Applicability of Resource-based Machine Translation to Airplane Manuals

Eiko YAMAMOTO
Kobe University / NICT

AkiraTERADA
Japan Airlines

Hitoshi ISAHARA
National Institute of Information and Communications Technology (NICT)

Abstract

Machine translation (MT) has been studied and developed since the advent of computers, and yet is rarely used in actual business. For business use, rule-based MT has been developed, but it requires rules and a domain-specific dictionary that have been created manually. On the other hand, as huge amounts of text data have become available, corpus-based MT has been actively studied, particularly corpus-based statistical machine translation (SMT). In this study, we tested and verified the usefulness of SMT for aviation manuals. Manuals tend to be similar and repetitive, so SMT is powerful even with a small amount of training data. Although our experiments with SMT are at the preliminary stage, the BLEU score is high. SMT appears to be a powerful and promising technique in this domain.

1 Introduction

Various services such as information retrieval and information extraction using natural language processing technologies trained by huge corpora have become available. In the field of translation, corpus-based machine translation, such as statistical machine translation (SMT) and example-based machine translation (EBMT), are typical applications of such large volumes of data in actual business situations.

But is machine translation really applicable to actual business situations? One study examined for what types of people current machine translation (MT) systems are useful (Fuji et al., 2001), by simulation of reading of web pages in a language different from one's mother tongue. However, there has been little research to verify the usefulness of MT systems in specific business situations.

In this paper we verify the usefulness of resource-based, or corpus-based, translation in the real-world business domain of aviation. This study is important from both a business perspective and an academic perspective. Even though the research is still at a preliminary stage, the BLEU scores in our experiments are already around 50, proving that corpus-based machine translation is applicable to real business situations.

2 Corpus-based MT

Generally speaking, corpus-based MT requires huge amounts of text data. When we treat pairs of languages whose linguistic features are rather different, the amount of parallel text required increases. Compared with rule-based approaches which use the concept of classification, e.g. abstraction, SMT normally considers a sentence as a string of characters and performs poorly when the grammatical features of the input language and the output language differ widely. In comparison, EBMT uses dictionary lookup and even shallow syntactic analysis, and so is considered to be more robust than SMT in such situations.

We are therefore developing a syntactically augmented EBMT system between Japanese and Chinese for scientific documents[1] (Isahara et al., 2007). The use of syntax analysis means that the head of a clause is treated not as a word, i.e. a character string, but as an instance of a class.

[1] The mechanism adopted in this project is widely applicable to other domains and languages. We applied the same mechanism to a system for translating tourist information in Beijing and for translating between Japanese and Thai.

One of the problems in machine translation is selecting proper equivalents. Although researchers claim that SMT and EBMT can cope with this problem provided there is sufficient training data, it is still not known whether SMT or EBMT, which mainly use not semantic (class) information but surface strings, can choose proper equivalents of fundamental terms in particular.

If we consider translation in a business context, i.e. translation used in a company and/or translation as a business, there is frequent repetition. Documents such as reports and announcements are often written with a similar style, vocabulary and contents, and are translated repeatedly. Other examples of this type are manuals, which are revised frequently, and there may be similar manuals for similar products, with some parts of the manuals being identical.

Intuitively, manuals for similar products, or manuals for different versions of the same product, are likely to resemble each other. Therefore, even with only a small amount of training data, a corpus-based MT system can output useful translations. The corpus-based approach is powerful when the target is repetitive.

3 MT in the Real World

Corpus-based MT needs a huge text corpus, particularly when handling languages whose linguistic features differ widely. Furthermore, because SMT does not utilize grammatical information, there is doubt as to whether SMT is applicable to business translation.

However, we believe that if we can restrict the domain and usage properly, even current MT technology is useful for actual translation needs.

In this paper, we use aviation manuals and verify the usefulness of SMT as an example of corpus-based machine translation. Manuals are characterized by similar sentences appearing repeatedly, and by the restricted number and usage of vocabulary. Therefore, even with a comparatively small amount of training data, we can expect high-quality translation results. The output of the MT system is not used directly as the final translation text, but is manually checked by plural post-editors. Therefore, provided post-editing is performed, SMT can be used as a translation aid. When we restrict the target domain, variations of linguistic expressions are restricted and therefore statistical processing can be performed.

For the sake of consistency, manuals must use the same expressions to convey the same meanings. In this respect, because SMT, i.e. machine translation utilizing statistical information, uses examples translated previously, such consistency of representation is easily maintained.

Aircraft manufacturers produce series of aircraft, but the manuals for each series are written independently. Airline companies whose official language is other than English must first write aviation manuals in their own language, by editing contents based on procedures defined in their operation policy and the optional systems they select, for each type of aircraft in their fleet, and then translate them into English for overseas use.

