

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Results Comparison of Machine Translation by Direct Translation and by Through Intermediate Language

Phan Thi Le Thuyen¹

University of Danang
Danang, Vietnam

Vo Trung Hung²

University of Danang
Danang, Vietnam

Abstract: *Accurate evaluation of machine translation is a very interesting problem in the field of automatic translation. The translation tools are mainly supported language pairs having been used in the world and very limited for the poor language pairs about bilingual corpus or linguistics resources. However, we can use Interlingua language which has a large bilingual corpus (e.g. English) to translate. In this paper, we use the free online Google Translation tool for testing the two methods: direct translation from English → Vietnamese and interlingual translation using French, German. We evaluated and compared results of the two translation methods by BLUE and NIST scores.*

Keywords: *machine translation, direct translation, interlingual translation, natural language processing, evaluation of machine translation*

I. INTRODUCTION

Automatic translation or machine translation is the process of translating a text from this language (source language) into another language (target language). In this process, there can appear human intervention or can be fully automatic performance.

Supposing that we use the method of direct translation for each language pair, so with n language we have developed $n*(n-1)$ of the translation pairs. Currently, there are more than 5,000 spoken languages and about 1,500 written languages in the world. Therefore, it's really difficult to develop automatic translation pair of all languages in use.

One solution for translating a text from a source language into a target language is using intermediate language. For example, to translate a Vietnamese text into Chinese (but not the translation pairs Vietnamese → Chinese), we will be use English as an intermediary and then perform 2 times of translation from Vietnamese into English and from English into Chinese. Thus, with n languages, instead of $n*(n-1)$ translation pairs, we require only $2*n$ translation pairs.

However, has the quality of the translation version been affected when using an intermediate language, and how influential level?

In order to answer this question, we conducted a study to evaluate the different levels of direct translation version and one by using an intermediate language.

To evaluate, we use bilingual data warehouse of English - Vietnamese including a set of sentences written in Vietnamese and English. Then we conduct automatic translation (using Google translate) under two ways. The first method is the direct translation (English - Vietnamese), and the second method is using French as an intermediate language (English - French and French - Vietnamese). The next step, we compare the results of the two translation versions with Vietnamese content available in the data warehouse to find the differences. Different levels are evaluated by three methods: direct methods (strict comparison), BLEU method [1] and NIST method [6]. The results show mainly differences in the results received when translate directly and via intermediate language.

In this paper, we present the methods of evaluating the quality of the translation, the suggested model and the evaluation experiment. The results of this study is to help us with finding more other solutions in multilingual automatic translation, instead of using a particular natural language (e.g. English) to make an intermediate language in the translation system.

II. RELATED WORK

The first approach on how to support query operations on encrypted data with bucketization, after the data is encrypted, the ciphertext is concatenated to a bucket number, which is assigned to a specific range that includes the data. When a user requests a query operation, the server uses the bucket numbers to execute the query operation. For example, if a client program wants to

1. BLEU

BLEU (Bilingual Evaluation Understudy), available standardized results a method of evaluating machine translation results, is based on comparing the results a machine translation with results available standards which are checked by humans. This is the method proposed by IBM in 2001 and then used widely. With this method, BLEU enables automatic evaluation (through computer programs) the degree of equivalence between the available standardized translations with automatic translation.

In BLEU method, a comparison was conducted through statistical coincidence of words in two translation taking into account their order in the sentence (N-gram method per word) from the data warehouse of translating results and high quality reference translation.

To compare a machine translation and standardized translation, BLEU method offers 2 steps. The first step is counting the minimum number of clusters N-grams appear in each sentence pairs (the machine translation sentence and the sample sentence), and then the total is divided by the total number of clusters on N-grams in machine translation. Scores are calculated according to the following formula:

$$scoreBLEU = \exp \left\{ \sum_{i=1}^N w_i \log(p_i) - \max\left(\frac{L_{ref}}{L_{tra}} - 1, 0\right) \right\}$$

In which:

$$P_i = \frac{\sum_j NR_j}{\sum_j NT_j}$$

- NR_j : is the number of N-grams in the segment j of translation text using for reference

- NT_j : is the number of N-grams in the segment j of machine translation

- $w_i = N-1$

- L_{ref} : is the number of words in the reference translation, its length is usually equal to the length of the machine translation

- L_{tra} : is the number of words in the machine translation

We can see the value of BLEU score ranges from 0 to 1. The lower BLEU score the translation has, the more coincident and accurate between machine translation and translation samples.

