

## Research Article

# Model and Simulation of Maximum Entropy Phrase Reordering of English Text in Language Learning Machine

**Weifang Wu** 

*Department of Foreign Language, Anhui Jianzhu University, Hefei 230601, China*

Correspondence should be addressed to Weifang Wu; [ahjzuwuf@ahjzu.edu.cn](mailto:ahjzuwuf@ahjzu.edu.cn)

Received 9 November 2020; Revised 9 December 2020; Accepted 12 December 2020; Published 21 December 2020

Academic Editor: Wei Wang

Copyright © 2020 Weifang Wu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a feature extraction algorithm based on the maximum entropy phrase reordering model in statistical machine translation in language learning machines. The algorithm can extract more accurate phrase reordering information, especially the feature information of reversed phrases, which solves the problem of imbalance of feature data during maximum entropy training in the original algorithm, and improves the accuracy of phrase reordering in translation. In the experiment, they were combined with linguistic features such as parts of speech, words, and syntactic features extracted by using the syntax analyzer, and the maximum entropy classifier was used to predict translation errors, and the experimental verification was performed on the Chinese-English translation data set and compared. The experimental results show that different word posterior probabilities have a significant impact on the classification error rate, and the combination of linguistic features based on the word posterior probability can significantly reduce the classification error rate and improve the translation error prediction performance.

## 1. Introduction

Phrase-based statistical machine translation is one of the current mainstream methods of machine translation. The basic unit of translation transitions from word to phrase, and the continuous word string is processed as a whole in the translation process, which solves the problem of word context dependence [1]. When translating, the input sentence is matched with the phrase dictionary, the best phrase division is selected, and the obtained phrase translation is reordered to obtain the best translation. Among them, reordering at the phrase level is an important research problem based on phrase machine translation. Many systems use the distortion model probability to adjust the order between the target language phrases [2]. The distortion probability of each target phrase can be based on the current target phrase's source language phrase starting position and the previous target phrase. The distance between the last position of the phrase in the source language of the phrase is calculated. Obviously, this simple strategy based on penalty length will affect the accuracy of the phrase reordering model. The introduction of syntactic knowledge into the

machine translation system can effectively improve the accuracy of reordering [3].

In recent years, with the development of machine translation (SMT) based on statistical methods, many different types of machine translation (MT) systems have emerged, such as phrase-based, hierarchical phrase-based, and syntax-based machine translation systems, and translate performance has been significantly improved [4]. Automatic translation quality evaluation is a hot spot in statistical machine translation research. It can be divided into two types: automatic evaluation with parameters and automatic evaluation without parameters [5]. In the field of software localization, the latter refers to automatically giving a confidence score to the translation quality or identifying and classifying translation errors in the translation without reference to the answer, so as to help the translation editors to quickly locate the translation error position and improve work efficiency. In order to improve the quality of machine translation, automatic error detection and classification play a vital role in the postprocessing of MT output. On the one hand, it can help posteditors improve work efficiency, and on the other hand, it can analyze the translation of the

corresponding source language based on translation errors. We could redecode by transforming the source language input, thereby improving the translation performance [6].

Among them, the bracket transcription grammar proposed by Alkazemi et al. [7] has also been widely used in the field of machine translation. However, because the bracket transfer grammar does not contain linguistic knowledge, it cannot predict the combination order of two adjacent target phrases well. Wang et al. [8] used the boundary words of the bilingual phrase as a feature on the basis of the bracket transcription grammar to perform maximum entropy training to obtain the reordering model and obtain the probabilities in order preservation and reverse order by calculating the characteristics of adjacent bilingual phrases; it can better predict the order between adjacent phrases, thereby effectively improving the translation results of the translation system. By observing the features of maximum entropy training based on the maximum entropy phrase reranking model, it is found that the number of instance features of order-preserving phrases is much greater than the number of instance features of reversed phrases, because the word order of Chinese and English is roughly the same [9]. The use of maximum entropy to achieve reordering of phrases can also be regarded as a classification problem, that is, the order-preserving class and the reverse-ordering class, and the feature data used to train the classifier have a data imbalance problem, which may affect the classifier actual classification effect. For example, if FBIS is selected as the training corpus, the baseline feature extraction system extracts 4,839,390 feature instances, of which order-preserving feature instances account for 82.7%, while reverse-order feature instances account for only 17.3% [10]. Taking 100,000 sentences in all feature instances as the open test set of the reranking model and the remaining data as the maximum entropy training set, the test results show that the reranking model has a judgment accuracy of 97.55% for order-preserving features [10]. The judgment accuracy rate of the reverse-order feature is only 72.03% [11]. In addition, based on the bracket transcription grammar, it is assumed that the source language end phrase is adjacent and the target language phrase is also adjacent, but there are adjacent source language phrases in the actual Chinese-English sentence pair. In view of the above situation, this paper improves the maximum entropy feature extraction algorithm from three aspects: order-preserving example selection strategy, introduction of combined features, and addition of new phrase order to improve the judgment accuracy of the reranking model and finally achieve the effect of improving translation quality.

## 2. Statistical Machine Translation Based on the Maximum Entropy Phrase Reordering Model

Wang et al. [8] proposed a statistical translation model based on bracket transcription grammar. The simplified bracket transcription grammar contains only the following two rules:

$$P^i: X \longrightarrow \frac{\alpha}{\beta}, \quad (1)$$

$$P^j: X \longrightarrow [x_1, x_2] \langle x_1, x_2 \rangle.$$

Among them,  $P^i$  is the vocabulary rule, which means that the source language phrase  $\alpha$  is translated into the target language phrase  $\beta$ .  $P^j$  is the merging rule, and the order of the source language phrase and the target language phrase can be expressed as preserving order and reverse order. In the process of phrase reordering, a priori preserving and reversing probabilities can be set for the two different orders in the merging rule. This method ignores the differences between different source language target language phrase pairs.

