IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal
SPECIAL SECTION ON COLLABORATION FOR INTERNET OF THINGS

Received September 1, 2018, accepted September 20, 2018, date of publication September 28, 2018,
date of current version October 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2872773

Minimum Bayes-Risk Phrase Table Pruning for Pivot-Based Machine Translation in Internet of Things

XIAONING ZHU, MUYUN YANG, TIEJUN ZHAO, AND CONGHUI ZHU
School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001 , China

Corresponding author: Tiejun Zhao (tjzhao@hit.edu.cn)

This work was supported by the National Key Technologies R&D Program of China under Grant 2017YFB1002102.

ABSTRACT Machine translation, which will be used widely in human–computer interaction services to Internet of Things (IoT), is a key technology in artificial intelligence field. This paper presents a minimum Bayes-risk (MBR) phrase table pruning method for pivot-based statistical machine translation (SMT). The SMT system requires a great amount of bilingual data to build a high-performance translation model. For some language pairs, such as Chinese–English, massive bilingual data are available on the web. However, for most language pairs, large-scale bilingual data are hard to obtain. Pivot-based SMT is proposed to solve the data scarcity problem: it introduces a pivot language to bridge the source language and the target language. Therefore, a source-target translation model based on well-trained source-pivot and pivot-target translation models can be derived with the pivot-based approach. However, due to the ambiguities of the pivot language, source and target phrases with different meanings may be wrongly matched. Consequently, the derived source-target phrase table may contain incorrect phrase pairs. To alleviate this problem, we apply the MBR method to prune the phrase table. The MBR pruning method removes the phrase pairs with the lowest risk from the phrase table. Experimental results on Europarl data show that the proposed method can both reduce the size of phrase tables and improve the performance of translations. This study also gives a useful reference to many IoT research field and smart web services.

INDEX TERMS Internet of Things, smart services, minimum Bayes risk, pivot-based SMT, phrase table pruning, statistical machine translation.

I. INTRODUCTION

Statistical machine translation (SMT) uses a statistical translation model to translate from one language to another language. It will be used widely in many future Internet of things (IoT) field, such as smart city, smart home and smart campus [1]–[3]. To build such a high-performance translation model requires large amounts of parallel corpora in both the source language and the target language. For widely used language pairs, e.g., Chinese-English, massive amounts of bilingual data are available on the web and easy to obtain. Unfortunately, large-scale corpora of bilingual data are hard to obtain for most language pairs. It is challenging to develop a well-trained translation model with limited parallel corpora.

Various methods have been proposed to overcome the data shortage of machine translation [8]–[14], [16]. Among these methods, phrase-based SMT with a pivot language is a represent work, which connects the source language and the target language with a pivot language as the "bridge" [15]–[18], [20]–[24]. The premise of the pivot approach is that a large number of source-pivot and pivot-target parallel data are available.

The triangulation method is the most common of the pivot-based approaches [16], [25]. If the source-pivot and pivot-target translation models are well trained with plenty of high-quality source-pivot and pivot-target bilingual corpora, a source-target translation model is generated by matching source phrases and target phrases via the appropriate pivot phrases.

The phrase table with phrase translations and their probabilities is the main knowledge source of both traditional SMT model and the pivot-based SMT model. Fig. 1 is an example of the traditional SMT phrase table. The Chinese source phrase is mapping to the English target phrase in the phrase table with a translation probability. And, the mapping

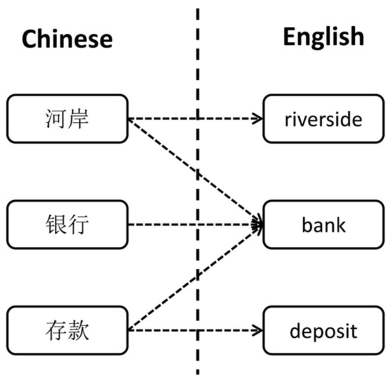


FIGURE 1. An example of a phrase table in traditional SMT.

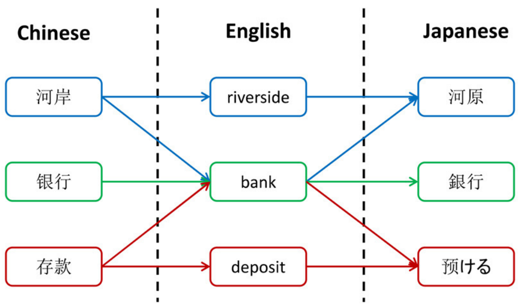


FIGURE 2. An example of a phrase table in pivot-based SMT.

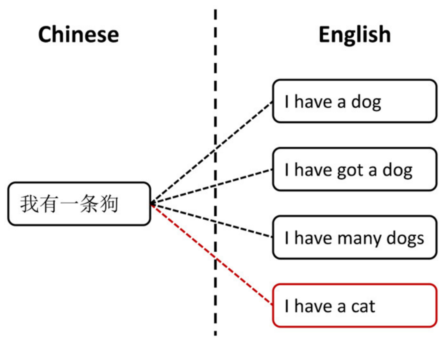


FIGURE 3. An example of minimum Bayes-based phrase table pruning method.

relationship in the phrase table is many-to-many. As shown in Fig. 1, Chinese phrase "河岸 (the river bank)" maps to English phrase "riverside" or "bank." At the same time, both the Chinese phrase "银行 (the bank for money)" and "存款 (deposit money)" map to English phrase "bank." Fig. 2 is an example of the pivot-based SMT phrase table. The Chinese phrases are mapping to the Japanese phrases via the English phrases. Fig. 1 and Fig. 2 indicate that the ambiguities of the language often lead to a wrong translation and introduce a large amount of noises into the phrase table, especially when translating from source language to target language via pivot language. Take Fig. 2 as an example, because the English word "bank" has different meanings (E.g. the river bank; the bank for money and deposit money), given a Chinese word "银行 (the bank for money)," it may be wrongly translated to the Japanese word "河原 (the river bank)" or "预ける (deposit money)."

To solve this problem, we present a minimum Bayes-risk (MBR) translation phrase table pruning method to select the proper rules and discard the redundant translation rules. The idea of using MBR to prune the phrase table are inspired by system combination in machine translation [26], [28], [29]. Some related paper also give us useful references [30]–[34]. The MBR phrase table pruning method aims to remove the translation rule that has the least expected loss under a probability model. The motivation relays on the hypothesis that the reasonable translation rules are similar, and the noisy translation rules differ widely. Take Fig. 3 as an example; if most people are talking about the dog, the one who is talking about the cat or some other things is noise. Thus, the phrase pair with the English phrase "I have a cat" should be removed from the phrase table.

The remainder of this paper is organized as follows. In section II, we brief introduce Internet of Things, phrase-based SMT and pivot-based SMT. Section III provides a review of the related work. We describe the proposed MBR phrase table pruning approach in section IV. Section V provides the experimental setup and results, and the analysis of performance. Finally, we conclude this paper in section VI.

II. BACKGROUND AND MOTIVATION

For much future Internet of things, machine translation will be a useful technique to connect people and things. To improve the performance of machine translation, we propose MBR pruning method. The proposed MRB pruning method mainly relies on the phrase-based SMT and pivot-based SMT. In this section, we will first describe the traditional phrase-based SMT and the phrase-based SMT via a pivot language. Then, we describe the motivation of phrase table pruning for pivot-based SMT.

A. INTERNET OF THINGS

IoT as an emerging service model brings us much convenience; it connects all the staff we use to communicate today, all the household products, and many other "things" together to the Internet without human interference. With IoT, your intelligent home manager can communicate with your smartphone, fridge, air conditioner, telling you should buy some fruits back, current temperature of your house is a bit high, thus you could know the status of your home. These advantages are based on the incorporation of IoT and various technologies, such as web service [38], complex system [39], artificial intelligence [34], natural language processing [35]–[37], and machine translation.