Airline companies need to prepare many manuals by hand for the different aircraft they produce. Whenever an aircraft is revised or enlarged, the manuals must be created again in a similar way.

Although manuals are usually created separately for each aircraft, they may be very similar in style and contents. Furthermore, a series of aircraft sold by an aircraft manufacturer is likely to be similar in operation, and so the manuals will also be similar. However, airline companies have traditionally translated each manual from scratch. Machine translation technology could be usefully tailored to such companies.

In this study, we used a collection of aviation manuals in Japanese and English. For example, we translated the text in a manual by using all the text in other manuals as resources.

4 Some Interesting Features

This study has several interesting aspects. Firstly, previously translated documents are often used as the basis for making new translations in the translation process. Therefore, the process conducted in this study simulates actual translation tasks.

Nowadays, although most researches on corpus-based machine translation use the BLEU score for evaluation, it is very hard to generate many reference translations in different styles for evaluation. In contrast, in our study the target translation already exists in the form of manuals in English and Japanese, so we need no other reference translations.

Another criticism of the BLEU score is that it is calculated as the average of correct n-grams and there is no concrete relation between the BLEU score and usefulness in a real-world task. Again, this problem does not exist in our study. There are at least two ways to use the output of MT systems, i.e. read the MT output for information acquisition such as personal translation of web pages, and use the MT output as a preliminary translation for post-editing and publishing. Our situation is the latter, so one of the evaluation criteria is the cost of post-editing, e.g., number of key strokes, or edit distance, for post-editing. As shown in the following section, the BLEU scores in our experiments are around 50; the BLEU score reflects how much of the MT output can be used in the final translation and therefore is related to the cost of post-editing.

Other features of our study were as follows.

There are many identical or similar sentences, syntaxes and terms in manuals, making them suitable for a corpus-based MT system. Therefore, even a small amount of parallel text effectively covers linguistic phenomena. In our smallest experiment, 7,000 sentence training data can generate translations with a BLEU score of about 50.

Generally speaking, corpus-based MT, or automatic MT in general, has the merit that the same equivalents are always selected in the same context. Especially, terms in manuals have few variants for a concept, so it is easy to acquire proper equivalents from corpora.

Normally, it is very hard for MT systems to do free translation, i.e. not word-for-word translation, however, if we limit the target documents to specific manuals, the system can learn such free translation from corpora, i.e. phrase-for-phrase translation. In addition, translation of double negatives into positive expressions, tuning for domains, and acquisition of long words are possible.

5 Statistical Machine Translation (SMT) Experiments

To verify the above-mentioned issues in corpus-based MT, we attempted to translate aviation manuals as a real-world example.

We used two different types of manuals, type 1 and type 2. There were five manuals for each of five series of aircraft in type 1 and three manuals in type 2. The total number of sentences in type 1 manuals was about 16,000 and the number in type 2 manuals was about 40,000.

Using these manuals of five different series of aircraft in type 1 and three different series of aircraft in type 2, we conducted translation experiments, both from English to Japanese and Japanese to English, with combinations of manuals and compared the results, in order to evaluate the effect of the amount of training data for this specific domain and these documents.

Japanese sentences were segmented by a Japanese morphological analyzer, ChaSen, and English sentences were tokenized and lowercased.

We used the Moses SMT system (Koehn et al., 2007), which is a phrase-based SMT system, to train and test our dataset. We made a phrase table from the training data. Usually, the lengths of phrases are limited to 7, for example. However, we included all phrases regardless of their length (except Experiment 6 in Figure 2), because sentences in a manual are often reused in other manuals. By retaining long phrases, we can successfully translate such recycled sentences.

In the experiments described in this paper, we translated Japanese sentences into English sentences and vice versa. We used 20% of the sentences in the test data for tuning parameters and translated the remaining 80% of sentences.

The BLEU score was rather high given the small size of the training data, so SMT appears to be very promising for this task. We also reviewed the translations and concluded that the outputs of the SMT system were relatively good. Some example translations and their references are shown in Section 6.

Our experiments on translating manuals are still at an early stage, yet the BLEU scores from a very small number of training sentences are already high. We believe that corpus-based machine translation (in this study, SMT) offers excellent potential in this domain.

The experiments we conducted are listed in Figure 1.

We firstly tried to translate manuals with a small amount of training data (Experiment 1). With only 7,000 sentences for training, the case-insensitive BLEU scores (Papineni et al., 2002) for the test data were about 50.

