Improving Pivot-Based Statistical Machine Translation Using Random Walk

Xiaoning Zhu[1*], Zhongjun He[2], Hua Wu[2], Haifeng Wang[2], Conghui Zhu[1], and Tiejun Zhao[1]
Harbin Institute of Technology, Harbin, China[1]
Baidu Inc., Beijing, China[2]
{xnzhu, chzhu, tjzhao}@mtlab.hit.edu.cn
{hezhongjun,wu_hua,wanghaifeng}@baidu.com

Abstract

This paper proposes a novel approach that utilizes a machine learning method to improve pivot-based statistical machine translation (SMT). For language pairs with few bilingual data, a possible solution in pivot-based SMT using another language as a "bridge" to generate source-target translation. However, one of the weaknesses is that some useful source-target translations cannot be generated if the corresponding source phrase and target phrase connect to different pivot phrases. To alleviate the problem, we utilize Markov random walks to connect possible translation phrases between source and target language. Experimental results on European Parliament data, spoken language data and web data show that our method leads to significant improvements on all the tasks over the baseline system.

1 Introduction

Statistical machine translation (SMT) uses bilingual corpora to build translation models. The amount and the quality of the bilingual data strongly affect the performance of SMT systems. For resource-rich language pairs, such as Chinese-English, it is easy to collect large amounts of bilingual corpus. However, for resource-poor language pairs, such as Chinese-Spanish, it is difficult to build a high-performance SMT system with the small scale bilingual data available.

The pivot language approach, which performs translation through a third language, provides a possible solution to the problem. The triangulation method (Wu and Wang, 2007; Cohn and Lapata, 2007) is a representative work for pivot-based machine translation. With a triangulation pivot approach, a source-target phrase table can be obtained by combining the source-pivot phrase table and the pivot-target phrase table. However, one of the weaknesses is that some corresponding source and target phrase pairs cannot be generated, because they are connected to different pivot phrases (Cui et al., 2013). As illustrated in Figure 1, since there is no direct translation between “很可口 henkekou” and “really delicious”, the triangulation method is unable to establish a relation between “很可口 henkekou” and the two Spanish phrases.

To solve this problem, we apply a Markov random walk method to pivot-based SMT system. Random walk has been widely used. For example, Brin and Page (1998) used random walk to discover potential relations between queries and documents for link analysis in information retrieval. Analogous to link analysis, the aim of pivot-based translation is to discover potential translations between source and target language via the pivot language.

[*] This work was done when the first author was visiting Baidu.

Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 524–534,
Seattle, Washington, USA, 18-21 October 2013.

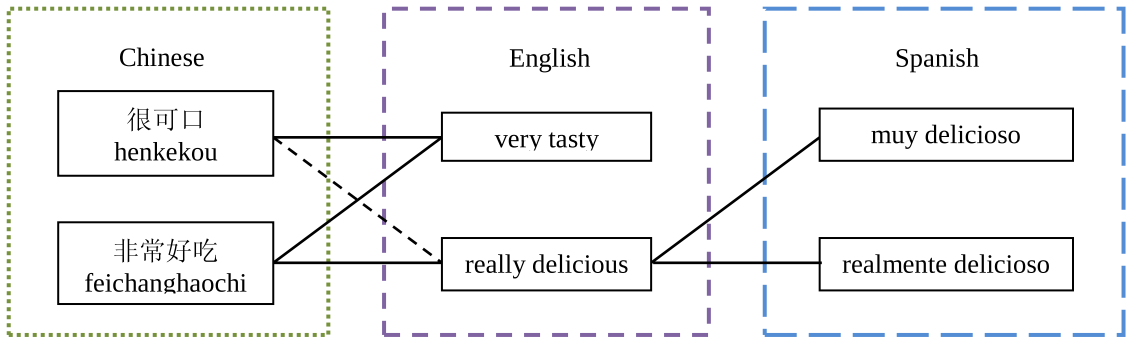

Figure 1: An example of random walk on phrase table. The dashed line indicates an implicit relation in the phrase table.

The goal of this paper is to extend the previous triangulation approach by exploring implicit translation relations using random walk method. We evaluated our approach in several translation tasks, including translations between European languages; Chinese-Spanish spoken language translation and Chinese-Japanese translation with English as the pivot language. Experimental results show that our approach achieves significant improvements over the conventional pivot-based method, triangulation method.

The remainder of this paper is organized as follows. In section 2, we describe the related work. We review the triangulation method for pivot-based machine translation in section 3. Section 4 describes the random walk models. In section 5 and section 6, we describe the experiments and analyze the performance, respectively. Section 7 gives a conclusion of the paper.

