Translation of Unseen Bigrams by Analogy Using an SVM Classifier

Hao Wang **Lu Lyu** **Yves Lepage**
Graduate School of Information, Production and Systems
Waseda University
2-7 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0135, Japan
{oko_ips@ruri., lulv90@ruri., yves.lepage@}waseda.jp

Abstract

Detecting language divergences and predicting possible sub-translations is one of the most essential issues in machine translation. Since the existence of translation divergences, it is impractical to straightforward translate from source sentence into target sentence while keeping the high degree of accuracy and without additional information. In this paper, we investigate the problem from an emerging and special point of view: bigrams and the corresponding translations. We first profile corpora and explore the constituents of bigrams in the source language. Then we translate unseen bigrams based on proportional analogy and filter the outputs using an Support Vector Machine (SVM) classifier. The experiment results also show that even a small set of features from analogous can provide meaningful information in translating by analogy.

1 Introduction

Over the last decade, phrase-based statistical machine translation (Koehn et al., 2003) systems have demonstrated that they can produce reasonable quality when ample training data is available, especially for language pairs with similar word order. However, the PB-SMT model has not yet been capable of satisfying the various translation tasks for very different languages (Isozaki et al., 2010). The existence of translation divergences makes the straightforward transfer from source sentences into target sentences hard. Though many previous pieces of work (Dorr, 1994; Habash et al., 2002; Dorr et al., 2004) have attempted to take account for divergences and to deal with this linguistic problem using various translation approaches. This paper further inquires the topic.

Since sentence consists of bigrams, instead of analysing the syntactic structures of the whole sentence or part of the sentence as in (Ding and Palmer, 2005), we explore the possibilities of translating unseen bigrams based on an analogy learning method. We investigate the coverage of translated bigrams in the test set and inspect the probability of translating a bigram using analogy. Analogical learning has been investigated by several authors. To cite a few, Lepage et al. (2005) showed that proportional analogy can capture some syntactic and lexical structures across languages. Langlais et al. (2007) investigated the more specific task of translating unseen words. Bayoudh et al. (2007) explored generating new learning examples from very scarce original learning data using analogy to train an SVM classifier. Dandapat et al. (2010) performed transliteration by analogical learning for English-to-Hindi.

In the issue of translation using analogy, one of the main drawbacks should be addressed is the problem of ”over-generative”. Analogy is able to capture the most divergences of translation in the most cases, yet it generates a great number of solutions that are ungrammatical and incorrect. In this paper, we propose to translate useen bigrams as reconstructing with the principle of analogy learning. In machine learning, SVMs have been shown that it is efficient in performing a non-linear classification. By specifying features used in experiment, we employ an SVM classifier to fast filter the solutions output by the analogy solver. The final goal of this research is to explore the possibility of translation

29th Pacific Asia Conference on Language, Information and Computation pages 16 - 25
Shanghai, China, October 30 - November 1, 2015

using analogy and point out a feasible way to solve the problem of "over-generative".

The remainder of this paper is organized as follows: Section 2 describes basic notions in alignment and analogy. In Section 3, we explore the classification of bigrams and their contributions to the whole corpus and report some profiling results. Section 4 presents our approach, depending on the analogous, and describes how to processing the data and extract examples for training an SVM classifier. We also evaluate the result using the some standard measures. Finally, in Section 5, conclusions and perspectives are presented.

2 Basic notions

2.1 Alignment classification

In this section, from a theoretical point of view, we study the categories of word alignment in translating. Given a sentence, various alignments of bigram exist. The following is an example of non-monotonic alignments where alignment links are crossing between parallel sentences (Japanese and English):

e: *He*$_1$ *saw*$_2$ *a cat*$_3$ *with a long*$_4$ *tail*$_5$.
j: *Kare_ha*$_1$ *nagai*$_4$ *sippo_no*$_5$ *neko_wo*$_3$ *mita*$_2$.
$\tilde{e}$: *He long tail_of cat saw*

In this example, e means an original English sentence in parallel texts, j means a Japanese sentence, and $\tilde{e}$ means an amended English sentence which is better for translation parameter training with j. The phrases with the same index are aligned. Based on these two sentences, different categories of alignments have been identified. For each category, examples are given:

