Using Context Vectors in Improving a Machine Translation System with Bridge Language

Samira Tofighi Zahabi   Somayeh Bakhshaei   Shahram Khadivi
Human Language Technology Lab
Amirkabir University of Technology
Tehran, Iran

`{Samiratofighi,bakhshaei,khadivi}@aut.ac.ir`

Abstract

Mapping phrases between languages as translation of each other by using an intermediate language (pivot language) may generate translation pairs that are wrong. Since a word or a phrase has different meanings in different contexts, we should map source and target phrases in an intelligent way. We propose a pruning method based on the context vectors to remove those phrase pairs that connect to each other by a polysemous pivot phrase or by weak translations. We use context vectors to implicitly disambiguate the phrase senses and to recognize irrelevant phrase translation pairs.

Using the proposed method a relative improvement of 2.8 percent in terms of BLEU score is achieved.

1   Introduction

Parallel corpora as an important component of a statistical machine translation system are unfortunately unavailable for all pairs of languages, particularly in low resource languages and also producing it consumes time and cost. So, new ideas have been developed about how to make a MT system which has lower dependency on parallel data like using comparable corpora for improving performance of a MT system with small parallel corpora or making a MT system without parallel corpora. Comparable corpora have segments with the same translations. These segments might be in the form of words, phrases or sentences. So, this extracted information can be added to the parallel corpus or might be used for adaption of the language model or translation model.

Comparable corpora are easily available resources. All texts that are about the same topic can be considered as comparable corpora. Another idea for solving the scarce resource problem is to use a high resource language as a pivot to bridge between source and target languages. In this paper we use the bridge technique to make a source-target system and we will prune the phrase table of this system. In Section 2, the related works of the bridge approach are considered, in Section 3 the proposed approach will be explained and it will be shown how to prune the phrase table using context vectors, and experiments on German-English-Farsi systems will be presented in Section 4.

2   Related Works

There are different strategies of bridge techniques to make a MT system. The simplest way is to build two MT systems in two sides: one system is source-pivot and the other is pivot-target system, then in the translation stage the output of the first system is given to the second system as an input and the output of the second system is the final result. The disadvantage of this method is its time consuming translation process, since until the first system's output is not ready; the second system cannot start the translation process. This method is called *cascading of two translation systems*.

In the other approach the target side of the training corpus of the source-pivot system is given to the pivot-target system as its input. The output of the pivot-target system is parallel with the source side of the training corpus of the source-pivot system. A source-to-target system can be built by using this noisy parallel corpus which in it each source sentence is directly translated to a target sentence. This method is called the *pseudo corpus approach*.

Another way is combining the phrase tables of the source-pivot and pivot-target systems to directly make a source-target phrase table. This combination is done if the pivot phrase is

identical in both phrase tables. Since one phrase has many translations in the other language, a large phrase table will be produced. This method is called *combination of phrase tables approach*.

Since in the bridge language approach two translation systems are used to make a final translation system, the errors of these two translation systems will affect the final output. Therefore in order to decrease the propagation of these errors, a language should be chosen as pivot which its structure is similar to the source and target languages. But even by choosing a good language as pivot there are some other errors that should be handled or decreased such as the errors of ploysemous words and etc.

For making a MT system using pivot language several ideas have been proposed. Wu and Wang (2009) suggested a cascading method which is explained in Section 1.

Bertoldi (2008) proposed his method in bridging at translation time and bridging at training time by using the cascading method and the combination of phrase tables.

Bakhshaei (2010) used the combination of phrase tables of source-pivot and pivot-target systems and produced a phrase table for the source-target system.

Paul (2009) did several experiments to show the effect of pivot language in the final translation system. He showed that in some cases if training data is small the pivot should be more similar to the source language, and if training data is large the pivot should be more similar to the target language. In Addition, it is more suitable to use a pivot language that its structure is similar to both of source and target languages.

Saralegi (2011) showed that there is not transitive property between three languages. So many of the translations produced in the final phrase table might be wrong. Therefore for pruning wrong and weak phrases in the phrase table two methods have been used. One method is based on the structure of source dictionaries and the other is based on distributional similarity.

Rapp (1995) suggested his idea about the usage of context vectors in order to find the words that are the translation of each other in comparable corpora.