2 Related Work

Several methods have been proposed for pivot-based translation. Typically, they can be classified into 3 kinds of methods:

Transfer Method: Within the transfer framework (Utiyama and Isahara, 2007; Wang et al., 2008; Costa-jussà et al., 2011), a source sentence is first translated to *n* pivot sentences via a source-pivot translation system, and then each pivot sentence is translated to *m* target sentences via a pivot-target translation system. At each step (source to pivot and pivot to target), multiple translation outputs will be generated, thus a minimum Bayes-risk system combination method is often used to select the optimal sentence (González-Rubio et al., 2011; Duh et al., 2011). A problem with the transfer method is that it needs to decode twice. On one hand, the time cost is doubled; on the other hand, the translation error of the source-pivot translation system will be transferred to the pivot-target translation.

Synthetic Method: A synthetic method creates a synthetic source-target corpus using source-pivot translation model or pivot-target translation model (Utiyama et al., 2008; Wu and Wang, 2009). For example, we can translate each pivot sentence in the pivot-target corpus to source language with a pivot-source model, and then combine the translated source sentence with the target sentence to obtain a synthetic source-target corpus, and vice versa. However, it is difficult to build a high quality translation system with a corpus created by a machine translation system.

Triangulation Method: The triangulation method obtains source-target model by combining source-pivot and pivot-target translation models (Wu and Wang, 2007; Cohn and Lapata 2007), which has been shown to work better than the other pivot approaches (Utiyama and Isahara, 2007). As we mentioned earlier, the weakness of triangulation is that the corresponding source and target phrase pairs cannot be connected in the case that they connect to different pivot phrases.

3 The Triangulation Method

In this section, we review the triangulation method for pivot-based translation.

With the two additional bilingual corpora, the source-pivot and pivot-target translation models can be trained. Thus, a pivot model can be obtained by merging these two models. In the translation model, the phrase translation probability and the lexical weight are language dependent, which will be introduced in the next two sub-sections.

3.1 Phrase Translation Probability

The triangulation method assumes that there exist translations between phrases $\bar{s}$ and phrase $\bar{p}$ in source and pivot languages, and between phrase $\bar{p}$ and phrase $\bar{t}$ in pivot and target languages. The phrase translation probability ϕ between source and target languages is determined by the following model:

$$\begin{aligned}\phi(\bar{s} \mid \bar{t}) &= \sum_{\bar{p}} \phi(\bar{s} \mid \bar{p}, \bar{t}) \phi(\bar{p} \mid \bar{t}) \\ &= \sum_{\bar{p}} \phi(\bar{s} \mid \bar{p}) \phi(\bar{p} \mid \bar{t})\end{aligned} \quad (1)$$

3.2 Lexical Weight

Given a phrase pair $(\bar{s}, \bar{t})$ and a word alignment a between the source word positions $i = 1, \ldots, n$ and the target word positions $j = 0, 1, \ldots, m$, the lexical weight of phrase pair $(\bar{s}, \bar{t})$ can be calculated with the following formula (Koehn et al. 2003) :

$$p_{\xi}(\bar{s} \mid \bar{t}, a) = \prod_{i=1}^{n} \frac{1}{|\{j \mid (i,j) \in a\}|} \sum_{\forall (i,j) \in a} \omega(s_i \mid t_j) \quad (2)$$

In formula 2, the lexical translation probability distribution $\omega(s \mid t)$ between source word s and target word t can be estimated with formula 3.

$$\omega(s \mid t) = \frac{count(s,t)}{\sum_{s'} count(s',t)} \quad (3)$$

Thus the alignment a between the source phrase $\bar{s}$ and target phrase $\bar{t}$ via pivot phrase $\bar{p}$ is needed for computing the lexical weight. The alignment a can be obtained as follows:

$$a = \{(s,t) \mid \exists p : (s,p) \in a_1 \,\&\, (p,t) \in a_2\} \quad (4)$$

where a_1 and a_2 indicate the word alignment between the phrase pair $(\bar{s}, \bar{p})$ and $(\bar{p}, \bar{t})$, respectively.

The triangulation method requires that both the source and target phrases connect to the same pivot phrase. Otherwise, the source-target phrase pair cannot be discovered. As a result, some useful translation relations will be lost. In order to alleviate this problem, we propose a random walk model, to discover the implicit relations among the source, pivot and target phrases.

4 Random Walks on Translation Graph

For phrase-based SMT, all source-target phrase pairs are stored in a phrase table. In our random walk approach, we first build a translation graph according to the phrase table. A translation graph contains two types of nodes: source phrase and target phrase. A source phrase $\bar{s}$ and a target phrase $\bar{t}$ are connected if exists a phrase pair $(\bar{s}, \bar{t})$ in the phrase table. The edge can be weighted according to translation probabilities or alignments in the phrase table. For the pivot-based translation, the translation graph can be derived from the source-pivot phrase table and pivot-target phrase table.

Our random walk model is inspired by two works (Szummer and Jaakkola, 2002; Craswell and Szummer,2007). The general process of random walk can be described as follows:

Let $G = (V, E)$ be a directed graph with n vertices and m edges. For a vertex $v \in V$, $\Gamma(v)$ denotes the set of neighbors of v in G. A random walk on G follows the following process: start at a vertex v_0, chose and walk along a random neighbor v_1, with $v_1 \in \Gamma(v_0)$. At the second step, start from v_1 and chose a random neighbor v_2, and so on.

