications can lead to difficulties in retrieval, impacting the progress of translation. This study proposes a translation ambiguity elimination method based on recurrent neural networks, which can effectively improve translation results. Nevertheless, it requires abundant translation resources and falls short in applications with scarce resources. Future research is expected to make more breakthroughs in this area.

References

Bahdanau, D., Cho, K., and Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. In Proceedings of the International Conference on Learning Representations (ICLR) (Accepted as Oral Presentation).

Clark, H.H., 2021. Language and Ambiguity: Exploring the Dynamics of Meaning. Cambridge, MA: Harvard University Press.

Engels. 1995. Selected Works of Marx and Engels Volume IV. Beijing: People's Publishing House.

Lu, R., and Hou, B., 2018. A study on language ambiguity and translation strategies in Chinese-English tourist texts. Journal of Tianshui Normal University, 38 (02), Pp. 112-117.

Pilault, J., Garcia, X., Bražinskas, A., 2023. Interactive-chain-prompting: Ambiguity resolution for crosslingual conditional generation with interaction. arxiv preprint arxiv: 2301.10309.

Robinson, D., 1979. The Ambiguity of Translation. Translation Review, 6 (1), Pp. 2-9.

Tokowicz, N., 2023. The translation ambiguity disadvantage in language processing. In Translation in Transition: Human and machine intelligence, pp. 183.

Wang, J., 2021. On the characteristics of ambiguity in legal English and Chinese translation strategies. Foreign Language Education and Application, (00), Pp. 308-313.

Wang, L., 2006. The ambiguity of language and translation (Doctoral dissertation). Shanghai Maritime University.

Wen, X., 1996. Semantic ambiguity and translation. Chinese Translation, 1996 (02), Pp. 6-9.

Wu, T., 1979. An initial exploration of fuzzy language. Foreign Languages (Journal of Shanghai International Studies University), (04), Pp. 41-46.

Wu, Y., 1996. Ambiguity: a perspective in the study of literary translation. Foreign Languages (Journal of Shanghai International Studies University), (05), Pp. 18-23.

Zadeh, L.A., 1965. Fuzzy Sets. Information and Control, 8 (3), Pp. 338-353.

Zhang, Z., 2008. Language ambiguity and translation. Chizhou University Journal, (04), Pp. 76-78.

Zhao, J., 1999. On the fuzzy information and processing techniques in economic and trade translation. Chinese Science and Technology Translation, (02), Pp. 5-7+23.

Zhong, S.N., 1991. On Ambiguity in Translation. Fujian Foreign Languages, (Z1), Pp. 67-70.

Zhu, S., 2019. A Chinese-Uyghur neural network machine translation model integrating multiple features. Computer Engineering and Design, 40 (05), Pp. 1484-1488+1501.