Maučec and Donaj [12] improved the ordering model of the abovementioned bracket transcription grammar model and proposed a phrase ordering model based on the maximum entropy bracket transcription grammar, that is, using the maximum entropy model for phrase ordering:

$$\chi = p_\omega(\delta|x_1, x_2) = \frac{\exp(\sum_i \omega_i f_i(\delta, x_1, x_2))}{\sum_j \exp(\sum_i \omega_i f_i(\delta, x_1, x_2))}. \quad (2)$$

Among them,  $f$  is the feature function,  $\omega$  is the feature weight, and the value of  $\delta$  is order preserving or reverse ordering, and the ending word of the phrase is selected as the feature of the maximum entropy model training. Experiments show that the performance of the phrase ordering model based on the maximum entropy bracket transcription grammar is significantly better than the traditional distortion-based short intonation ordering model and the bracketing transcription grammar-based ordering model. However, it can be seen from experiments that the number of order-preserving instances is much higher than the number of reverse-order instances, which may affect the performance of the maximum entropy model. This paper cuts in from two aspects of the reranking instance extraction algorithm and feature selection, aiming to solve the problem of maximum entropy training data imbalance. In the experiment, the statistical machine translation system [13] based on the maximum entropy ordering model will be used as the baseline system. The maximum entropy phrase reordering model is shown in Figure 1.

*2.1. Reordering Instance Extraction Algorithm.* The extraction algorithm of reordering examples in the maximum entropy phrase reordering system in this paper is more flexible and concise in implementation and easy to expand, which can meet different extraction strategies in the experiment. The input of the reranking instance extraction algorithm is a word alignment matrix that has been GIZA bidirectionally aligned, and the output is the order-preserving phrase instance and the reverse-order phrase instance [14]. The extraction algorithm first traverses all consecutive word sequences in the source language and extracts the maximum span of the target language that is aligned with this continuous sequence. Then, the target language word sequence and the source language word

sequence that do not satisfy the alignment consistency are filtered, that is, the span of the target language is scanned in reverse order to check whether the corresponding source language span is within the range of the original continuous word sequence. Finally, according to the given different extraction strategies, reordering examples are extracted.

*2.1.1. Variable Definition.* Before introducing the reordering example extraction algorithm, first define the variables related to the algorithm:

- (1) Align set: store all alignment matrices from the source language to the target language
- (2) Straight set: store a collection of instances of the target language phrase preserving order
- (3) Inverted set: store a collection of instances in reverse order of target language phrases
- (4) Else set: store the instances where the source language phrase is adjacent and the target language phrase is not adjacent
- (5) Sec\_span[  $i, j$ ]: a sequence of consecutive words from  $i$  to  $j$  in the source language
- (6) Span[  $i, j$ ]: record the sequence of consecutive words from  $i$  to  $j$  in the source language and the sequence of consecutive words in the corresponding target language

*2.1.2. Algorithm Implementation.* This algorithm first obtains the largest alignment matrix span [  $i, j$  ] corresponding to any source language span [  $i, j$  ] and then filters out illegal span [  $i, j$  ]. Finally, it classifies the reordering examples and extracts the characteristics of the examples; see Algorithm 1 for specific steps. The specific algorithm is shown in Figure 2.

The last lines of the algorithm describe the framework of the improved algorithm for extracting examples. Based on this framework, it is convenient to formulate various extraction rules. Among them, the 10<sup>th</sup> step pairs the extracted bilingual word alignment matrix, checks whether it can be split into two adjacent bilingual phrase pairs, and judges the combination order of the split adjacent bilingual phrase pairs. In the final step, the algorithm introduces a new classification, namely, nonadjacent bilingual phrase pairs.

*2.2. Reordering Instance Selection Strategy.* The baseline system uses a simple method to control the number of reordering instances, that is, only the smallest block is reserved in the order-preserving instances and only the largest block is reserved for the reverse-order instances. Obviously, some phrase boundary features will be lost in this way, and the number of preserving instances still far exceeds the number of reversed instances. This imbalance of feature data will affect the judgment accuracy of the maximum entropy reordering model, especially the judgment of the features of the reverse-order instance [15]. Open-ended testing is performed with 100,000 instances, of which the number of reverse-order instances is 17,286, and the test accuracy of

reverse-order instances is only 72.03%. In this paper, under the algorithm framework proposed in Section 3, the following three attempts are made in sequence for the reordering instance selection strategy:

- (1) In order to solve the imbalance of feature data during the maximum entropy training process, the most direct idea is to adopt a certain selection strategy to directly limit the number of preserving instances [16, 17]. Compared with the minimum block in the order-preserving example selected by the baseline system, this paper uses a random algorithm to select the number of order-preserving examples, which avoids the loss of long phrase boundary features that may be caused by the previous method.
- (2) In bilingual sentences, the source language phrases are adjacent but the target language phrases are not adjacent. In response to this situation, this article adds a new classification based on (1) to reduce the imbalance of feature data to a certain extent. If the extracted instance does not belong to the order-preserving and reverse-ordering categories, the instance can be classified into one category [18].
- (3) Because of the misalignment in the alignment results, extending the unaligned words to the examples will improve the recall rate of phrase feature extraction. Here, we define the order-preserving and reverse-ordering rules,  $i = \{0, 1\}$ , where  $i = 0$ , it means that the extracted instance is not expanded by unaligned words, and  $i = 1$ , it means that the extracted instance is expanded by unaligned words.