B. PHRASE-BASED SMT

The traditional phrase-based SMT model is based on the noisy channel model [45], [46]. Given a source sentence s, the best target translation t_{best} can be obtained with the

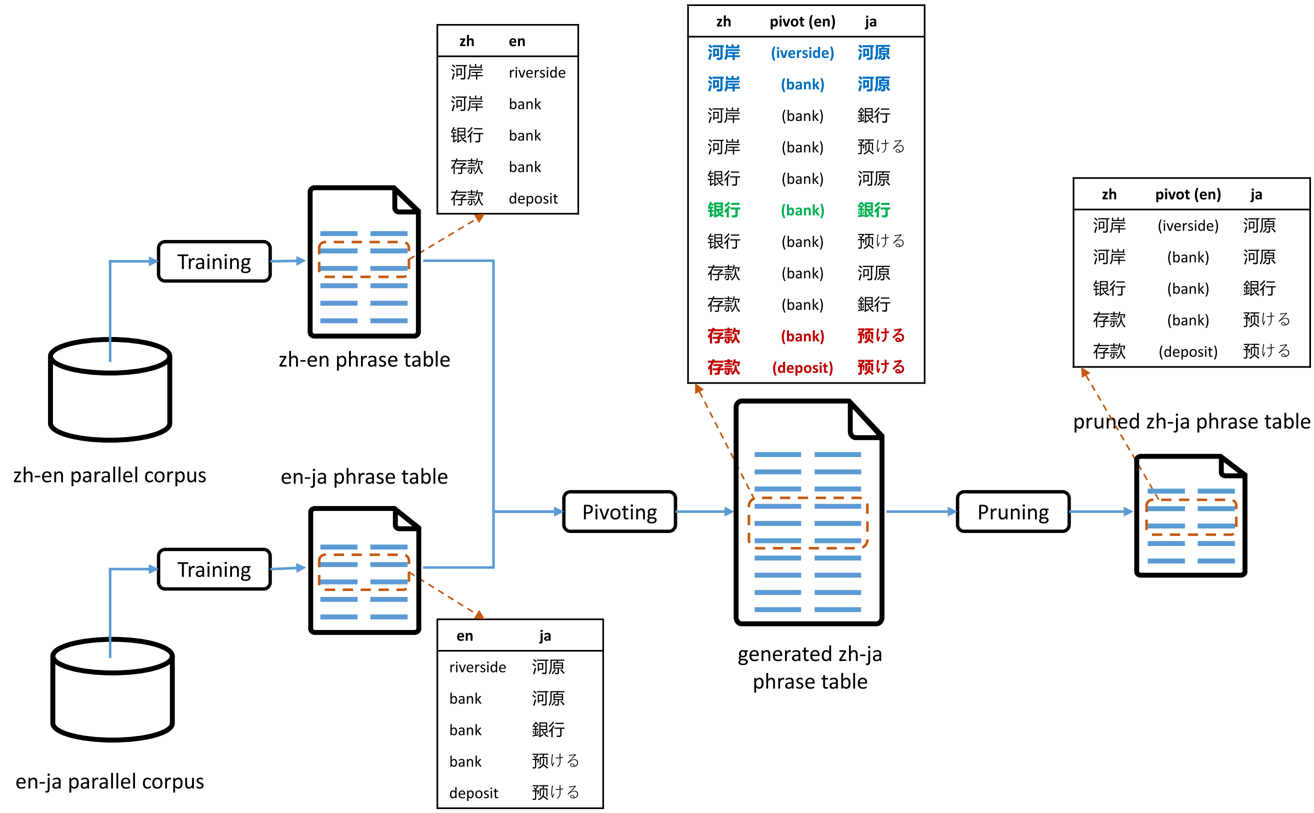


FIGURE 4. High-level overview of the MBR phrase table pruning framework.

following equation.

$$
\begin{aligned}
t_{best} &= \arg\max_t p(t|s) \\
&= \arg\max_t p(s|t)p(t)
\end{aligned}
\tag{1}
$$

where $p(s|t)$ is a translation model and $p(t)$ is the language model.

More often, a log-linear model is used to integrate more features instead of the noisy-channel model with the following equation:

$$
\begin{aligned}
t_{best} &= \arg\max_t p(t|s) \\
&= \arg\max_t \sum_{m=1}^{M} \lambda_m h_m(t, s)
\end{aligned}
\tag{2}
$$

where $h_m(t, s)$ represents a feature function, and λ_m represents the weight of the feature.

C. PHRASE-BASED SMT VIA A PIVOT LANGUAGE

It is hard to build a high-performance SMT system with the traditional phrase-based SMT model for language pairs with limited bilingual data. A pivot language is introduced, since there are large amounts of source-pivot and pivot-target bilingual data. These two sources of additional bilingual data make up for the data shortage of the original language pair.

Given the source-pivot and pivot-target translation models trained with the traditional phrase-based SMT model, the source-target translation model can be formulated with the following equation:

$$
\begin{aligned}
p(s|t) &= \sum_p p(s|t, p)p(p|t) \\
&\approx \sum_p p(s|p)p(p|t)
\end{aligned}
\tag{3}
$$

because $p(s|t, p)$ is estimated only in source-pivot bilingual corpora, it can be approximated to $p(s|p)$.

D. MOTIVATION OF PHRASE TABLE PRUNING IN PIVOT-BASED SMT

As we mentioned in section I, the ambiguities of the pivot language often lead to a wrong translation. In Fig. 4, we show the generation of the noises when pivoting the source-pivot and pivot-target phrase table.

We take Chinese-Japanese translation via English as an example. At first step, we train Chinese-English and English-Japanese phrase tables with large-scale Chinese-English parallel corpus and English-Japanese parallel corpus. The training method is the traditional phrase-based SMT described in subsection II.A. From Figure 4 we can find that the Chinese phrases can be correctly mapped to English phrases and the English phrases can also be correctly mapped to Japanese phrases.

Based on well-trained Chinese-English and English-Japanese phrase table, at second step, we merge the well-trained Chinese-English and English-Japanese phrase table with the pivot-based method described in subsection II.B. Thus, a huge and noisy Chinese-Japanese phrase table is generated. From the figure we can find that, due to the ambiguity of the English phrase "bank," many irrelevant Chinese-Japanese phrase pairs are mapped together. And more than 50% of the phrase pairs are improper. It is clear that we cannot obtain a high-quality translation with the noisy Chinese-Japanese phrase table. It is necessary to prune the generated phrase table to improve the quality of the translation.

III. RELATED WORK

A. MACHINE TRANSLATION IN IOT

Machine translation is a key technology in artificial intelligence and has been widely used in IoT. For example, Nakamura *et al.*, and Yun *et al.*, introduce the multilingual speech-to-speech translation system for mobile consumer devices [4], [5]. Khadivi and Ney [6] integration the speech recognition and machine translation in computer-assisted translation. Lavie *et al.* [7] proposed a multilingual speech communication over the Internet. These represent works show that machine translation, especially the multilingual machine translation plays an important role in IoT.

B. PIVOT-BASED SMT

According to the granularity of the pivot language, the pivot-based machine translations can be divided into 3 categories: corpus-level, sentence-level and phrase-level.