2. NIST

NIST [1] is developed based on the BLEU method. However, there is a difference in evaluation approaches; it is choosing N-gram and information on each N-gram to evaluate the quality of the translation.

If we change the position of elements on the same N-gram, evaluation score will be changed. This shows that the scores will change if we change the position of N-grams in the same segment. This change also greatly affects the quality translation evaluation. NIST offers a higher score evaluation showed in the N-gram containing more information while BLEU gives the same score if they have the same number of words. With this standard, if the score is calculated by NIST method higher and higher, the translation system is considered better.

The formula for calculating score of the NIST method is as follows:

$$score_{NIST} = \sum_{i=1}^N \left\{ \frac{\sum_{\forall w_1 \dots w_n} \inf(w_1 \dots w_n)}{\sum_{\forall w_1 \dots w_n \in D_{tra}} \log(p_i)} \right\} \cdot \exp \left\{ \beta \log^2 \left[\min \left(\frac{L_{tra}}{L_{ref}}, 1 \right) \right] \right\}$$

In which:

The information parameter used to calculate the score of N-grams in the set of all reference translation follows the equation below:

$$\inf(w_1 \dots w_n) = \log_2 \left(\frac{N_1}{N_2} \right)$$

N_1 = the corresponding number of words: $w_1 \dots w_{n-1}$

N_2 = the corresponding number of words: $w_1 \dots w_n$

β is the coefficient that was chosen equal 0.5 when the number of words in the translation machine less than or equal to 2/3 the number of words in the reference translation, whereas the $\beta = 1$

$N = 5$ (used for all phrases that is 5 in length)

L_{tra} : the number of words in machine translation

L_{ref} : the number of words in the reference translation

3. UD CORPUS

Parallel Linguistics Corpus is a repository of bilingual text pairs represented in electronic format, in which each text of this language is a translation of the other languages [3].

Currently, there are many bilingual corpus which are used for many different purposes such as: dictionaries, automatic translation, searching information through language, the study of linguistics, language learning, etc. For example, the British National Corpus (BNC) with 100,000,000 words; Canadian Hansard Corpus with 90 million from the English - France; JENAAD Japanese - English parallel corpus containing 150,000 sentence pairs; PKU 863 3066435 from English and Chinese [2];... In Vietnam, if are required 1-1 requires translations, it could be the famous parallel multilingual corpus Bible with hundreds of languages, Harry Potter (Chinese-Korean- Japanese-English), bilingual websites are English-Vietnamese VOV Online, thanhnien, VietnamNet, NhanDan, ect. As for bilingual corpus (1-1) of translation English – Vietnamese, It is labeled and EVC is the most popular one.

To evaluate the translation system, we prepare the bilingual data with the best quality to make text translation and reference translation. We inherit data warehouse taken from the results of [2] which was extracted from various sources. The data is an original in Vietnamese data, English and French with 58719 sentences in XML format to make the input data for the translation system and evaluation system.

```

<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
- <root>
  <sen id="1">The earth is round</sen>
  <sen id="2">He is a teacher</sen>
  <sen id="3">There is a concert today</sen>
  <sen id="4">Are you often in town?</sen>
  <sen id="5">To be or not to be, that is the question</sen>
  <sen id="6">They'll be linguists in some years</sen>
  <sen id="7">When is the wedding to be?</sen>
  <sen id="8">This book is five pence</sen>
  <sen id="9">What time am i to come?</sen>
  <sen id="10">He is to leave for hanoi tomorrow</sen>