*2.3. Feature Extraction.* Features from reranking instances are extracted for maximum entropy training. The reordering instance can be represented by  $\langle a_1, a_2 \rangle$ , where  $a = \langle b, c \rangle$ ,  $b$  represents the source language phrase,  $c$  represents the target language phrase, and  $a_1$  and  $a_2$  represent adjacent or non-adjacent phrases. Here,  $b.f$  is used to denote the first word of the source language phrase and  $b.l$  is the last word of the source language phrase. The same definition is used for the target phrase  $c$ . The baseline system takes into account the scale of feature extraction and uses only the tail words in the rearrangement examples. In the feature extraction experiment, in addition to the above four tail word features, the first word feature and combination feature are added [19]. Because of the different grammatical structures between Chinese and English, the corresponding English translation of the phrase or clause before and after the Chinese punctuation marks may express the phrase or clause in reverse order [20]. The decoding method of the baseline system is that if punctuation marks are searched in the reordering window, this window will not perform the reverse-order operation. This method is quite effective for symmetric symbols, such as  $\langle \langle \rangle \rangle \{ \}$ . However, the “.” cannot be simply judged based on this. In this paper, on the basis of increasing the first word feature and combination feature of the reranking instance, punctuation feature is added for maximum entropy training. The characteristics of reordering examples are shown in Table 1.

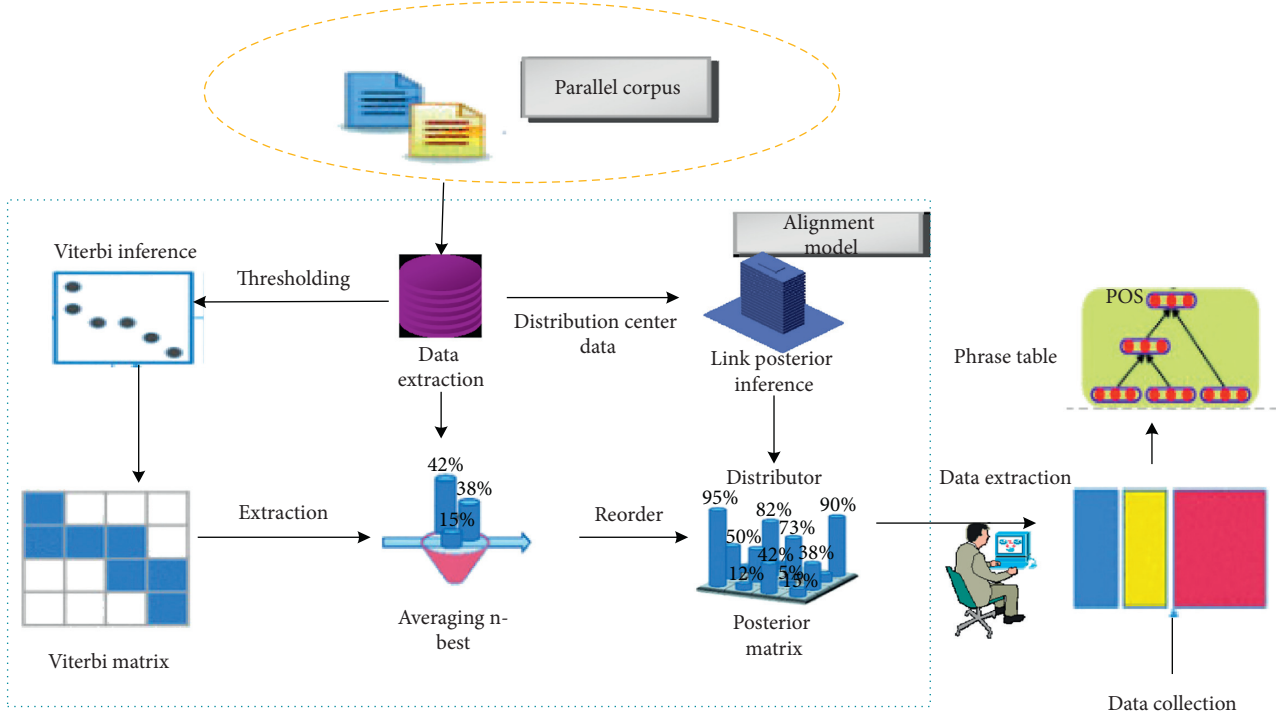


FIGURE 1: Maximum entropy phrase reordering model.

### 3. English Translation System Evaluation Criteria

The evaluation criteria for the effectiveness of the error detection method adopted are classification error rate (CER), accuracy rate (AR), recall rate (RR), and  $F$  criterion. The classification error rate is calculated as follows:

$$\text{CER} = \frac{\text{the number of words whose classification category is wrong}}{\text{the total number of words}}. \quad (3)$$

In the Chinese-to-English translation error detection and classification task, because the number of true categories in the translation hypothesis “incorrect” is greater than the number of “correct,” so when determining the baseline level of the classification error rate, the usual approach is as follows: the evaluation criterion score is obtained when all the “correct” words are marked as “incorrect”, namely, the baseline level of classification error rate = the number of “correct” samples/the total number of samples.