Let *S* be the set of source phrases, and *P* be the set of pivot phrases. Then the nodes *V* are the union of *S* and *P*. The edges *E* correspond to the relations between phrase pairs.

Let *R* represent the binary relations between source phrases and pivot phrases. Then the 1-step translation R_{ik} from node *i* to node *k* can be directly obtained in the phrase table.

Define operator $\otimes$ to denote the calculation of relation *R*. Then 2-step translation R_{ij} from node *i* to node *j* can be obtained with the following formula.

$$R_{ij} = R_{ik} \otimes R_{kj} \quad (4)$$

We use $R_{t|0}(k \mid i)$ to denote a *t*-step translation relation from node *i* to node *k*. In order to calculate the translation relations efficiently, we use a matrix *A* to represent the graph. A *t* step translation probability can be denoted with the following formula.

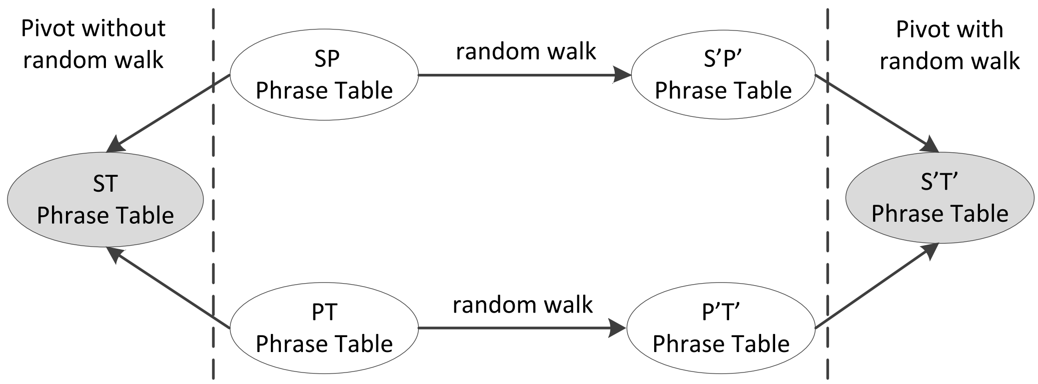

Figure 2: Framework of random walk based pivot translation. The ST phrase table was generated by combining SP and PT phrase table through triangulation method. The phrase table with superscript "'" means that it was enlarged by random walk.

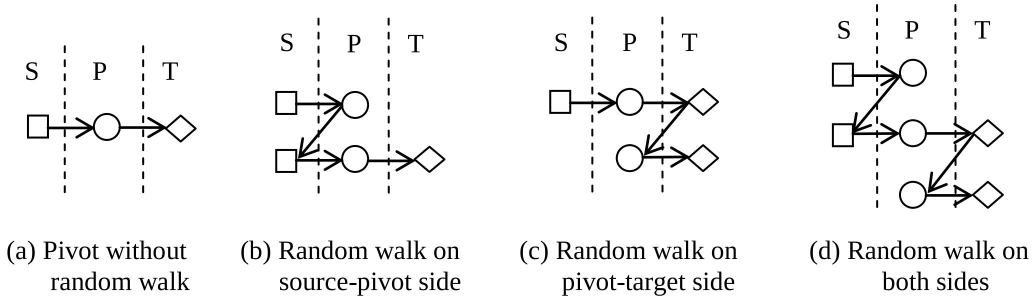

Figure 3: Some possible decoding processes of random walk based pivot approach. The □ stands for the source phrase (S); the ○ represents the pivot phrase (P) and the ◇ stands for the target phrase (T).

$$P_{t|0}(k \mid i) = [A^t]_{ik} \qquad (5)$$

where $\boldsymbol{A}$ is a matrix whose i,k-th element is R_{ik} .

4.1 Framework of Random Walk Approach

The overall framework of random walk for pivot-based machine translation is shown in Figure 2. Before using random walk model, we have two phrase tables: source-pivot phrase table (SP phrase table) and pivot-target phrase table (PT phrase table). After applying the random walk approach, we can achieve two extended phrase table: extended source-pivot phrase table (S'P' phrase table) and extended pivot-target phrase table (P'T' phrase table). The goal of pivot-based SMT is to get a source-target phrase table (ST phrase table) via SP phrase table and PT phrase table.

Our random walk was applied on SP phrase table or PT phrase table separately. In next 2 subsections, we will explain how the phrase translation probabilities and lexical weight are obtained with random walk model on the phrase table.

Figure 3 shows some possible decoding processes of random walk based pivot approach. In figure 3-a, the possible source-target phrase pair can be obtained directly via a pivot phrase, so it does not need a random walk model. In figure 3-b and figure 3-c, one candidate source-target phrase pair can be obtained by random walks on source-pivot side or pivot-target side. Figure 3-d shows that the possible source-target can only by obtained by random walks on source-pivot side and pivot-target side.