Next, to see the effect of adding sentences from different types of manuals, we conducted Experiments 2, 3, 4 and 5. Comparing Experiments 2 and

3, the BLEU score for translating a type 2 manual was not improved by adding sentences from type 1 manuals. Comparing Experiments 4 and 5, the BLEU score for translating a type 1 manual was decreased by adding sentences from type 2 manuals. Thus, we should not add sentences from different types of manuals to the training data. When the types of manuals are different, even manuals within a domain differ widely, so learning with different types of manuals may cause overfitting.

The effect of the amount of training data can be seen by comparing Experiments 1 and 2. Adding sentences from the same type of manual to training data increases the BLEU score.

We conducted Experiment 6 in order to grasp the effect of the lengths of phrases in a phrase table made from training data. Compared with Experiment 5, Experiment 6 yields a better BLEU score by retaining long phrases.

Generally speaking, corpus-based MT requires huge amounts of text data, e.g. 1 million parallel sentences. However, the numbers of sentences used in our experiments were rather small, i.e. 7,000 (Experiment 1), 43,000 (Experiment 2), 27,000 (Experiment 3), 53,000 (Experiment 4) and 12,000 (Experiments 5 and 6). The case-insensitive BLEU scores ranged from 45.9 to 60.3. These scores are very high given the small size of the training data, and so SMT is very promising for this task.

6 Examination of Actual Translation

We also reviewed the translations and concluded that the outputs of the SMT system were relatively good. In this section, we will discuss some of the typical patterns with actual examples of Japanese-English translations (Figure 2).

The first line of each example (line with "Input") in Figure 2 is input sentences for the MT system. The second line (also with "Input") shows the same input sentence with word number. The third line (with "Output") is the output from the MT system with the corresponding word number. The fourth line (with "Reference") is the correct (or reference) translation. In this case, this is the sentences in aviation manuals that have been translated manually. The sentences in Figure 2 have been modified to conceal the contents of the aviation manuals.

(1) Experiment 1
We used two type 1 manuals as training data and used a third type 1 manual as test data. The BLEU scores in this experiment were 47.8 (Japanese to English) and 50.1 (English to Japanese).

(2) Experiment 2
We used five type 1 manuals and two type 2 manuals as training data and used the remaining type 2 manual as test data. The BLEU scores in this experiment were 45.9 (Japanese to English) and 47.2 (English to Japanese).

(3) Experiment 3
We used two type 2 manuals as training data and used the remaining type 2 manual as test data. The BLEU score in this experiment was 47.0 (both Japanese to English and English to Japanese).

(4) Experiment 4
We used four type 1 manuals and three type 2 manuals as training data and used the remaining type 1 manual as test data. The BLEU scores in this experiment were 52.0 (Japanese to English) and 55.9 (English to Japanese).

(5) Experiment 5
We used four type 1 manuals as training data and used the remaining type 1 manual as test data. The BLEU scores in this experiment were 60.3 (Japanese to English).

(6) Experiment 6
We used the same test data and training data as in Experiment 5. However, the length of phrases in the phrase table made from the training data was limited to 7. The BLEU scores in this experiment were 54.0 (both Japanese to English and English to Japanese).

Figure 1: List of Experiments

In Example 1, a set of English words in an input sentence is translated into a set of Japanese words. The order of sets of words is different between input and output. In Example 2, a whole Japanese sentence is translated into one English sentence. This shows the merit of retaining long phrases. In Example 3, the parallel structure, which is hard for rule-based MT systems, is translated properly. In Example 4, the sentence is grammatically incorrect, but we can use "the bulk cargo compartment" for post-editing. This shows the possibility of using the output of SMT for post-editing. Example 5 shows the disagreement of singular and plural expressions. Because the current system does not consider grammatical properness, it is hard to handle this kind of phenomena.

7 Conclusion

Corpus-based machine translation, such as EBMT and SMT, looks very promising in this area. Even with a small training corpus and without additional development of the system, we can generate proper output. Using similar manuals of five different series of aircraft, we conducted translation experiments with combinations of manuals and compared the results, in order to evaluate the effect of the amount of training data for this specific domain and these documents.

Corpus-based MT is attractive because we simply need to prepare parallel text as examples. Especially, SMT can be used even with a small amount of training data, provided it is sufficiently repetitive.

We intend to apply the method described in this paper to a new plane for which there is no translation yet. We will try English-to-Japanese translation for this task. To apply this method in the real world, we need to develop post-editing tools that allows the post-editor to easily copy-and-paste usable parts of the output of machine translation systems. This might be different from the existing tools for translation aids such as translation memories. We may be able to use the confidence ratio which the MT system outputs in order to point out to the post-editors some doubtful translations among the MT output. We could also try post- and/or pre-processing which exchanges some specific expressions which should be translated differently.