The accuracy is the ratio of the number  $m_n$  that the classifier accurately classifies the words that are actually in category  $i$  to the number  $t_n$  of words that the classifier marks as  $i$ , that is,

$$\text{AR} = \frac{m_n}{t_n}. \quad (4)$$

The recall rate is the ratio of the number  $m_n$  that the classifier accurately classifies the words of the real category  $i$  to the total number of words  $g_n$  of the real category  $i$ :

$$\text{RR} = \frac{g_n}{t_n}. \quad (5)$$

The  $F$  criterion is the trade-off between accuracy and recall, namely,

$$F = \frac{2 \times \text{AR} \times \text{RR}}{\text{AR} + \text{RR}}. \quad (6)$$

**3.1. Experimental Results and Analysis.** In the experiment, the language model uses the  $N$ -gram statistical language model, the monolingual corpus for training the English language model, and the mature open-source language model training tool recognized in the field of statistical machine translation for the  $N$ -gram language model training. The experiment uses a four-gram language model with a scale of 518 M [15]. Based on the reranking instance extraction algorithm, we designed 7 comparison experiments to compare the impact of different feature extraction strategies on the maximum entropy training and the effect on the final translation result score BLEU. Choosing training corpus, extracting phrase lists, and reordering examples, the corpus size is about 239,000 sentence pairs. Take NIST-MT 02 as the experimental development set and NIST-MT-05 as the test set.

**3.2. The Impact of Feature Extraction Strategies on the Results of Reranking.** The 100,000 records in the feature data of the reranking instance are selected as the open test set of the maximum entropy reranking model. Table 2 shows the scale of the reranking instance extracted from the training data, the sorting category, the proportion of each category, and the test accuracy and extracted features. Among them, the test accuracy is the ratio of the number of

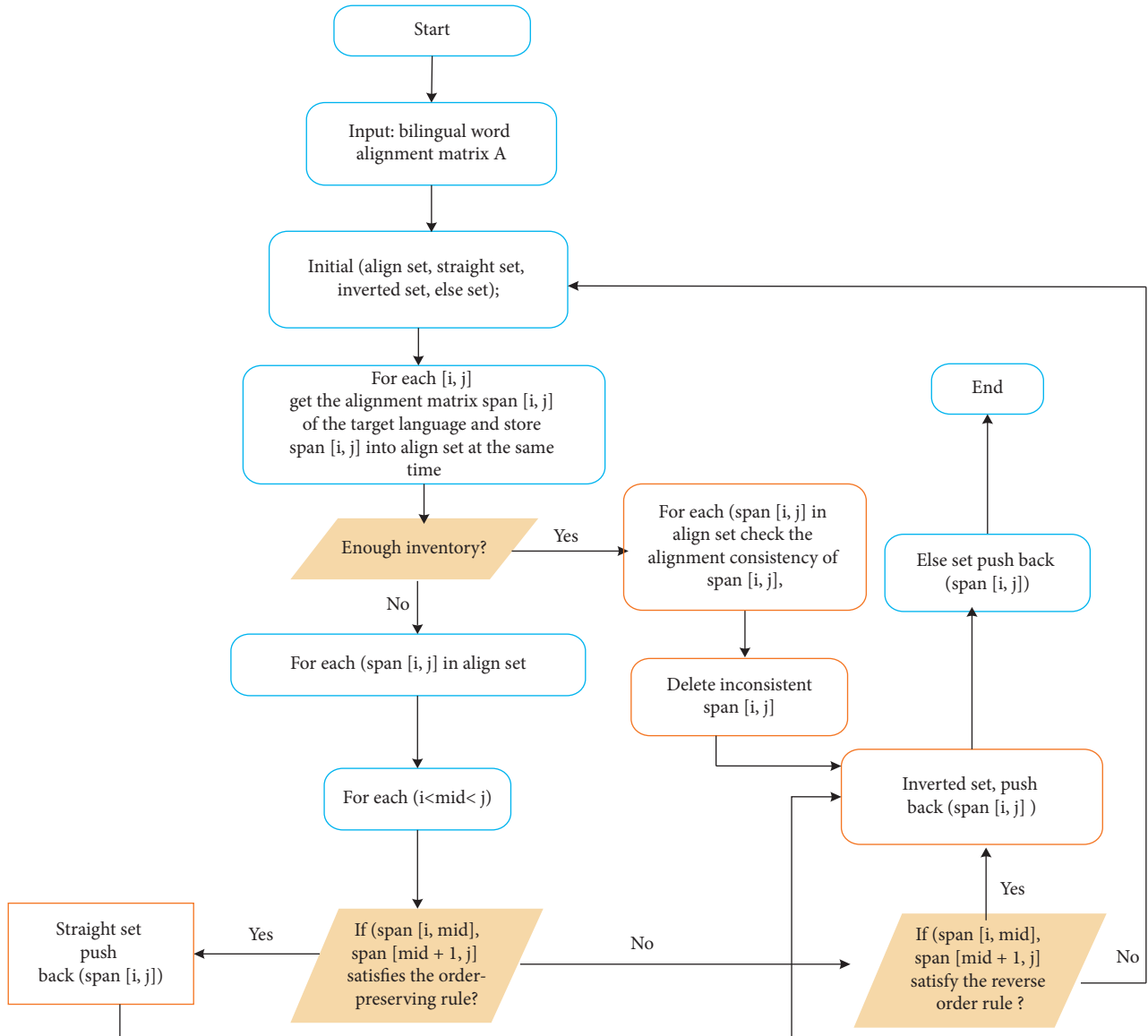


FIGURE 2: Algorithm implementation.

TABLE 1: Characteristics of reordering examples.

