Study of Practical Effectiveness for Machine Translation Using Recursive Chain-link-type Learning

Hiroshi Echizen-ya
Dept. of Electronics and Information
Hokkai-Gakuen University
S 26-Jo, W 11-Chome, Chuo-ku
Sapporo, 064-0926 Japan
echi@eli.hokkai-s-u.ac.jp

Kenji Araki
Division of Electronics and Information
Hokkaido University
N 13-Jo, W 8-Chome, Kita-ku
Sapporo, 060-8628 Japan
araki@media.eng.hokudai.ac.jp

Yoshio Momouchi
Dept. of Electronics and Information
Hokkai-Gakuen University
S 26-Jo, W 11-Chome, Chuo-ku
Sapporo, 064-0926 Japan
momouchi@eli.hokkai-s-u.ac.jp

Koji Tochinai
Division of Business Administration
Hokkai-Gakuen University
4-Chome, Asahi-machi, Toyohira-ku
Sapporo, 060-8790 Japan
tochinai@econ.hokkai-s-u.ac.jp

Abstract
A number of machine translation systems based on the learning algorithms are presented. These methods acquire translation rules from pairs of similar sentences in a bilingual text corpora. This means that it is difficult for the systems to acquire the translation rules from sparse data. As a result, these methods require large amounts of training data in order to acquire high-quality translation rules. To overcome this problem, we propose a method of machine translation using a Recursive Chain-link-type Learning. In our new method, the system can acquire many new high-quality translation rules from sparse translation examples based on already acquired translation rules. Therefore, acquisition of new translation rules results in the generation of more new translation rules. Such a process of acquisition of translation rules is like a linked chain. From the results of evaluation experiments, we confirmed the effectiveness of Recursive Chain-link-type Learning.

1 Introduction
Rule-Based Machine Translation (MT) (Hutchins and Somers, 1992) requires large-scale knowledge to analyze both source language (SL) sentences and target language (TL) sentences. Moreover, it is difficult for a developer to completely describe large-scale knowledge that can analyze various linguistic phenomena. Therefore, Rule-Based MT is time-consuming and expensive. Statistical MT and Example-Based MT have been proposed to overcome the difficulties of Rule-Based MT. These approaches correspond to Corpus-Based approach. Corpus-Based approach uses translation examples that keep including linguistic knowledge. This means that the system can improve the quality of its translation only by adding new translation examples. However, in Statistical MT (Brown et al., 1990), large amounts of translation examples are required in order to obtain high-quality translation. Moreover, Example-Based MT (Sato and Nagao, 1990; Watanabe and Takeda, 1998; Brown, 2001; Carl, 2001) which relies on various knowledge resources results in the same difficulties as Rule-Based MT. Therefore, Example-Based MT, which automatically acquires the translation rules from only bilingual text corpora, is very effective. However, existing Example-Based MT systems using the learning algorithms require large amounts of translation pairs to acquire high-quality translation rules.

In Example-Based MT based on analogical reasoning (Malavazos, 2000; Guvenir, 1998), the different parts are replaced by variables to generalize translation examples as shown in (1) of Figure 1. However, the number of different parts of the two SL sentences must be same as the number of different parts of the two TL sentences. This means that the condition of acquisition of translation rules is very strict because this method allows only $n:n$ mappings in the number of the different parts between the
SL sentences and the TL sentences. As a result, many translation rules cannot be acquired. (McTait, 2001) generalizes both the different parts and the common parts as shown in Figure 1(2). This means that (McTait, 2001) allows m:n mappings in the number of the different parts, or the number of the common parts. However, it is difficult to acquire the translation rules that correspond to the lexicon level. On the other hand, we have proposed a method of Machine Translation using Inductive Learning with Genetic Algorithms(GA-ILMT)(Echizenya et al., 1996). This method automatically generates the similar translation examples from only given translation examples by applying genetic algorithms(Goldberg, 1989) as shown in (3a) of Figure 1. Moreover, the system performs Inductive Learning. By using Inductive Learning, the abstract translation rules are acquired by performing phased extraction of different parts as shown in Figure 1(3b) and (3c). In all methods shown in Figure 1, the condition of acquisition of translation rules is that two similar translation examples must exist. As a result, the systems require large amounts of translation examples.

We propose a method of MT using Recursive Chain-link-type Learning as a method to overcome the above problem. In our method, the system acquires new translation rules from sparse data using other already acquired translation rules. For example, first, translation rule B is acquired by using translation rule A when the translation rule A exists in the dictionary. Moreover, translation rule C is acquired by using the translation rule B. Such a process of acquisition of translation rules is like a chain where each ring is linked. Therefore, we call this mechanism Recursive Chain-link-type Learning(RCL). This method can effectively acquire many translation rules from sparse data without depending on the different parts of similar translation pairs. In this paper, we describe the effectiveness of RCL through evaluation experiments.

