Effectiveness of Layering Translation Rules based on Transition Networks in Machine Translation using Inductive Learning with Genetic Algorithms

Hiroshi Echizen-ya†, Kenji Araki††, Yoshio Momouchi† and Koji Tochinai††

†Dept. of Electronics and Information, Hokkai-Gakuen University
††Division of Electronics and Information, Hokkaido University

echi@eli.hokkai-s-u.ac.jp   araki@media.eng.hokudai.ac.jp
momouchi@eli.hokkai-s-u.ac.jp tochinai@media.eng.hokudai.ac.jp

Abstract

In a previous study, we proposed a method of machine translation using inductive learning with genetic algorithms (GA-ILMT) based on learning capability. In GA-ILMT, all translation rules are acquired from given translation examples only, without using analytical knowledge, and the system uses these rules for translation. In the present study, we propose a method which performs translation by recognizing the structure of whole sentences without the need for language-dependent knowledge. In this new method, the system layers multiple translation rules to formulate a translation rule that represents the basic structure of the whole sentence by automatically producing a translation transition network (TTN) constructed from multiple translation rules.

1 Introduction

The development of machine translation (MT) systems has increased rapidly in recent years. However, the current level of MT systems is still inadequate for practical use. Mainstream MT systems are rule-based (Hutchins and Somers, 1992). In rule-based MT, a grammar writer is required to provide a limited rule set to initialize the system. The obvious problem with this approach is that it is difficult to perfectly describe rules that can deal with all linguistic phenomena. To solve this problem, Corpus-based or example-based MT (Sato and Nagao, 1990) and statistics-based MT (Brown et al., 1993) have been proposed. In these methods, users are only required to supply a collection of source and target sentences pairs (translation examples). However, in order to realize practical MT systems, these methods require the input of many translation examples, including examples that contain analytical knowledge. Pattern-based MT (Watanabe and Takeda, 1998) incorporates both rule-based MT and example-based MT. However, this method is problematic in that context-free grammar (CFG) for analytical knowledge must be provided by a grammar writer as in rule-based MT.

A computer system based on learning capability is effective as a solution to these problems because it does not require initial analytical knowledge. In a

5-1

system based on learning capability, the rules for translation are acquired automatically from given translation examples. Such a system does not require a large number of examples because the system generates general translation rules from the examples. However, in order to realize such a MT system, the system itself must have high learning capability.

We previously proposed machine translation using inductive learning with genetic algorithms (Echizen-ya et al., 1996), called GA-ILMT. In the present study, we propose a method in which the system translates by recognizing the structure of whole sentences without the need for language-dependent knowledge. In order to formulate a translation rule that represents the basic structure of the whole sentence, the system layers multiple translation rules by automatically producing a translation transition network (TTN), constructed from multiple translation rules. As a result, the system can perform translation using translation rules more effectively.

2   GA-ILMT

2.1   Outline of GA-ILMT

In GA-ILMT, translation rules are acquired from translation examples alone by inductive learning, and many translation examples are automatically produced from only a small number of translation examples by applying genetic algorithms. Thus, a GA-ILMT-based system translates based on learning capability alone.
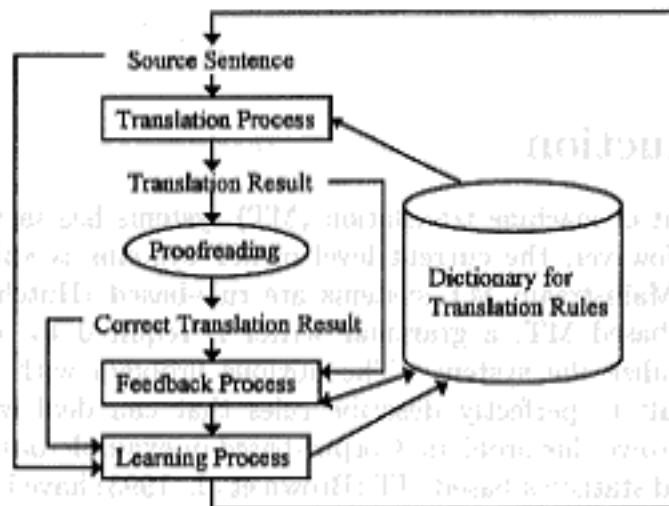

Figure 1: Outline of GA-ILMT

An outline of GA-ILMT is shown in figure 1. In the case of English-to-Japanese translation, users first input a source sentence in English. In the translation process, the system produces translation results using translation rules acquired in the learning process. The users then proofread the translated sentences to check for errors. In the feedback process, the system determines the fitness value for the translation rules used in the translation process and performs a selection

process for erroneous translation rules. In the learning process, new translation examples are automatically produced by crossover and mutation. In addition, various translation rules are acquired from the translation examples by inductive learning.

