Hindawi
Scientific Programming
Volume 2022, Article ID 5076637, 9 pages
https://doi.org/10.1155/2022/5076637

Research Article

Research on Cross-cultural Text Reconstruction of Urban Publicity Translation Based on Computer Corpus

Zhenli Li [1] and Jian Tang [2]

[1] *School of Foreign Languages, Fuyang Normal University, Fuyang 236037, Anhui, China*
[2] *School of Mathematics and Statistics, Fuyang Normal University, Fuyang 236037, Anhui, China*

Correspondence should be addressed to Zhenli Li; zhenlili@fynu.edu.cn

Received 8 December 2021; Revised 26 December 2021; Accepted 28 December 2021; Published 2 February 2022

Academic Editor: Tongguang Ni

Urban publicity translation, as a cross-cultural communication activity, should aim for communication, employ various translation strategies, adapt to the target language's expression habits, overcome cultural differences, and make the translation easy to accept for target readers. In order to achieve the goal of external promotion, publicity texts should respect and conform to the target culture's language expression as well as the psychology of the audience during the initial stage of urban publicity translation. This paper analyzes the causes of cultural vacancies in the translation of urban publicity materials, starting with the classification and sorting of cultural vacancies in the translation of publicity materials. This paper focuses on using a computer corpus to reconstruct cross-cultural text for urban publicity translation. An automatic corpus expansion method combined with the EM (expectation-maximization) algorithm is proposed to solve this problem. The model is iteratively trained after the generated single corpus is combined with the original data set to create a parallel corpus. Finally, as another important feature of words, the word cooccurrence degree is incorporated into the interword relationship extraction model to create a new word translation evaluation index. Finally, the experiment demonstrates that the EIWR (extraction of interword relations) has higher accuracy than the VSM (vector space model).

1. Introduction

The external image of a city is a comprehensive material and cultural impression that the city shows to the public, and it is an important part of the city's competitiveness. With the continuous improvement of China's international influence and economic strength, China has closer ties with other countries. As an important means of foreign communication, the translation of city publicity is an important means to show the traditional culture, regional features, and development characteristics of the region to the international community, and plays a vital role in improving the image of the city, enhancing the international reputation and strengthening the opening to the outside world [1]. In recent years, scholars at home and abroad have increasingly studied the translation of urban publicity, and the translation of urban publicity has made great progress.

Corpus linguistics is a new interdisciplinary field that combines linguistics, computer science, applied linguistics, and cognitive linguistics. It is still in its early stages of development. Corpus linguistics uses actual language facts as the research object and performs macroscopic and microscopic, qualitative and quantitative statistics and analysis on a large number of corpora using computer tools, revealing the objective laws of language use and the complexity of natural language [2, 3]. The parallel corpus has strong text alignment and high translation accuracy because it is made up of source language texts and translated texts that correspond to the source language texts. However, the parallel corpus's construction costs are high, and parallel corpus resources are scarce and difficult to come by, making it difficult to cover all fields of research. At the same time, the artificial translation quality has a significant impact on translation accuracy [4]. China is a vast country with many ethnic groups, and each region has its own distinct local characteristics and national culture. Chinese cities want a place on the international stage, and language texts are particularly important when they show their own culture and characteristics to foreign audiences [5, 6].

Many different types of corpora have been built and used in the past, depending on the purpose and nature of the corpus. A Chinese-English parallel corpus, as well as an analogy corpus [7], were used in this study. This paper proposes a model training method combined with the EM(expectation-maximization) algorithm to solve the problem of cross-cultural text reconstruction in urban publicity translation. The joint EM optimization method is used to learn translation models from source language to target language and from target language to source language and to complete the bilingual dictionary extraction based on the word relation matrix. Finally, experiments are used to verify the feasibility of this extraction method, which is compared to VSM (vector space model), and the impact of variables such as context window size, corpus size, dictionary size, and word frequency on the final experimental results of the two models is examined.

2. Related Work

The translation is a medium form of cross-cultural communication, and publicity is a cross-language and cross-cultural information exchange and communication with nations, regions, and countries as the main body. The essence and basic task of urban publicity translation are cross-language and cross-cultural information dissemination, which is an important way of external communication [8, 9]. If a city wants to go global, it must first create an international environment conducive to city publicity, which will be recognized and supported by the international community. Literature [10] clearly puts forward the stage theory of urban publicity translation, that is, in the first stage, due to the dominant target language culture, the purpose of publicity is to win the recognition and understanding of the international community, mainly adopting the translation strategy based on the target language; Literature [11] points out that the differences between Chinese and western cultures directly lead to the great differences in vocabulary system and meaning expression between Chinese and English. Because the audience of urban publicity translation is foreign readers, and there are many differences between Chinese and English languages and cultures, it is difficult to achieve the expected publicity effect if mechanical literal translation is adopted. Literature [12, 13] proposed a method of dependency constraint on the target language part of translation knowledge, which effectively improved the translation accuracy. Literature [14] deals with translation knowledge in a semistructured way and proposes an example-based machine translation method. In their experiments, the effect of this method is significantly better than that of a statistical machine translation system. Translation knowledge automatically acquired from corpus usually contains a lot of noise, which affects the translation process. Literature [15] proposes a method to filter monotone combination phrase pairs and a method to filter combination phrase pairs by using a logarithmic linear model.

