Machine Translation Method Using Inductive Learning with Genetic Algorithms

Hiroshi Echizen-ya   **Kenji Araki**   **Yoshio Momouchi**

Department of Electronics and Information Engineering
Faculty of Engineering, Hokkai-Gakuen University
S 26-Jo, W 11-Chome, Chuo-ku, Sapporo, 064 Japan
E-mail:{echi,araki,momouchi}@eli.hokkai-s-u.ac.jp

Koji Tochinai

Department of Electronics and Information Engineering
Faculty of Engineering, Hokkaido University
N 13-Jo, W 8-Chome, Kita-ku, Sapporo, 060 Japan
E-mail:tochinai@hudk.hokudai.ac.jp

Abstract

We have proposed a method of machine translation, which acquires translation rules from translation examples using inductive learning, and have evaluated the method. And we have confirmed that the method requires many translation examples. To resolve this problem, we applied genetic algorithms to the method. In this paper, we describe our method with genetic algorithms and evaluated it by some experiments. We confirmed that the accuracy rate of translation increased from 52.8% to 61.9% by applying genetic algorithms.

1 Introduction

A practical and high quality method of machine translation is important for the internationalization of Japanese society. Many studies have been carried out on machine translation. The rule-based machine translation (John and Harold, 1992) could not deal adequately with various linguistic phenomena due to use only limited rules. To resolve this problem, example-based machine translations (Sato and Nagao, 1990; Akama and Ichikawa, 1979; Stanfill and Waltz, 1986; Sumita et al., 1993) have recently been proposed. However, these methods require many translation examples to realize a practical and high quality translation.

The goal of our research is to design a computer system with the same capability of language and knowledge acquisition as human beings (Araki and Momouchi, 1994; Araki et al., 1995).

In this paper, we propose a method of machine translation using inductive learning with genetic algorithms. The genetic algorithms (Goldberg, 1989) imitate the evolutionary process which repeats generational replacement to adapt to the environment. The purposes are to establish various high quality translation rules from only a small amount of data, and produce high quality translation results. The system is expected to continuously evolve to higher learning and translation capability.

In this paper, we describe a method of machine translation using inductive learning with genetic algorithms, and show through the results of evaluation experiments that genetic algorithms are effective for the example-based machine translation.

2 Processes in the New Method

2.1 Outline of the Translation Method

Figure 1 shows the outline of the new translation method. In this paper, we describe the process of English-Japanese translation as one possible application of this method. First, a user inputs a source sentence in English. Second, in the translation process, the system produces several candidates of translation results using translation rules extracted in the learning process. Third, the user proofreads the translated sentences. Fourth, in the feedback process, the system determines the fitness value and performs the selection process of translation rules. In the learning process, new translation examples are automatically produced by crossover and mutation, and various translation rules are extracted from the translation examples by inductive learning. A translation example includes the source sentence and a translated
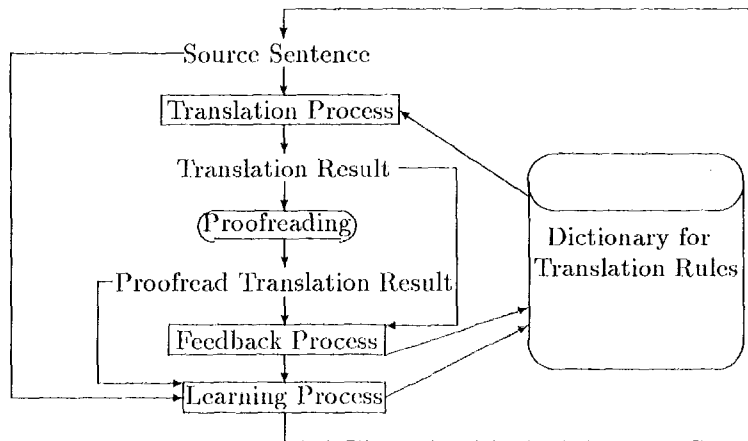
Figure 1: Outline of the new method of machine translation

sentence. There are two kinds of translation rules: those for sentences and those for words. The former are called sentence translation rules and the latter word translation rules. Repetition of the above mentioned processes corresponds to generational replacement of the whole of system, and the system continuously evolves to higher quality translation.