1) CORPUS LEVEL

For the corpus level, a pseudo source-target corpus is generated using the source-pivot and pivot-target translation systems [48]. There are two ways to obtain a pseudo source-target corpus by translating the pivot sentences. One is to translate all the pivot sentences in the pivot-target corpus to source sentences with the pivot-source translation system. Another is to translate all the pivot sentences in the source-pivot corpus to target sentences with the pivot-target translation system. If necessary, we can merge the two corpora into one pseudo corpus. The disadvantage of this method is the low quality of the translated sentences. It is difficult to generate a high-quality translation system with a corpus filled with machine-translated sentences.

2) SENTENCE LEVEL

At the sentence level, the source-pivot translation system and the pivot-target translation system are connected into a single source-target translation system [15]. Given a source sentence, the connected translation system works with two steps: 1) It translates the source sentence into the pivot sentence; 2) It translates the pivot sentence into the target sentence. At each step, the translation system may generate n-best translation results. The optimal translation result can be selected with the minimum Bayes-risk system combination method [26], [47]. Because the connected translation system needs to translate at least twice, the speed of this method is slower than that of the other methods. Another disadvantage is that the translation errors of the source-pivot system will be transferred to the pivot-target system, and the best translation of the source-pivot system may not be the best one to produce an adequate target language output.

3) PHRASE LEVEL

The triangulation method is a representative work at the phrase level [16], [25]. It directly connects the source-pivot phrase pairs and the pivot-target phrase pairs with identical pivot phrases. The probabilities of source-target phrase pairs are induced by multiplying the corresponding probabilities of the source-pivot and pivot-target phrase pairs. The triangulation method has been shown to work better than the other pivot approaches [15].

C. PHRASE TABLE PRUNING

Phrase table pruning is an important issue in statistical machine translation, especially for pivot-based SMT. Because the phrase table of pivot-based SMT is derived from the source-pivot and pivot-target phrase tables, it is unavoidable that many wrong and redundant translation rules will be generated in the phrase table.

Many algorithms have been proposed to deal with this problem [19]. In this subsection, we will introduce some typical pruning methods.

1) ABSOLUTE PRUNING

Absolute pruning methods rely on the statistical information in the phrase table. The translation rule table can be described as a mapping of phrase s in a source language to phrase t in a target language. Given a phrase pair (s, t), the phrase translation probability $p(t|s)$, and the reverse translation probability $p(s|t)$, the co-occurrence counts $c(s, t)$ can be derived from the bilingual corpus. If the translation probability or the co-occurrence count is below an empirical threshold, the phrase pair (s, t) is pruned.

The absolute pruning method is simple and effective. Zhu *et al.* [49] uses this method to prune a phrase table in pivot-based SMT.

2) SIGNIFICANCE PRUNING

Johnson *et al.* [50] proposed the significance pruning method; Tomeh *et al.* [51] extended the work by considering the complexity of the phrase pairs.

The idea of significance pruning is to test whether a source phrase s and a target phrase t co-occur more frequently in a bilingual corpus than they would just by chance. Given a bilingual corpus, the count of a source phrase $c(s)$, the count of a target phrase $c(t)$, and the co-occurrence count of the source and target phrases $c(s, t)$ can be derived. Thus, a *p-value*, which can represent the spuriousness of the phrase pair, is calculated. The phrase pairs with a high *p-value* will be pruned.

3) RELEVANCE PRUNING

The relevance pruning method [52] aims to prune the phrase pairs that are least used when translating the corpus. At the translation stage, the translator selects suitable phrases to generate output sentences. Thus, the relevance pruning method counts the number of times that the phrases occurred during the translation and prunes the less used phrases.

4) ENTROPY-BASED PRUNING

The main idea of the entropy-based pruning method [53], [54] is to remove the phrase pairs that can be derived using smaller phrase pairs with similar probability. The goal of the entropy-based pruning method is to remove redundant phrases, while the other pruning methods usually try to remove low-quality or unreliable phrases.

Formally, the entropy-based pruning method aims to maximum the similarity between the pruned model $p'(t|s)$ and the original model $p(t|s)$. The conditional relative entropy method measures the model similarity with the following equation:

$$\begin{aligned} D(p(t|s)||p'(t|s)) &= \sum_{s} p(s) \sum_{t} p(t|s) \log[\frac{p(t|s)}{p'(t|s)}] \\ &= \sum_{s,t} p(t,s)[\log p(t|s) - \log p'(t|s)] \end{aligned} \tag{4}$$

Thus, given an threshold τ_E, a phrase pair (s, t) is pruned if:

$$p(t,s)[\log p(t|s) - \log p'(t|s)] < \tau_E \tag{5}$$

5) CONTEXT-BASED PRUNING

Due to the ambiguities of the pivot language, source and target phrases with different meanings may be wrongly matched. To solve the ambiguities of the languages, many context-based method were proposed [40]–[44]. In pivot-based machine translation, context-based pruning method uses a context vector to identify the exact meaning of the pivot language [24].

Given source-pivot and pivot-target corpora, the source context vector S, the target context vector T and the pivot context vector P can be calculated following Rapp [63]. Because the length of source context vectors and the length of target context vectors may not be equal, so they are not comparable. To solve this problem, we can map the source context vector and target context vector to pivot context vector with the same length. Hence, the cosine similarity can be used to calculate the similarity between source context vector and target context vector. The phrase pairs with a lower similarity will be deleted.

D. MINIMUM BAYES-RISK

Minimum Bayes-risk related methods are widely used in IoT fields [55], [56], automatic speech recognition [57]–[61], text classification, and some other fields. In statistical machine translation, the minimum Bayes-risk is usually used in system combination [62].

Given a source sentence s, the MT system generates *n-best* target sentences. With the minimum Bayes-risk approach, the hypothesis output with the lowest Bayes risk would be selected as the final translation.

IV. MINIMUM BAYES-RISK TRANSLATION RULE PRUNING

The minimum Bayes-risk has usually been used for system combination in machine translation. To the best of our knowledge, no one has used it for phrase table pruning or any other rule pruning. The MBR translation rule pruning method aims to prune the translation rule that has the least expected loss under a probability model. The motivation is the hypothesis that reasonable translation rules are similar while noisy translation rules differ widely.

A. THE FRAMEWORK OF MINIMUM BAYES-RISK PRUNING

The translation rule table can be described as a mapping of phrases s in a source language to phrases t in a target language, which is a many-to-many relationship.

Given a phrase s, if there exists a reference phrase (a right translation of word sequence s) t, the quality of the hypothesis phrase t' can be measured by the loss function $L(t, t')$. With the loss function $L(t, t')$ and the phrase translation probability $p(t|s)$, we can prune the hypothesis phrase t with the following equation.

$$\begin{aligned} \hat{t} &= \arg\min_{t' \in T} R(t') \\ &= \arg\min_{t' \in T} \sum_{t \in T} P(t|s) \cdot L(t,t') \end{aligned} \tag{6}$$

where $R(t')$ denotes the Bayes risk of candidate translation t' under loss function $L(t, t')$, and T represents the space of translations.

B. BI-DIRECTIONAL MINIMUM BAYES-RISK PRUNING

As we mentioned above, one word sequence s can map to many word sequence t, and vice versa. Thus, with the underlying probability models $p(t|s)$ and $p(s|t)$, we can modify the MBR pruning to a bi-directional MBR pruning with the following equation.

$$\begin{aligned} \hat{t} &= \arg\min_{t' \in T, s' \in S} R(t')R(s') \\ &= \arg\min_{t' \in T, s' \in S} \sum_{t \in T} P(t|s) \sum_{s \in S} P(s|t) \cdot L(t,t') \cdot L(s,s') \end{aligned} \tag{7}$$

where $R(t')$ denotes the Bayes risk of candidate translation t' and $R(s')$ denotes the Bayes risk of reverse translation s' under loss functions $L(t, t')$ and $L(s, s')$. T and S represent the spaces of translation and reverse translation.