The subject of translation is the translator, and the translation theory of urban publicity places higher demands on translators. Not only linguistic equivalence but also cultural equivalence and communication equivalence should be considered when translating publicity texts. With the help of the concept of intermediate language, literature [16] proposes a many-to-one translation mechanism. These methods solve the problem of data sparsity, but they also complicate the model and increase the training costs. By introducing tags into the input, literature [17] proposes a method of training translation models. Despite improvements in training efficiency and translation performance, data scarcity remains a problem. Literature [18] proposes a multilingual translation model with an incremental self-learning strategy, which solves the problem of data scarcity by generating pseudobilingual data automatically, but the pseudobilingual data may have noise issues, lowering translation quality. The cooccurrence rule of words in the target language in both parallel and comparable corpora is essentially the same as in the source language, according to the literature [19]. The calculation assumes a one-to-one correspondence between the words in the source language and the target language text. According to the literature [20], a third-party intermediate language could be used to complete the construction of a bilingual dictionary. Its basic concept is to use a common intermediate language, such as English, and a multilingual vocabulary to first translate source-language words into intermediate language words, and then to translate the translated intermediate language words into target language words using the vocabulary to complete the construction of a bilingual dictionary. According to the literature [21] there is a random translation matrix from the source language to the target language, which can translate the source language to the target language, assuming a one-to-many relationship in translation. The local ambiguity problem is solved by this method, but the global ambiguity problem is not. Literature [22] proposes a dependency tree-based method for completing the construction of a context vector and extracting bilingual dictionaries. Literature [23] proposed a method for extracting parallel resources from comparable corpora based on document-level alignment. The basic idea is to use alignment information instead of word context information.

3. Research Method

3.1. Selection and Design of Translation Model for Urban Publicity. The translation of urban publicity is aimed at foreign readers, so that foreign readers can understand and accept the information conveyed by China. In the process of urban publicity translation, the common problems are that translators lack awareness of cross-cultural communication, do not know enough about the thinking patterns of foreign audiences, and cannot translate according to the thinking habits of target audiences. To solve this problem, literature [24] puts forward the principle of "three closeness" in the translation of urban publicity, that is, "the translation of urban publicity should be close to the reality of China's development, the needs of foreign audiences for Chinese information and the thinking habits of foreign audiences".

To maximize the cross-cultural communication effect of urban publicity translation, it should follow the principle of "communication priority" and begin with the foreign audience to truly understand and accept cross-cultural information. In this paper, powerful models are not required for initialization training, but there are two popular models: statistical machine translation and neural machine translation. Because the neural machine translation model is prone to overfitting when the data is sparse, it performs worse in low-resource situations than the statistical machine translation model. The transformer model [25], whose structural design is shown in Figure 1 was chosen among many neural machine translation models for this study.

In this paper, two transformer models in opposite directions are initialized and pretrained, that is, one-way tasks from the source language to the target language and one-way tasks from the target language to the source language.

The pretraining process is completed by the traditional method based on the maximum likelihood principle. The general approach of this method is to maximize the logarithmic conditional probability of correct translation, give the model parameter θ of the source language sentence, and its goal is to get θ^* which satisfies the following formula:

$$\theta^* = \arg\max_{\theta} \sum_{i=1}^{N} \sum_{j=1}^{|t_i|} \log\ p\left(t_{i,j} \mid t_{i,0\longrightarrow j-1}, s_i\right), \tag{1}$$

where N is the scale of training corpus, $|t_i|$ is the length of the target language sentence t_i.

After pretraining, the iterative training process of the model is carried out. Given the original parallel corpus (real corpus) $D = \{S_i, T_i\}_{i=1}^{N}$ and the target language monolingual corpus $T' = \{t_i\}_{i=1}^{P}$, the training goal of the model is to maximize the possibility of bilingual data and monolingual data, namely

$$L^*\left(\theta_{S\longrightarrow T}\right) = \sum_{i=1}^{N} \log p\left(t_i|s_i\right) + \sum_{i=1}^{P} \log p\left(t_i\right), \tag{2}$$

where the first part represents the conditional probability of generating the target language T for the source language S; the second part represents the language model of T, whose main function is to maximize the possibility of sentences.

The degree of vocabulary variation is defined as the number of different words in a given length corpus that reflect the diversity of vocabulary. The number of different words is referred to as the shape symbol number, whereas the number of different word shapes is referred to as the class symbol number. The amount of information and difficulty of a text can be reflected in lexical density. The higher the vocabulary density, the more meaningful words there are, and the more information and difficulty the text contains. The lower the vocabulary density, the less information the text contains and the easier it is to understand. The purpose of calculating instance similarity is to determine the degree of similarity between the source language phrase of the translation instance pair and the input sentence phrase fragment. The expression form of the translation instance is closely related to the similarity judgment. Language surface matching is a type of character-based matching. Two parts to be matched are treated as strings in this method. In general, it only takes into account the length of common substrings in two strings, and the common substrings must be ordered.