2.2 Chromosome and gene

As shown in Figure 2, a chromosome corresponds to a translation example which consists of English and Japanese sentence, and a gene corresponds to a word. In this paper, Japanese words are written in italics.
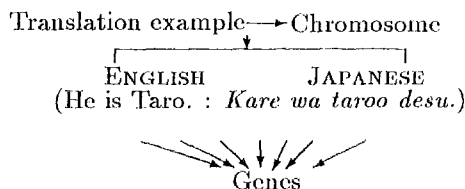

Figure 2: Chromosome and gene

2.3 Feedback Process

First, the system evaluates the translation results using the translated sentences which have been proofread. The system adds one to the correct translation frequency when the translation results have the same character strings as the proofread translation results, and adds one to the erroneous translation frequency when the translation rules have different character strings to the proofread translation results. Second, the system determines the fitness value of the rules used in translation using these correct and erroneous translation frequencies. The fitness value is calculated by the fitness function as follows:

Fitness value(%) =
$$\frac{\text{The correct translation frequency}}{\text{The number of uses}} \times 100 \quad (1)$$

Third, the system performs the selection process using the fitness value. The conditions of the selection process are that the number of uses is over 5 and the fitness value is under 25%. These thresholds were determined by preliminary experiment.

2.4 Learning Process

In this process, new translation examples are automatically produced by crossover and mutation. In crossover, two translation examples which have common parts are selected. Crossover positions are the common parts, and one-point crossovers for these two translation examples are performed. These one-point crossovers use each common part of the English and Japanese sentences. Figure 3 shows examples of crossover. In Figure 3, "likes" is the common part in the two English sentences, and "wa" and "ga suki desu" are the common parts in the two Japanese sentences. Therefore, "likes" and "wa" are the crossover positions. One-point crossovers are performed, producing two translation examples. Next, one-point crossovers are performed for "likes" and "ga suki desu". However, the translation examples which are produced have the same character strings as the source sentences, and therefore, these translation examples are not inputted into the dictionary. Translation examples are randomly changed by mutation, at a rate of 2%. New translation examples are also produced by replacing the words of translation examples with those of translation rules.

The system extracts the common and different parts from the character strings of all translation examples. These common and different parts are used as translation rules.

(1)Selection from translation examples
    ENGLISH   JAPANESE
  (He likes tennis.
    :*Kare wa tenisu ga suki desu.*)
  (She likes tea.
    :*Kanojyo wa ocha ga suki desu.*)
(2)The crossover of the English sentence
  He likes⟍/tennis.⟶He likes tea.
  She likes/⟍tea.⟶She likes tennis.
(3)The crossover of the Japanese sentence
  Kare wa ⟍ /*tenisu ga suki desu.*
  Kanojyo wa⟍ *ocha ga suki desu.*
    ⟶*Kare wa ocha ga suki desu.*
    ⟶*Kanojyo wa tenisu ga suki desu.*
(4)The translation examples produced
    ENGLISH   JAPANESE
  (He likes tea.
    :*Kare wa ocha ga suki desu.*)
  (She likes tennis.
    :*Kanojyo wa tenisu ga suki desu.*)

Figure 3: Crossover example

2.5 Translation Process

In this process, the system produces several candidates of translation results for a source sentence using extracted translation rules. This process also uses genetic algorithm. The details of this process are as follows:

1. Initial population

The system selects the translation rules which can be applied to the source sentence. The set of selected translation rules is called the initial population.

2. Determination of fitness value

The system calculates the fitness value of the translation rules by the fitness function (1).

3. Selection process

The method of the selection process was described in the section on feedback process.

4. Crossover

The method of crossover was described in the section on learning process.

5. Mutation

The method of mutation was described in the section on learning process.

6. Evaluation of population

The system substitutes the words in the word translation rules for the variables in the sentence translation rules. A translation rule includes a Japanese sentence or words corresponding to an English sentence or words. The system produces a Japanese sentence for the English sentence when the English sentence has the same character string as the source sentence. The Japanese sentence which is produced is the translation result. Figure 4 shows an example of how the translation result is produced.

The system selects the correct translation result according to two criteria when there are several candidates of translation results: one criterion

(1)The input sentence
  I am your teammate.
(2)The initial group
  ENGLISH   JAPANESE
  (I am @0.   :*Watashi wa @0 desu.*)
  (@0 am your @1. :*@0 wa kimi no @1 desu.*)
  (@0 am @1.   :*@0 wa @1 desu.*)
(3)The translation rule by crossover
  ENGLISH   JAPANESE
  (I am your @1.:*Watashi wa kimi no @1 desu.*)
(4)The evaluation of population
    ENGLISH   JAPANESE
    (teammate :*chiimumeito*)

  (I am your @1.:*Watashi wa kimi no @1 desu.*)

(I am your teammate.
  :*Watashi wa kimi no chiimumeito desu.*)
(5)The translation result
    JAPANESE
  Watashi wa kimi no chiimumeito desu.

Figure 4: Example of how the translation result is produced

is the translation rule which has a higher fitness value and the other is the translation rule which is more similar to the source sentence.