```

Fig. 1 For example of the English data

```

<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
- <root>
  <sen id="1">Quả đất thì tròn</sen>
  <sen id="2">Anh ta là giáo viên</sen>
  <sen id="3">Hôm nay có một buổi hoà nhạc</sen>
  <sen id="4">Anh thường có ở tỉnh không</sen>
  <sen id="5">Sống hay là chết, đó mới là vấn đề</sen>
  <sen id="6">Vài năm nữa họ sẽ trở thành những nhà ngôn ngữ học</sen>
  <sen id="7">Bao giờ đám cưới sẽ cử hành?</sen>
  <sen id="8">Cuốn sách này giá năm xu</sen>
  <sen id="9">Mấy giờ tôi phải đến</sen>
  <sen id="10">Ngày mai nó sẽ đi hà nội</sen>

```

Fig. 2 For example of the Vietnamese data

4. GOOGLE TRANSLATE

Google Translate was launched in 2001 and now has become the popular tool in the Internet environment.

Google translate is a free translation tool for word, phrase, sentence, file or URL with about 80 different languages. This tool with visual interface is both capable of automatic translation with relatively precise meaning and investigating the meaning of the word with fast speed.

In 2013, according to statistics of <http://www.cnet.com>, there are more than 200 million users of Google Translate per day with over 1 billion translation participations.

Currently, Google Translate can not yet supported for automatic translation of languages all over the world. However, in some cases, we can translate via intermediate language (e.g. English), but the quality of such translations can be reduced.

In this paper, to compare the quality of the translation directly and translation through intermediate language, we use the Google Translate tool for testing.

III. PROPOSED SOLUTIONS

Proposed model enables automatic evaluation of quality of the translation when directly translated from this language to the other language and translation through intermediate language.

With direct translation, the system simply sends the text to be translated to the online translation tool once, Google Translate will perform a direct translation between the source language and the target language. As for the intermediate translation, the system will have to send the text in online translation tool twice. For the first time, translation tools will translate from the source language into the intermediate language (French) and gives a temporary version. Next, the system will send it to the tool again. Here, translation tools will translate from the intermediate language into the target language.

The system becomes a tool for comparison of results based on the quality translation evaluation of NIST and BLEU method

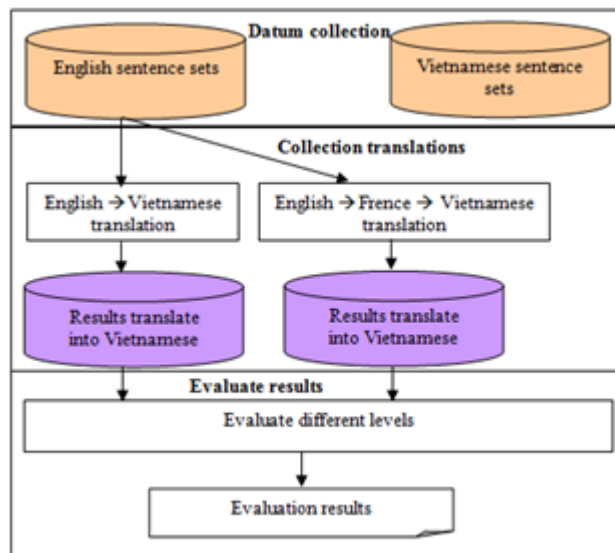


Fig. 3 Evaluation system model

With this model, we need to have a set of original English sentences and a corresponding set of the Vietnamese (These standardized sets sentences must be test by experts in the language). Through automatic translation function the evaluate system offers to two versions. One version is translated directly from the source language into the target language. Another version is translated from a source language into the target language through an intermediate language. Then, the system will assess the accuracy of two text machine translation based on reference of translation.

The function of evaluation the translation quality will be used by both BLEU and NIST methods. Based on the comparison of overlap words with words between evaluation translation and a reference translation the two methods will give the evaluation. In case there is more than one reference translation, the system will be conducted to compare translation need evaluating with a reference translations is the better translation will get higher scores.