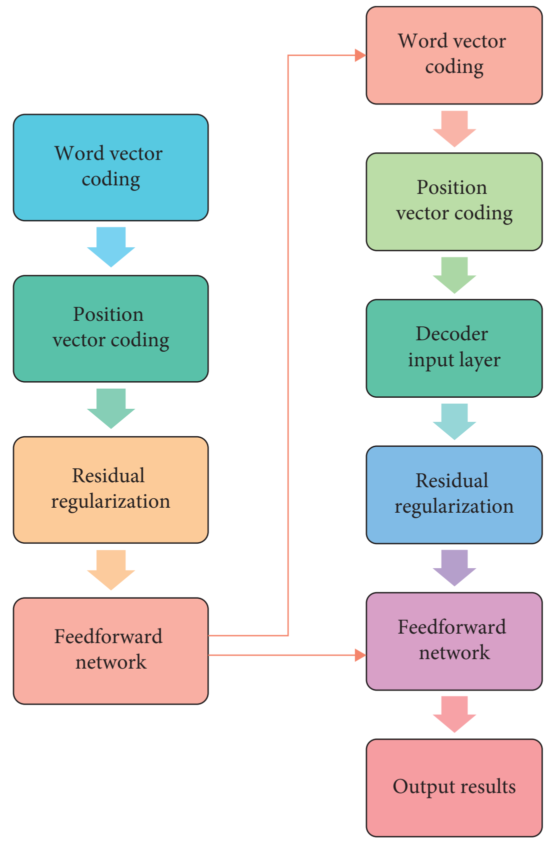

Figure 1: Transformer model architecture diagram.

Among the algorithms based on character matching, the method of editing distance is the most widely used one. This method can be used to calculate the minimum cost $D(m, n)$ of transforming from one string $S = s_1, s_2, \ldots, s_m$ to another string $T = t_1, t_2, \ldots, t_m$.

The algorithm is based on a dynamic programming algorithm, and its complexity is $O(m * n)$. The following are the core parts of the algorithm:

$$D(i,j) = \begin{cases} 0, & \text{if } i = j = 0, \\ D(0, j-1) + w\left(t_j\right), & \text{if } \ i = 0, j! = 0, \\ D(i-1, 0) + w\left(s_i\right), & \text{if } \ i! = 0, j = 0, \\ \text{Min}\begin{bmatrix} D(i-1, j) + w\left(s_i\right) \\ D(i, j-1) + w\left(t_j\right) \\ M(i, j) \end{bmatrix}, & \text{other}, \end{cases}$$

$$M(i,j) = \begin{cases} \text{Max}\left(w\left(s_i\right), w\left(t_j\right)\right), & s_i \longrightarrow t_j, \\ 0, & s_i = t_j. \end{cases} \tag{3}$$

Among them, the three transformation operations are replacement, insertion, and deletion. $w(t_j)$ is the cost of inserting t_j , $w(s_i)$ represents the cost of deleting s_i in S, and $\mathrm{Max}(w(s_i), w(t_j))$ represents the cost of replacing s_i with t_j.

In word-based matching, matching is investigated in terms of words. Because the semantics of words should be considered, semantic dictionaries play a very important role. Commonly used semantic dictionaries include WordNet and synonym word forest. In English, apart from semantic dictionaries, word-building analysis is usually required.

At the same time, in order to reduce the workload of the algorithm, a stoplist is often used to filter out these words. Then, judge the similarity between sentences according to the remaining meaningful words. There are many kinds of VSM, among which the most well-known one is $tf-idf$ value.

3.2. Cross-Cultural Text Reconstruction of Urban Publicity Translation. The ideal state of translation is to be able to translate word for word along with the original text. However, in the translation of urban publicity, this situation is very rare. Because English is hypotaxis structure and Chinese is parataxis structure, the two languages are completely different in writing structure and writing habits and are difficult to be compatible at the syntactic level. Therefore, in Chinese-English translation, in order to accurately express the meaning of Chinese and achieve good translation effect, the method of "reorganizing the original text" can be adopted, reasonably change sentence patterns or adjust word order in some parts of the original text, and if necessary, can "reinvent the stove" and reintegrate to reduce the influence of Chinese grammar and sentence patterns, so as to make the translation richer in English charm.

Translators must explore the cultural differences between English and Chinese and cultivate a cross-cultural perspective when working on translation projects. Translators must be culturally aware, observe idiomatic expressions and thinking differences in English, use English expression habits and thinking styles, and cannot create words out of thin air or break the horizontal combination relationship between words and words. They should take responsibility for themselves, their culture, and the target language users in translation, as well as contribute their own efforts to help foreign friends and export China's excellent culture. In general, a single word cannot express a complete theme, it must be combined with other words to do so. Different words are combined to express various theme contents. This word combination reflects the semantic information of words as well as the correlation between them. Because only nouns, verbs, adjectives, and adverbs in comparable corpora are studied in this paper, the corpus must be preprocessed.

In this paper, based on the selection rules of seed word pairs, the seed dictionary is extracted from the general dictionary, but the uniqueness of words in the seed dictionary needs to be ensured in the process of mapping relations between words, so this paper adopts certain selection rules to extract seed word pairs from the general dictionary, among which the selection rules of the seed dictionary are shown in Figure 2.

In this paper, the words with low similarity in the source language are selected first to ensure the differentiation of the correlation between words in the seed dictionary, among which is the formula for calculating the similarity of words in the source language.

$$S\left(W_{s_i}, W_{s_j}\right) = \sum_{1<k<m} V_{s_{ik}} \times V_{s_{jk}}, W_{s_i}, W_{s_j} \in \text{Set}. \tag{4}$$

In the formula, V_s represents the word vector of the word, k is the component of the k-th dimension in the word vector, and m is the dimension value of the VSM trained word by the source language corpus.

If the frequency of the word W_{t_i} in the target language corpus is too low, the relevance with the target words will be lower, which can not reflect the relevance between the words in the target language. If the frequency of W_{t_i} is too high, it may be related to the words in the whole target language corpus, which can not achieve the purpose of taking the relationship between words as an important distinguishing feature of words.