In this paper the combination of phrase tables approach is used to make a source-target system. We have created a base source-target system just similar to previous works. But the contribution of our work compared to other works is that here we decrease the size of the produced phrase table and improve the performance of the system. Our

pruning method is different from the method that Saralegi (2011) has used. He has pruned the phrase table by computing distributional similarity from comparable corpora or by the structure of source dictionaries. Here we use context vectors to determine the concept of phrases and we use the pivot language to compare source and target vectors.

3   Approach

For the purpose of showing how to create a pruned phrase table, in Section 3.1 we will explain how to create a simple source-to-target system. In Section 3.2 we will explain how to remove wrong and weak translations in the pruning step. Figure 1 shows the pseudo code of the proposed algorithm.

In the following we have used these abbreviations: **f, e** *stands for source and target phrases*. **pl, src-pl, pl-trg, src-trg** respectively *stand for pivot phrase, source-pivot phrase table, pivot-target phrase table and source-target phrase table*.

3.1   Creating source-to-target system

At first, we assume that there is transitive property between three languages in order to make a base system, and then we will show in different ways that there is not transitive property between three languages.

```
for each source phrase f
    pls = {translations of f in src-pl }
    for each pl in pls
      Es ={ translations of pl in pl-trg }
      for each e in Es
        p(e|f) =p(pl|f)*p(e|pl) and add (e,f) to src-
trg
create source-to-destination system with src-trg
create context vector V for each source phrase f
using source corpora
create context vector V' for each target phrase e
using target corpora
convert Vs to pivot language vectors using src-pl
system
convert V's to pivot language vectors using pl-trg
system
for each f in src-trg
Es = {translations of f in src-trg}
For each e in Es calculate similarity of its context
vector with f context vector
    Select k top similar as translations of f
    delete other translations of f in src-trg
```

Figure 1. Pseudo code for proposed method

For each phrase **f** in **src-pl** phrase table, all the phrases **pl** which are translations of **f,** are considered. Then for each of these **pl**s every phrase **e** from the **pl-trg** phrase table that are translations of **pl,** are found. Finally **f** is mapped to all of these **e**s in the new **src-trg** phrase table.

The probability of these new phrases is calculated using equation (1) through the algorithm that is shown in figure 1.

$$p(e \mid f) = p(pl \mid f) \times p(e \mid pl) \qquad (1)$$

A simple **src-trg** phrase table is created by this approach. **Pl** phrases might be ploysemous and produce target phrases that have different meaning in comparison to each other. The concept of some of these target phrases are similar to the corresponding source phrase and the concept of others are irrelevant to the source phrase.

The language model can ignore some of these wrong translations. But it cannot ignore these translations if they have high probability.

Since the probability of translations is calculated using equation (1), therefore wrong translations have high probability in three cases: first when p(**pl**|**f**) is high, second when p(**e**|**pl**) is high and third when p(**pl**|**f**) and p(**e**|**pl**) are high.

In the first case **pl** might be a good translation for **f** and refers to concept c, but **pl** and **e** refer to concept c' so mapping **f** to **e** as a translation of each other is wrong. The second case is similar to the first case but **e** might be a good translation for **pl**. The third case is also similar to the first case, but **pl** is a good translation for both **f** and **e**.

The pruning method that is explained in Section 3.2, tries to find these translations and delete them from the **src-trg phrase table**.

3.2   Pruning method

To determine the concept of each phrase (p) in language L at first a vector (V) with length N is created. Each element of V is set to zero and N is the number of unique phrases in language L.

In the next step all sentences of the corpus in language L are analyzed. For each phrase p if p occurs with p' in the same sentence the element of context vector V that corresponds to p' is pulsed by 1. This way of calculating context vectors is similar to Rapp (1999), but here the window length of phrase co-occurrence is considered a sentence. Two phrases are considered as co-occurrence if they occur in the same sentence. The distance between them does not matter. In other words phrase p might be at the beginning of the sentence while p' being at

the end of the sentence, but they are considered as co-occurrence phrases.

For each source (target) phrase its context vector should be calculated within the source (target) corpus as shown in figure 1.

The number of unique phrases in the source (target) language is equal to the number of unique source (target) phrases in the **src-trg** phrase table that are created in the last Section.