C. TRILINGUAL-CONSTRAINED MINIMUM BAYES-RISK PRUNING

For the pivot-based SMT, the mapping of a phrase s in a source language to a phrase t in a target language is generated from a phrase p in a pivot language. Thus, we can modify the bi-directional MBR pruning framework to a trilingual-constrained minimum Bayes-risk pruning framework with the following equation.

$$\begin{aligned}\hat{t} &= \underset{t'\in T, s'\in S}{\arg\min}\; R(t')R(s') \\ &= \underset{t'\in T, s'\in S}{\arg\min} \sum_{t\in T} P(t|s) \sum_{s\in S} P(s|t)\cdot L(t,t')\cdot L(s,s')\cdot L(p,p')\end{aligned} \tag{8}$$

where $L(p, p')$ is the loss function of pivot phrases.

D. LOSS FUNCTION

In the previous section, we assume that a reference word sequence t is ready for the calculation of the loss function. However, in fact, we do not have the reference word sequence t. To find the reference word sequence t, we can make two reasonable inferences:

1. Most of the translation rules in the phrase table are correct.
2. Reasonable translation rules are similar, while the noisy translation rules differ widely.

Fig. 3 is a simple example of the two reasonable inferences. If most people are talking about a dog, then talk about a cat or another thing is noise. Thus, if we want to evaluate the quality of the hypothesis word sequence t', we can use all the other word sequences as the reference word sequence t.

1) BLEU (BILINGUAL EVALUATION UNDERSTUDY)

In machine translation, BLEU is always used to evaluate the quality of the translation. The BLEU score value ranges from 0 to 1, and a larger value reflects a higher similarity.

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-\frac{r}{c})} & \text{if } c \le r \end{cases} \tag{9}$$

$$BLEU = BP \cdot \exp(\sum_{n=1}^{N} w_n \log p_n) \tag{10}$$

where BP is the brevity penalty, r is the length of the reference word sequence, c is the length of the hypothesis word sequence, and p_n is the geometric average of the modified n-gram precisions.

In this paper, we use $BLEU$ as the loss function to calculate the similarity of the hypothesis word sequence t' and the reference word sequence t.

2) WER (WORD ERROR RATE)

The word error rate is a metric that is commonly used to evaluate the performance of a machine translation system. It measures the Levenshtein distance of the hypothesis word sequence t' and the reference word sequence t with the following equation:

$$WER = \frac{S+D+I}{N} = \frac{S+D+I}{S+D+C} \tag{11}$$

where S is the number of substitutions, D is the number of deletions, I is the number of insertions, C is the number correct words, and N is the total number of words in the reference.

E. LETTER-BASED LOSS FUNCTION

The traditional word-based BLEU and WER are used for system-level machine translation evaluation; they are ineffective for sentence-level machine translation evaluation. Thus, we also apply letter-based BLEU and letter-based WER to evaluate the similarity of the hypothesis word sequence and the reference word sequence.

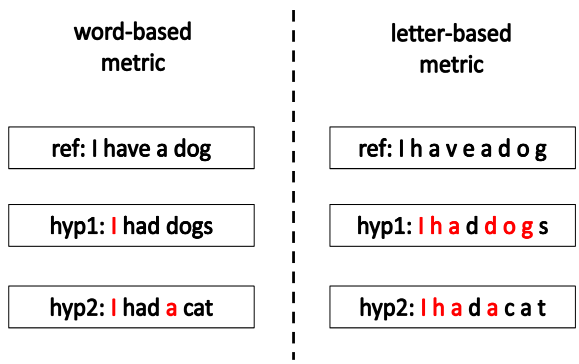


FIGURE 5. A comparison of word-based loss function and letter-based loss function.

Fig. 5 is a comparison of word-based loss function and letter-based loss function. From the figure, we can see that the letter-based loss function can capture fine-grained information in the phrase table. With the word-based metric, the phrase "I had a cat" is closer to the reference phrase "I have a dog" than the phrase "I had dogs." However, we know that, in fact, "I had dogs" is better than "I had a cat" in this example. If we use the letter-based metric, the result is exactly the opposite.

V. EXPERIMENTS

A. TRANSLATION SYSTEM AND EVALUATION METRIC

The translation system used in this paper is a phrase-based translation system under a log-linear framework that is analogous to the widely used Moses [66]. The system contains a phrase translation model, a lexical translation model, a lexical reordering model and a language model. The word alignment is generated by an in-house word alignment system that is similar to GIZA++ [64]; the heuristic "grow-diag-final" refinement rule is used during the word aligning stage [65].

In this paper, we pay attention to two metrics, one is the compression ratio of the phrase table, and the other is the quality of the translation. Our aim is to reduce the quantity of the phrase table while maintaining the performance of the translation. The translation quality is evaluated by case-insensitive BLEU-4 [67]. The paired bootstrap resampling method [68] is used to test the statistical significance using 95% confidence intervals.

B. DATA SETS

The Europarl[1] corpus [69] is a multilingual corpus that is always used in WMT[2] translation tasks. It contains 21 European languages, and 11 languages with more than 1 million sentences are used in our experiments. They are Danish (da), German (de), Greek (el), English (en), Spanish (es), Finnish (fi), French (fr), Italian (it) Dutch (nl) Portuguese (pt) and Swedish (sv). Because the Europarl corpus is a multilingual parallel corpus, it is unfair to train the source-pivot and pivot-target translation model under a trilingual scenario. We divide the training data according to the year of the data: odd years for training the source-pivot translation model and even years for training the pivot-target translation model.

To verify the effectiveness of our pruning method more thoroughly, we first compare three MBR frameworks and compare the word-based loss function with letter-based loss function in subsection V.D. Based on the conclusion of V.D, we designed two sets of experiments. One is translating between different language pairs with English as the pivot language. The performance of the phrase table pruning method may be affected by many factors (e.g., the ambiguity of the pivot language). To avoid any influence of the pivot language, another experimental set was translated from Portuguese to Swedish via different pivot languages.

Given that we are limited by the length of the paper, we cannot list all the language pairs of the Europarl corpus in all experimental sets. Some common language pairs were considered in subsection V.D; they are German, French and Spanish. And in subsection V.E, we list all the language pairs on Europarl data based on subsection V.D.

The detailed training data are listed in Table 1 and Table 2.

TABLE 1. Training data for experiments using English as the pivot language.

Language Pairs (src-pvt)	Sentence Pairs #	Language Pairs (src-pvt)	Sentence Pairs #
da-en	974,189	en-da	953,002
de-en	983,411	en-de	905,167
el-en	609,315	en-el	596,331
es-en	968,527	en-es	961,782
fi-en	998,429	en-fi	903,689
fr-en	989,652	en-fr	974,637
it-en	934,448	en-it	938,573
nl-en	982,696	en-nl	971,379
pt-en	967,816	en-pt	960,214
sv-en	960,631	en-sv	869,254

Several test sets have been released for the Europarl corpus. In our experiments, we use WMT2007[3] as our development data and WMT2008[4] as our test data. The original test data includes 4 languages, and extended versions with 11 languages of these test sets are available from the EuroMatrix[5] project. Table 3 summarizes the test sets.

[1] http://www.statmt.org/europarl/
[2] http://www.statmt.org/wmt17/translation-task.html
[3] http://www.statmt.org/wmt07/shared-task.html

TABLE 2. Training data for experiments using different pivot language.