2.2 Acquisition of Translation Rules by Inductive Learning

In the learning process, translation rules are acquired by inductive learning. In GA-ILMT, inductive learning is the process in which the system automatically acquires general translation rules that are present in the translation examples. The system first extracts the parts that differ between two similar translation examples (see Fig. 2). Parts are extracted when there is only one part that differs between the translation examples. In case there are several parts that differ between two translation examples, the system extracts the parts that differ between two translation examples if it can decide only one part that differs by using translation rules acquired. In this way, the system can determine the correspondence between English and Japanese. The system further extracts common parts by replacing the previously extracted parts with variables. An example of the acquisition of translation rules is shown in figure 2.
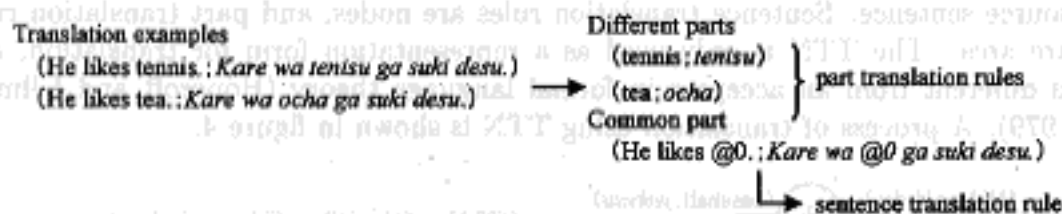
Translation examples
(He likes tennis. ; *Kare wa tenisu ga suki desu.*)
(He likes tea. ; *Kare wa ocha ga suki desu.*)
$\longrightarrow$

Different parts
(tennis ; *tenisu*)
(tea ; *ocha*) } part translation rules

Common part
(He likes @0. ; *Kare wa @0 ga suki desu.*)
$\longrightarrow$ sentence translation rule

Figure 2: Example of acquisition of translation rules

In figure 2, the different parts are (tennis; *tenisu*) and (tea; *ocha*), in the form (English; Japanese). The common part is (He likes @0.; *Kare wa @0 ga suki desu.*). The common part and differing parts are used as the translation rules. There are two kinds of translation rules; those for sentences, called sentence translation rules, and those for parts of sentences, called part translation rules. The system also performs phased extraction of the differing and common parts from the character strings of translation rules. As a result, as translation rules are generated, more are acquired. Examples of general sentence translation rules are shown in figure 3.

(He likes @0. ; *Kare wa @0 ga suki desu.*)
(She likes @0. ; *Kanojyo wa @0 ga suki desu.*)
(@0 likes tennis. ; *@0 wa tenisu ga suki desu.*)
(@0 likes tea. ; *@0 wa ocha ga suki desu.*)
(@0 likes @1. ; *@0 wa @1 ga suki desu.*)

Figure 3: Examples of general sentences translation rules

In building block translation memory (Langè et al., 1997), skeleton sentences are used for translation. These skeleton sentences are acquired based on machine-aided human translation. However, in GA-ILMT, the system acquires translation rules by extracting the differing and common parts from the character strings of translation examples. This means that GA-ILMT system acquires the translation rules based on learning capability.

3 Basic Concept

In the present study, we propose a method which performs translation by recognizing the structure of whole sentences. This is one of the important translation processes that humans possess. Humans first recognize a general translation rule that represents the basic structure of the whole sentence. The GA-ILMT system produces a translation by combining part translation rules with this general translation rule. The system also formulates a translation rule that represents the basic structure of the whole sentence by layering multiple translation rules. In order to layer translation rules, the system uses a translation transition network (TTN), which represents the process of translation. In a TTN, the translation rules are layered from the general translation rules representing the basic structure of the whole sentence, to concrete translation rules that are similar to the source sentence. Sentence translation rules are nodes, and part translation rules are arcs. The TTN is only used as a representation form for translation, and is different from an acceptor in formal language theory (Hopcroft and Ullman, 1979). A process of translation using TTN is shown in figure 4.