According to whether the translation is continuous or not, we divide the alignments into 2 categories: 1. both the n-gram and its translation in the target language are continuous. 2. the translation in the target language contains gaps because of *syntactic divergence* (Dorr et al., 2004). We define "[X]" to stand for gaps in the target side as denoted by (Chiang, 2005) in syntax-based MT and we can have the following classifications:

- **Continuous Alignment**
  - **Bigram-to-ngram** the translation in the target language is continuous ngram, e.g., (1) *long tail* to *nagai sippo*.
  - **Bigram-to-unigram** the bigram corresponds to a unigram, e.g., (3) *a cat* to *neko*.
  - **Crossing-N-gram** the translation is continuous, but in a different order, e.g., (2) *cat with* to *no neko*.

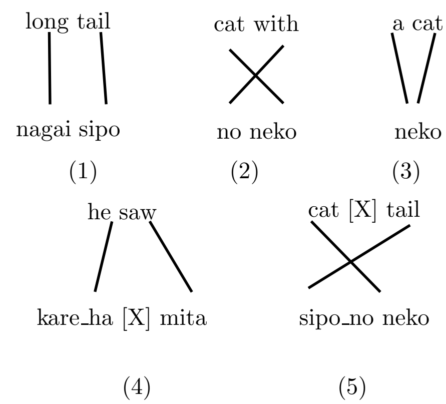

Figure 1: Various Alignments found in the experiment corpus, "[X]" stands gaps between words.

- **Discontinuous Alignment**
  - **Bigram-to-N-gram-with-gaps** a large number of translations in the target language are not continuous. This is a common phenomenon is illustrated by (4). *he saw* to *kara_wa [X] mita*.
  - **Crossing-N-gram-with-gaps** the bigram was aligned with dis-continuous words with gaps in the middle, at same time, the translation is in a different order, e.g., (5). *sipo_no neko* to *cat [X] tail*.

2.2 Proportional analogy

In this section, we describe employing analogy to deal with diverse alignments for bigram translation. We follow (Turney, 2006) to describe the basic notions of proportional analogy used in this work. Verbal analogies are often written *A : B :: C : D*. They meaning *A* is to *B* as *C* is to *D*. For example:

annual taxes : *annual statistics* :: *the taxes* : *the statistics*

The above example can be understood as follows: we reconstruct an unseen bigram *annual taxes* by a triple of known bigrams. All the elements in the unseen bigram is taken by similarity from the second (*annual statistics*) and third (*the taxes*) known bigrams and put together by difference with the fourth known bigram (*the statistics*). The definition of proportional analogy that we use in this paper is drawn from (Lepage, 1998) and we focus in this study on formal proportional analogies. A 4-tuple of n-grams *A, B, C* and D is said to be a proportional analogy if the following 3 constraints are verified. The lengths of the n-grams may be different, but should meet the following constraints:

1. $|A|_a + |D|_a = |C|_a + |B|_a, \forall a$
2. $d(A, B) = d(C, D)$
3. $d(A, C) = d(B, D)$

where d is the edit distance that counts the minimal number of insertions and deletions that are necessary to transform a string into another string. $|A|_a$ is the number of occurrences of the word a in the n-gram A. This approach still works well on different length of n-grams in fact. However, this method is a necessary condition but not sufficient when applying to translation issue.

As for bilingual translation using analogy, Denoual et al. (2007) presented a parallelopiped view on translating unknown words using analogy, we expand it to bigrams (see Figure 2). Suppose that we want to translate the following bigram (English): *annual taxes* into French, in order to translate the unknown bigram, bilingual proportional analogy requires a triple of source bigrams and corresponding translations. This procedure can be splitted into 2 steps:

1. reconstruct unseen bigram with a triple of source bigrams
2. translate using analogy

2.3 Bigram reconstruction

Given a bigram, it can be reconstructed using other n-grams via different reconstruction patterns. For instance, we can rebuild the bigram: *annual taxes* in following several ways:

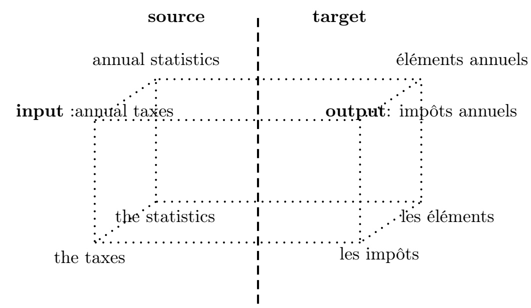

Figure 2: View of the harmonization parallelopiped: four terms in each language form a monolingual proportional analogy.