Language Pairs (src-pvt)	Sentence Pairs #	Language Pairs (src-pvt)	Sentence Pairs #
pt-da	941,876	da-sv	865,020
pt-de	939,932	de-sv	814,678
pt-el	591,429	el-sv	558,765
pt-es	934,783	es-sv	827,964
pt-fi	950,588	fi-sv	872,182
pt-fr	954,637	fr-sv	860,272
pt-it	900,185	it-sv	813,000
pt-nl	945,997	nl-sv	864,675

TABLE 3. Statistics of test sets.

Test Set	Sentence #	Reference #
WMT07	2,000	1
WMT08	2,000	1

C. OUR PHRASE TABLE PRUNING SYSTEM AND THE BASELINE SYSTEM

To make a comprehensive comparison, we reimplement two baseline systems to compare with our system. The basic baseline system is the triangulation method based on the pivot approach with no pruning method [16]. The triangulation method in combination with the entropy-based pruning method is another baseline system.

Under the three minimum Bayes-risk pruning framework, we develop several phrase table pruning methods with different loss functions including: 1) word-based WER, 2) word-based BLEU, 3) letter-based WER, and 4) letter-based BLEU. In addition, we also try to combine the 4 loss functions above and the entropy-based pruning method under a log-linear framework.

The pruning procedure runs following three steps: 1)fix the source phrases and prune the target phrases, 2)fix the target phrase and prune the source phrases, and 3)merge the two pruned phrase table.

D. COMPARISON OF DIFFERENT MBR FRAMEWORK AND DIFFERENT LOSS FUNCTION

We performed our experiments among 3 languages with English as the pivot language. Under the traditional MBR framework, the bi-directional MBR framework and the trilingual-constrained MBR framework, we test word-based loss function and letter-based loss function.

Table 4 shows the details of the experiments. The compression ratio of the experiments is set to 40% by experience. The trends with other compression ration is similar with 40%.

[4] http://www.statmt.org/wmt08/shared-task.html
[5] http://matrix.statmt.org/test_sets/list

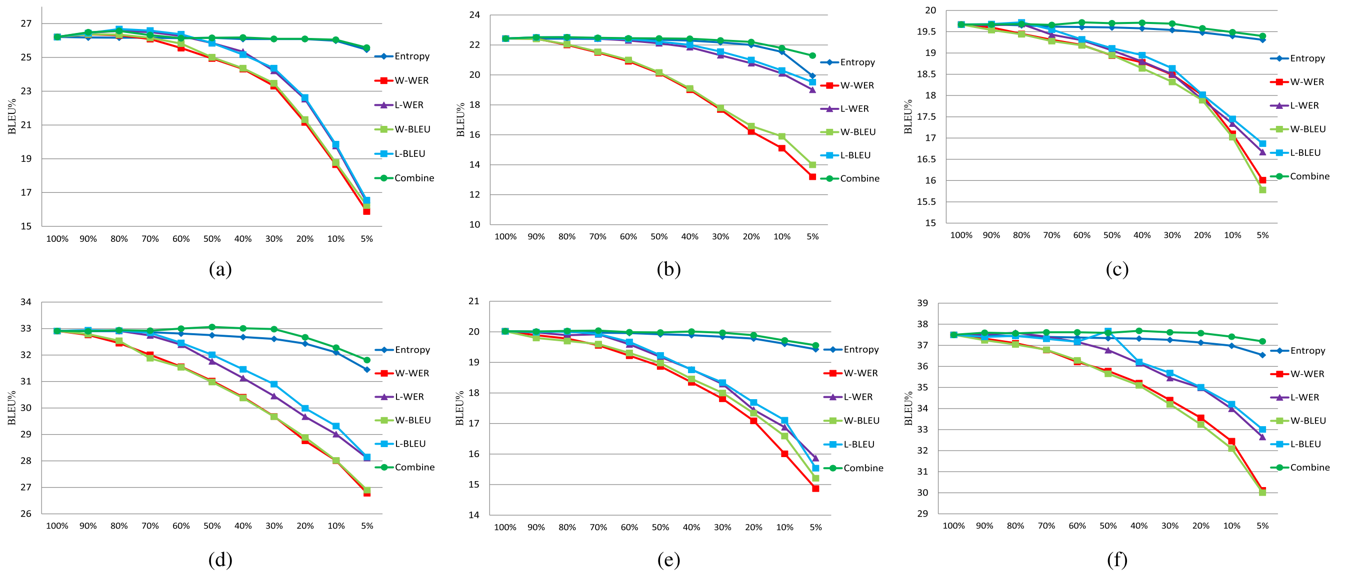


FIGURE 6. **Results using one pivot language under different compress ratio. (a) Translating from German to Spanish. (b) Translating from German to French. (c) Translating from Spanish to German. (d) Translating from Spanish to French. (e) Translating from French to German. (f) Translating from French to Spanish.**

TABLE 4. **Comparison of different MBR framework and different granularity of loss function.**

		W-WER	L-WER	W-BLEU	L-BLEU
de-es	MBR	23.89	24.71	23.81	24.81
	Bi-MBR	24.01	24.87	23.91	24.79
	Tri-MBR	**24.32***	**25.34***	**24.36***	**25.17***
de-fr	MBR	18.92	21.50	18.45	21.43
	Bi-MBR	18.91	21.71	19.07	21.78
	Tri-MBR	19.01	**21.85***	**19.10***	**22.01***
es-de	MBR	18.14	18.35	18.21	18.44
	Bi-MBR	18.65	18.72	18.53	18.79
	Tri-MBR	**18.78***	**18.80***	**18.64***	**18.95***
es-fr	MBR	30.17	30.79	30.14	30.98
	Bi-MBR	30.13	31.1	30.21	31.29
	Tri-MBR	30.41	31.13	30.38	**31.46***
fr-de	MBR	18.01	18.52	18.11	18.31
	Bi-MBR	18.21	18.67	18.3	18.59
	Tri-MBR	18.35	18.77	18.46	**18.76***
fr-es	MBR	34.71	35.83	34.87	35.79
	Bi-MBR	34.89	35.81	35.01	36.01
	Tri-MBR	**35.21***	36.15	35.10	**36.21***

From the table, we can draw several conclusions:

1. For all language pairs, the trilingual-constrained MBR framework (Tri-MBR in table 4) performs better than bi-directional MBR framework (Bi-MBR in table 4) and traditional MBR framework (MBR in table 4).
2. For all language pairs, the letter-based loss function performs better than the word-based loss function.
3. There is no obvious difference between WER and BLEU.

E. RESULTS USING ONE PIVOT LANGUAGE UNDER DIFFERENT COMPRESS RATIO

As it is the most commonly spoken language in the world, English was used as the pivot language in this subsection.

Fig. 6 show trends in the translation performance with the reduction of the compression ratio. From the figures, we can see that the entropy-based pruning method and the combined method can maintain the performance of the translation with the reduction of the compression ratio. The MBR pruning method can improve the performance of the translation when only a few phrase pairs are deleted. However, with the reduction of the compression ratio, the performance dropped rapidly. The combined method performed better than all other methods when most of the phrase pairs were deleted.

F. RESULTS ON ALL EUROPARL DATA

According to subsection V.D, 40% is the calculated compression ratio for phrase table pruning. Because the combination of the entropy-based pruning method and the MBR pruning method is better than the MBR pruning method alone, we only used the combined method in this subsection.

In this subsection, we list all the language pairs in the Europarl Corpus under 40% compression ratios with English as the pivot language.

Several observations can be made from Table 5.

1. In all 90 language pairs, our combined method achieves general improvements over the entropy-based method.
2. Among the 90 language pairs, our combined method is significantly better than the entropy-based method in 24 language pairs. It indicates that the translation direction may affect the performance of the method.
3. The pruning method, including the entropy-based method and the combined methods, can reduce the size of the phrase table with the performance remaining almost unchanged.