Therefore, if the word frequencies are the same, the word pairs with the smallest index value are selected according to the index values in the dictionary, and if the number of selected word pairs can not meet the calculation requirements, then the word pairs are selected from low frequency to high frequency.

In this paper, the correlation between a word and other words is regarded as an important distinguishing feature of the word, and the correlation is quantified by the similarity of word vectors. The construction of the interword relationship matrix is completed by the seed dictionary, and at the same time, the correlation between the interword relationship VSM of the source language and the interword relationship VSM of the target language is also completed. The specific steps are shown in Figure 3.

Firstly, the rules are extracted from the general dictionary through the seed dictionary, and the seed dictionary is extracted. The number of seed word pairs is N, and the seed set formed is expressed as $\left\{W_{s_i}, W_{t_i}\right\}, i \in \{1, 2, \ldots, N\}$, W_s is the source language word, W_t is the translation word corresponding to W_s in the target language, and i is the index value of W_s in the seed dictionary.

Then, the correlation degree between each word in the source language and the words in the source language in the seed dictionary is constructed by the word vector. For the quantification of the correlation between the unknown word W_{s_x} and each word in the seed word set, this paper adopts the calculation method of vector inner product

$$M\left(V_{s_x}, V_{s_i}\right) = \sum_{1<j<m} V_{s_{xj}} \times V_{s_{ij}}, i \in \{1, 2, \ldots, k\}. \tag{5}$$

In the formula $V_{s_i} \in \left\{V_{s_1}, V_{s_2}, \ldots, V_{s_k}\right\}$, $V_{s_{xj}}, V_{s_{ij}}$, j represents the component of the j-th dimension of the word vector V_{s_x}, V_{s_i}.

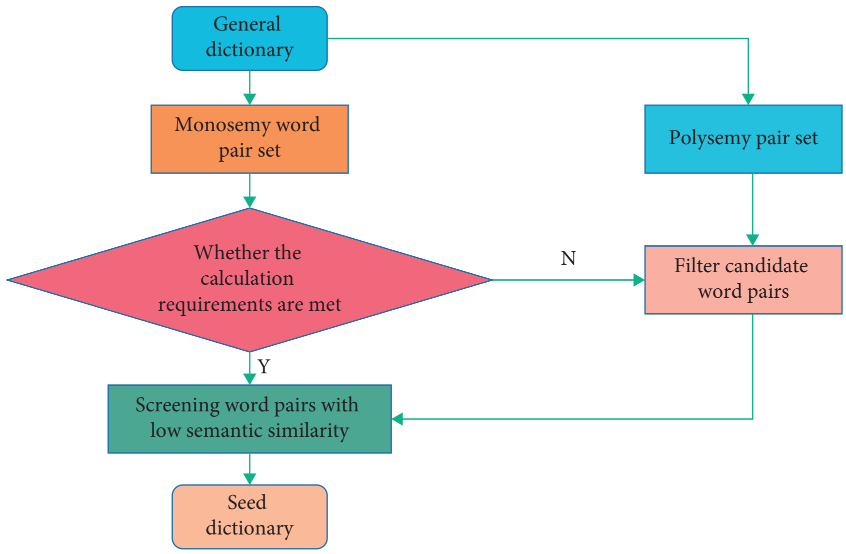

FIGURE 2: Seed dictionary extraction rules.

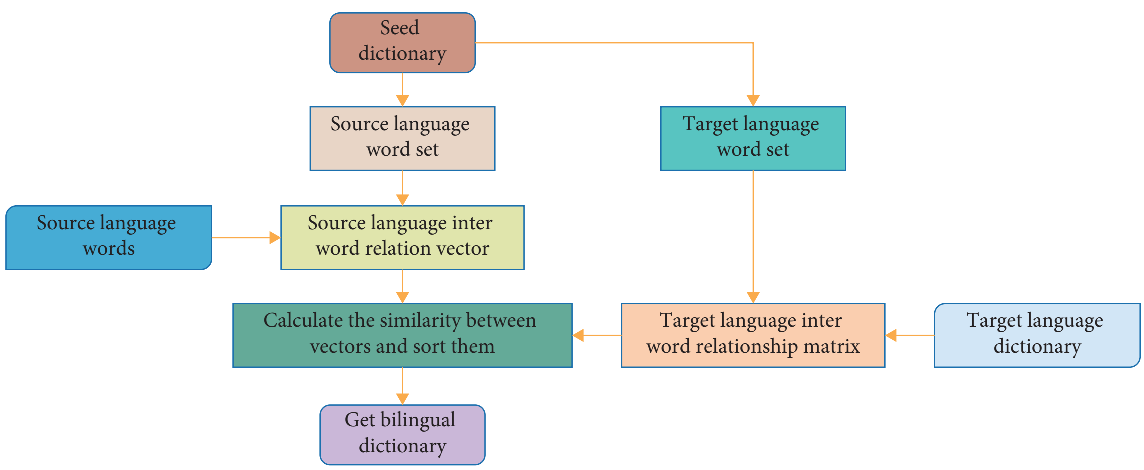

FIGURE 3: Extracting bilingual dictionaries based on the relationship between words.