q_0: (@0 likes @1. ; @0 wa @1 ga suki desu.)
q_1: (Hideo likes @0. ; Hideo wa @0 ga suki desu.)
q_2: (@0 likes baseball. ; @0 wa yakyuu ga suki desu.)
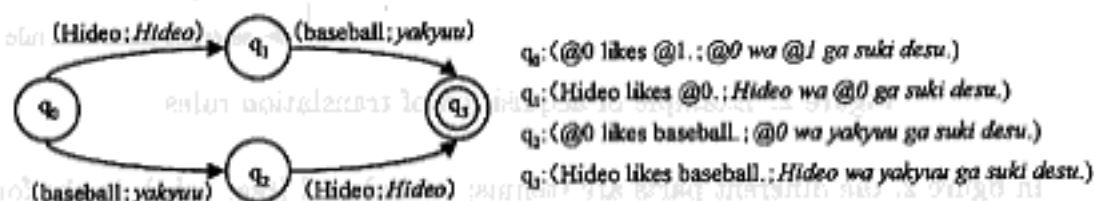q_3: (Hideo likes baseball. ; Hideo wa yakyuu ga suki desu.)

Figure 4: TTN in translation examples "(Hideo likes baseball.; *Hideo wa yakyuu ga suki desu.*)"

In figure 4, node q_0 is the translation rule representing the basic structure of the whole sentence. In other systems, a state transition network is used to control the application of grammatical rules, such as in Mu-machine translation (Nakamura et al., 1984). However, such control is based on language-dependent knowledge given to the system by a grammar writer. In our proposed method, the system controls translation rules by producing a TTN automatically. In this study, we provide GA-ILMT system with the capability to recognize the structure of the whole sentence without the need for language-dependent knowledge.

4 GA-ILMT based on TTN

The system generates a TTN for a correct translation result in the learning rocess. The TTN of a correct translation result is shown in figure 5.
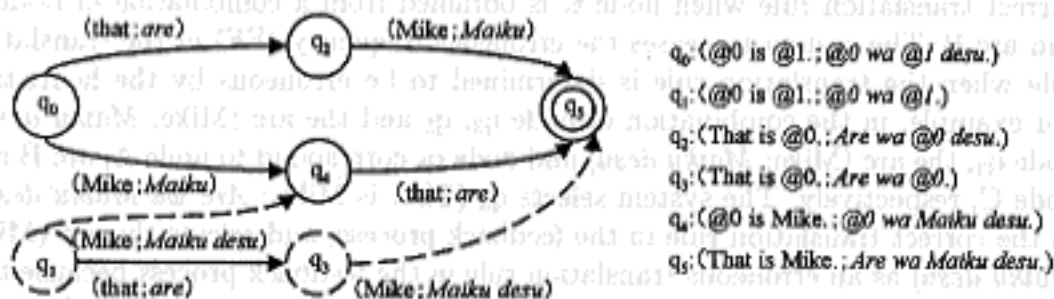

q_0: (@0 is @1.; @0 wa @1 desu.)
q_1: (@0 is @1.; @0 wa @1.)
q_2: (That is @0.; Are wa @0 desu.)
q_3: (That is @0.; Are wa @0.)
q_4: (@0 is Mike.; @0 wa Maiku desu.)
q_5: (That is Mike.; Are wa Maiku desu.)

Figure 5: TTN of correct translation result "(That is Mike.;*Are wa Maiku desu*.)"

A correct translation rule is a translation rule for which the English and Japanese corresponds perfectly, whereas an erroneous translation rule is a translation rule for which the English and Japanese does not correspond. The system memorizes the combinations of translation rules by using the TTN of a correct translation result which are decided in the feedback process. For example, in the TTN in figure 5, node q_5 is a combination of node q_4 and the arc (that; *are*), and node q_4 is a combination of node q_0 and the arc (Mike; *Maiku*). The system compares each node with the correct translation result. For example, in node q_4, the character strings other than the variables (is Mike; *wa Maiku desu*) are perfectly matched in the correct translation result (That is Mike.; *Are wa Maiku desu*.). Therefore, the system chooses node 4 as one of nodes in the TTN. Nodes q_0, q_2, q_4 and q_5 are correct translation rules. However, erroneous translation rules are included in the TTN in figure 5 even though this TTN is the TTN of correct translation result. Nodes q_1 and q_3, and the arc (Mike; *Maiku desu*), shown in dotted lines, are erroneous translation rules. For example, *Maiku desu* in the arc (Mike; *Maiku desu*) means "Mike is" in English. This indicates that there exists a case in which a correct translation rule is obtained from a combination of erroneous translation rules. However, the system can not recognize whether these translation rules are correct translation rules or not because GA-ILMT system are not given language-dependent knowledge.

