

遺伝的アルゴリズムを適用した

5H-10

帰納的学習による機械翻訳手法の実用化に向けて

越前谷 博[†] 荒木 健治[†] 桃内 佳雄[†] 栃内 香次^{††}
[†]北海学園大学工学部 ^{††}北海道大学工学部

1. はじめに

実用的な機械翻訳システムの実現に向け、これまでに多くの翻訳手法の研究が行われている。文法解析による解析型の機械翻訳手法¹⁾は、代表的な翻訳手法である。しかし、解析型の機械翻訳手法は有限個の文法規則を用いるため、多様な言語現象を持つ文章の良質な翻訳の困難さとその際の辞書の作成や改良の困難さが問題点として指摘されている。また、近年、盛んに研究されている実例や用例による学習型の機械翻訳手法^{2, 3)}では、十分な翻訳を行うために大量の学習データが必要となる。我々は従来より与えられた実例から帰納的に翻訳ルールを学習し、翻訳を行う機械翻訳手法⁴⁾に遺伝的アルゴリズムを適用することにより良質な翻訳を行うことを目的とした研究を行っている^{5, 6)}。その結果、与えられた少量の実例を最大限に活用した学習と翻訳が可能となったが、中学1年生レベルの翻訳であり実用的なものとはなっていない。本稿では、より実用レベルのデータを用いた実験結果に基づき、実用化に向けての遺伝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の有効性と問題点について考察する。

2. 処理過程

本手法では、翻訳例を含む翻訳ルールを個体として、また、翻訳ルールを構成している単語を遺伝子として位置付けている。このような状況で遺伝的アルゴリズムを適用することにより、システムが最適な翻訳を行うための世代交代を繰り返し、より良い学習型の機械翻訳システムへと成長し続ける。図1に処理過程を示す。また、()は遺伝的アルゴリズムの適用範囲を示している。

2.1 学習部

学習部では、入力された翻訳例に対し、交叉と突然変異を行うことで、与えられたデータを最大限に活用したより多くの翻訳ルールを生成する。まず、英文とその日本語訳文のそれぞれにおいて、共通部分を持つ2つの翻訳例を選択する。以下に選択された翻訳例の具体例を示す。

(Makoto is older than Osamu.
 ; 真/は/修/より/年/を/とって/いる。)

Toward a Practical Machine Translation Method That Uses Inductive Learning to Apply Genetic Algorithms
 Hiroshi Echizen-ya[†] Kenji Araki[†] Yoshio Momouchi[†] and Koji Tochinal^{††}

[†]Faculty of Engineering, Hokkai-Gakuen University
^{††}Faculty of Engineering, Hokkaido University

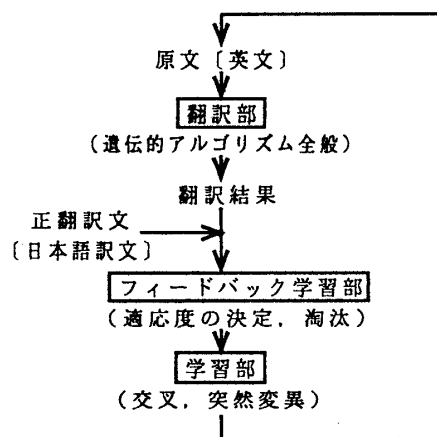


図1 処理過程

(Judy is the oldest of the three.
 ; ジュディ/は/3人/で/いちばん/年/を/とって/いる。)
 次に、共通部分の「is」と「は」を交叉位置とし、英文とその日本語訳文のそれぞれにおいて一点交叉を行う。交叉後に生成された新たな翻訳例を以下に示す。
 (Makoto is the oldest of the three.
 ; 真/は/3人/で/いちばん/年/を/とって/いる。)
 (Judy is older than Osamu.
 ; ジュディ/は/修/より/年/を/とって/いる。)

また、日本語訳文の「年/を/とって/いる」を交叉位置とした交叉は新たな翻訳例を作り出さないので、翻訳ルールとしての登録は行わない。このようにして新たに生成された翻訳例を加え、我々が先に提案した共通部分と差異部分を多段階に抽出する手法⁷⁾を用いることで多様な翻訳ルールを生成する。

突然変異は、翻訳例の単語に対し突然変異率を2%としたランダムな変更を行う。

(I like baseball better.
 ; 私/は/野球/の/ほう/が/好き/です。)
 ↓ 変数との置換 ↓
 (I like @0 better.
 ; 私/は/@0/の/ほう/が/好き/です。)