4.2 Phrase Translation Probabilities

For the translation probabilities, the binary relation R is the translation probabilities in the phrase table. The operator $\otimes$ is multiplication. According to formula 5, the random walk sums up the probabilities of all paths of length t between the node i and k.

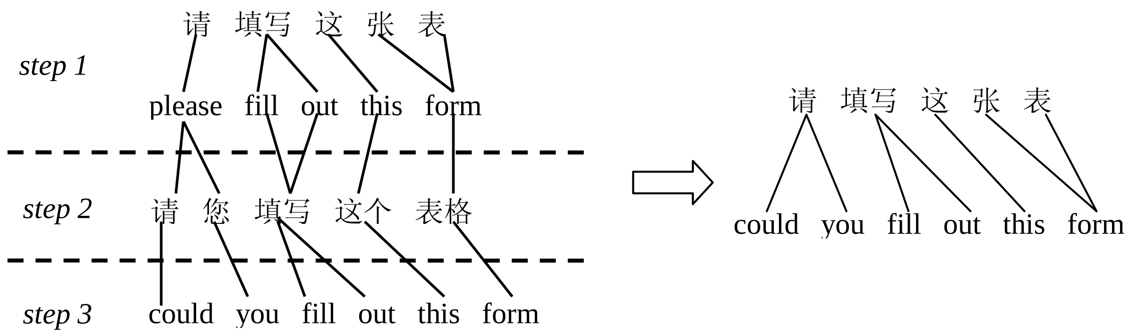

Figure 4: An example of word alignment induction with 3 steps random walks

Take source-to-pivot phrase graph as an example; denote matrix A contains $s+p$ nodes (s source phrases and p pivot phrases) to represent the translation graph.

$$A = \left[g_{ij} \right]_{(s+p)\times(s+p)} \tag{6}$$

where g_{ij} is the ***i***,***j***-th elements of matrix A.

We can split the matrix A into 4 sub-matrixes:

$$A = \begin{bmatrix} 0_{s\times s} & A_{sp} \\ A_{ps} & 0_{p\times p} \end{bmatrix} \tag{7}$$

where the sub-matrix $A_{sp} = [p_{ik}]_{s\times p}$ represents the translation probabilities from source to pivot language, and A_{ps} represents the similar meaning.

Take 3 steps walks as an example:

Step1:

$$A = \begin{bmatrix} 0_{s\times s} & A_{sp} \\ A_{ps} & 0_{p\times p} \end{bmatrix}$$

Step2:

$$A^2 = \begin{bmatrix} A_{sp}\times A_{ps} & 0_{s\times p} \\ 0_{p\times s} & A_{ps}\times A_{sp} \end{bmatrix}$$

Step3:

$$A^3 = \begin{bmatrix} 0_{s\times s} & A_{sp}\times A_{ps}\times A_{sp} \\ A_{ps}\times A_{sp}\times A_{ps} & 0_{p\times p} \end{bmatrix}$$

For the 3 steps example, each step performs a translation process in the form of matrix's self-multiplication.

1. The first step means the translation from source language to pivot language. The matrix *A* is derived from the phrase table directly and each element in the graph indicates a translation rule in the phrase table.
2. The second step demonstrates a procedure: S-P-S'. With 2 steps random walks, we can find the synonymous phrases, and this procedure is analogous to paraphrasing (Bannard and Callison-Burch, 2005). For the example shown in figure 1 as an example, the hidden relation between "*很可口 henkekou*" and "*非常好吃 feichanghaochi*" can be found through Step 2.
3. The third step describes the following procedure: S-P-S'-P'. An extended source-pivot phrase table is generated by 3-step random walks. Compared with the initial phrase table in Step1, although the number of phrases is not increased, the relations between phrase pairs are increased and more translation rules can be obtained. Still for the example in Figure 1 , the hidden relation between "*很可口 henkekou*" and "*really delicious*" can be generated in Step 3.

4.3 Lexical Weights

To build a translation graph, the two sets of phrase translation probabilities are represented in the phrase tables. However, the two lexical weights are not presented in the graph directly. To deal with this, we should conduct a word alignment random walk model to obtain a new alignment *a* after *t* steps. For the computation of lexical weights, the relation *R* can be expressed as the word alignment in the phrase table. The operator $\otimes$ can be induced with the following formula.

$$a = \{(x,y) \mid \exists p : (x,z) \in a_1 \,\&\, (z,y) \in a_2\} \tag{8}$$

where a_1 and a_2 represent the word alignment information inside the phrase pairs $(\overline{x},\overline{y})$ and $(\overline{y},\overline{z})$ respectively. An example of word alignment inducing is shown in Figure 4. With a new word alignment, the two lexical weights can be calculated by formula 2 and formula 3.

5 Experiments

5.1 Translation System and Evaluation Metric

In our experiments, the word alignment was obtained by GIZA++ (Och and Ney, 2000) and the heuristics "grow-diag-final" refinement rule. (Koehn et al., 2003). Our translation system is an in-house phrase-based system using a log-linear framework including a phrase translation model, a language model, a lexicalized reordering model, a word penalty model and a phrase penalty model, which is analogous to Moses (Koehn et al., 2007). The baseline system is the triangulation method based pivot approach (Wu and Wang, 2007).