3 Experiments for Performance Evaluation

3.1 Method of Evaluation

The effective translation results are grouped into two categories:

(1)The translation result has the same character string as the proofread translation result.

(2)The translation result has the same structure as the proofread translation result.

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

The ineffective translation results are grouped into three categories:

(3)The translation result has a different character string than the proofread translation result without unregistered words.

(4)The translation result has a different character string than the proofread translation result with unregistered words.

(5)A failed translation.

The system ranks ten candidates of translation results for the user. The method for determining optimal translation results was described in Subsection 2.5.

3.2 Method of Experiments

In the experiments, 1,810 translation examples were used as data, of which 1,010 examples were taken from a textbook (Hasegawa et al., 1991) for first grade junior high school students, and

Table 1: Results of experiments using genetic algorithms

Rank		Rate	Total
Effective translation	(1)	47.0%	61.9%
	(2)	14.9%	
Ineffective translation	(3)	14.7%	38.1%
	(4)	17.1%	
	(5)	6.3%	

800 examples from another textbook (Ota et al., 1991) for first grade junior high school students in Japan. All of these translation examples were processed by the method outlined in Figure 1. First, 1,010 translation examples were used for the learning process, and 800 translation examples were used for evaluation of the translation. Experiments were carried out with and without genetic algorithms. In the experiments without genetic algorithms, crossover, mutation and the selection process were not performed.

3.3 Results of Experiments

The accuracy rate of translation increased from 52.8% to 61.9% by applying genetic algorithms. Table 1 shows the results of experiments using genetic algorithms. In this table, (1) $\sim$ (5) correspond to (1) $\sim$ (5) in Subsection 3.1. Table 2 shows examples of translation results using genetic algorithms.

3.4 Discussion

In the experiments without genetic algorithms, high quality translation results could not be obtained due to the requirement of a very large amount of translation examples which are similar to other translation examples. Therefore, we applied genetic algorithms to a method of machine translation using inductive learning to automatically produce new translation examples which are similar to other translation examples. By using genetic algorithms, the accuracy rate of translation increased from 52.8% to 61.9%.

4 Conclusion

In this paper, we proposed a new method of machine translation using inductive learning with genetic algorithms. The results of evaluation experiments showed that the accuracy rate of translation increased from 52.8% to 61.9% by using genetic algorithms. Thus, we consider that this new method can get the higher accuracy rate of translation and produce the higher quality translation results than that of other machine translations.

Table 2: Examples of translation results using genetic algorithms

Input sentences	Translation results
ENGLISH He is three years old.	JAPANESE *Kare wa san sai desu.*
ENGLISH Mike, is this your brother?	JAPANESE *Maiku, kochira wa kimi no ani desu ka?*
ENGLISH We are playing baseball.	JAPANESE *@0 wa yakyuu o shite imasu.*
ENGLISH Yumi speaks English very well.	JAPANESE *@0 wa totemo joouzu ni eigo o hanasu.*

References

W. John Hutchins and Harold L. Somers. 1992. An Introduction to Machine Translation. *ACADEMIC PRESS.* (London)

Sato, S. and Nagao, M. 1990. Toward Memory-based Translation. In *Proceedings of the Coling'90*, pages 247–252, Helsinki, Finland, August.

Akama, K. and Ichikawa, A. 1979. A Basic Model for Learning Systems. In *Proceedings of IJCAI-79*, pages 4–6, Tokyo, Japan, August.

Stanfill C. and Waltz D. 1986. Toward Memory-Based Reasoning. *Communications of the ACM*, Vol.29, No.12, pages 1213–1228.

Sumita, E., Oi, K., Furuse, O,. Iida, H., Higuchi, T., Takahashi, N. and Kitano, H. 1993. Example-Based Machine Translation on Massively Parallel Processors. In *Proceedings of IJCAI-93*, pages 1283–1288, Chambery, France, August–September.

Araki, K. and Momouchi, Y. 1994. Concept Learning from Japanese Copular Sentences using Heuristics. *World Scientific*, In *Proceedings of AI'94*, pages 466–473, Armidale, New South Wales, Australia, November.

Araki, K., Momouchi, Y. and Tochinai, K. 1995. Evaluation for Adaptability of Kana-Kanji Translation of Non-Segmented Japanese Kana Sentences using Inductive Learning. In *Proceedings of PACLING-II*, pages 1–7, Brisbane, Australia, April.

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

Hasegawa, K. et al., 1991. One World English Course 1 New Edition. *Kyoiku Shuppan.* (Tokyo).

Ota, A. et al., 1991. New Horizon English Course 1. *Tokyo Shoseki.* (Tokyo).