Pattern 1: *ab* : *ac* :: *db* : *dc*

annual taxes : *annual statistics* :: *the taxes* : *the statistics*

Pattern 2: *ab* : *b* :: *ac* : *c*

annual taxes : *taxes* :: *annual statistics* : *statistics*

Pattern 3: *ab* : *a* :: *db* : *d*

annual taxes : *annual* :: *the taxes* : *the*

Pattern 4: *ab* : *db* :: *ac* : *dc*

annual taxes : *the taxes* :: *annual statistics* : *the statistics*

Pattern 5: *ab* : *aeb* :: *ac* : *aec*

annual taxes : *annual income taxes* :: *annual statistics* : *annual income statistics*

annual taxes is reconstructed with different n-grams extracted from the training corpus. Beside these 5 Patterns, analogy in general can capture other various patterns in natural language.

We restrict to Pattern 1 in reconstructing of source bigrams because this Pattern contains more information of context and crossing-language alignment. On the contrary, we allow all Patterns in the target side as we want to collect as many translations as possible.

2.4 Translation by analogy

The problem that we define is, given an unseen bigram A in the source languages, supposing we have known an alignment between n-gram and its translation which is represented by a, we want to find the appropriate template T_i, to adapt the synchronous analogy and finally generate the target $\widetilde{A}'$ successfully. We formalize analogical deduction as following:

$$A : B_i :: C_j : x \qquad (1)$$

Assume the previous analogical equation has a solution x. We define the case when x belongs to the training set as "*reconstructible*". $\varphi(.)$ is the translation function, bidirectional analogical deduction also requires to repeat this operation with all target translations corresponding to the source bigrams in the opposite direction. In other words, satisfies following equation:

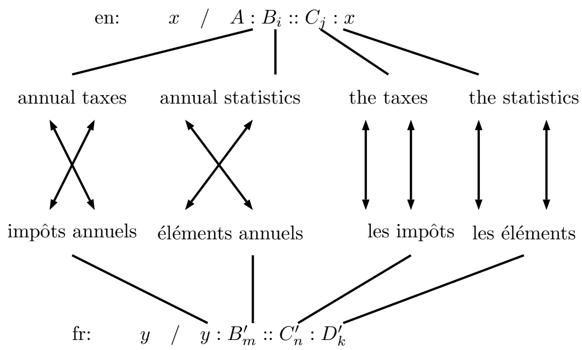

Figure 3: Bilingual analogical reduction for the bigram from the input *annual taxes* (English) to the output *impôts annuels* (French), the related analogous and its translation are indicated in the figure.

$$\exists(B'_m, C'_n, D'_k) \in \varphi(B_i) \times \varphi(C_i) \times \varphi(x)/ \qquad (2)$$

$$\exists y / y : B'_m :: C'_n : D'_k \qquad (3)$$

We define "*bidirectional reconstructible*" as when input an unseen bigram and finally it outputs the solution as y. In this model, a stands alignment between source language bigram and its translation in target language, $a \Leftrightarrow (X, X')$, if the alignment (A, y) appears in the test set (as $\exists y \in \varphi(A)$), we recognize the output as the translation, called "*attested translation*".

The Figure 3 describes this procedure and Figure 4 shows the details about constituents of bigrams. Since the proceeding of the whole produce of analogical derivation is very time-consuming, in order to evaluate the ceiling coverage of "*attested translation*", we conduct the synchronous parsing for fast obtaining the examples. It is easy to obtain the alignments between A and A' in the test set with some automatic aligners. From a bigram A and its translation A', for each elements in source side and with all relevant of bigrams $\widehat{B}, \widehat{C}$ from the source part of the bicorpus, if there also exists the translations $\widehat{B'}$, $\widehat{C'}$, we can reduce the remaining D and D' which is described as following formula:

$$(A, A'){:}(B_i, B'_m){::}(C_j, C'_n){\Rightarrow}(D, D')$$

If finally we find D and D' at the end of this equation are linked, we consider that from A it can arrive to A' successfully.