To evaluate the translation quality, we used BLEU (Papineni et al., 2002) as our evaluation metric. The statistical significance using 95% confidence intervals were measured with paired bootstrap resampling (Koehn, 2004).

5.2 Experiments on Europarl

5.2.1. Data sets

We mainly test our approach on Europarl[1] corpus, which is a multi-lingual corpus including 21 European languages. Due to the size of the data, we only select 11 languages which were added to Europarl from 04/1996 or 01/1997, including Danish (da), German (de), Greek (el), English (en), Spanish (es), Finnish (fi), French (fr), Italian (it) Dutch (nl) Portuguese (pt) and Swedish (sv). In order to avoid a trilingual scenario, we split the training corpus into 2 parts by the year of the data: the data released in odd years were used for training source-pivot model and the data released in even years were used for training pivot-target model.

We perform our experiments on different translation directions and via different pivot languages. As a most widely used language in the world (Mydans, 2011), English was used as the pivot language for granted when carrying out experiments on different translation directions. For translating Portuguese to Swedish, we also tried to perform our experiments via different pivot languages. Table 1 and Table 2 summarized the training data.

Language Pairs (src-pvt)	Sentence Pairs #	Language Pairs (pvt-tgt)	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

Table1. Training data for experiments using English as the pivot language. For source-pivot (src-pvt; xx-en) model training, the data of odd years were used. Instead the data of even years were used for pivot-target (pvt-src; en-xx) model training.

Language Pairs (src-pvt)	Sentence Pairs #	Language Pairs (pvt-tgt)	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

Table2. Training data for experiments via different pivot languages. For source-pivot (src-pvt; pt-xx) model training, the data of odd years were used. Instead the data of even years were used for pivot-target (pvt-src; xx-sv) model training.

Test Set	Sentence #	Reference #
WMT06	2,000	1
WMT07	2,000	1
WMT08	2,000	1

Table3. Statistics of test sets.

[1] http://www.statmt.org/europarl/

	TGT / SRC	da	de	el	es	fi	fr	it	nl	pt	sv
Baseline	**da**	-	19.83	20.46	27.59	14.76	24.11	20.49	22.26	24.38	28.33
RW			**20.15***	**21.02***	**28.29***	**15.63***	**24.71***	**20.82***	**22.57***	**24.88***	**28.87***
Baseline	**de**	23.35	-	19.83	26.21	12.72	22.43	18.82	23.74	23.05	21.17
RW		**23.69***		20.05	**26.70***	**13.57***	**22.78***	**19.32***	**24.11***	**23.35***	21.27
Baseline	**el**	23.24	18.12	-	32.28	13.31	27.35	23.19	20.80	27.62	22.70
RW		**23.82***	**18.49***		32.48	**14.08***	**27.67***	**23.63***	**21.26***	27.86	**23.15***
Baseline	**es**	25.34	19.67	27.24	-	13.93	32.91	27.67	22.37	34.73	24.83
RW		**26.07***	**20.17***	27.52		**14.61***	33.16	27.92	**22.85***	34.93	**25.50***
Baseline	**fi**	18.29	13.20	14.72	20.17	-	17.52	14.76	15.50	17.30	16.63
RW		**18.63***	13.40	**15.00***	**20.48***		**17.84***	15.01	**16.04***	**17.68***	16.79
Baseline	**fr**	25.67	20.02	26.58	37.50	13.90	-	28.51	22.65	33.81	24.64
RW		**26.51***	**20.45***	26.75	**37.80***	**14.75***		28.71	**23.33***	33.93	**25.59***
Baseline	**it**	22.63	17.81	24.24	34.36	13.20	30.16	-	21.37	30.84	22.12
RW		**23.27***	**18.40***	**24.66***	**35.42***	**14.11***	**30.48***		**21.81***	**30.92***	**22.64***
Baseline	**nl**	22.49	19.86	18.56	24.69	11.96	21.48	18.36	-	21.71	19.83
RW		22.76	**20.45***	**19.10***	**25.19***	**12.63***	**22.05***	**18.67***		**22.13***	**22.17***
Baseline	**pt**	24.08	19.11	25.30	36.59	13.33	32.47	28.08	21.52	-	22.90
RW		**25.29***	**19.83***	**26.20***	**37.13***	**14.21***	**32.78***	**28.44***	**22.46***		**23.90***
Baseline	**sv**	31.24	20.26	22.06	29.21	15.39	25.63	21.25	22.30	25.60	-
RW		**31.75***	**20.74***	**22.59***	**29.87***	**16.13***	**26.18***	**21.81***	**22.62***	**26.09***	

Table4. Experimental results on Europarl with different translation directions (BLEU% on WMT08). RW=Random Walk. * indicates the results are significantly better than the baseline ($p<0.05$).