TABLE 5. Experimental results on Europarl with different translation directions.

	TGT \ SRC	da	de	el	es	fi	fr	it	nl	pt	sv
No-prune			19.83	20.46	27.59	14.76	24.11	20.49	22.26	24.38	28.33
Entropy	**da**	-	19.7	20.28	27.47	14.61	24.09	20.34	22.22	24.21	28.19
Combine			**20.01***	20.48	27.56	14.79	24.1	20.58	22.39	24.28	28.35
No-prune		23.35		19.83	26.21	12.72	22.43	18.82	23.74	23.05	21.17
Entropy	**de**	23.24	-	19.83	26.09	12.42	22.29	18.8	23.44	22.97	20.96
Combine		**23.68***		19.91	26.18	**12.79***	22.42	18.81	**23.79***	23.1	21.16
No-prune		23.24	18.12		32.28	13.31	27.35	23.19	20.8	27.62	22.7
Entropy	**el**	23.19	17.88	-	31.96	13.24	27.2	22.92	20.59	27.6	22.38
Combine		23.22	18.14		32.11	13.29	27.28	**23.21***	20.79	27.85	**22.69***
No-prune		25.34	19.67	27.24		13.93	32.91	27.67	22.37	34.73	24.83
Entropy	**es**	25.1	19.58	27.01	-	13.78	32.68	27.51	22.2	34.39	24.57
Combine		25.33	19.71	27.23		13.94	**33.01***	27.67	22.22	**34.68***	24.69
No-prune		18.29	13.2	14.72	20.17		17.52	14.76	15.5	17.3	16.63
Entropy	**fi**	18.21	13.11	14.7	19.89	-	17.48	14.68	15.34	17.13	16.32
Combine		18.35	13.19	14.81	20.05		17.59	14.86	15.56	17.28	16.58
No-prune		25.67	20.02	26.58	37.5	13.9		28.51	22.65	33.81	24.64
Entropy	**fr**	25.54	19.89	26.39	37.32	13.7	-	28.41	22.31	33.56	24.49
Combine		**25.68***	20.01	26.59	**37.69***	13.93		28.55	22.58	**33.92***	24.7
No-prune		22.63	17.81	24.24	34.36	13.2	30.16		21.37	30.84	22.12
Entropy	**it**	22.38	17.7	24.23	34.01	13.17	29.82	-	21.11	30.65	22.01
Combine		22.54	17.82	24.46	**34.32***	13.21	**30.23***		**21.43***	30.9	22.23
No-prune		22.49	19.86	18.56	24.69	11.96	21.48	18.36		21.71	19.83
Entropy	**nl**	22.38	19.78	18.43	24.44	11.91	21.21	18.1	-	21.45	19.76
Combine		22.65	19.95	18.65	24.58	11.95	**21.58***	19.29		**21.89***	19.8
No-prune		24.08	19.11	25.3	36.59	13.33	32.47	28.08	21.52		22.9
Entropy	**pt**	23.89	18.97	25.09	36.01	13.12	32.12	27.78	21.19	-	22.72
Combine		24.1	18.99	25.31	**36.43***	13.31	**32.50***	**28.06***	**21.54***		22.91
No-prune		31.24	20.26	22.06	29.21	15.39	25.63	21.25	22.3	25.6	
Entropy	**sv**	30.81	20.11	21.73	29.01	15.21	25.23	21.09	22.01	25.38	-
Combine		**31.21***	20.2	**22.01***	**29.32***	15.34	**25.64***	21.26	**22.42***	25.58	

4. The improvements of our approach are not equal for different language pairs. The improvement ranges from 0.01(de-it) to 0.44(de-da, nl-pt).

G. RESULTS USING DIFFERENT PIVOT LANGUAGES

The performance of the phrase table pruning method may be affected by many factors e.g., the ambiguity of the pivot language. To avoid the influence of the pivot language, we also test our approach on translating from Portuguese to Swedish via different pivot languages.

TABLE 6. Experimental results on Portuguese-Swedish via different pivot languages.

	S-P-T	BLEU		S-P-T	BLEU
No-prune		22.49	**No-prune**		20.26
Entropy	pt-da-sv	22.31	**Entropy**	pt-fi-sv	20.12
Combine		22.5	**Combine**		20.35
No-prune		21.76	**No-prune**		22.89
Entropy	pt-de-sv	21.51	**Entropy**	pt-fr-sv	22.46
Combine		21.65	**Combine**		**22.78***
No-prune		21.37	**No-prune**		22.79
Entropy	pt-el-sv	21.03	**Entropy**	pt-it-sv	22.41
Combine		**21.32***	**Combine**		**22.80***
No-prune		22.8	**No-prune**		21.36
Entropy	pt-es-sv	22.68	**Entropy**	pt-nl-sv	21.12
Combine		22.92	**Combine**		21.29

Table 6 summarizes the results using different pivot languages. From the table, we can find that although we changed the pivot language when translating from Portuguese to Swedish, the performance of the pruning method is similar to when using English as the pivot language. Thus, considering that English is the most common used language in the world, it is better to use English as the pivot language.

VI. CONCLUSION

In this paper, we present a minimum Bayes-risk phrase table pruning method for pivot-based SMT. Under a minimum Bayes-risk framework, we apply word-based and letter-based machine translation metrics as the loss functions. We also try to combine the MBR-based phrase table pruning method and the entropy-based phrase table pruning method to achieve a better performance. Experimental results on Europarl data show that our method can both reduce the scalar of the phrase table and improve the performance of the translation. As a widely use method, the minimum Bayes-risk can also be used in many IoT fields, and the conclusion of this paper will give a useful reference to many IoT research fields.

REFERENCES

[1] J. Peng, C. Wang, Y. Shao, and J. Xu, "Visual search efficiency evaluation method for potential connected vehicles on sharp curves," *IEEE Access*, vol. 6, pp. 41827–41838, 2018.

[2] J.-Q. Li, F. R. Yu, G. Deng, C. Luo, Z. Ming, and Q. Yan, "Industrial Internet: A survey on the enabling technologies, applications, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1504–1526, 3rd Quart., 2017.

[3] J. Wu, Y. D. Zhang, M. G. Amin, and M. Uysal, "Multiple-relay selection in amplify-and-forward cooperative wireless networks with multiple source nodes," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, pp. 256–268, 2012.

[4] S. Nakamura *et al.*, "The ATR multilingual speech-to-speech translation system," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 2, pp. 365–376, Mar. 2006.

[5] S. Yun, Y. J. Lee, and S. H. Kim, ''Multilingual speech-to-speech translation system for mobile consumer devices,'' *IEEE Trans. Consum. Electron.*, vol. 60, no. 3, pp. 508–516, Aug. 2014.

[6] S. Khadivi and H. Ney, ''Integration of speech recognition and machine translation in computer-assisted translation,'' *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 8, pp. 1551–1564, Nov. 2008.

[7] A. Lavie, F. Pianesi, and L. Levin, ''The NESPOLE! System for multilingual speech communication over the Internet,'' *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 5, pp. 1664–1673, Sep. 2006.

[8] H. Nakayama and N. Nishida, ''Zero-resource machine translation by multimodal encoder–decoder network with multimedia pivot,'' *Mach. Transl.*, vol. 31, nos. 1–2, pp. 49–64, 2017.

[9] A. Karakanta, J. Dehdari, and J. van Genabith, ''Neural machine translation for low-resource languages without parallel corpora,'' *Mach. Transl.*, vol. 32, nos. 1–2, pp. 167–189, 2018.