IV. EXPERIMENT

We conducted translation experiment by two different methods and evaluated the quality of two translations. The system gave the results as follows:

TABLE I
1st evaluation results

Evaluation method Translation	BLEU	NIST
Direct:English→Vietnamese	0.00071	2.15090
Indirect:English→Frence→Vietnamese	0.00075	1.97572

When comparing the statistics with the evaluation results, it showed that the direct method will offered better results than an intermediate language method.

When using German as an intermediate language to test the program once more time, we had the following results:

TABLE II
2nd evaluation results

Evaluation method Translation	BLEU	NIST
Direct:English→Vietnamese	0.00071	2.15090
Indirect:English→German→Vietnamese	0.00072	1.85885

Although BLEU and NIST scores when using German as intermediate language was different from using French as an intermediate language, the direct method showed better result than that of an intermediate language method.

Besides, the statistic obtained from the two translation methods also produced similar results with the results of the two methods of assessing the quality of translation of NIST and BLEU.

TABLE III
The rate similar / different between two machine translations

Machine translation	Reference translation	
	The same sentences	Different sentences
Direct translation	18155(31%)	40564(69%)
Through the intermediate language translation	17602(29%)	41117(71%)

V. CONCLUSION

From the results of three tables, which show comparing result of the translation quality between direct method and intermediate language method, we reach some suggestion as follows:

- » With Google Translate tool, translation quality through an intermediate language is not as equal as that of direct translation;
- » In the inter-language translation system, the use of natural language as an intermediary language is not a good solution to translate between different languages;
- » To have a higher quality translation for multilingual translation system, we need to select a language that allows performing all the knowledge of natural language without encountering of problems multi-semantics of the word, semantic ambiguity, grammar and context dependence.

In the future, we will study a system of multilingual automatic translation that uses UNL (Universal Networking Language) as intermediate language. The UNL is a declarative formal language specifically designed to represent semantic data extracted from natural language texts. It can be used as a pivot language in interlingual machine translation systems or as a knowledge representation language in information retrieval applications. We will conduct tests to compare quality between using natural language (English, French,...) and using UNL as intermediate language.

References

1. V. T. Hung (2007), "Methods and Tools for automatic evaluation of online machine translation system", Journal of Science and Technology, the University of Danang, Volumn 1 (18), p.p. 37-42.
2. V. T. Hung, N.T. Hoa (2013), "Building a multilingual corpus for English – French – Vietnamese", Journal of Science and Technology, the University of Danang, Volumn 4 (7), p.p. 24-31
3. H. Q. Bao, D. Dien, D. B. Van, L. V. Minh (2008), "Technical report of project SP.74", <http://www.jaist.ac.jp/~bao/VLSP text/March2008/>
4. L. K. Hung (2003), "Multiligual translation", Proceeding ICT.rda'03, p.p. 282-291.
5. V. T. Hung (2004), "Reuse of Free Online MT Engines to Develop a Meta-system of Multilingual Machine Translation", Proceeding ESTAL 2004, Morocco.
6. K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu(2002), "BLEU: a Method for Automatic Evaluation of Machine Translation", In Processing of 40th Annual Meeting of the Association for Computational Linguistics (ACL), pages 311-318.
7. G. Leusch, N. Ueffing, and H. Ney (2003), "String-to-String Distance Measure with Applications to Machine Translation Evaluation", In Processing of MT Summit IX. New Orleans, LA. Sept. 2003. pp. 240-247.
8. D. Dien (2006), "Natural Language Processing", Published by Ho Chi Minh publisher.
9. S. Tripathi and J. K. Sarkhel (2010), "Approaches to machine translation", Annals of Library and Information Studies, vol. 57, pages 388-393.
10. Y. Zhou, p. Liu, and C. Zong (2011), "Approaches to Improving Corpus Quality for Statistical Machine Translation", International Journal of Computer Processing Of Languages, vol. 23, pages 327-348.