3 Data profiling

We first profile the test set by exploring the proportion of unseen bigrams in the source language. Then we investigate the reconstructiblility/bidirectional reconstructibility of unseen bigrams in the source language. Finally, we estimate the maximum of attested translation bigrams using this analogy-based approach.

3.1 Data preprocessing

We use the Europarl Corpora[1] (Koehn, 2005) to prepare the classification examples used to train and test the SVM classifier. We split the corpus into two parts: a training set and a test set. A set of 100,000 sentences which lengths less than 30 with the French translation are extracted as the training set. We also sample a set of 10,000 sentences from the remaining corpus not contained in training set as the test set. This corpus only offers aligned texts, however, it does not provide word alignment information for each language pair. Table 1 shows some statistic of bigrams and the proportion of unseen bigrams in the experiment data.

3.2 Word-to-word alignment

Before reconstructing, we preprocess to obtain word-to-word alignments. Our work is based on the dominant method to obtain word alignment, which trained from the Expectation Maximization (EM) algorithm. To extract the word alignment, EM algorithm will be utilized to train the bilingual corpus for several iterations, and then phrase pairs that are consistent with this word alignment will be extracted. We align the words automatically relying on the

[1] http://www.statmt.org/europarl/archives.html#v3

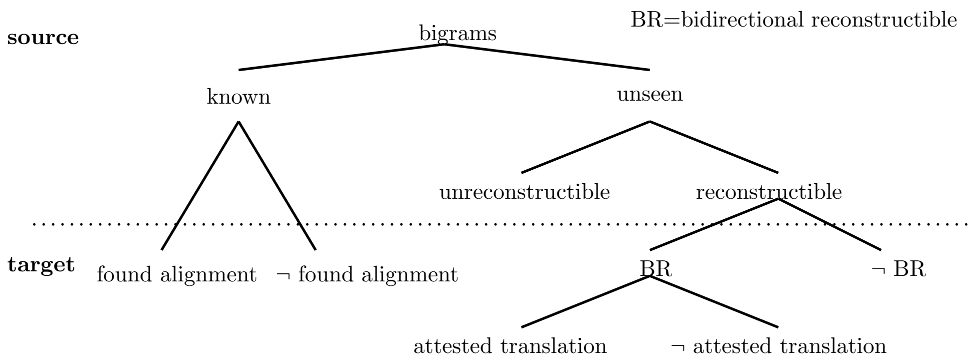

Figure 4: Logic binary tree for the problem of analogy and bidirectional analogy in the source language, "not found alignment" means the known bigrams that have not been aligned in the training set.

		English	**French**
Test	sentences	10k	10k
	words	177,890	202,418
	avg.(words/sentence)	17.79	20.24
	stdev.(words/sentence)	±6.24	±7.17
	bigrams (unique)	68,600	73,126
Training	sentences	100k	100k
	words	1,780,128	2,027,369
	avg.(words/sentence)	17.80	20.27
	stdev.(words/sentence)	±6.25	±7.16
	bigrams (unique)	345,384	336,995
Unseen	bigrams	22,078	23,251
	Proportion	32.18%	31.80%

Table 1: Statistics on the English-French parallel corpus used for the training and test sets, it also indicates the statistics of unseen bigrams in the test set.

		bigrams	**proportion**
Test	aligned	63,537	92.68%
	unaligned	5,063	7.38%
Training	aligned	320,983	92.94%
	unaligned	24,401	7.06%

Table 2: Statistics on the aligned and unaligned bigrams in data, it also indicates GIZA++ can not align all words in the source language after grow-diag-final-and.

		bigrams	**proportion**
known	¬ found alignment	995	1.45%
	found alignment	45,527	66.37%
unseen	reconstructible	20,056	29.14%
	unreconstructible	2,022	2.95%
Total		68,600	100.00%

Table 3: Distribution of bigrams, e.g., unaligned and aligned in the training data. More than 90% of unseen bigrams can be reconstructed.

GIZA++[2] (Ochet al., 2003) implementation of the IBM Models in Moses toolkit (Koehn et al., 2007), running the algorithm in both directions, source to target and target to source.

