An Improvement in the Selection Process of Machine Translation Using Inductive Learning with Genetic Algorithms

Hiroshi Echizen-ya  Kenji Araki
Division of Electronics and Information  Dept. of Electronics and Information
Hokkaido University  Hokkai-Gakuen University
Sapporo, 060 Japan  Sapporo, 064 Japan
echi@hudk.hokudai.ac.jp  araki@eli.hokkai-s-u.ac.jp

Yoshikazu Miyanaga  Koji Tochinai
Division of Electronics and Information  Division of Electronics and Information
Hokkaido University  Hokkaido University
Sapporo, 060 Japan  Sapporo, 060 Japan
{miyanaga, tochinai}@hudk.hokudai.ac.jp

Abstract
We proposed a method of machine translation using inductive learning with genetic algorithms, and confirmed the effectiveness of applying genetic algorithms. However, the system based on this method produces many erroneous translation rules that cannot be completely removed from the dictionary. Therefore, we need to improve how to apply genetic algorithms to be able to remove erroneous translation rules from the dictionary. In this paper, we describe this improvement in the selection process and the results of evaluation experiments.

1 Introduction
Many studies have been carried out on machine translation and a number of problems has been recognized. Rule-based machine translation (Hutchins and Somers, 1992) could not deal adequately with various linguistic phenomena due to the use of limited rules. To resolve this problem, Example-based machine translation (Sato and Nagao, 1990) has recently been proposed. However, this method requires many translation examples to achieve a practical and high-quality translation.

Echizen-ya and others previously proposed a method of Machine Translation using Inductive Learning with Genetic Algorithms (GA-ILMT), and this method has been evaluated (Echizen-ya et al., 1996). By applying genetic algorithms, we consider that our proposed method can effectively solve problems that Example-based machine translation would require many translation examples. However, the results of the evaluation experiments show that this method has some problems. The main problem is that many erroneous translation rules are produced and these rules cannot be completely removed from the dictionary. Therefore, we need to improve how to apply genetic algorithms to be able to remove erroneous translation rules. In this paper, we describe an improvement in the selection process of GA-ILMT, and confirm the effectiveness of improvement in the selection process of GA-ILMT.

2 Outline of Translation Method
Figure 1 shows the outline of our proposed translation method. First, the 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 if they include some errors. Fourth, in the feedback process, the system determines the fitness value of translation rules used in the translation process and performs the selection process of erroneous 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.

3 Improvement in Selection Process
In the previous method of selection process described in Section 2, translation rules are evaluated only when they are used in the translation process. These translation rules are part of all the translation rules in the dictionary. Therefore, many erroneous
translation rules cannot be completely removed from
the dictionary.

To resolve this problem, we propose an improve-
ment in the selection process. Our proposed im-
provement does not require any analytical knowl-
edge as initial condition. Methods that use analyti-
cal knowledge have some problems, such as difficulty
in dealing with unregistered words. We consider
that this problem can be resolved by the learning
method without any analytical knowledge. There-
fore, we consider that our proposed improvement
can remove many erroneous translation rules by uti-
llizing only the given translation examples without
the requirement of analytical knowledge.

The system evaluates the translation rules by
utilizing the given translation examples directly.
Namely, it determines whether a combination of the
English word and the Japanese word in a translation
rule is true or false by utilizing the given translation
examples. The combination may be true when it ex-
ists in a given translation example. For example, the
combination of words which are "I" in English and
"Watashi" (In Japanese "I") in Japanese is true
when this combination exists in a given translation
example. On the other hand, the combination of
words which are "volleyball" in English and "Ba-
suketoboru" (In Japanese "basketball") in Japanese
is false when this combination does not exist in all
given translation examples. In the all combinations
of words in a translation rule, the system determines
whether the each combination of words is true or
false. And the system determines the rate of error
based on the number of erroneous combinations, and
removes the translation rules for which the rate of
error is high.

4 Experiments

In the experiments, 461 translation examples were
used as data. The examples were taken from a text-
book (Hasegawa et al., 1991) for first-grade junior

1 Italic means pronunciation of Japanese

high school students. All of the translation ex-
amples were processed by the method outlined in Fig-
ure 1. The initial dictionary was empty. The exper-
iments were carried out with and without the im-
provement for the selection process described in Sec-
ion 3. In the experiments, the precision increased
from 87.5% to 93.7% and the recall increased from
4.5% to 56.0%.

5 Conclusion

In the previous selection process, the translation
rules are evaluated only when they are used in the
translation process. Therefore, the translation rules
which are not used in the translation process are
never removed from the dictionary. However, the
proposed improvement can evaluate all of the pro-
duced translation rules by utilizing only the given
translation examples.

6 Acknowledgements

The part of this research has been done during the
second author Kenji Araki's stay at CSLI of Stan-
ford University. We would like to express our special
appreciation to the stuffs of CSLI.

References

W. John Hutchins and Harold L. Somers. 1992.
An Introduction to Machine Translation. ACA-
DEMIE PRESS. (London)

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

Echizen-ya, H., Araki, K., Momouchi, Y. and Tochi-
nai, K. 1996. Machine Translation Method Us-
ing Inductive Learning with Genetic Algorithms.
In Proceedings of the Coling’96, pages 1020–1023,
Copenhagen, Denmark, August.

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