Several test sets have been released for the Europarl corpus. In our experiments, we used WMT2006[2], WMT2007[3] 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 by the EuroMatrix[5] project. Table 3 shows the test sets.

5.2.2. Experiments on Different Translation Directions

We build 180 pivot translation systems[6] (including 90 baseline systems and 90 random walk based systems) using 10 source/target languages and 1 pivot language (English).

The baseline system was built following the traditional triangulation pivot approach. Table 4 lists the results on Europarl training data. Limited by the length of the paper, we only show the results on WMT08, the tendency of the results on WMT06 and WMT07 is similar to WMT08.

Several observations can be made from the table.

1. In all 90 language pairs, our method achieves general improvements over the baseline system.

2. Among 90 language pairs, random walk based approach is significantly better than the baseline system in 75 language pairs.

3. The improvements of our approach are not equal in different translation directions. The improvement ranges from 0.06 (it-es) to 1.21 (pt-da). One possible reason is that the performance is related with the source and target language. For example, when using Finnish as the target language, the improvement is significant over the baseline. This may be caused by the great divergence between Uralic language (Finnish) and Indo-European language (the other European language in Table4). From the table we can find that the translation between languages in different language family is worse than that in some language family. But our random walk approach can im-

[2] http://www.statmt.org/wmt06/shared-task/
[3] http://www.statmt.org/wmt07/shared-task.html
[4] http://www.statmt.org/wmt08/shared-task.html
[5] http://matrix.statmt.org/test_sets/list
[6] Given N languages, a total of N*(N-1) SMT systems should be build to cover the translation between each language.

prove the performance of translations between different language families.

5.2.3. Experiments via Different Pivot Languages

In addition to using English as the pivot language, we also try some other languages as the pivot language. In this sub-section, experiments were carried out from translating Portuguese to Swedish via different pivot languages.

Table 5 summarizes the BLEU% scores of different pivot language when translating from Portuguese to Swedish. Similar to Table 4, our approach still achieves general improvements over the baseline system even if the pivot language has been changed. From the table we can see that for most of the pivot language, the random walk based approach gains more than 1 BLEU score over the baseline. But when using Finnish as the pivot language, the improvement is only 0.02 BLEU scores on WMT08. This phenomenon shows that the pivot language can also influence the performance of random walk approach. One possible reason for the poor performance of using Finnish as the pivot language is that Finnish belongs to Uralic language family, and the other languages belong to Indo-European family. The divergence between different language families led to a poor performance. Thus how to select a best pivot language is our future work.

The problem with random walk is that it will lead to a larger phrase table with noises. In this sub-section, a pre-pruning (before random walk) and a post-pruning (after random walk) method were introduced to deal with this problem.

We used a naive pruning method which selects the top N phrase pairs in the phrase table. In our experiments, we set N to 20. For pre-pruning, we prune the SP phrase table and PT phrase table before applying random walks. Post-pruning means that we prune the ST phrase table after random walks. For the baseline system, we also apply a pruning method before combine the SP and PT phrase table. We test our pruning method on pt-en-sv translation task. Table 6 shows the results.

With a pre- and post-pruning method, the random walk approach is able to achieve further improvements. Our approach achieved BLEU scores of 25.11, 24.69 and 24.34 on WMT06, WMT07 and WMT08 respectively, which is much better than the baseline and the random walk approach with pruning. Moreover, the size of the phrase table is about half of the no-pruning method. When adopting a post-pruning method, the performance of translation did not improved significantly over the pre-pruning, but the scale of the phrase table dropped to 69M, which is only about 2 times larger than the triangulation method.

Phrase table pruning is a key work to improve the performance of random walk. We plan to explore more approaches for phrase table pruning. E.g. using significance test (Johnson et al., 2007) or monolingual key phrases (He et al., 2009) to filter the phrase table.

	trans language	**WMT 06**	**WMT 07**	**WMT 08**
Baseline RW	pt-da-sv	23.40 **24.47***	22.80 **24.21***	22.49 **23.75***
Baseline RW	pt-de-sv	22.72 **23.12***	22.21 **23.26***	21.76 **22.35***
Baseline RW	pt-el-sv	22.53 **23.75***	22.19 **23.22***	21.37 **22.40***
Baseline RW	pt-en-sv	23.54 **24.66***	23.24 **24.22***	22.90 **23.90***
Baseline RW	pt-es-sv	23.58 **24.65***	23.37 **24.10***	22.80 **23.77***
Baseline RW	pt-fi-sv	21.06 21.17	20.06 **20.42***	20.26 20.28
Baseline RW	pt-fr-sv	23.55 **24.75***	23.09 **24.15***	22.89 **23.96***
Baseline RW	pt-it-sv	23.65 **24.74***	22.96 **24.18***	22.79 **24.02***
Baseline RW	pt-nl-sv	21.87 **23.06***	21.83 **22.76***	21.36 **22.29***

Table5. Experimental results on translating from Portuguese to Swedish via different pivot language. RW=Random Walk. * indicates the results are significantly better than the baseline ($p<0.05$).