2 Basic Idea

RCL is a method with an ability that automatically acquires translation knowledge in a computer without any analytical knowledge, such as GA-ILMT. This is the ability to extract corresponding parts from pairs of objects with which it corresponds. In this paper, we apply this ability to a translation example that consists of TL and SL sentences. A system with RCL can acquire translation rules from sparse translation examples. Figure 2 shows how translation rules are acquired using this method1.

Figure 2 shows the process where translation rules B, C and D are acquired one after another using RCL. In this paper, source parts are those parts that are extracted from the SL sentences of translation examples, and target parts are those parts that are extracted from the TL sentences of translation examples. Moreover, part translation rules are pairs of source parts and

---

1In Figure 2, the use of a Greek character means that all language characters correspond to unknown character strings for a computer.
3 Outline

Figure 2: Schema in process of acquisition of translation rules using RCL.

Figure 3: Process flow.

The GA-ILMT and RCL system with RCL acquires part translation rules using RCL.

A system with RCL acquires part translation rules using RCL.

Therefore, the system that acquires the part translation rules, and can extract some parts of source and target parts, can excite some parts of translation examples of the sentence. For example in Figure 3 as a bootstrapping system, and we implemented a new system based on the learning process, the translation rules acquired by RCL, and the translation rules acquired by GA-ILMT are used in Inductive Learning of GA-ILMT. In this paper, we implemented a new system based on the learning process, the translation rules acquired by RCL, and the translation rules acquired by GA-ILMT. In the process of translation examples, or the source information that the system can extract from the source sentence or the target parts. By using this in the right of "δ" in the SL sentence, the right of "θ" in the TL sentence. The acquired translation rule B has rules that are used as starting points in the acquisition process, the translation rules acquired by RCL, and the translation rules acquired by GA-ILMT. In the process of translation examples, or the source information that the system can extract from the source sentence or the target parts. By using this in the right of "δ" in the SL sentence, the right of "θ" in the TL sentence. The acquired translation rule B has

Moreover, in process 2, translation rule D is acquired by using the translation rules acquired in process 1 as a starting point in the SL sentence. In the process of translation examples, or the source information that the system can extract from the source sentence or the target parts. By using this in the right of "δ" in the SL sentence, the right of "θ" in the TL sentence. The acquired translation rule B has

As a result, a chain reaction causes the acquisition process, the translation rules acquired by RCL, and the translation rules acquired by GA-ILMT. In the process of translation examples, or the source information that the system can extract from the source sentence or the target parts. By using this in the right of "δ" in the SL sentence, the right of "θ" in the TL sentence. The acquired translation rule B has
4 Process

4.1 Translation process
In the translation process, the system generates translation results using acquired translation rules. First, the system selects the sentence translation rules that can be applied to the SL sentence. Second, the system generates the translation results by replacing the variables in the sentence translation rules with the part translation rules.

4.2 Feedback process
In the feedback process, the system evaluates the translation rules used. First, the system evaluates the translation rules without variables by using the processes of combinations between the translation rules with variables and the translation rules without variables (Echizen-ya et al., 1996). Next, the system evaluates translation rules with variables by using the processes of combinations between the translations rules with variables and the translation rules without variables (Echizen-ya et al., 2000). As a result, the system increases the correct translation frequencies, or the erroneous translation frequencies, of the translation rules by using these evaluation methods for the translation rules.

4.3 Learning process

4.3.1 GA-ILMT
In this paper, by using the process of acquisition of translation rules in GA-ILMT, the system acquires both sentence and part translation rules. These rules are then used as starting points when the system performs RCL.

4.3.2 Recursive Chain-link-type Learning (RCL)
In this section, we describe the process of acquisition of translation rules using RCL. The details of the process of acquisition of part translation rules are as follows.

(1) The system selects translation examples that have common parts with the sentence translation rules.

(2) The system extracts the parts that correspond to the variables in the source parts and in the target parts of the sentence translation rules from the SL sentences, and the TL sentences of the translation examples.

(3) The system registers pairs of the parts extracted from the SL sentences and the parts extracted from the TL sentences, as the part translation rules.

(4) The system gives the correct and erroneous frequencies of sentence translation rules to the acquired part translation rules.