Finally, the cosine similarity is used to calculate the relation vector between two words, which can be expressed as

$$S\left(V_{m_s}, V_{m_t}\right) = \frac{\sum_{1<i<k} V_{m_{si}} \times V_{m_{ti}}}{\sqrt{\sum_{1<i<k} V_{m_{si}}^2 \times V_{m_{ti}}^2}}. \tag{6}$$

The translation appears to be a process of changing one linguistic sign to another, but in reality, it is a cross-language and cross-cultural communication activity. As a result, translation is more than just a language conversion; it is also a social and cultural transformation process. Different scholars advocate for various translation standards and methods in the translation process, and various translation methods and standards have reached a theoretical level. The original text has more nouns and adjectives than the translated text, but the latter has slightly fewer verbs and adverbs. The difference in vocabulary density between the two subdatabases can be attributed to a large number of nouns and adjectives in the original text.

4. Results Analysis and Discussion

4.1. Comparison of Parameters and Results. In this paper, the transformer model is used as the baseline model of translation training. After several rounds of parameter combination optimization, the relevant important parameters of the transformer model are set in the experiment to achieve the best performance of the model.

In order to explore the corpus generation of the new method in this paper in more detail, according to the convergence of the iterative model of EM algorithm, Figure 4 shows the curve of the performance of the model changing with each iteration process from different translation tasks.

As can be seen from Figure 4, when the EM algorithm is iterated for 10 times, it tends to converge, and the performance of the model is hard to be improved, which also proves that the method in this paper makes the model tend to be stable after a certain iteration process.

Different versions are born as a result of the subjectivity of the translators. The translator's cultural mentality is full of contradictions, both submissive and treacherous, in the translated text. As target readers, it may be difficult to imagine the translator's confusion and cultural references in the translated text. The cultural oddity in the translation demonstrates that the target readers generally lack the background knowledge necessary to appreciate the source culture, and the translator must perform some cultural filtering in order for the translation to be accepted. Figure 5 depicts the relationship between the confusion of generating corpus and the EM algorithm iteration times in order to see the benefits and drawbacks of the transformer model's corpus as the EM algorithm iterates.

As can be seen from Figure 5, after 10 iterations, the confusion of the generated corpus will not decrease significantly, which is consistent with the convergence of the model and avoids the possibility of the model falling into local optimum.

The phenomenon and activity of translation is a cross-ethnic cultural phenomenon and activity. Cultures of different countries are at different stages of development, with different characteristics and cultural positions, due to the imbalance and asymmetry of cultural development. When universal things are localized, they frequently face local opposition, and when local things are globalized, they are often misunderstood or suppressed. More importantly, when different cultures collide and clash, it is difficult to come to an agreement on values, behavior patterns, problem-solving methods, and procedures, and cultural friction or conflict can quickly spiral out of hand, escalating into difficult-to-adjust disputes.

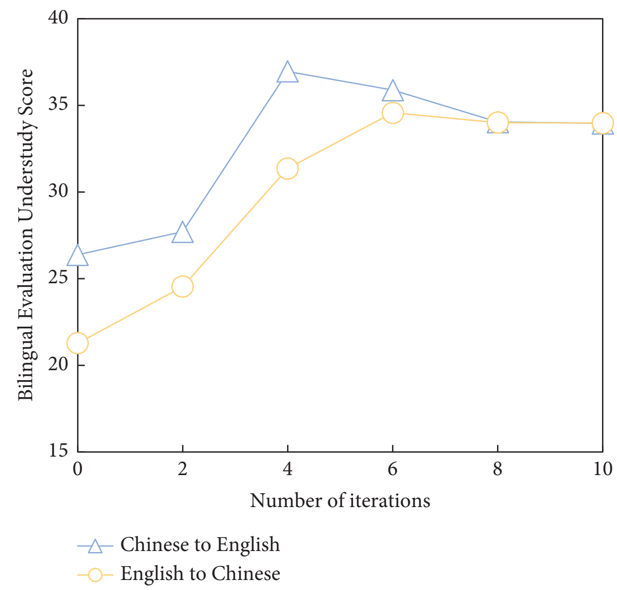

FIGURE 4: Variation curve of model performance.

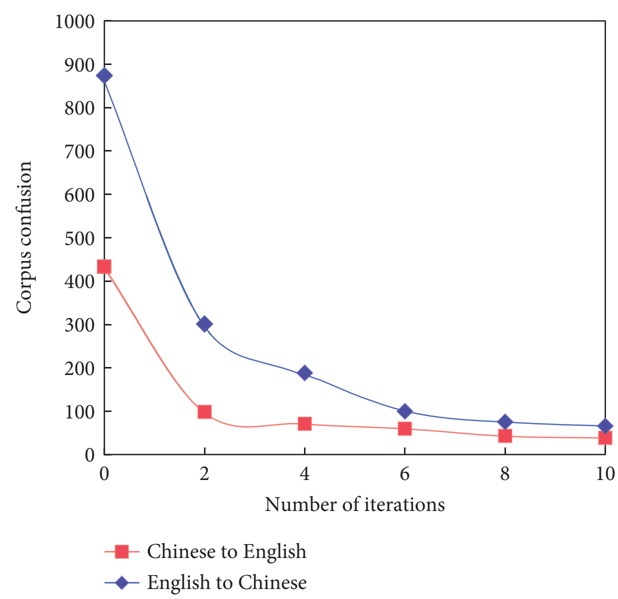

FIGURE 5: Curve of the confusion degree of corpus generated by the model with the number of iterations.