	WMT 06	**WMT 07**	**WMT 08**	**Phrase Pairs #**
Baseline +pruning	23.54 **24.05 ***	23.24 **23.70 ***	22.90 **23.59 ***	46M **32M**
RW +pre-pruning +post-pruning	24.66 25.11 **25.19 ***	24.22 24.69 **24.79 ***	23.90 24.34 **24.41 ***	215M 109M **69M**

Table6. Results of Phrase Table Filtering

5.3 Experiments on Spoken Language

The European languages show various degrees of similarity to one another. In this sub-section, we consider translation from Chinese to Spanish with English as the pivot language. Chinese belongs to Sino-Tibetan Languages and English/Spanish belongs to Indo-European Languages, the gap between two languages is wide.

A pivot task was included in IWSLT 2008 in which the participants need to translate Chinese to Spanish via English. A Chinese-English and an English-Spanish data were supplied to carry out the experiments. The entire training corpus was tokenized and lowercased. Table 7 and Table 8 summarize the training data and test data.

Table 9 shows the similar tendency with Table 4. The random walk models achieved BLEU% scores 32.09, which achieved an absolute improvement of 2.08 percentages points on BLEU over the baseline.

Corpus	Sentence pair #	Source word #	Target word #
CE	20,000	135,518	182,793
ES	19,972	153,178	147,560

Table 7: Training Data of IWSLT2008

Test Set	Sentence #	Reference #
IWSLT08	507	16

Table8. Test Data of IWSLT2008

System	BLEU%	phrase pairs #
Baseline	30.01	143,790
+pruning	30.25	108,407
RW	31.57	2,760,439
+pre-pruning	31.99	1,845,648
+post-pruning	**32.09***	**1,514,694**

Table9. Results on IWSLT2008

5.4 Experiments on Web Data

The setting with Europarl data is quite artificial as the training data for directly translating between source and target actually exists in the original data sets. The IWSLT data set is too small to represent the real scenario. Thus we try our experiment on a more realistic scenario: translating from Chinese to Japanese via English with web crawled data.

All the training data were crawled on the web. The scale of Chinese-English and English-Japanese is 10 million respectively. The test set was built in house with 1,000 sentences and 4 references.

System	BLEU%	phrase pairs #
Baseline	28.76	4.5G
+pruning	28.90	273M
RW	29.13	46G
+pre-pruning	29.44	11G
+post-pruning	**29.51***	**3.4G**

Table10. Results on Web Data

Table 10 lists the results on web data. From the table we can find that the random walk model can achieve an absolute improvement of 0.75 percentages points on BLEU over the baseline.

In this subsection, the training data contains parallel sentences with different domains. And the two training corpora (Chinese-English and English-Japanese) are typically very different. It means that our random walk approach is robust in the realistic scenario.

6 Discussions

The random walk approach mainly improves the performance of pivot translation in two aspects: reduces the OOVs and provides more hypothesis phrases for decoding.

6.1 OOV

Out-of-vocabulary (OOV[7]) terms cause serious problems for machine translation systems (Zhang et al., 2005). The random walk model can reduce the OOVs. As illustrated in Figure 1, the Chinese phrase "很可口henkekou" cannot be connected to any Spanish phrase, thus it is a OOV term.

We count the OOVs when decoding with triangulation model and random walk model on IWSLT2008 data. The statistics shows that when using triangulation model, there are 11% OOVs when using triangulation model, compared with 9.6% when using random walk model. Less OOV often lead to a better result.

[7] OOV refer to phrases here.

6.2 Hypothesis Phrases

To illustrate how the random walk method helps improve the performance of machine translation, we illustrate an example as follows:

- **Source: 我 想 要 枕头**
  wo xiang yao zhentou
- **Baseline trans: Quiero almohada**
- **Random Walk trans: Quiero una almohada**

For translating a Chinese sentence "我想要枕头 wo xiang yao zhentou" to Spanish, we can get two candidate translations. In this case, the random walk translation is better than the baseline system. The key phrase in this sentence is "枕头 zhentou", figure 5 shows the extension process. In this case, the article "*a*" is hidden in the source-pivot phrase table. The same situation often occurs in articles and prepositions. Random walk is able to discover the hidden relations (hypothesis phrases) among source, pivot and target phrases.

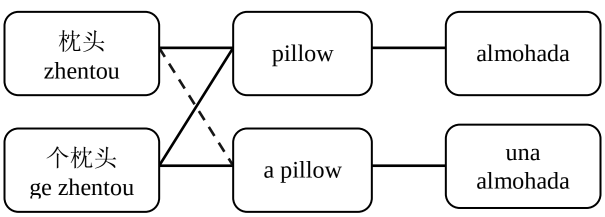

Figure 5: Phrase extension process. The dotted line indicates an implicit relation in the phrase table.

7 Conclusion and Future Work

In this paper, we proposed a random walk method to improve pivot-based statistical machine translation. The random walk method can find implicit relations between phrases in the source and target languages. Therefore, more source-target phrase pairs can be obtained than conventional pivot-based method. Experimental results show that our method achieves significant improvements over the baseline on Europarl corpus, spoken language data and the web data.