As for evaluation, we will consider a method of calculating the edit distance automatically.

References

Fuji, M. et al. (2001). Evaluation Method for Determining Groups of Users Who Find MT Useful. In *Proceedings of the Machine Translation Summit VIII*.

Isahara, H. et al. (2007). Development of a Japanese-Chinese Machine Translation System. In *Proceedings of the Machine Translation Summit XI*.

Utiyama, M., Isahara, H. (2003). Reliable Measures for Aligning Japanese-English News Articles and Sentences. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL).*

Utiyama, M., Isahara, H. (2007). A Japanese-English Patent Parallel Corpus. In *Proceedings of the Machine Translation Summit XI.*

Koehn, P. et al. (2007). Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session.*

Papineni, K. et al. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL).*

==
Example 1: Phrase-to-phrase translation

Input: on the ground with both fuel pump switches off , the first fuel pump switch positioned on arms the fuel pump in the takeoff go-around (to/ ga) on mode .
Input: on(0) the(1) ground(2) with(3) both(4) fuel(5) pump(6) switches(7) off(8) ,(9) the(10) first(11) fuel(12) pump(13) switch(14) positioned(15) on(16) arms(17) the(18) fuel(19) pump(20) in(21) the(22) takeoff(23) go-around(24) ((25) to(26) /(27) ga(28))(29) on(30) mode(31) .(32)
Output: 地上で両方の|0-4| Fuel Pump |5-6| Switch が OFF の場合，|7-10|最初の Fuel Pump Switch |11-14|を|17-17| ON 位置|15-15|にする|16-16|と，Fuel Pump は Takeoff|18-23|Go-around|24-24|(TO/ GA)ON|25-30| Mode にして飛行する ． |31-32|
Reference: 地上で両方の Fuel Pump Switch が OFF の場合，最初の Fuel Pump Switch を ON 位置にすると，Fuel Pump の On Mode は Takeoff Go - Around (TO / GA) Mode に Arm される ．
==
Example 2: Whole input sentence is translated in to an English sentence

Input: 本Boxは，必要によりエンジン部品，工具及び Oilを運ぶ為のBoxである.
Input: 本(0) Box (1)は(2) , (3)必要(4)により(5) エンジン(6) 部品(7) , (8) 工具(9) 及び(10) Oil(11) を(12) 運ぶ(13) 為(14) の(15) Box (16)で(17)ある(18) ． (19)
Output: the purpose of this box is to carry engine parts , tools and oil as required . |0-19|
Reference: The purpose of this box is to carry engine parts , tools and oil as required .
==
Example 3: Translation of parallel structure
Input: この場合，NAV Display Message INVALID ROUTE UPLINK LまたはINVALID ROUTE UPLINK Rが表示される ．
Input: この(0)場合(1) , (2) NAV (3) Display (4) Message（5）INVALID (6)ROUTE（7）UPLINK（8）L（9）または（10）INVALID（11）ROUTE（12）UPLINK（13）R（14）が（15）表示（16）さ（17）れる（18）．（19）
Output: in this case , |0-2| NAV Display Message |3-5| INVALID |6-6| ROUTE |7-7| UPLINK |8-8| l or |9-10| INVALID |11-11| ROUTE |12-12| UPLINK |13-13| r |14-14| will be displayed . |15-19|
Reference: In this case , NAV Display Message INVALID ROUTE UPLINK L or R will be displayed .
==
Example 4: Grammatically incorrect translation

Input: 気圧管理が必要な貨物をBulk Cargo Compartmentに搭載してはならない.
Input: 気圧(0)管理(1)が(2)必要(3)な(4)貨物(5)を(6) Bulk (7) Cargo (8) Compartment (9)に(10)搭載(11)し(12)て(13)は(14)なら(15)ない(16) . (17)
Output: inform |1-1| the bulk cargo |7-8| Compartment |9-9| must not be used |12-16| to |6-6| pressure |0-0| is required |2-4| to |10-11| freighter |5-5| . |17-17|
Reference: Pressure sensitive cargo should not be carried in the bulk cargo compartment .
==
Example 5: Disagreement of singular and plural expression

Input: c)両方のFuel Valveが正常に作動すること.
Input: c(0))(1)両方(2)の(3) Fuel Valve (5)が(6) 正常(7)に(8)作動(9)する(10)こと(11) . (12)
Output: c) both |0-3| fuel valve |4-5| operates normally , and |6-12|
Reference: c) Both fuel valves operate normally ,

Figure 2: Example of Translation