End word features	$a_1 . b . f, a_2 . c . f, a_1 . b . f, a_2 . c . f$
First word features	$a_1 . b . l, a_2 . c . l, a_1 . b . l, a_2 . c . l$
Combination features	$a_1 . b . f \& a_2 . b . f, a_1 . c . f \& a_2 . c . f, a_1 . b . f \& a_2 . b . f, a_1 . c . f \& a_2 . c . l, a_1 . c . f \& a_2 . c . l, a_1 . c . l \& a_2 . c . l$

samples correctly judged by the maximum entropy classifier to the total number of samples in the test set. Among them, experiment 1 is the baseline system, and there is no restriction on the number of order-preserving instances. Experiments 2–6 limit the number of order-preserving instances to twice the number of reverse-ordering instances, experiments 2–4 did not expand the unaligned words when extracting examples, experiments 5–7 all carry out unaligned word expansion, and

experiments 4 and 5 add a new category. Because of the inconsistency of the characteristics required by different experiments, only the number of test sets can be determined, but the consistency of the content of the test sets cannot be ensured. Therefore, the test accuracy of the maximum entropy reordering model cannot be simply reflected as the level of translation performance. The test accuracy of the maximum entropy reordering model can still be used as a reference indicator.

TABLE 2: The scale, sort classification, test accuracy, and extracted features of the reranking instance.

Experiment	Total number of instances	Guaranteed order (%)	Reverse order (%)	Others (%)	Test accuracy (%)	Experimental program
1	4353250	81.34	16.48	0.00	92.48	End word features
2	1424597	65.38	32.49	0.00	85.38	Limit the number of instances of order preservation
3	1424597	67.39	32.49	0.00	91.39	First word feature Combination features
4	3242782	35.83	20.38	38.49	75.38	Third category
5	3124891	35.38	20.81	38.51	74.30	Extended unaligned words
6	3144248	74.28	34.21	0.00	91.39	Punctuation features

It can be seen from Figure 3 that the test accuracy of experiment 1 reached the highest value of 92.48%. Because of the limitation of the number of preserving instances in experiment 2, the total number of extracted instances was reduced by 60% compared with experiment 1, resulting in a maximum. The amount of data for entropy training is insufficient, so the test accuracy is only 85.38%. Considering that when the number of instances is reduced, the amount of feature data generated by a single instance needs to be increased, so in experiment 3, the first word feature and combination feature are added to the instance, and the test accuracy reaches 91.39%. However, the adjacent source language phrases do not indicate that the target language phrases are adjacent, so experiment 4 introduces the third category, namely, the target language.

The phrase is not adjacent. The test accuracy of experiment 4 dropped to 75.38% because a new category also increased the uncertainty of the maximum entropy reordering model judgment. Experiment 5 is based on experiment 4 and expands unaligned words to increase the number of examples, but the result of experiment 5 is slightly lower than that of experiment 4. Both experiments 4 and 5 are based on experiment 3. The introduction of the third category leads to a large decrease in test accuracy. To a certain extent, it shows that the introduction of the third category will not improve the accuracy of the maximum entropy model judgment. Therefore, this paper designs experiment 6 to expand the unaligned words on the basis of experiment 3 and to quote on the basis of experiment 6.

The test accuracy of these two experiments is only slightly lower than experiment 1. This paper pays more attention to the accuracy of the feature extraction strategy for the maximum entropy model to judge the inverted instance. Figure 4 shows the test accuracy of the maximum entropy reranking model on the preserving subset and the inverted subset (Invert) of the test set. A subset of the order-preserving examples in the test set is tested. Except for the introduction of new classifications in experiments 4 and 5, the uncertainty in the judgment of order-preserving features increases. The test accuracy of experiments 2, 3, and 6 is not different from that of experiment 1 more than 4%. The test results of a subset of reverse-ordered instances in the test set are observed. In experiment 2, because the amount of training data for reverse-ordered features is small, the test accuracy on the reverse-ordered instance subset is lower, and the test accuracy of experiments 3, 4, 5, and 6 are all

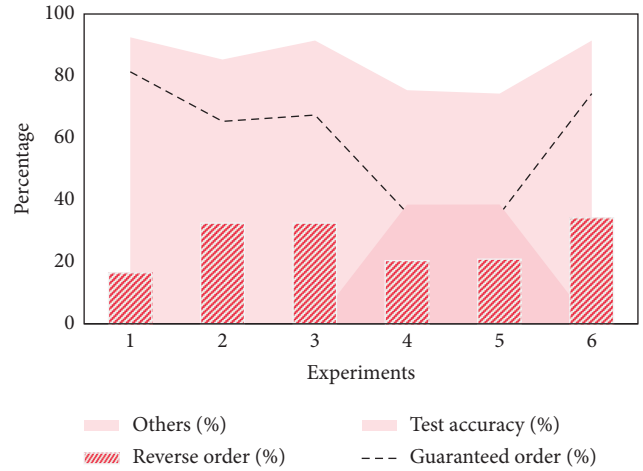


FIGURE 3: The test accuracy and extracted features of the reranking instance.

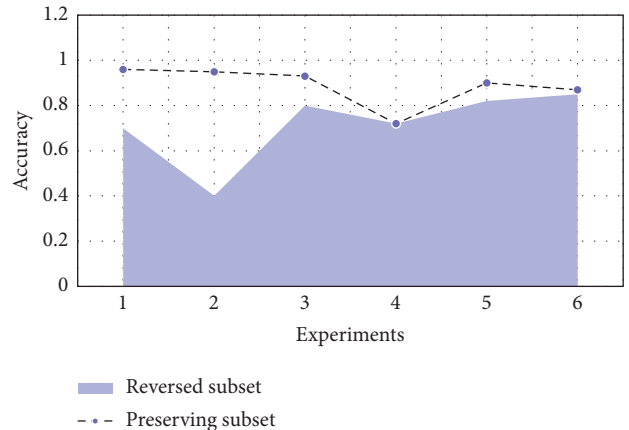


FIGURE 4: Test accuracy of the preserving subset and reversed subset.

better than that of experiment 1. The accuracy of the subset of instances in reverse order is high. Among them, the test accuracy of experiment 6 is 6% higher than that of experiment 1. It can be seen from the above experimental data that the maximum entropy reordering model feature extraction algorithm proposed in this paper solves the inaccurate judgment of the reverse-order feature caused by the imbalance of feature data.