The heuristics applied to obtain a symmetrized alignment in this step is *grow-diag-final-and*, it starts with the intersection of directional word alignments and enrich it with alignment points from the union. We employ this algorithm to obtained alignment, and from that we extract the continuous bigrams and their aligned targets directly from the alignment files. At same time, an aligned test set was build as the golden reference using the same approach. "aligned" means it is aligned by GIZA++.

3.3 Reconstructiblity

Though the most of bigrams are reconstructible, not all bigrams belonging to this set can really generate a solution (case of BR) as same as the aligned translations in the target language. That is a quiet interesting and rifeness phenomenon in the most cases (case of $\neg BR$). We implement bilingual synchronizing parsing to quickly search the reusable and useful templates (case of *attested translation*). As the matter of fact, though not all final solution are acceptable, we are aiming at to bound the mount of successful analogy in total. The statistics are provided in the following.

[2] http://www.statmt.org/moses/giza/GIZA++.html

Negative Examples		Templates T_s: (B_i, B'_m) ,(C_j, C'_n), (D, D')		
Input: **Output:** **ref:**	joint development débattues [X] codes développement communautaire	joint talks débattues [X] pourparlers	the development des codes	the talks des pourparlers
Input: **Output:** **ref:**	rates within des taux [X] au sein [X] des taux de [X] au sein de	rates will des taux [X] permettra	areas within domaines [X] au sein	areas will domaines permettra
Input: **Output:** **ref:**	military security [X] de sécurité [X] militaires militaires [X] sécurité	military interests intérêts [X] militaires	our security nos [X] de sécurité	our interests nos intérêts
Input: **Output:** **ref:**	common set limites communes une série	common institutions institutions communes	the set les limites	the institutions les institutions
Positive Examples		Templates T_s: (B_i, B'_m) ,(C_j, C'_n), (D, D')		
Input: **Output:** **ref:**	this renegotiation cette renégociation cette renégociation	this transition cette transition	the renegotiation la renégociation	the transition la transition
Input: **Output:** **ref:**	accounts procedure procédure [X] comptes procédure [X] comptes	accounts for comptes de	voting procedure procédure [X] vote	voting for vote de
Input: **Output:** **ref:**	efficient legal judiciaire [X] efficace judiciaire [X] efficace	efficient european européen efficace	of legal judiciaire [X] de	of european européen de
Input: **Output:** **ref:**	bold measures des mesures audacieuses des mesures audacieuses	bold proposals des propositions audacieuses	various measures diverses mesurese	various proposals diverses propositions

Table 5: Samples of bigrams and related analogical templates, according $(B_i, B'_m), (C_j, C'_n), (D, D')$, the translation A' is produced. Both positive and negative examples are presented in the table.

	reconstructible			
	BR			$\neg BR$
	attested	unattested	total	
bigrams	7,659	10,347	18,006	2,050
proportion	11.16%	15.09%	26.25%	2.99%

Table 4: Distribution of bigrams, e.g., attested translation and unattested translation using analogy, it means more than 3/4 (66.37%+11.16%) of bigrams are attested translation only referring to the training data.

3.4 SVM Classifier

Since the proportional analogy for translation mapping is the necessary condition but not sufficient, identifying the correct translation via proportional analogy with some machine learning approaches is very necessary. In the following, we will describe how we collect the examples and from them to extract the features to train the SVM classifier. It implements the estimating-processing by using the specified features: independent features from (A, A') as well as relative features from analogical templates of (B_i, B'_m), (C_j, C'_n), (D, D').

3.4.1 Features

For classifying the outputs as correct translation or not, the software LIBSVM[3] (Chang et al., 2012) in used, which is an integrated software comes with scripts that automate normalization of the features and optimization of the γ and C parameters. We still need to restrict the features to feed it for training.

- **Independent Features**

  Lexical Weighting: the direct lexical weighting $P_{lex}(e|f)$ and inverse lexical weighting $P_{lex}(f|e)$ for (A, A'). Given a word alignment a, we apply the formula of IBM Model 1 to compute the lexical translation probability of a phrase e given the foreign phrase f as (Koehn

[3]https://www.csie.ntu.edu.tw/ cjlin/libsvm/

et al., 2003):

$$P_{lex}(e|f,a) = \prod_{i=1}^{I} \frac{1}{\{j|(i,j) \in a\}} \sum_{\forall(i,j)\in a} w(e_i|f_j) \quad (4)$$

Here, we compute the score as the following equation without the word alignment:

$$P_{lex}(e|f) = \frac{1}{I} \sum_{i=1}^{I} \log \max_{\{j|\forall(i,j)\in a\}} \{w(e_i|f_j)\} \quad (5)$$

Length: the lengths of A' in words, '[X]' should not be recognized as a word, because it can be ε.

