

4 B-8 遺传的アルゴリズムを適用した帰納的学習による 機械翻訳手法（GA-ILMT）における淘汰処理の改良

越前谷博† 荒木健治†† 宮永喜一† 枋内香次†
北海道大学工学部† 北海学園大学工学部††

1 はじめに

より実用的な機械翻訳システムの構築に向け、これまでに多くの研究が行なわれてきた。その主流となっているのが文法解析に基づく解析型の機械翻訳手法 [1] である。しかし、これらの手法は、有限個の文法規則や固定化された辞書を使用することによるロバストネスの問題が指摘されている。この問題点を解決するために、近年、コーパスに基づく学習型の機械翻訳手法 [2][3] の研究が盛んに行なわれている。しかし、これらの手法においては、良質な翻訳を行なうために、大量のコーパスが必要となるのが問題となっている。我々は、このような状況下において、与えられた少量の翻訳例から、より多くの翻訳ルールを自動的に抽出し、翻訳を行なう、遺传的アルゴリズムを適用した帰納的学習による機械翻訳手法（以下、GA-ILMT と呼ぶ）の提案とその性能評価 [4][5] を行なってきた。その結果、翻訳例をより効率的に活用した翻訳が可能となったが、いくつかの問題点も明らかとなった。最も大きな問題点は、淘汰処理の精度が低いために、誤った翻訳ルールを十分に辞書中から消去できないことである。本稿では、淘汰処理の精度向上のために行なった改良についてその手法と考察結果について述べる。

2 処理過程

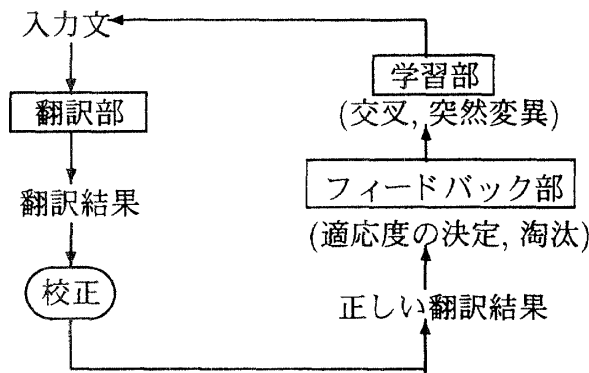


図1 処理過程

GA-ILMT に基づき構築した英日機械翻訳システムの処理過程を図1に示す。まず、入力文として、英文を入力する。すると、翻訳部において、辞書中の翻訳ルールを用い、最適な翻訳結果を生成する。生成された翻訳結果に対しては、必要に応じて、人手による校正を行ない、正しい翻訳結果を得る。次いで、フィードバック部において、使用された翻訳ルールに対する適応度の決定と淘汰を行なう。そして、学習部において、翻訳例に対し、交叉と突然変異を行ない、多くの翻訳ルールを抽出する。

3 淘汰手法の改良

3.1 問題点

図1に示した処理過程に基づき構築したシステムを用いて行なった性能評価の結果 [4]、問題点として、誤った翻訳ルールに対する淘汰処理が不十分であることが明らかとなった。従来の淘汰手法では、翻訳に使用された翻訳ルールのみが評価の対象となる。しかし、そうした翻訳ルールは、辞書中に存在する翻訳ルールにおいて、少数である。また、従来手法では、淘汰される翻訳ルールは、少ない単語数より構成されている翻訳ルールである。これは、構成されている単語数の多い翻訳ルールほど、翻訳時に、適用可能な翻訳ルールを選択する場合、入力文と表層レベルで一致する可能性が低くなり、使用される回数が少なくなるためである。したがって、多くの単語数より構成されている翻訳ルールを淘汰することが困難となる。

3.2 改良手法

3.1 節で述べた問題点を解決するために、我々は、淘汰手法の改良を試みた。改良手法では、遺传的アルゴリズムの基本的な枠組に沿って、与えられた環境を有効利用することにより淘汰処理の精度向上を行なう。また、解析的な知識をシステムに与えることは、例外的なものに対処することが困難になるというロバストネスの問題を抱え込むことになると考えられるため、初期条件としての解析的な知識は取り入れていない。こうした点から、我々は、GA-ILMT において環境となる翻訳例のみを利用することにより、淘汰処理の精度向上を試みた。本改良手法では、翻訳例を直接的に参照することにより、翻訳ルールの正誤を判定する。翻訳例の英文とその日本語訳文を構成している単語の組合せを利用することにより、翻訳ルールを構成している単語の組合せが、翻訳例中に存在しているかどうかを検索する。存在している場合、その単語の組合せは確率的に正しい組合せと考えられる。存在

An Improvement in the Selection Process of Machine Translation Using Inductive Learning with Genetic Algorithms (GA-ILMT)