The system distinguishes between correct and erroneous translation rules in the TTN for the correct translation result from several heuristics focused on the combination of translation rules. These heuristics are based on preliminary experiments and are provided beforehand. One of the heuristics is that node A is a correct translation rule if arc B and node C are correct translation rules when node C is obtained from a combination of node A and arc B. The system increases the correct frequency (CF) of the translation rule when the translation rule is determined as a correct translation rule from the heuristics. For example, in the combination of node q_4, q_5 and the arc (that; are), node q_4, arc (that; are) and node q_5 correspond to node A, arc B and node C, respectively. The system selects q_5 (That is Mike.; *Are wa Maiku desu*.) and the arc (that; *are*) as correct translation rules in the feedback process because these translation rules

are often used to produce correct translation results. Therefore, the selects q_4 (@0 is Mike.; *@0 wa Maiku desu.*) as the correct translation rule by applying the heuristics, and increases the CF of q_4. Another heuristic is that node A is an erroneous translation rule if arc B is an erroneous translation rule and node C is correct translation rule when node C is obtained from a combination of node A and arc B. The system increases the erroneous frequency (EF) of the translation rule when the translation rule is determined to be erroneous by the heuristics. For example, in the combination of node q_3, q_5 and the arc (Mike; *Maiku desu*), node q_3, the arc (Mike; *Maiku desu*) and node q_5 correspond to node A, arc B and node C, respectively. The system selects q_5 (That is Mike.; *Are wa Maiku desu.*) as the correct translation rule in the feedback process, and selects the arc (Mike; *Maiku desu*) as an erroneous translation rule in the feedback process because this translation rule often produces only erroneous translation results. Therefore, the system selects q_3 (That is @0.; *Are wa @0.*) as an erroneous translation rule by applying the heuristics and increases the EF of q_3.

In the translation process, the system produces a TTN for the source sentence and evaluates each sentence translation rule as follows:

$$Correctness(\%) = \alpha \times CF - \beta \times EF + \gamma \qquad (1)$$

The system decides one transition for which the total of correctness is the highest and uses the final node (sentence translation rule) in the selected transition for translation.

5 Experiments for Performance Evaluation

5.1 Standards of Evaluation

The correct translation results are grouped into two categories:

① A correct translation that does not include an unregistered word

This means that the translation result has the same character string as the proofread translation result by combining correct translation rules.

② A correct translation that includes one unregistered word

This means that the proofread translation result has the same character string as the translation result with nouns or adjectives substituted for the variables.

5.2 Procedure

In the experiment, 1,010 translation examples, taken from a textbook (Hasegawa et al., 1991) for first-grade junior high school students, were used as learning data. The average number of words in English translation examples was 4.6. A further 800 translation examples were taken from another textbook (Ota et al., 1991) for first-grade junior high school students, as evaluation data. The average number of words in English translation examples was 5.4. All of these translation examples

were processed by the method outlined in figure 1. Also, the initial dictionary was empty. These experiments were carried out using GA-ILMT with and without a TTN. In the experiments with TTN, α, β and γ in function (1) are 2.0, 1.0 and 5.0, respectively. These values are based on the preliminary experiments. We only evaluated the translation result ranked as the first translation result in cases where the system produced several translation results.

5.3 Results

The correct translation rates in GA-ILMT with and without TTN are listed in Table 1, where ① and ② correspond to ① and ② in subsection 5.1 and the values in parentheses are number of correct translation results. The correct translation rate was increased from 38.3% to 46.6% by using TTN.