3.3. *Comparison of Translation Results.* The case-sensitive BLEU value was tested on NIST-MT 05. Figure 5 shows the impact of 6 groups of maximum entropy reordering models trained with different feature data on the final translation effect. The BLEU value of baseline system experiment 1 is 0.2283. As can be seen from Figure 5, except for experiment 2, the performance of the maximum entropy reordering model has been greatly reduced during the translation process due to too little feature training data. Experiments 3, 4, 5, and 6 are all based on experiment 2 to add feature information, and the performance of the reranking model while limiting the number of preserving instances is higher than that of the baseline system. In experiment 4, the translation performance of the nonadjacent classification is reduced but the BLEU value is still higher than that of the baseline system. Experiment 6 adds punctuation features and the translated BLEU. The value reaches the highest value of 0.243. The reranking instance extraction and feature extraction algorithms proposed in this paper can significantly improve the performance of the reranking model and improve the translation quality by limiting the number of preserving instances and increasing the number of features.

#### 4. Misclassification Experiment

The feature function of the maximum entropy classifier is the feature vector considering the context; that is, in addition to each current feature variable, it also considers its front and back. Experimental design: (1) perform classification experiments on 3 typical word posterior probability features and compare and analyze their performance; (2) perform maximum entropy model classification experiments on individual linguistic features and analyze them; (3) combine three typical word posterior probability features with linguistic features, perform classification experiments, and compare and analyze them.

4.1. *Classification Experiment Based on Word Posterior Probability Features.* Table 3 shows the classification experiment results based on the posterior probability of 3 typical words. In Table 3, Dir represents a word posterior probability feature based on a fixed position, Win represents a word posterior probability feature based on a sliding window, sliding window  $t=2$ , Lev represents a word posterior probability feature based on Levenshtein alignment. When aligning the 1-best translation hypothesis in the  $N$ -best list with other translation hypotheses, the open-source toolkit TER [13] is used, and its “shift” function is turned off, which is WER alignment. The abovementioned three posterior probabilities have been discretized before use [10].

It can be seen from Figure 6 that, in terms of CER, compared with the baseline system, the features Dir, Win, and Lev are reduced by 2.34%, 3.97%, and 34% (relative values), respectively, and Win performs best. Analyzing the above results, we can obtain the following: (1) the Win feature changes the fixed position into a sliding window, which has higher alignment flexibility, so it is more in line with the ordering phenomenon of the source language and the target

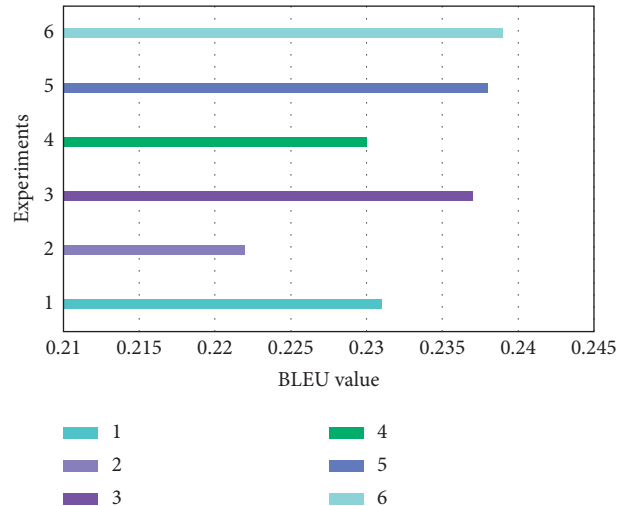


FIGURE 5: The influence of different maximum entropy reordering models on the BLEU value.

TABLE 3: Error detection results of 3 typical WPP features.

Features	CER (%)	AR (%)	RR (%)	F (%)
Baseline	42.24	—	—	—
Dir	38.49	62.39	71.93	66.82
Win	39.29	62.38	73.17	67.91
Lev	38.92	58.01	91.39	72.91

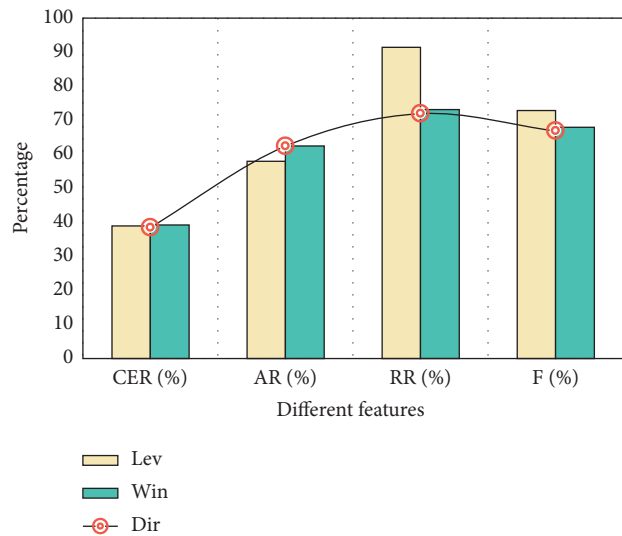


FIGURE 6: Detection results of 3 typical features.

language due to different word orders, but sliding window is limited to limited local ordering; (2) the Lev feature is based on Levenshtein alignment, so the alignment is better, but it also introduces too many editing operations, such as insertion, deletion, and replacement, and because there is no word order, although the alignment is better than Dir, the flexibility is lower than Win. From the above analysis and data, it can be known that combining CER and  $F$  value, the characteristic Win has the best comprehensive performance.