Hiroshi Echizen-ya†, Kenji Araki††, Yoshikazu Miyanaga† and Koji Tochinai†

†Faculty of Engineering, Hokkaido University

††Faculty of Engineering, Hokkai-Gakuen University

していない場合には、誤った組合せと考えられる。このように、生成された翻訳ルールに対して、過去に与えられた翻訳例を直接的に参照することにより、翻訳ルールがどの程度誤ったものなのかを評価する。以下に詳細な処理手順を示す。

1. 翻訳ルールから単語の組合せを抽出する。
2. 単語の組合せが翻訳例中に存在しているかどうかを検索する。
3. 存在している場合には true を与え、存在していない場合には false を与える。
4. false となった単語の組合せにおいて、組合せ誤りを検索し、消去する。
 - ① 全ての false となった単語の組合せ同士で、英単語と日本語単語の組合せを変更する。
 - ② 変更後の単語の組合せが、辞書中に翻訳ルールとして存在しているかどうかを検索する。
 - ③ 存在していた場合、変更前の単語の組合せを false から true に変更する。
5. 翻訳ルールの誤り率を決定する。以下にその計算式を示す。

$$\text{誤り率 (\%)} = \frac{\text{誤った組合せの数}}{\text{組合せ総数}} \times 100 \quad (1)$$

6. 閾値 x を設け、以下の条件を満たす翻訳ルールを辞書中から削除する。

$$\text{誤り率} \geq x \quad (2)$$

4 評価実験

4.1 実験方法

実験は、最初に辞書の初期状態を空にし、中学1年生用教科書ガイド・ワンワールド [6] に掲載されている英文とその日本語訳文 400 組を用いて、図 1 の処理過程に従い、翻訳ルールを抽出し、淘汰を行なった。また、閾値には、予備実験により、最適と考えられる 5% を用いて行なった。

4.2 実験結果

表 1 淘汰処理における適合率と再現率

	適合率	再現率
従来手法	85.6%	4.6%
改良手法	91.8%	56.1%

従来手法と改良手法のそれぞれの淘汰処理における適合率と再現率を表 1 に示す。適合率は、淘汰された翻訳ルールにおける、正しく淘汰されたものの占める割合である。再現率は、辞書中に存在する誤った翻訳ルールにおける、正しく淘汰されたものの占める割合である。

4.3 考察

実験の結果、従来手法に比べ、適合率は 6.2%、再現率は 51.5% 増加した。再現率の大幅な向上は、改良手法が、生成された翻訳ルールの全てに対して評価することが可能となったことに依るものと考えられる。また、適合率においては、6.2 ポイントの向上であったが、正しく淘汰された翻訳ルールの数は、約 11 倍であった。したがって、多くの誤った翻訳ルールを淘汰でき、その中で、比較的高い精度を示していることを確認できた。そして、改良手法では、多くの単語数より構成されている翻訳ルールに対する淘汰が可能となった。従来手法において、淘汰された翻訳ルールの平均構成単語数は 4.4 であった。それに対し、改良手法の場合では、7.5 であった。また、生成された全翻訳ルールの場合には、7.3 であった。したがって、改良手法では、比較的多くの単語数より構成されている翻訳ルールに対する淘汰が可能になったことを確認できた。また、処理時間は、改良手法の導入により、約 42.3% 減少した。

5 おわりに

本稿では、GA-ILMT における淘汰処理の精度向上のための改良手法とその有効性について述べた。評価実験の結果、解析的な知識を初期条件として与えずに、翻訳例のみから、学習的手法を用いて淘汰の精度を向上させることが可能であることを確認できた。今後は、GA-ILMT において、翻訳ルールの階層化を行ない、翻訳処理の性能向上を図る。そして、より実用的な学習型の機械翻訳システムの実現に向けての研究を進める予定である。

参考文献

- [1] 野村浩郷 (編): 言語処理と機械翻訳, 講談社 (1991).
- [2] 佐藤理史: MBT 2: 実例に基づく翻訳における複数翻訳例の組合せ利用, 人工知能学会誌, Vol. 6, No. 6, pp. 861-871 (1991).
- [3] 北村美穂子, 松本裕治: 対訳コーパスを利用した翻訳規則の自動獲得, 情報処理学会論文誌, Vol. 37, No. 6, pp. 1030-1040 (1996).
- [4] 越前谷博, 荒木健治, 桃内佳雄, 枅内香次: 実例に基づく帰納的学習による遺伝的アルゴリズムの適用とその有効性, 情報処理学会論文誌, Vol. 37, No. 8, pp. 1565-1579 (1996).
- [5] Echizen-ya, H., Araki, K., Momouchi, Y. and Tochinnai, K. 1996. Machine Translation Method Using Inductive Learning with Genetic Algorithms. In *Proceedings of the Coling'96*, pages 1020-1023, Copenhagen, Denmark, August.
- [6] 教科書ガイド 教育出版版ワンワールド 1, 日本教材, 東京 (1991).