A critical problem in the approach is the noise that may bring in. In this paper, we used a simple filtering to reduce the noise. Although the filtering method is effective, other method may work better. In the future, we plan to explore more approaches for phrase table pruning.

Acknowledgments

We would like to thank Jianyun Nie, Muyun Yang and Lemao Liu for insightful discussions, and three anonymous reviewers for many invaluable comments and suggestions to improve our paper. This work is supported by National Natural Science Foundation of China (61100093), and the Key Project of the National High Technology Research and Development Program of China (2011AA01A207).

References

Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with Bilingual Parallel Corpora. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 597-604

Sergey Brin and Lawrence Page. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *Proceedings of the Seventh International World Wide Web Conference*

Trevor Cohn and Mirella Lapata. 2007. Machine Translation by Triangulation: Make Effective Use of Multi-Parallel Corpora. In *Proceedings of 45th Annual Meeting of the Association for Computational Linguistics*, pages 828-735.

Marta R. Costa-jussà, Carlos Henríquez, and Rafael E. Banchs. 2011. Enhancing Scarce-Resource Language Translation through Pivot Combinations. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 1361-1365

Nick Craswell and Martin Szummer. 2007. Random Walks on the Click Graph. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 239-246

Yiming Cui, Conghui Zhu, Xiaoning Zhu, Tiejun Zhao and Dequan Zheng. 2013. Phrase Table Combination Deficiency Analyses in Pivot-based SMT. In *Proceedings of 18th International Conference on Application of Natural Language to Information Systems*, pages 355-358.

Kevin Duh, Katsuhito Sudoh, Xianchao Wu, Hajime Tsukada and Masaaki Nagata. 2011. Generalized Minimum Bayes Risk System Combination. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 1356–1360

Jesús González-Rubio, Alfons Juan and Francisco Casacuberta. 2011. Minimum Bayes-risk System

Combination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 1268–1277

Zhongjun He, Yao Meng, Yajuan Lü, Hao Yu and Qun Liu. 2009. Reducing SMT Rule Table with Monolingual Key Phrase. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 121-124

Howard Johnson, Joel Martin, George Foster, and Roland Kuhn. 2007. Improving translation quality by discarding most of the phrase table. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 967–975.

Philipp Koehn, Franz J. Och, and Daniel Marcu. 2003. Statistical Phrase-Based Translation. In *HLT-NAACL: Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 127-133

Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In P*roceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 388–395.

Philipp Koehn. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of MT Summit X*, pages 79-86.

Philipp Koehn, Hieu Hoang, Alexanda Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, demonstration session*, pages 177–180.

Franz Josef Och and Hermann Ney. 2000. A comparison of alignment models for statistical machine translation. In *Proceedings of the 18th International Conference on Computational Linguistics*, pages 1086–1090

Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computation Linguistics*, pages 311-319

Karl Pearson. 1905. The Problem of the Random Walk. *Nature*, 27(1865):294

Mydans, Seth. 2011. Across cultures, English is the word. *New York Times*.

Martin Szummer and Tommi Jaakkola. 2002. Partially Labeled Classification with Markov Random Walks. In *Advances in Neural Information Processing Systems*, pages 945-952

Kristina Toutanova, Christopher D. Manning and Andrew Y. Ng. 2004. Learning Random Walk Models for Inducting Word Dependency Distributions. In *Proceedings of the 21st International Conference on Machine Learning*.

Masao Utiyama and Hitoshi Isahara. 2007. A Comparison of Pivot Methods for Phrase-Based Statistical Machine Translation. In *Proceedings of Human Language Technology: the Conference of the North American Chapter of the Association for Computational Linguistics*, pages 484-491

Masao Utiyama, Andrew Finch, Hideo Okuma, Michael Paul, Hailong Cao, Hirofumi Yamamoto, Keiji Yasuda, and Eiichiro Sumita. 2008. The NICT/ATR speech Translation System for IWSLT 2008. In *Proceedings of the International Workshop on Spoken Language Translation*, pages 77-84

Haifeng Wang, Hua Wu, Xiaoguang Hu, Zhanyi Liu, Jianfeng Li, Dengjun Ren, and Zhengyu Niu. 2008. The TCH Machine Translation System for IWSLT 2008. In *Proceedings of the International Workshop on Spoken Language Translation*, pages 124-131

Hua Wu and Haifeng Wang. 2007. Pivot Language Approach for Phrase-Based Statistical Machine Translation. In *Proceedings of 45th Annual Meeting of the Association for Computational Linguistics*, pages 856-863.

Hua Wu and Haifeng Wang. 2009. Revisiting Pivot Language Approach for Machine Translation. In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th IJCNLP of the AFNLP*, pages 154-162

Ying Zhang, Fei Huang, Stephan Vogel. 2005. Mining translations of OOV terms from the web through cross-lingual query expansion. In *Proceedings of the 27th ACM SIGIR*. pages 524-525