Figure 4\(^2\) shows an example of the acquisition of a part translation rule using the sentence translation rule. In Figure 4, (thirty; 30\([sanju]\)) as the part translation rule is acquired because “thirty” corresponds to the variable in the source part of sentence translation rule and “30\([sanju]\)” corresponds to the variable in the target part of sentence translation rule.

Figure 4: Example of the acquisition of a part translation rule using the sentence translation rule.

The details of the process of acquisition of sentence translation rules are as follows:

(1) The system selects the part translation rules in which the source parts are included in the SL sentences of the translation example or in the source parts of sentence translation rules, and in which the target parts are included in the TL sentences of the translation examples or in the target parts of sentence translation rules.

(2) The system acquires new sentence translation rules by replacing the parts which are same as the part translation rules with the variables to the translation examples or the sentence translation rules.

(3) The system gives the correct and erroneous frequencies of the part translation rules to the acquired sentence translation rules.

\(^2\) Italic are the pronunciation in Japanese.
Figure 5 shows examples of the acquisition of the sentence translation rules using the part translation rules. In Figure 5, the system acquires (It starts in @0 minutes. ; それ/は/@0/分/たて/ば/始まります. [Sore wa @0 pun tate ba hajimari masu,] ) as a sentence translation rule by using the part translation rule (thirty;30[sanjū]) acquired in Figure 4, and ( @1 starts in @0 minutes. ; @1/は/@0/分/たて/ば/始まります. [ @1 wa @0 pun tate ba hajimari masu,] ) as the sentence translation rule, that is more abstracted, is acquired by using the part translation rule (it; それ [sore]).

The correct translation results are grouped into two categories:

1. The correct translation
   This means that the translation results correspond to the correct translation results taken from textbooks respectively (Bunri, 2001; Sinko Shuppan, 2001).

2. A correct translation which includes unknown words
   This means that the translation results with substituted nouns or adjectives as variables correspond to the correct translation results taken from textbooks respectively (Bunri, 2001; Sinko Shuppan, 2001).

In this paper, the effective translation results are the translation results that correspond to (1) and (2), and the ineffective translation results are the translation results that do not correspond to (1) and (2). Moreover, the effective translation rate is the rate of the effective translation results in all the evaluation data. The translation results are ranked when several translation results are generated. The translation results using the translation rules whose rate of correct translation frequency is high, are ranked at the top. We evaluated the translation results that are ranked from No.1 to No.3.
Table 1: Examples of effective translation results.

| SL sentences                                                                 | TL sentences                                                                 |
|-------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| This bag was made in France.                                                  | このバックはフランス製です。 [Kono baggu wa furansu sei desu.]                   |
| We went there to play baseball.                                               | わたしたちは野球をするためにそこへ行きました。 [Watashi tachi wa yakyu wo suru tame soko e iki mashi ta.] |

Examples of the correct translation results which includes the unknown words

| SL sentences                                                                 | TL sentences                                                                 |
|-------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| Shall I take you to the amusement park?                                      | @0 へ連れていてあげましょうか？ [@0 e tsure te itte age masho ka?]              |
| How far is it from Kyoto to Hiroshima?                                       | @0 から広島までどのくらいの距離がありますか？ [@0 kara hiroshima made dono kurai no kyori ga ari masu ka?] |

Table 2: Results of comparative experiments.

| System     | Effective translation rates | Details (1) | Details (2) |
|------------|----------------------------|-------------|-------------|
| Our system | 85.0%                      | 41.6%       | 58.4%       |
| system A   | 85.8%                      | 84.0%       | 16.0%       |
| system B   | 81.7%                      | 83.7%       | 16.3%       |
| system C   | 76.9%                      | 82.7%       | 17.3%       |

Table 3: Comparison of effective translation rates based on quality.

| System     | Effective translation rates | Details (1) | Details (2) |
|------------|----------------------------|-------------|-------------|
| Our system | 73.7%                      | 7.5%        | 52.5%       |
| system A   | 70.3%                      | 84.2%       | 15.8%       |
| system B   | 63.8%                      | 85.0%       | 15.0%       |
| system C   | 58.7%                      | 82.8%       | 17.2%       |

Moreover, we evaluated translation results more strictly in terms of the quality of translation. Meaning that only translation results that had almost the same character strings as the correct translation results taken from the textbooks (Bunri, 2001; Sinko Shuppan, 2001) were effective translation results. For example, “それは約 10 分かかります [Sore wa yaku juppun kakari masu.]” is an ineffective translation result because of the correct translation results for “It takes about ten minutes.” is “約 10 分かかります [Yaku juppun kakari masu.]” in textbook (Bunri, 2001; Sinko Shuppan, 2001). In this Japanese sentence, phrase “それは [sore wa]” results in needlessly long. Therefore, we evaluate the translation results that have different phrases to the correct translation results as the ineffective translation results in terms of the quality of translation. Table 3 shows a comparison of effective translation rates based on quality. In Table 3, we confirmed that the system with RCL can generate more high-quality translation results than the three other Rule-Based MT systems.