[10] H. Setiawan, Z. Huang, and R. Zbib, ''BBN's low-resource machine translation for the LoReHLT 2016 evaluation,'' *Mach. Transl.*, vol. 32, nos. 1–2, pp. 45–57, 2018.

[11] N. Malandrakis, A. Ramakrishna, V. Martinez, T. Sorensen, D. Can, and S. Narayanan, ''The ELISA situation frame extraction for low resource languages pipeline for LoReHLT'2016,'' *Mach. Transl.*, vol. 32, nos. 1–2, pp. 127–142, 2018.

[12] U. Hermjakob *et al.*, ''Incident-driven machine translation and name tagging for low-resource languages,'' *Mach. Transl.*, vol. 32, nos. 1–2, pp. 59–89, 2018.

[13] M. S. Rasooli, N. Farra, A. Radeva, T. Yu, and K. McKeown, ''Cross-lingual sentiment transfer with limited resources,'' *Mach. Transl.*, vol. 32, nos. 1–2, pp. 143–165, 2018.

[14] R. Gabbard, J. DeYoung, C. Lignos, M. Freedman, and R. Weischedel, ''Combining rule-based and statistical mechanisms for low-resource named entity recognition,'' *Mach. Transl.*, vol. 32, nos. 1–2, pp. 31–43, 2018.

[15] M. Utiyama and H. Isahara, ''A comparison of pivot methods for phrase-based statistical machine translation,'' in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, Rochester, NY, USA, Apr. 2007, pp. 484–491.

[16] H. Wu and H. Wang, ''Pivot language approach for phrase-based statistical machine translation,'' *J. Mach. Transl.*, vol. 21, no. 3, pp. 165–181, 2007.

[17] N. Bertoldi, M. Barbaiani, M. Federico, and R. Cattoni, ''Phrase-based statistical machine translation with pivot languages,'' in *Proc. Int. Workshop Spoken Lang. Transl. (IWSLT)*, Honolulu, HI, USA, Oct. 2008, pp. 143–149.

[18] A. El Kholy, N. Habash, G. Leusch, E. Matusov, and H. Sawaf, ''Language independent connectivity strength features for phrase pivot statistical machine translation,'' in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, Sofia, Bulgaria, vol. 2, Aug. 2013, pp. 412–418.

[19] M. Tu, Y. Zhou, and C. Zong, ''Exploring diverse features for statistical machine translation model pruning,'' *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 11, pp. 1847–1857, Nov. 2015.

[20] B. J. Dorr, ''Machine translation divergences: A formal description and proposed solution,'' *Comput. Linguistics*, vol. 20, no. 4, pp. 597–633, 1994.

[21] M. Paul, A. Finch, and E. Sumita, ''How to choose the best pivot language for automatic translation of low-resource languages,'' *ACM Trans. Asian Lang. Inf. Process.*, vol. 12, no. 4, 2013, Art. no. 14.

[22] T. Shimizu, Y. Ashikari, E. Sumita, J. Zhang, and S. Nakamura, ''NICT/ATR Chinese-Japanese-English speech-to-speech translation system,'' *Tsinghua Sci. Technol.*, vol. 13, no. 4, pp. 540–544, Aug. 2008.

[23] A. Kumaran, M. M. Khapra, and P. Bhattacharyya, ''NICT/ATR Chinese-Japanese-English speech-to-speech translation system,'' *ACM Trans. Asian Lang. Inf. Process.*, vol. 9, no. 4, 2010, Art. no. 13.

[24] S. T. Zahabi, S. Bakhshaei, and S. Khadivi, ''Using context vectors in improving a machine translation system with bridge language,'' in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, Sofia, Bulgaria, vol. 2, Aug. 2013, pp. 318–322.

[25] T. Cohn and M. Lapata, ''Machine translation by triangulation: Making effective use of multi-parallel corpora,'' in *Proc. 45th Annu. Meeting Assoc. Comput. Linguistics*, Prague, Czech Republic, Jun. 2007, pp. 728–735.

[26] J. González-Rubio, A. Juan, and F. Casacuberta, ''Minimum Bayes-risk system combination,'' in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, Hum. Lang. Technol.*, Portland, OR, USA, vol. 1, Jun. 2011, pp. 1268–1277.

[27] N. Ehling, R. Zens, and H. Ney, ''Minimum Bayes risk decoding for BLEU,'' in *Proc. 45th Annu. Meeting ACL Interact. Poster Demonstration Sessions*, Prague, Czech Republic, Jun. 2007, pp. 101–104.

[28] H. Xu, D. Povey, L. Mangu, and J. Zhu, ''Minimum Bayes risk decoding and system combination based on a recursion for edit distance,'' *Comput. Speech Lang.*, vol. 25, no. 4, pp. 802–828, 2011.

[29] K. C. Sim, W. J. Byrne, M. J. F. Gales, H. Sahbi, and P. C. Woodland, ''Consensus network decoding for statistical machine translation system combination,'' in *Proc. ICASSP*, Honolulu, HI, USA, Apr. 2007, pp. 105–108.

[30] R.-H. Zhang, Z.-C. He, H.-W. Wang, F. You, and K.-N. Li, ''Study on self-tuning tyre friction control for developing main-servo loop integrated chassis control system,'' *IEEE Access*, vol. 5, pp. 6649–6660, 2017.

[31] X. J. Sun, H. Zhang, W. J. Meng, R. H. Zhang, K. L. Li, and T. Peng, ''Primary resonance analysis and vibration suppression for the harmonically excited nonlinear suspension system using a pair of symmetric viscoelastic buffers,'' *Nonlinear Dyn.*, vol. 94, no. 4, pp. 1–23, 2018.

[32] H. Xiong, X. Zhu, and R. Zhang, ''Energy recovery strategy numerical simulation for dual axle drive pure electric vehicle based on motor loss model and big data calculation,'' *Complexity*, vol. 2018, Aug. 2018, Art. no. 4071743, doi: 10.1155/2018/4071743.

[33] L. Wang, L. Zhang, H. Li, Y. Ma, and R. Zhang, ''High selective production of 5-hydroxymethylfurfural from fructose by sulfonic acid functionalized SBA-15 catalyst,'' *Compos. B, Eng.*, vol. 156, pp. 88–94, Jan. 2019.

[34] Z. Huang, J. Peng, H. Lian, J. Guo, and W. Qiu, ''Deep recurrent model for server load and performance prediction in data center,'' *Complexity*, vol. 2017, Nov. 2017, Art. no. 8584252.

[35] B. J. Grosz, ''Smart enough to talk with us? Foundations and challenges for dialogue capable AI systems,'' *Comput. Linguistics*, vol. 44, no. 1, pp. 1–15, 2018.

[36] E. Shutova, L. Sun, E. D. Gutiérrez, P. Lichtenstein, and S. Narayanan, ''Multilingual metaphor processing: Experiments with semi-supervised and unsupervised learning,'' *Comput. Linguistics*, vol. 43, no. 1, pp. 71–123, 2017.

[37] P. Jansen, R. Sharp, M. Surdeanu, and P. Clark, ''Framing QA as building and ranking intersentence answer justifications,'' *Comput. Linguistics*, vol. 43, no. 2, pp. 407–449, 2017.

[38] Y. Liu, D. Pi, and L. Cui, ''Mining community-level influence in microblogging network: A case study on Sina Weibo,'' *Complexity*, vol. 2017, Dec. 2017, Art. no. 4783159.

[39] S. Hou, J. Lin, S. Zhou, M. Qin, W. Jia, and Y. Zheng, ''Deep hierarchical representation from classifying logo-405,'' *Complexity*, vol. 2017, Oct. 2017, Art. no. 3169149.