4.2. Corpus-Based Dictionary Extraction Analysis. The translation of urban publicity is like a fish swimming between the ponds of two languages, and this state just provides the prerequisite for cultural hybridity, internalization, transformation, and reconstruction. The essence of translation is to translate meaning. However, the symbols of cultural meaning contained in the source text are the most difficult to control, copy, and reconstruct in the process of translation. Many texts, reluctantly translated, cannot make the target readers feel similar to the source readers. Cultural or aesthetic failure shows that such translation of city publicity is a failure.

The corpus of this experiment is English and Chinese. English-Chinese bilingual dictionaries are extracted with English as the source language and Chinese as the target language. Dictionaries are another important resource of this experiment, including English dictionary, Chinese dictionary, English-Chinese bilingual dictionary, English stoplist, and Chinese stoplist. In the experiment, 10% of words in the English dictionary were randomly selected as the test set, and the corresponding translations of English words were obtained from the English-Chinese bilingual dictionary as the verification set to calculate the accuracy of bilingual dictionary extraction.

Accuracy is the most direct index to evaluate a model. Firstly, this paper compares the overall accuracy of VSM and EIWR (extraction of interword relations), in which the context window size is set to 10, and the experimental results are shown in Figure 6.

From Figure 6, it can be seen that it is feasible to quantify the relationship between words by using word vectors in a comparable corpus and to apply the relationship between words as the extraction feature of words in information extraction.

Whether for VSM or EIWR, the window will affect the expression of word context to a certain extent. For VSM, if the window is too small, the context semantics of the current word cannot be accurately expressed. Therefore, this paper takes the window as a parameter and takes the value of n in P@N as 20 to study the influence of the window on the final

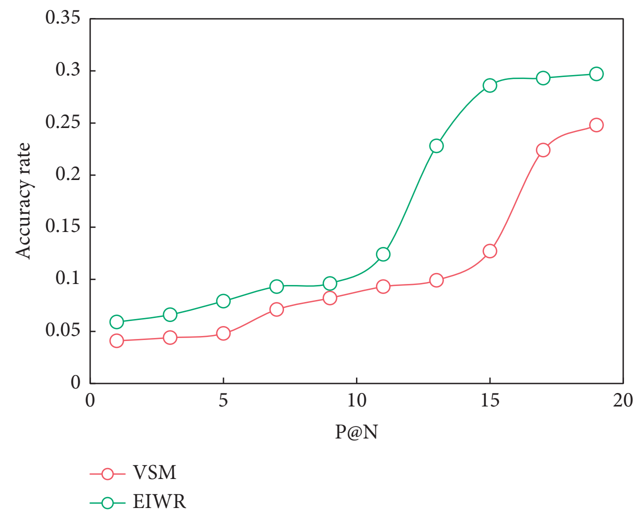

Figure 6: Model accuracy.

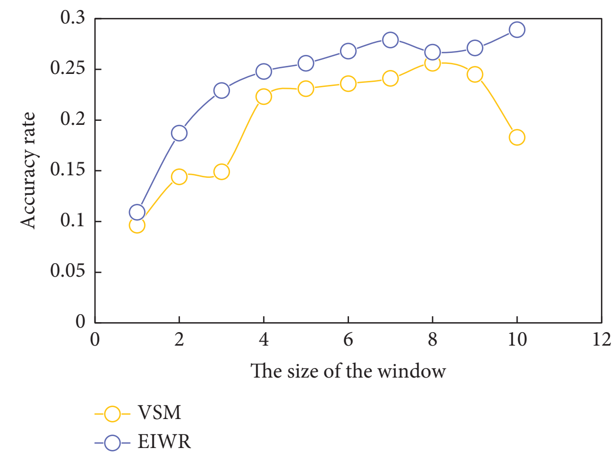

Figure 7: Influence of window on the accuracy of two models.

extraction effect of the two models. The experimental results are shown in Figure 7.

From Figure 7, it can be seen that the window size has a certain influence on the experimental results of the two models. Especially for VSM, the window size has a particularly obvious influence on the final extraction results. At the beginning of the experiment, the accuracy of VSM is positively correlated with the window size, and when the window size exceeds 10, the accuracy is negatively correlated with the window size, which also verifies the hypothesis that VSM introduces more useless information with the increase of window size.

Experiments show that choosing different sizes of windows in VSM or VSM will affect the final extraction effect, and choosing the correct window value will help to improve the accuracy of bilingual dictionary extraction.

In VSM, the size of the corpus will directly affect the calculation of word frequency and document frequency. In Fan model, the size of the corpus will also directly affect the generation of word vectors, so the size of the corpus will directly or indirectly affect the final dictionary extraction results. Therefore, this paper selects different corpus sizes for experiments. The experimental results are shown in Figure 8.

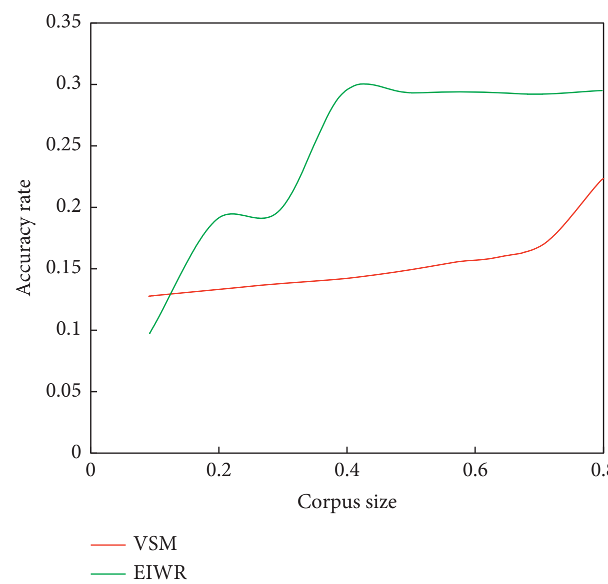

Figure 8: The influence of corpus size on the accuracy of the two models.