#### 4.2. Classification Experiment Based on Linguistic Features.

Table 4 shows the error detection results based on linguistic features, namely, word entity (Word), part-of-speech tagging (POS), and syntactic relationship (Link).

Compared with the baseline system, as shown in Figure 7, Word, POS, and Link have reduced CER by 5.36%, 4.98%, and 1.72% (relative values), respectively, among which Word performs best. In terms of  $F$  value, Link performs better than the other two features and POS is better than Word. Analyzing the above results and comparing them with Table 3, we can obtain the following: (1) except for the Link feature, the classification error rates of Word and POS in the linguistic features are lower than the classification error rates of the 3 word posterior probability features; (2) Link feature has the highest recall rate and the lowest accuracy rate. This is mainly due to the relatively small number of Link features. Therefore, when classifying, the classification result is more inclined to mark the target word as category  $i$ , resulting in category  $c$ . The number is relatively small, so that the recall rate is high and the accuracy rate is low; (3) the classification result of the Word feature is better than that of the POS feature. The reason may be that the development set and the test set are more relevant (both in the news field), and the number of features is much more than the number of POS features, so in terms of classification ability, its tendency (or probability) to predict the target word as category  $i$  is lower than POS with a relatively small number of features, resulting in a lower recall rate, but the accuracy is better.

#### 4.3. Combination Feature Classification Experiment.

In the classification task of natural language processing research, feature combination can often reduce the classification error rate more effectively. Table 5 lists the classification experiment results based on the maximum entropy model after combining the three typical word posterior probability features described in this paper and the three linguistic features. It can be seen from Figure 8 that, in terms of CER, compared with the baseline system, the CER of the three combined features has been reduced by 13.14%, 14.25%, and 13.92% (relative values), respectively, and the  $F$  value has also been significantly improved. Although the classification performance of the three feature combinations is not significant, the classification characteristics of the three feature combinations are consistent with those of a single WPP feature; that is, the combination “Win + Word + POS + Link” has the lowest classification error rate and the combination “Dir + Word + POS + Link” has the highest  $F$  value, indicating that the word posterior probability feature based on the sliding window position can capture more contextual information, so that its ability to distinguish translation errors is stronger than the word posterior probability feature based on a fixed position. This ability is manifested not only in the comparison of individual features but also in combined features. While comparing the combined effects of three different WPP features, Table 5 also reveals the contribution of linguistic features to error detection, indicating that linguistic features can effectively

TABLE 4: Experimental results of linguistic feature error detection.

Linguistic features	CER (%)	P (%)	R (%)	F (%)
Baseline	40.24	—	—	—
Word	38.29	63.18	76.83	68.48
POS	39.26	60.72	85.82	70.93
Link	40.27	58.72	92.37	72.74

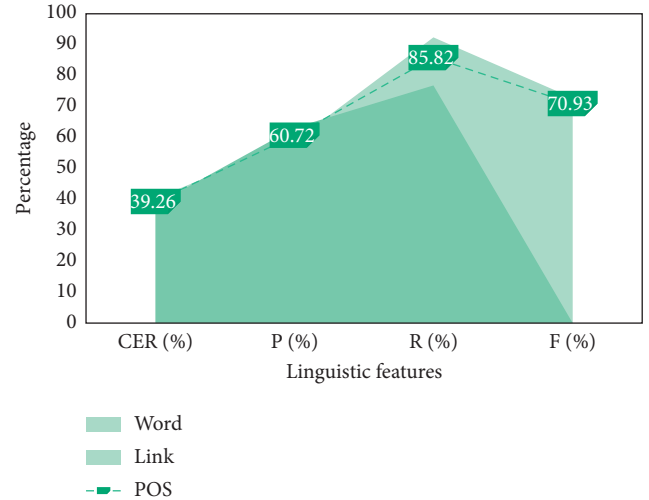


FIGURE 7: Experimental results of linguistic feature error detection.

TABLE 5: Error detection experiment results of 3 different combinations of WPP features and linguistic features.

Feature combination	CER (%)	P (%)	R (%)	F (%)
Dir + Word + POS + Link	34.58	62.49	87.38	72.84
Win + Word + POS + Link	34.58	63.81	83.48	73.19
Lev + Word + POS + Link	35.38	64.29	82.74	73.16

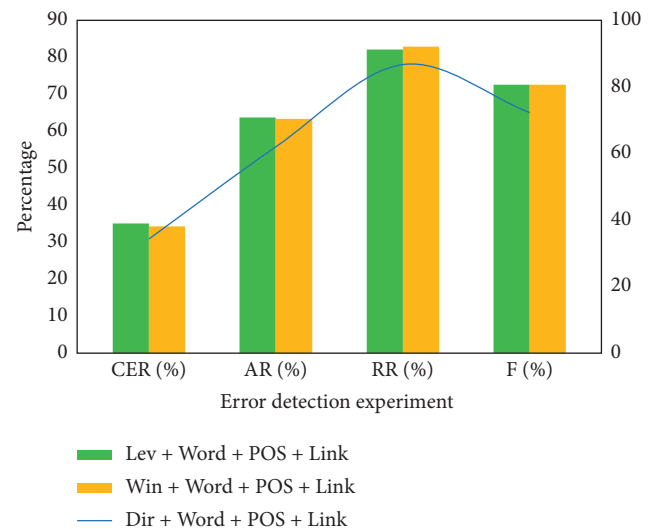


FIGURE 8: Error detection experiment results of 3 different combinations of WPP features and linguistic features.

reduce the classification error rate and improve the ability of error prediction.