So, the length of source context vectors is **m** and the length of target context vectors is **n**. These variables (**m** and **n)** might not be equal. In addition to this, source vectors and target vectors are in two different languages, so they are not comparable.

One method to translate source context vectors to target context vectors is using an additional source-target dictionary. But instead here, source and target context vectors are translated to pivot context vectors. In other words if source context vectors have length **m** and target context vectors have length **n**, they are converted to pivot context vectors with length **z**. The variable **z** is the number of unique pivot phrases in **src-pl** or **pl-trg** phrase tables.

To map the source context vector $S(s_1, s_2, \ldots, s_m)$ to the pivot context vector, we use a fixed size vector V_1^z. Elements of vector $V_1^z = (v_1, v_2, \ldots, v_z)$ are the unique phrases extracted from **src-pl** or **pl-trg** phrase tables.

$$V_1^z = (v_1, v_2, \ldots, v_z) = (0, 0, \ldots, 0)$$

In the first step v_is are set to 0. For each element, s_i, of vector S if $s_i > 0$ it will be translated to k pivot phrases. These phrases are the output of k-best translations of s_i by using the **src-pl** phrase table.

$$s_i \underset{\text{src-pl} \, phrase \, table}{\longmapsto} \left\{ V_1'^k = (v_1', v_2', \ldots, v_k') \right\}$$

For each element v' of $V_1'^k$ its corresponding element v of V_1^z which are equal, will be found, then the amount of v will be increased by s_i.

$$\forall \, v' \in V_1'^k \, find \, (v \in V_1^z) \ni v = v'$$
$$val(v) \leftarrow val(v) + s_i$$

Using K-best translations as middle phrases is for reducing the effect of translation errors that cause wrong concepts. This work is done for each target context vector. Source and target context vectors will be mapped to identical length vectors and are also in the same language (pivot language).Now source and target context vectors are comparable, so with a simple similarity metric their similarity can be calculated.

Here we use cosine similarity. The similarity between each source context vector and each

target context vector that are translations of the source phrase in **src-trg, are calculated**. For each source phrase, the **N**-most similar target phrases are kept as translations of the source phrase. These translations are also similar in context. Therefore this pruning method deletes irrelevant translations form the **src-trg** phrase table. The size of the phrase table is decreased very much and the system performance is increased. Reduction of the phrase table size is considerable while its performance is increased.

4　Experiments

In this work, we try to make a German-Farsi system without using parallel corpora. We use English language as a bridge between German and Farsi languages because English language is a high resource language and parallel corpora of German-English and English-Farsi are available.

We use Moses[1] (Koehn et al., 2007) as the MT decoder and IRSTLM[2] tools for making the language model. Table 1 shows the statistics of the corpora that we have used in our experiments. The German-English corpus is from Verbmobil project (Ney et al., 2000). We manually translate 22K English sentences to Farsi to build a small Farsi-English-German corpus. Therefore, we have a small English-German corpus as well.

With the German-English parallel corpus and an additional German-English dictionary with 118480 entries we have made a German-English (**De-En**) system and with English-Farsi parallel corpus we have made a German-Farsi (**En-Fa**) system. The BLEU score of these systems are shown in Table 1.

Now, we create a translation system by combining phrase tables of De-En and En-Fa systems. Details of creating the source-target system are explained in Section 3.1. The size of this phrase table is very large because of ploysemous and some weak translations.

	Sentences	BLEU
German-English	58,073	40.1
English-Farsi	22,000	31.6

Table 1. Information of two parallel systems that are used in our experiments.

The size of the phrase table is about 55.7 MB. Then, we apply the pruning method that we

explained in Section 3.2. With this method only the phrases are kept that their context vectors are similar to each other. For each source phrase the 35-most similar target translations are kept. The number of phrases in the phrase table is decreased dramatically while the performance of the system is increased by 2.8 percent BLEU. The results of these experiments are shown in Table 2. The last row in this table is the result of using small parallel corpus to build German-Farsi system. We observe that the pruning method has gain better results compared to the system trained on the parallel corpus. This is maybe because of some translations that are made in the parallel system and do not have enough training data and their probabilities are not precise. But when we use context vectors to measure the contextual similarity of phrases and their translations, the impact of these training samples are decreased. In Table 3, two wrong phrase pairs that pruning method has removed them are shown.