The experimental results in the graph above show that the corpus size has an impact on the final experimental results of the two models. The accuracy of VSM is higher than that of EIWR in a small-scale corpus, but as the corpus grows larger, the experimental results of EIWR clearly outperform those of VSM. Whether using the VSM or Pa model, the seed dictionary is an important component of bilingual dictionary extraction because it serves as a link between the source and target languages. The effect of seed dictionary size on experimental results is investigated in this paper.

It can be seen from Figure 9 that the size of the seed dictionary has a certain influence on the extraction results of both models, but it is more significant for VSM, and its accuracy is positively correlated with the size of the seed dictionary. On the other hand, it also shows that using a small-scale bilingual seed dictionary in the interword relationship model can achieve higher accuracy and greatly reduce computational complexity.

Figure 10 shows that the VSM is better than EIWR for low-frequency words. However, the accuracy rate of using EIWR for our daily use of high-frequency words has increased from 31.4% to 48.8%, and the improvement effect is obvious, which also shows that the relationship features of our daily use of high-frequency words are more distinguishable than the contextual features.

Accurate comprehension is the foundation of urban publicity translation. It is impossible to talk about style or aesthetic transmission of translation if the meaning is not understood correctly. Information transmission is not automatic or superficial; it necessitates decoding and interpretation. The meaning of reading is directly related to the context in the translation transformation process, and the context plays an important role in the generation of

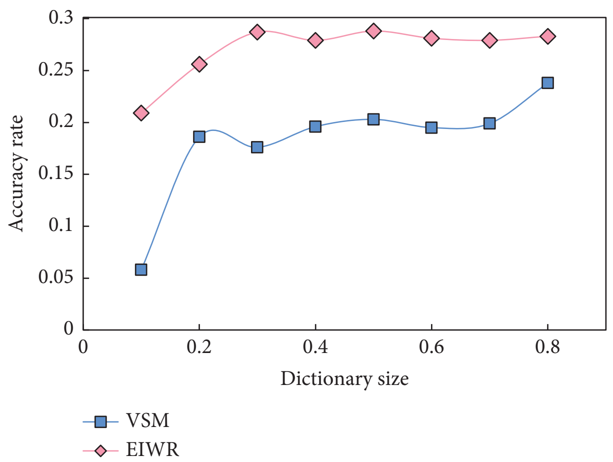

Figure 9: Influence of dictionary size on accuracy of two models.

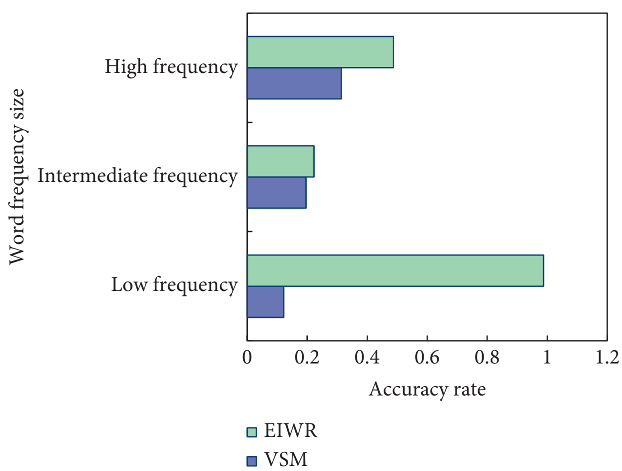

Figure 10: Accuracy of two models with different word frequencies.

meaning. Aside from accurate interpretation, the translated object's publicity effect cannot be overlooked. After analyzing the composition and characteristics of various components in the target language, one of the challenges of urban publicity translation is maximizing their inclusion and synthesis in the target language. The ecological environment, religious beliefs, national customs, aesthetic taste, way of thinking and values, and so on differ between nationalities. The same thing will inevitably produce different associative meanings based on the concept of vocabulary itself under the influence of their own unique cultural traditions. This associative meaning is not always related to the word's actual meaning, but it can evoke a familiar feeling in a particular culture.

5. Conclusion

The translation of city publicity is an important tool for a city to project its image to the rest of the world, as well as a cross-cultural communication medium. Starting with the translation of publicity materials, this paper focuses on using a computer corpus to reconstruct cross-cultural text for urban publicity translation. The two translation models are optimized using a joint EM training algorithm to improve their translation performance. On Chinese-English machine translation tasks, the experimental results show that the machine translation model based on this method outperforms the current popular strong baseline system. Experiments show that the interword relation model, when compared to the basic model, can significantly improve the extraction effect of bilingual dictionaries in comparable corpora, particularly for the translation words of high-frequency words, with an accuracy rate of 48.8%.

The assumption that the correlation between words is symmetrical is used by EIWR, but this assumption does not apply to all situations in real life. If future work can use linguistics' grammar and semantic knowledge to analyze the correlation between words, the relationship between words can be evaluated and quantified more thoroughly, and the final extraction effect can be improved.

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.

Acknowledgments

The study was supported by the "National Natural Science Foundation of China (Grant no. 11801081)," "University Humanities and Social Sciences Project of Anhui Province, China (Grant nos. SK2019A0330, SK2019A0320, and SK2021A0389)," and "Teaching and Research Projects in Anhui Province, China (Grant no. 2019jyxm0281)".