Frequency: we compile the data with the suffix array for fast searching (Lopez, 2007). We calculate the frequency of occurrence for each n-gram generated by analogy in French (with/without gaps). The complete French subset of Europarl corpus is used as the reference.

	Reference (French)
sentences	386,237
words	12,175,424
avg.(words/sentence)	31.52
stdev.(words/sentence)	±6.24

Table 6: Statistics on the French monolingual corpus used as reference.

MutualInformation: It is considered as the most widely used measure in extraction of collocations. We only compute the score only for A' as following:

$$I(X) = \log \frac{p(w_1, w_2, .., w_m)}{\prod_{i=1}^{m} p(w_i)} \quad (6)$$

- **Relative Features**

LexicalWeight: the lexical weightings of (B_i, B'_m), (C_j, C'_n) and (D, D') in both directions (direct lexical weighting $P_{lex}(e|f)$ and inverse phrase translation probabilities $P_{lex}(f|e)$. Blue triangles stand positive examples and red circles stand negative examples. We found that the output with the balanced template in lexical weighting does not mean it has the larger probability to be a positive examples.

Length: the lengths of B'_m, C'_n and D' in words, "[X]" should not be recognized as a word, because it can be ε.

Frequency: the occurrences of B_i, C_j and D and same to targets.

Dice's coefficient: Dice coefficient measures the presence/absence of data between to phrases, where $|X|$ and $|Y|$ are the number of words in set X and Y, respectively, and $|X \cap Y|$ is the number of words shared by the two set. We import the following formula to compute the score of Dice coefficient among B', C' and D', e.g.:

$$Dice(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|} \quad (7)$$

MutualInformation: This measures the co-occurrence phrases mutual dependence. x stands the word in source bigram and y stands the word in the solution of analogy. $p(x, y)$ is the word-to-word translation probability. $p(.)$ is the probability distribution function.

$$I(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (8)$$

3.4.2 Problem formulation

As we treat verifying analogy output as a binary classification problem, we obtained various outputs from analogy engine for each bigram. $\varphi(.)$ is the translation function, we label the training examples as in (5):

$$y = \begin{cases} 1, & if\ A' \in \varphi(A) \\ 0, & if\ A' \notin \varphi(A) \end{cases} \quad (9)$$

Each instance is associated with a set of features that have been discussed in the previous section.

3.4.3 Experimental settings

The bilingual-crossing examples are generated by the previous script depends on the alignment output by GIZA++. During training of the SVM classifier, positive and negative instances of examples are generated from the subset of *attested translation* and unusable templates in the middle of analogy proceeding. We also build a test set to validate the accuracy of such a classifier.

	Negative	Positive	Total
Test	1k	1k	2k
Training	5k	5k	10k

Table 7: Size of the examples used as the test set and training set in the experiment.

3.4.4 Evaluation

To test the performance of our approach we focus on the accuracy of the results. We first sample 2k examples as test data (as in Table 7). During training the SVM classifier determines a maximum margin hyperplane between the positive and negative examples. We measure the quality of the classification by precision and recall. Let C be the set of output predictions. We standardly define precision P, recall R and F-measure as in (10):

$$P = \frac{C_{tp}}{C_{tp} + C_{fp}}, R = \frac{C_{tp}}{C_{tp} + C_{fn}}, F = \frac{2PR}{P + R} \quad (10)$$

It should be noted that the number of examples for training are different for the systems of different language pairs. Because we are interested in the possibilities of found translation, we used the standard accuracy measure to evaluate the performance of classifier on the test set:

$$accuracy = \frac{C_{tp} + C_{tn}}{C} \quad (11)$$

where C_{tp} is the counts of true-positive and C_{tn} is the counts of true-negative. C is the total counts of candidates. We show the details of evaluation scores in Table 8.