	BLEU	# of phrases
Base bridge system	25.1	500,534
Pruned system	27.9	26,911
Parallel system	27.6	348,662

Table 2. The MT results of the base system, the pruned system and the parallel system.

German phrase	Wrong translation	Correct translation
vorschlagen , wir	به تاتر	پیشنهاد میکنیم
um neun Uhr morgens	ده	ساعت نه صبح

Table 3. Sample wrong translations that the prunning method removed them.

In Table 4, we extend the experiments with two other methods to build German-Farsi system using English as bridging language. We see that the proposed method obtains competitive result with the pseudo parallel method.

System	BLEU	size (MB)
Phrase tables combination	25.1	55.7
Cascade method	25.2	NA
Pseudo parallel corpus	28.2	73.2
Phrase tables comb.+prune	27.9	3.0

Table 4. Performance results of different ways of bridging

Now, we run a series of significance tests to measure the superiority of each method. In the first significance test, we set the pruned system as our base system and we compare the result of the pseudo parallel corpus system with it, the significance level is 72%. For another significance test we set the combined phrase table system without pruning as our base system and we compare the result of the pruned system with it, the significance level is 100%. In the last significance test we set the combined phrase table system without pruning as our base system and we compare the result of the pseudo system with it, the significance level is 99%. Therefore, we can conclude the proposed method obtains the best results and its difference with pseudo parallel corpus method is not significant.

5 Conclusion and future work

With increasing the size of the phrase table, the MT system performance will not necessarily increase. Maybe there are wrong translations with high probability which the language model cannot remove them from the best translations.
By removing these translation pairs, the produced phrase table will be more consistent, and irrelevant words or phrases are much less. In addition, the performance of the system will be increased by about 2.8% BLEU.
In the future work, we investigate how to use the word alignments of the source-to-pivot and pivot-to-target systems to better recognize good translation pairs.

References

Somayeh Bakhshaei, Shahram Khadivi, and Noushin Riahi. 2010. Farsi-German statistical machine translation through bridge language. *Telecommunications (IST) 5th International Symposium on*, pages 557-561.

Nicola Bertoldi, Madalina Barbaiani, Marcello Federico, and Roldano Cattoni. 2008. Phrase-Based Statistical Machine Translation with Pivot Language. In *Proc. Of IWSLT,* pages 143-149, Hawaii, USA.

Philipp Koehn. 2004. Statistical Significance Tests for Machine Translation Evaluation. In *proc. of EMNLP*, pages 388-395, Barcelona, Spain.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris C. Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, WadeShen, Christine Moran, RichardZens, Chris Dyer, Ondrei Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proc. of ACL Demonstration Session*, pages 177–180, Prague.

Hermann Ney, Franz. J. Och, Stephan Vogel. 2000. Statistical Translation of Spoken Dialogues in the Verbmobil System. In Workshop on Multi Lingual Speech Communication, pages 69-74.

Michael Paul, Hirofumi Yamamoto, Eiichiro Sumita, and Satoshi Nakamura. 2009. On the Importance of Pivot Language Selection for Statistical Machine Translation. In *proc. Of NAACL HLT*, pages 221-224, Boulder, Colorado.

Reinhard Rapp. 1995. Identifying Word Translations in Non-Parallel Texts. In *proc. Of ACL*, pages 320-322, Stroudsburg, PA, USA.

Reinhard Rapp. 1999. Automatic Identification of Word Translations from Unrelated English and German Corpora. In *Proc. Of ACL*, pages 519-525, Stroudsburg, PA, USA.

Xabeir Saralegi, Iker Manterola, and Inaki S. Vicente, 2011.Analyzing Methods for Improving Precision of Pivot Based Bilingual Dictionaries. In *proc. of the EMNLP*, pages 846-856, Edinburgh, Scotland.

Masao Utiyama and Hitoshi Isahara. 2007. A comparison of pivot methods for phrase-based SMT. In*Proc. of HLT*, pages 484-491, New York, US.

Hua Wu and Haifeng Wang. 2007. Pivot Language Approach for Phrase-Based SMT. In *Proc. of ACL*, pages 856-863, Prague, Czech Republic.