[40] K. Chen *et al.*, ''A neural approach to source dependence based context model for statistical machine translation,'' *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 2, pp. 266–280, Feb. 2018.

[41] 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 Trans. Audio, Speech, Language Process.*, vol. 26, no. 10, pp. 1727–1741, Oct. 2018.

[42] K. Chen, R. Wang, M. Utiyama, E. Sumita, and T. Zhao, ''Context-aware smoothing for neural machine translation,'' in *Proc. 8th Int. Joint Conf. Natural Lang. Process.*, Taipei, Taiwan, vol. 1, Nov./Dec. 2017, pp. 11–20.

[43] K. Chen *et al.*, ''Neural machine translation with source dependency representation,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Copenhagen, Denmark, Sep. 2017, pp. 2846–2852.

[44] K. Chen, T. Zhao, M. Yang, and L. Liu, ''Translation prediction with source dependency-based context representation,'' in *Proc. 31st AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, Feb. 2017, pp. 3166–3172.

[45] P. F. Brown, V. J. D. Pietra, S. A. D. Pietra, and R. L. Mercer, ''The mathematics of statistical machine translation: Parameter estimation,'' *Comput. linguistics*, vol. 19, no. 2, pp. 263–311, 1993.

[46] P. Koehn, F. J. Och, and D. Marcu, ''Statistical phrase-based translation,'' in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol.*, Edmonton, AB, Canada, May/Jun. 2003, pp. 48–54.

[47] K. Duh, K. Sudoh, X. Wu, H. Tsukada, and M. Nagata, ''Generalized minimum Bayes risk system combination,'' in *Proc. 5th Int. Joint Conf. Natural Lang. Process.*, Chiang Mai, Thailand, Nov. 2011, pp. 1356–1360.

[48] H. Wu and H. Wang, ''Revisiting pivot language approach for machine translation,'' in *Proc. Joint Conf. 47th Annu. Meeting ACL 4th Int. Joint Conf. Natural Lang. Process. (AFNLP)*, Suntec, Singapore, vol. 1, Aug. 2009, pp. 154–162.

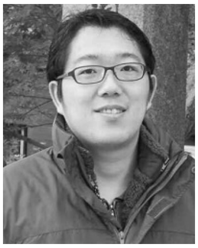
[49] X. Zhu, Z. He, H. Wu, H. Wang, C. Zhu, and T. Zhao, ''Improving pivot-based statistical machine translation using random walk,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Seattle, WA, USA, Oct. 2013, pp. 524–534.

[50] J. H. Johnson, J. Martin, G. Foster, and R. Kuhn, "Improving translation quality by discarding most of the phrasetable," in *Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn. (EMNLP-CoNLL)*, Prague, Czech Republic, Jun. 2007, pp. 967–975.
[51] N. Tomeh, C. Nicola, and D. Marc, "Complexity-based phrase-table filtering for statistical machine translation," in *Proc. Summit XII*, Ottawa, ON, Canada, Aug. 2009, pp. 144–151.
[52] M. Eck, S. Vogel, and A. Waibel, "Translation model pruning via usage statistics for statistical machine translation," in *Proc. Hum. Lang. Technol., Conf. North Amer. Chapter Assoc. Comput. Linguistics, Companion*, Rochester, NY, USA, Apr. 2007, pp. 21–24.
[53] W. Ling, J. Graça, I. Trancoso, and A. Black, "Entropy-based pruning for phrase-based machine translation," in *Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn.*, Jeju Island, South Korea, Jul. 2012, pp. 962–971.
[54] R. Zens, D. Stanton, and P. Xu, "A systematic comparison of phrase table pruning techniques," in *Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn.*, Jeju Island, South Korea, Jul. 2012, pp. 972–983.
[55] E. B. Flynn and M. D. Todd, "A Bayesian approach to optimal sensor placement for structural health monitoring with application to active sensing," *Mech. Syst. Signal Process.*, vol. 24, no. 4, pp. 891–903, 2010.
[56] W. Zhang, R. K. Mallik, and K. B. Letaief, "Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5761–5766, Dec. 2009.
[57] V. Goel and W. J. Byrne, "Minimum Bayes-risk automatic speech recognition," *Comput. Speech Lang.*, vol. 14, no. 2, pp. 115–135, 2000.
[58] V. Venkataramani, S. Chakrabartty, and W. Byrne, "*Gini*support vector machines for segmental minimum Bayes risk decoding of continuous speech," *Comput. Speech Lang.*, vol. 21, no. 3, pp. 423–442, 2007.
[59] P. C. Loizou, "Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 857–869, Sep. 2005.
[60] V. Goel and W. Byrne, "Minimum Bayes-risk methods in automatic speech recognition," in *Pattern Recognition in Speech and Language Processing*. Boca Raton, FL, USA: CRC Press, 2003, pp. 59–87.
[61] W. Byrne, "Minimum Bayes risk estimation and decoding in large vocabulary continuous speech recognition," *IEICE Trans. Inf. Syst.*, vol. E89-D, no. 3, pp. 900–907, 2006.
[62] M. Zhang, Y. Liu, H. Luan, and M. Sun, "Listwise ranking functions for statistical machine translation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 8, pp. 1464–1472, Aug. 2016.
[63] R. Rapp, "Automatic identification of word translations from unrelated English and German corpora," in *Proc. 37th Annu. Meeting Assoc. Comput. Linguistics Comput. Linguistics*, College Park, MD, USA, Jun. 1999, pp. 519–526.
[64] F. J. Och and H. Ney, "A comparison of alignment models for statistical machine translation," in *Proc. 18th Conf. Comput. Linguistics*, Saarbrücken, Germany, vol. 2, Jul./Aug. 2000, pp. 1086–1090.
[65] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol.*, Edmonton, AB, Canada, vol. 1, May/Jun. 2003, pp. 48–54.
[66] P. Koehn *et al.*, "Moses: Open source toolkit for statistical machine translation," in *Proc. 45th Annu. Meeting ACL Interact. Poster Demonstration Sessions*, Prague, Czech Republic, Jun. 2007, pp. 177–180.
[67] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics*, Philadelphia, PA, USA, Jul. 2002, pp. 311–318.
[68] P. Koehn, "Statistical significance tests for machine translation evaluation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Barcelona, Spain, Jul. 2004, pp. 3250–3257.
[69] P. Koehn, "Europarl: A parallel corpus for statistical machine translation," in *Proc. MT Summit X*, Phuket, Thailand, Sep. 2005, pp. 79–86.

XIAONING ZHU received the B.S. and M.S. degrees in computer science from the Harbin Institute of Technology, China, in 2007 and 2011, respectively, where he is currently pursuing the Ph.D. degree. His research interests include natural language processing and machine translation.

MUYUN YANG received the Ph.D. degree in computer science and engineering from the Harbin Institute of Technology. He is currently an Associate Professor with the School of Computer Science and Technology, Harbin Institute of Technology. His research interests include machine translation and information retrieval. He is a member of IEEE, CCF, and CIPSC.

TIEJUN ZHAO is currently a Professor of the School of Computer Science and Technology, Harbin Institute of Technology. He has published three academic books and 60 papers on journals and conferences in recent three years. His research interests include natural language understanding, content-based web information processing, and applied artificial intelligence. He has been a PC member on ACL, COLING in current five years and was also assigned an MT Track Co-chair on COLING 2014.



CONGHUI ZHU received the Ph.D. degree in computer science from the Harbin Institute of Technology. He is currently a Lecturer with the School of Computer Science and Technology, Harbin Institute of Technology. His research interests include Chinese medical information lexical processing, knowledge graph, machine translation, and machine learning.