4 Conclusion and Future works

In this paper we have performed an investigation on translating unseen bigrams in MT by employing an analogy-based method empirically, which has never been explored. We investigated the maximum possible coverage of bilingual reconstructible bigrams in the test and the probabilities when a bigram is attested translation by using the analogy.

As can be noticed from the presented results, after importing the features of templates which are used in analogy diveration, the performance of SVM classifier improves. In other words, it means that the analogous information has the positive effects on classification.

Though the accuracy is not as high as we expected, there are some reason can explain it, first, even the alignment output by GIZA++ is still so far from completely correct, and second, the used features are very simple. Moreover, without the contextual information, this result should be acceptable. The results suggest lexical weighting and mutual information contribute most to identifying the correct translation.

Another should be addressed that bigrams translation is the most difficult in analogy-based machine translation. If a bigram is attested translation, unquestionable, it will help the longer n-grams translation.

The future works should focus on identifying the proper longer chunk/phrase translations using the similar approach.

Acknowledgments

This work is supported in part by China Scholarship Council (CSC) under the CSC Grant No.201406890026 is acknowledged. We also thank the anonymous reviewers for their insightful comments.

References

Dice, L.R. 1945. Measures of the amount of ecologic association between species. *Ecology*, Vol.26, No.3, pp.297–302.

Bonnie J. Dorr. 1994. Machine translation divergences: A formal description and proposed solution. *Computational Linguistics*.20.4: pp.597–633.

Philippe Langlais and Alexandre Patry 2007. Translating unknown Words by Analogical Learning. *In EMNLP/CoNLL'07*, pages 877–886, Prague, Czech Republic.

Sandipan Dandapat, Sara Morrissey, Sudip Kumar Naskar, and Harold Somers. 2010. Mitigating problems in analogy-based ebmt with smt and vice versa: a case study with named entity transliteration. *In 24th Pacific Asia Conference on Language Information and Computation (PACLIC'10)*, pages 365–372, Sendai, Japan.

Ron Bekkerman and James Allan. Using bigrams in text categorization. Department of Computer Science, University of Massachusetts, Amherst 1003 (2004): 1-2.

Features Used		Precision	Recall	F-measure	Accuracy
Independent Features	**Length**	65.51%	71.60%	68.42%	66.95%
	LexicalWeight	68.32%	81.11%	**74.47%**	71.75%
	Freq	**82.76%**	2.40%	4.66%	50.95%
	MutualInfo	62.74%	**86.90%**	72.87%	67.65%
	Length+LexicalWeight+Freq+MutualInfo	69.92%	78.92 %	74.15%	**72.48%**
Relative Features	**Length**	65.64%	72.60%	68.95%	67.30%
	LexicalWeight	64.97%	71.60%	68.13%	66.50%
	Freq	**71.90%**	21.50%	33.10%	56.55%
	Dice	65.52%	72.20%	68.70%	67.10%
	MutualInfo	63.28%	85.30%	72.66%	**67.90%**
	Length+LexicalWeight+Freq+MutualInfo	62.74%	**86.90%**	**72.87%**	67.65%
Independent Features + Relative Features	**Length**	65.54%	73.40%	69.25%	67.40%
	LexicalWeight	68.71%	79.70%	73.80%	71.70%
	Dice+Length	65.10%	58.20%	61.46%	63.50%
	LexicalWeight+Length	63.18%	80.15%	70.66%	66.73%
	LexicalWeight+Length+Dice	70.01%	85.80%	77.10%	74.52%
	LexicalWeight+Length+Freq+MutualInfo	71.83%	86.20%	78.36%	76.20%
	LexicalWeight+Length+Dice+Freq+MutualInfo	**73.32%**	**87.33%**	**79.71%**	**77.78%**

Table 8: Classifier's performance on identification the successes of bigram translation.

Dekai Wu. 1997. Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. *Computational linguistics*.23.3: pp.377–403.

Yves Lepage. 1998. Solving analogies on words: an algorithm. *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics.

Philip Resnik, Douglas Oard and Gina Levow. 2001. Improved cross-language retrieval using backoff translation. *In Proceedings of the Proceedings of the First International Conference on Human Language Technology Research (HLT)*.

Nizar Habash and Bonnie Dorr. 2002. Handling translation divergences: Combining statistical and symbolic techniques in generation-heavy machine translation. *Springer Berlin Heidelberg*.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational linguistics*29.1, pp.19–51.