Table 1: Correct translation rates in GA-ILMT with and without TTN

Method	GA-ILMT without TTN	GA-ILMT based on TTN
①	31.4%(251)	39.4%(315)
②	6.9%(55)	7.2%(58)
Total	38.3%(306)	46.6%(373)

6 Discussion and Conclusions

The advantage of using the TTN is that the quality of the translation process in GA-ILMT is improved. This shows that the system can perform translation by recognizing a translation rule that represents the basic structure of the whole sentence. For example, in the translation process, the TTN produced by this system is shown in figure 6.
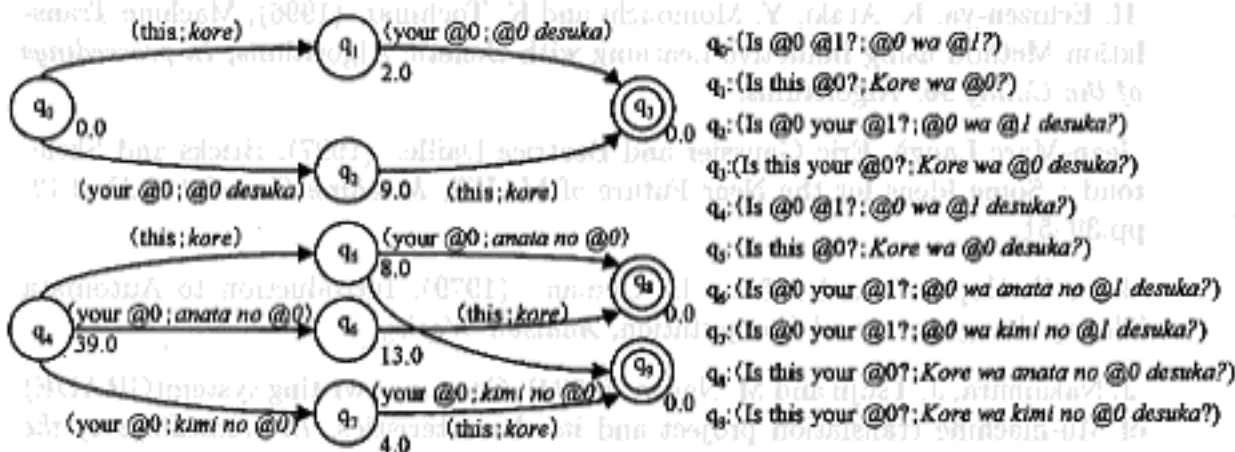

Figure 6: TTN for (Is this your camera?;*Kore wa anata no kamera desuka?*)

In figure 6, the number under each node is the correctness of that node. Three nodes; q_3, q_8 and q_9, are translation rules for the English sentence "Is this your

@0?.". The system randomly selects node q_3 "(Is this your @0?; *Kore wa @0 desu ka?*)" in GA-ILMT without TTN. As a result, "*Kore wa kamera desuka?*" is produced as an erroneous translation result. This Japanese sentence translates to "Is this a camera?" in English. With TTN, the system selects node q_8 "(Is this your @0?; *Kore wa anata no @0 desuka?*)" because the correctness for the transition $q_4 \rightarrow q_6 \rightarrow q_8$ (52.0) is the highest. As a result, "*Kore wa anata no kamera desuka?*" is produced as the correct translation result. Thus, the system recognizes that node q_4 is the sentence translation rule that represents the basic structure of the whole sentence. This indicates that the TTN causes the system to translate by recognizing the structure of whole sentences.

Acknowledgements

References

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

S. Sato and M. Nagao. (1990), Toward Memory-based Translation, *In proceedings of the Coling'90*.

P. F. Brown, S. A. Della Pietra, V. J. Della Pietra and R. L. Mercer. (1993), The Mathematics of Statistical Machine Translation: Parametric Estimation, *Computational Linguistics*, Vol.9, No.2.

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

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

Jean-Marc Langè, Èric Gaussier and Bèatrice Daille. (1997), Bricks and Skeletond : Some Ideas for the Near Future of MAHT, *Machine Translation*, Vol.12, pp.39–51.

John E. Hopcroft and Jeferey D. Ullman. (1979), Introduction to Automata Theory, Languages, and Computation, *Addison-Wesley*.

J. Nakamura, J. Tsujii and M. Nagao. (1984), Grammar writing system(GRADE) of Mu-machine translation project and its characteristics, *In proceedings of the Coling'84*.

K. Hasegawa, et al., (1991), One World English Course 1 new edition, *Kyoiku Shuppan*.

A. Ota, et al., (1991), New Horizon English Course, *Tokyo Shoseki*.