## 5. Conclusion

This paper proposes a new reordering instance extraction algorithm and adds new features on this basis to achieve better translation results. First, the problem of data imbalance in the maximum entropy training process is directly solved by limiting the number of order-preserving instances, and the translation performance is reduced due to too little feature information. On this basis, the addition of first word features and combination features improves translation performance. Second, the third type of phrase combination order is introduced, that is, nonadjacent cases other than order-preserving and reverse-ordering; although the BLEU value has decreased, it is still higher than the baseline system. Finally, this article attempts to expand the unaligned words in aligned phrases in the experiment, increase the amount of reranking example feature data, and achieve the best translation performance. In the next step, we will continue to study the impact of reordering instance features on translation performance, focusing on the integration of syntactic knowledge features, hoping to further improve translation performance. In addition, we will further explore the improvement of the decoder based on the bracketed transcription grammar framework, so that it can handle the situation where the source language phrases are adjacent but the target language phrases are not adjacent.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The author declares that there are no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the Key Humanities and Social Sciences Project in Anhui Province in 2019: A Study of Interpersonal Strategies in Intercultural Communication (no. SK2019A0662).

## References

- [1] X. Wang, Z. Tu, and M. Zhang, "Incorporating statistical machine translation word knowledge into neural machine translation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 12, pp. 2255–2266, 2018.
- [2] J. Kanis, "Digging language model-maximum entropy phrase extraction," *Text, Speech, and Dialogue*, vol. 9, no. 12, pp. 46–53, 2016.
- [3] J. Srivastava, S. Sanyal, and A. K. Srivastava, "Extraction of reordering rules for statistical machine translation," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 4809–4819, 2019.
- [4] F. Klubička, A. Toral, and V. M. Sánchez-Cartagena, "Quantitative fine-grained human evaluation of machine translation systems: a case study on English to Croatian," *Machine Translation*, vol. 32, no. 3, pp. 195–215, 2018.
- [5] S. Farzi, H. Faili, and S. Kianian, "A neural reordering model based on phrasal dependency tree for statistical machine translation," *Intelligent Data Analysis*, vol. 22, no. 5, pp. 1163–1183, 2018.
- [6] Y. Liu, C. M. Vong, and P. K. Wong, "Extreme learning machine for huge hypotheses re-ranking in statistical machine translation," *Cognitive Computation*, vol. 9, no. 2, pp. 285–294, 2017.
- [7] B. Alkazemi, M. Nour, A. Naseer, A. Natto, and G. Grami, "AccurIT: a prototype of a machine translation engine for English to Arabic translation," *International Journal of Innovation and Learning*, vol. 26, no. 2, pp. 115–130, 2019.
- [8] R. Wang, M. Utiyama, A. Finch, L. Liu, K. Chen, and E. Sumita, "Sentence selection and weighting for neural machine translation domain adaptation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 10, pp. 1727–1741, 2018.
- [9] H. Cuong and K. Sima'an, "A survey of domain adaptation for statistical machine translation," *Machine Translation*, vol. 31, no. 4, pp. 187–224, 2017.
- [10] S. Lyons, "A review of Thai-English machine translation," *Machine Translation*, vol. 34, no. 2-3, pp. 197–230, 2020.
- [11] M. Domingo, Á. Peris, and F. Casacuberta, "Segment-based interactive-predictive machine translation," *Machine Translation*, vol. 31, no. 4, pp. 163–185, 2017.
- [12] M. S. Maučec and G. Donaj, "Morphology in statistical machine translation from English to highly inflectional language," *Information Technology and Control*, vol. 47, no. 1, pp. 63–74, 2018.
- [13] Z. Yang, M. Li, L. Chen, and K. Sun, "A morpheme-based weighting for Chinese-Mongolian statistical machine translation," *IEICE Transactions on Information and Systems*, vol. E99.D, no. 11, pp. 2843–2846, 2016.
- [14] N. K. Jadoon, W. Anwar, U. I. Bajwa, and F. Ahmad, "Statistical machine translation of Indian languages: a survey," *Neural Computing and Applications*, vol. 31, no. 7, pp. 2455–2467, 2019.
- [15] Z. Tan, J. Su, B. Wang, Y. Chen, and X. Shi, "Lattice-to-sequence attentional neural machine translation models," *Neurocomputing*, vol. 284, no. 4, pp. 138–147, 2018.
- [16] B. Zhang, D. Xiong, J. Su, and Y. Qin, "Alignment-supervised bidimensional attention-based recursive autoencoders for bilingual phrase representation," *IEEE Transactions on Cybernetics*, vol. 50, no. 2, pp. 503–513, 2018.
- [17] Y. Liu, K. Wang, C. Zong, and K.-Y. Su, "A unified framework and models for integrating translation memory into phrase-based statistical machine translation," *Computer Speech & Language*, vol. 54, no. 4, pp. 176–206, 2019.
- [18] S. Singh, M. Anand Kumar, and K. P. Soman, "Attention based English to Punjabi neural machine translation," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 3, pp. 1551–1559, 2018.
- [19] S. Mall and U. C. Jaiswal, "Survey: machine translation for Indian language," *International Journal of Applied Engineering Research*, vol. 13, no. 1, pp. 202–209, 2018.
- [20] B. Ahmadnia and J. Serrano, "Employing pivot language technique through statistical and neural machine translation frameworks: the case of under-resourced Persian-Spanish language pair," *International Journal on Natural Language Computing*, vol. 6, no. 5, pp. 37–47, 2017.