Arul Menezes and Stephen D. Richardson. 2003. A best-first alignment algorithm for automatic extraction of transfer mappings from bilingua corpora. *Recent advances in example-based machine translation.* Springer Netherlands. pp.421–442.

Bonnie Dorr, Necip Fazil Ayan and Nizar Habash. 2004. Divergence Unraveling for Word Alignment of Parallel Corpora. *Natural Language Engineering,* 1 (1), pp.1–17.

Philipp Koehn, Franz Josef Och and Daniel Marcu 2003. Statistical phrase-based translation. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*.Association for Computational Linguistics

Michel Simard, Nicola Cancedda, Bruno Cavestro, Marc Dymetman, Eric Gaussier, Cyril Goutte, Kenji Yamada, Philippe Langlais and Arne Mauser. 2005. Translating with non-contiguous phrases *In Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP)*. Association for Computational Linguistics, pp.755–762.

Chris Callison-Burch, Colin Bannard and Josh Schroeder. 2005. Scaling phrase-based statistical machine translation to larger corpora and longer phrases. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics.

Yuan Ding and Martha Palmer. 2005. Machine translation using probabilistic synchronous dependency insertion grammars. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics.

Peter D. Turney and Michael L. Littman. 2005. Corpus-based learning of analogies and semantic relations. *Machine Learning*60.1-3 (2005): pp.251–278.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. *MT summit*. Vol. 5

J. M. Crego, M. R. Costa-Jussa, J. B. Mariño and J. A. Fonollosa. 2005. N-gram-based versus phrasebased statistical machine translation *In Proceedings of the International Workshop on Spoken Language Technology (IWSLT'05).*

Josep M. Crego, José B. Mariño and Adrià de Gispert. 2005. An Ngram-based statistical machine translation decoder. *Proc. of the 9th European Conference on Speech Communication and Technology (Interspeech'05).*

David Chiang. 2005. A hierarchical phrase-based model for statistical machine translation *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics.*Association for Computational Linguistics

Peter D. Turney 2006. Similarity of semantic relations *Computational Linguistics*.32(2):pp.379–416.

Adam Lopez and Philip Resnik. 2006. Word-based alignment, phrase-based translation: What's the link. *Proc. of AMTA*.

Etienne Denoual. 2007. Analogical translation of unknown words in a statistical machine translation framework. *Proceedings of Machine Translation Summit XI.* Copenhagen

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. *In Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*. Association for Computational Linguistics, pp.177–180

Sabri Bayoudh, Harold Mouchère, Laurent Miclet and E. Anquetil. 2007. Learning a classifier with very few examples: analogy based and knowledge based generation of new examples for character recognition. *Machine Learning: ECML 2007*, pp.527–534.

Adam Lopez. 2007. Hierarchical Phrase-Based Translation with Suffix Arrays. *EMNLP-CoNLL*, pp.976–985

Peter D Turney. 2008. A uniform approach to analogies, synonyms, antonyms, and associations. *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pp.527–534.

Maike Erdmann, Kotaro Nakayama, Takahiro Hara and Shojiro Nishio . 2009. Using an SVM Classifier to Improve the Extraction of Bilingual Terminology from Wikipedia. *User-Contributed Knowledge and Artificial Intelligence: An Evolving Synergy*, pp.15.

Yves Lepage and Etienne Denoual. 2005. The 'purest' EBMT system ever built: no variables, no templates, no training, examples, just examples, only examples. *Proceedings of the MT Summit X, Second Workshop on Example-Based Machine Translation*, pp.81–90.

Yves Lepage, Julien Gosme and Adrien Lardilleux. 2010. The structure of unseen trigrams and its application to language models: A first investigation. *Universal Communication Symposium (IUCS), 2010 4th International.* IEEE. pp.944–952.

Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. 2010. Automatic evaluation of translation quality for distant language pairs. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing *Association for Computational Linguistics*. pp.944–952.

Ahmet Aker, Yang Feng and Robert J. Gaizauskas. 2012. Automatic Bilingual Phrase Extraction from Comparable Corpora. *COLING*. pp.23–32.

Chang, C. C., and C. J. Lin. 2012. LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology*. 2: 27: 1-27: 27.