

帰納的学習による機械翻訳手法への 遺伝的アルゴリズムの適用

2K-1

越前谷 博 荒木 健治 桃内 佳雄

北海学園大学工学部

1. はじめに

我々は実例より翻訳ルールを学習し、その翻訳ルールを用いて翻訳を行う帰納的学習による機械翻訳手法¹⁾の提案とその性能評価実験²⁾を行ってきた。本稿では、この帰納的学習による機械翻訳手法に遺伝的アルゴリズムを適用することで更に良質な学習型の機械翻訳手法を構築することを試みた結果について述べる。遺伝的アルゴリズムの帰納的学習による機械翻訳手法への適用に際しては、その基本操作を導入することで良質な翻訳結果を導き出すために不可欠となる多様な翻訳ルールの生成を目的としている。

2. 処理過程

本手法は、個体に翻訳例を用い、また、その個体を構成する遺伝子には翻訳例を構成している単語を用いる。翻訳例は、翻訳ルールを抽出した後に、それ自体も翻訳ルールとして登録される。これを原文の翻訳ルールとよぶ。この様な状況で、遺伝的アルゴリズムの基本操作を行う。本手法は、環境に適応するための進化を繰り返し行うことによりシステムが成長し、最適な翻訳を行うことができる学習型の翻訳システムを実現するためのものである。学習部は翻訳ルールに対して選択交配処理を行い多くの新しい翻訳ルールを抽出する。翻訳部では学習部より抽出された翻訳ルールを用いて翻訳を行う。そして、フィードバック学習部で正しい日本語訳文から適応度の決定と淘汰を行い不必要な適応度の低い翻訳ル

ールを消滅させる。図1に処理の流れを示す。

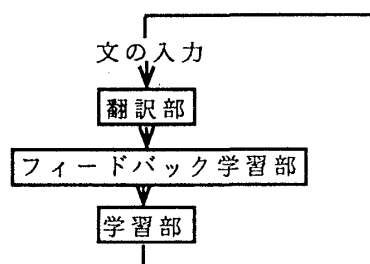


図1 処理の流れ

3. 学習部

学習部では生き残った個体同士が交叉により次の世代を作り出す。本手法において、このプロセスは以後の翻訳を行うための新たな翻訳ルールを生成する役割を持つ。その生成方法の例を以下に示す。まず、翻訳例において共通部分を持つものを選択する。

[1-1] (This is Hanako.
こちら/は/花子/です。)

[1-2] (She is my friend.
彼女/は/私の/友達/です。)

この2つの翻訳例を比較すると、共通部分として英文と日本語訳文からそれぞれ"is", "は"と"です"が抽出される。よって、英文においては"is", 日本語訳文においては"は"を交叉位置として一点交叉を行う。すると、次の様な新たな翻訳例が得られる。

[1-3] (This is my friend.
こちら/は/私の/友達/です。)

[1-4] (She is Hanako.
彼女/は/花子/です。)

また、"です"を交叉位置とする日本語訳文の交叉は新たな翻訳例を生成しないので行わない。この様にして交叉により得られた文に我々が先に提案した翻訳例の字面情報から差異部分と共通部分を多段階に抽出する手法¹⁾を適用することで、[1-1], [1-2]と[1-3], [1-4]よりそれぞれ以下の翻訳ルールが抽出される。

The Application of Genetic Algorithms to
Machine Translation Method by Inductive
Learning

Hiroshi Echizen-ya, Kenji Araki and
Yoshio Momouchi

Faculty of Engineering, Hokkai-Gakuen
University

- [2-1] (Hanako 花子)
 [2-2] (my friend 私の/友達)
 [2-3] (This is @0. こちら/は/@0/です。)
 [2-4] (This こちら)
 [2-5] (She 彼女)
 [2-6] (@0 is Hanako.@0/は/花子/です。)
 [2-7] (She is @0. 彼女/は/@0/です。)
 [2-8] (@0 is my friend.
 @0/は/私の/友達/です。)
 [2-9] (@0 is @1. @0/は/@1/です。)

4. 評価実験

4.1 実験方法

実験システムはワークステーション上に作成した。なお、今回作成した実験システムは学習部のみなので誤った原文の翻訳ルールから生成された翻訳ルールに対する淘汰は人手により行った。実験の手順として、中学1年用教科書ガイド・ワンワールド³⁾に掲載されている英文とその日本語訳文の組68文に対して1文ずつ学習させ翻訳ルールに対する淘汰を繰り返した。淘汰の例を以下に示す。

[3-1] (This is my classmate.
 これ/は/私の/クラスメイト/です。)

[3-2] (You are not my friend.
 あなた/は/私の/友達/ではありません。)

この2つの翻訳例に対し共通部分を交叉位置として交叉させると、次の翻訳例が得られる。

[4-1] (This is my friend.
 これ/は/私の/友達/ではありません。)

[4-2] (You are not my classmate.
 あなた/は/私の/クラスメイト/です。)

この様にして得られた誤った原文の翻訳ルールを親として新たに作り出される翻訳ルールは、誤った性質を継承することになる。よって、淘汰の対象となり削除される。

4.2 実験結果

得られた翻訳ルールの抽出結果を表1に示す。()の数字は正しい翻訳ルールの数と最終的に抽出された翻訳ルールの数である。

表1 正しい翻訳ルールの割合

	割合
変数なし	77.2% (401 / 519)
変数1つ	79.2% (145 / 183)
変数2つ	77.2% (34 / 44)
変数3つ	75.0% (3 / 4)

4.3 考察

実験結果から、遺伝的アルゴリズムを適用することで従来の手法に比べて3倍~7倍の正しい翻訳ルールの抽出が可能となった。特に、抽象度の低い翻訳ルールの絶対数が増加した。これは従来の帰納的学習による機械翻訳手法の問題点であった、抽象度の高い翻訳ルールの誤った適用による誤訳文の生成と翻訳例の不足による出現頻度の低い単語の抽出不能を解消するために大きな役割を果たすものと考えられる。

5. おわりに

本稿では従来の帰納的学習による機械翻訳手法に遺伝的アルゴリズムを適用する第1歩として、学習部における交叉による次世代への個体生成、つまり翻訳ルールの生成に対する評価実験を行った。その結果、良質な翻訳結果を得るうえで有効な手段となることが確認できた。しかし、その必須条件としてフィードバック学習部による淘汰を高い精度で行う必要がある。今後は、更に複雑な翻訳例を使用しながら、フィードバック学習の精度の向上を中心に研究を進める予定である。

参考文献

- (1) 荒木健治、栃内香次：多段階共通パターン抽出法を用いた翻訳例からの帰納的学習による翻訳、情報処理北海道シンポジウム'91, pp.47-49(1991).
- (2) 内山智正、荒木健治、宮永喜一、栃内香次：帰納的学習による機械翻訳手法の評価実験、情報処理学会研究報告, NL93-4, pp.23-30(1993).
- (3) 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).