References

[1] X. Chen, "Research on translation strategies of urban publicity from the perspective of cross-cultural communication——taking the brochure of xixiangtang district, nanning city, guangxi as an example," *Overseas English*, vol. 20, p. 3, 2020.

[2] S. Hua and Z. Liu, "A study on the English translation of publicity materials of yinshan rock paintings from a cross-cultural perspective," *Overseas English*, vol. 5, p. 2, 2018.

[3] N. Wang, S. Yao, and X. He, "New Ideas on Translation Studies of Tianjin External Propaganda in the New Media Era--Taking Innovative Research on Tianjin External Propaganda as an Example," *Publishing Wide Angle*, vol. 9, p. 3, 2021.

[4] Y. Yang and E. Wu, "Research on translation strategy of foreign propaganda texts from the cross-cultural perspective," *Masterpiece Appreciation: Review Edition (Mid)*, vol. 25, p. 3, 2019.

[5] Z. Wang, L. Ma, and H. Wang, "Analysis of Mistranslation in Chinese and Foreign Propaganda Translation from a Cross-Cultural Perspective," *English Square: Academic Research*, vol. 19, p. 3, 2021.

[6] J. Liu, J. Wang, and Y. Huang, "Research on the translation strategies of government external propaganda from the

perspective of memetic theory——taking the English version of some government websites of hebei Province as an example," *Journal of Cangzhou Normal University*, vol. 37, no. 3, p. 5, 2021.
[7] M. Su, "Strategies for translation of rural tourism industry publicity in the cross-cultural perspective——comment on Cross-Cultural tourism translation," *Edible Fungus*, vol. 39, no. 4, p. 1, 2020.
[8] L. Li, "A study on translation of tourism publicity from a cross-cultural perspective," *Overseas English*, vol. 18, p. 3, 2018.
[9] Q. Su, "Analysis on translation skills of foreign propaganda from the cross-cultural perspective," *Overseas English*, vol. 17, p. 2, 2020.
[10] Y. Zhang, "An analysis of translation strategies for foreign publicity from the perspective of cross-cultural pragmatics," *Educational Research*, vol. 2, no. 11, pp. 2630–4686, 2019.
[11] J. Guo, "Strategies for translation of foreign propaganda in cross-cultural communication——comment on research on translating propaganda in cross-cultural perspective," *Educational Theory and Practice*, vol. 38, no. 14, p. 1, 2018.
[12] Y. Zhu and T. Liu, "Research on translation strategies of foreign propaganda in cross-cultural context," *Tangshan Literature*, vol. 10, p. 1, 2017.
[13] W. Tan and X. Xiong, "Research on Chinese-English translation of external propaganda materials from the cross-cultural perspective," *Journal of Jilin Radio and Television University*, vol. 35, p. 3, 2017.
[14] J. Wang and Y. Zhang, "The cultural function of intangible cultural heritage translation from the perspective of cross-cultural communication," *Culture Monthly*, vol. 11, p. 2, 2021.
[15] K. Zhou, "Analysis on how to achieve cultural equivalence in translation of foreign propaganda," *Reading World*, vol. 13, p. 2, 2017.
[16] Yu Yang, "Cultivation of Cross-Cultural Awareness in the Translation of Foreign Propaganda News," *TV Guide*, vol. 23, p. 2, 2017.
[17] T. Nagata and H. Mori, "Defining laughter context for laughter synthesis with spontaneous speech corpus," *IEEE Transactions on Affective Computing*, vol. 11, 2018.
[18] W. Y. Xiao, M. W. Wang, Z. Weng, and Z Li-lin, "Corpus-based research on English word recognition rates in primary school and word selection strategy," *Frontiers of Information Technology & Electronic Engineering*, vol. 18, no. 3, pp. 362–372, 2017.
[19] R. Su, S. Shi, Z. Meng, and H. Heyan, "Utilizing crowdsourcing for the construction of Chinese-Mongolian speech corpus with evaluation mechanism," *Proceedings of the International Conference on Computer Frontier*, vol. 12, p. 3, 2017.
[20] L. Kong, Z. Han, H. Qi, and Z. Lu, "A ranking-based text matching approach for plagiarism detection," *IEICE - Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 101, no. 5, pp. 799–810, 2018.
[21] C Lhioui, Z Anis, and M Zrigui, "The constitution of an Arabic touristic corpus," *Procedia Computer Science*, vol. 142, pp. 14–25, 2018.
[22] F. Al-Raisi, W. Lin, and A. Bourai, "A monolingual parallel corpus of Arabic," *Procedia Computer Science*, vol. 142, pp. 334–338, 2018.
[23] M. A. Abuelwafa, "Legitimation and manipulation in political speeches: a corpus-based study," *Procedia Computer Science*, vol. 189, pp. 11–18, 2021.
[24] H Pérez-Espinosa, J Martínez-Miranda, I Espinosa-Curiel, J Rodríguez-Jacobo, L Villaseñor-Pineda, and H Avila-George, "IESC-child: an interactive emotional children's speech corpus," *Computer Speech & Language*, vol. 59, pp. 55–74, 2020.
[25] T Strak and M Tuszynski, "Quantitative analysis of a private tax rulings corpus - ScienceDirect," *Procedia Computer Science*, vol. 176, pp. 2445–2455, 2020.