In the system with RCL, the erroneous translation rules are also acquired like a linked chain. For example, in Figure 2, the translation rules B, C and D are acquired as the erroneous translation rules when the translation rule A is the erroneous translation rule. Namely, a chain reaction causes the acquisition of erroneous translation rules. In learning data, the rate of erroneous part translation rules to the acquired part translation rules was 47.9%, and the rate of erroneous sentence translation rules to the acquired sentence translation rules was 38.2%. However, such erroneous translation rules are automatically decided as being erroneous translation rules in the feedback process resulting from the ineffective translation results.
6 Conclusion

In existing Example-Based MT systems based on learning algorithms, similar translation pairs must exist to acquire high-quality translation rules. This means that the systems require large amounts of translation examples to acquire high-quality translation rules. On the other hand, a system with RCL can acquire many new translation rules from sparse translation examples because it uses other already acquired translation rules based on the learning algorithms described in section 2. As a result, the quality of the translation and the effective translation rate of our system is higher than other Rule-Based MT systems. However, our system still does not reach the level of a practical MT system and requires more translation rules to realize the goal of a practical MT system. Although our system is not a practical enough MT system, the system can effectively acquire the translation rules from sparse data by using RCL. Therefore, we consider that the quality of translation improves only by adding new translation examples without the difficulty of Rule-Based MT systems in which a developer must completely describe large-scale knowledge.

In the future, we plan to add a mechanism that effectively combines the acquired translation rules so that the system realizes the translation of practical SL sentences.

7 Acknowledgements

This work was partially supported by the Grants from the High-Tech Research Center of Hokkai-Gakuen University and a Government subsidy for aiding scientific research (No.14658097) of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Hutchins, W. J and H. L. Somers. 1992. An Introduction to Machine Translation. ACADEMIC PRESS.

Sato, S and M. Nagao. 1990. Toward Memory-based Translation. In proceedings of the Coling’90.

Brown, P., J. Cockett, S. Della Pietra, V. J. Della Pietra, F. Jelinek, J. D. Lafferty, R. L. Mercer and P. S. Roossin. 1990. A Statistical Approach to Machine Translation. Computational Linguistics Vol.16, No.2.

Watanabe, H, and K. Takeda. 1998. A Pattern-based Machine Translation System Extended by Example-based Processing. In proceedings of the Coling-ACL’98.

Brown, R.D. 2001. Transfer-Rule Induction for Example-Based Translation. In proceedings of the Workshop on EBMT, MT Summit VIII.

Carl, M. 2001. Inducing Translation Grammars from Bracketed Alignments. In proceedings of the Workshop on EBMT, MT Summit VIII.

Malavazos, C and S. Piperidis. 2000. Application of Analagical Modelling to Example Based Machine Translation. In proceedings of the Coling2000.

Güvenir, H.A and I. Cicekli. 1998. Learning Translation Templates from Examples. Information Systems Vol.23, No.6.

McTait, K. 2001. Linguistic Knowledge and Complexity in an EBMT System Based on Translation Patterns. In proceedings of the Workshop on EBMT, MT Summit VIII.

Echizen-ya, H., K. Araki, Y. Momouchi and K. Tochinai. 1996. Machine Translation Method using Inductive Learning with Genetic Algorithms. In proceedings of the Coling’96.

Goldberg, D. E. 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley.

Araki, K., Y. Momouchi and K. Tochinai. 1995. Evaluation for Adaptability of Kana-kanji Translation of Non-segmented Japanese Kana Sentences using Inductive Learning. In proceedings of the PACLING’95.

Echizen-ya, H., K. Araki, Y. Momouchi and K. Tochinai. 2000 Effectiveness of Layering Translation Rules based on Transition Networks in Machine Translation using Inductive Learning with Genetic Algorithms. In proceedings of the MT and Multilingual Applications in the New Millennium.

Nihon-Kyozai(1). 2001. One World English Course 1 new edition. Tokyo.

Nihon-Kyozai(2). 2001. One World English Course 2 new edition. Tokyo.

Hoyu Shuppan. 2001. System English Course 2 new edition. Tokyo.

Bunri. 2001. Work English Course 2 new edition. Tokyo.

Sinko Shuppan 2001. Training English Course 2 new edition. Osaka.