このように学習部では、遺伝的アルゴリズムの基本操作を適用し、より多くの翻訳ルールを生成する。

2.2 翻訳部

本手法では、より最適な翻訳結果を生成するため、翻訳部のみで1つの完結した遺伝的アルゴリズムのシステムを構築している。翻訳部における翻訳結果の生成過程を以下に示す。

(1) 学習部において生成された翻訳ルールから入力文に対し適用可能な翻訳ルールを選択し初期集団を

作り出す。

- (2) フィードバック学習部と同様の方法により、各翻訳ルールの適応度の決定と淘汰を行う。
- (3) 学習部と同様の方法により、翻訳ルールに対し交叉と突然変異を行う。
- (4) 生物集団の評価を行う。つまり、翻訳ルールを組み合わせ、入力文に対する翻訳結果を求める。
- (5) 最適な翻訳結果を求め(1)から(4)を繰り返す。

2.3 フィードバック学習部

フィードバック学習部では、翻訳部で使用された翻訳ルールに対し正しい日本語訳文より適応度の決定と淘汰を行う。適応度は以下の計算式により決定される。

$$\text{適応度 (\%)} = \frac{\text{正翻訳度数}}{\text{全翻訳度数}} \times 100$$

淘汰は適応度が25%以下、全翻訳度数が5以上の翻訳ルールに対し行う。

3. 評価実験

3.1 実験方法

まず、中学1年生レベルの文章1,810組より学習した翻訳ルールを辞書に登録することで初期辞書を作成した。次いで、中学2年生用教科書ガイド・ワンワールド⁹⁾に掲載されている英文とその日本語訳文の80組(160文)に対し、翻訳と学習を1文ずつ繰り返して行った。

3.2 実験結果

実験結果を表1に示す。ここで、有効な翻訳とは、未登録語を含まない正翻訳と未登録語を含む正翻訳である。未登録語は名詞句または形容詞など未登録語を与えることで容易に正翻訳が得られるものである。無効な翻訳とは、未登録語を含まない誤翻訳と未登録語を含む誤翻訳、及び全く翻訳が行えなかった翻訳不能である。

表1 実験結果

	割合(文数)
有効な翻訳	52.5% (42)
無効な翻訳	47.5% (38)

3.3 考察

実験の結果、中学1年生レベルの初期の翻訳結果⁵⁾と比較し、翻訳率は62.0%から52.5%に低下した。これは、中学1年生レベルの実験では名詞句または形容詞以外の未登録語を持った文章が25.0%を占めていたのに対し、本実験では35.0%を占めていたことから、学習不足が原因であったと考えられる。この初めて出現する名詞句または形容詞以外の未登録語において、本実験データ中に、再度、他の文章の中に出現したものは57.1%であった。さらに、その文章において有効な翻訳となったものは81.3%であった。これは、1度教わったことが効率良く学習され、それ以後の翻訳に十分に活用されていることを示している。つまり、遺伝的アルゴリズムの適用が

より実用レベルの文章の翻訳においても、与えられた事例を最大限に活用した翻訳ルールの学習とその翻訳ルールを使用したより最適な翻訳結果の生成を可能にしていることを示している。以下にその具体例を示す。

- (1) 入力文: Kyushu is larger than Shikoku.
- (2) 共通部分を持った翻訳例の選択
[1-1] (Honsyu is larger than Shikoku.
: 本州/は/四国/より/大きい。)
[1-2] (Hokkaido is very popular in Japan.
: 北海道/は/日本/中/で/とても/人気/が/あり/ます。)
- (3) (is:は)を交叉位置とした交叉から生成された翻訳例
[1-3] (Hokkaido is larger than Shikoku.
: 北海道/は/四国/より/大きい。)
- (4) 翻訳ルールの抽出
[1-1]と[1-3]より、(@0 is larger than Shikoku.
: @0/は/四国/より/大きい。)
- (5) 既存の単語の翻訳ルール (Kyushu:九州)の代入による翻訳結果の生成
翻訳結果:九州は四国より大きい。

また、問題点として、淘汰処理の能力の低さが挙げられる。特に、今回の実験のように初めて出現する単語が多く存在する場合、新たに生成される翻訳ルールが増加する。その中には、正しい翻訳ルールと同時に誤った翻訳ルールも存在しているため誤った翻訳ルールの絶対数が増加することになる。

4. おわりに

本稿では、これまでの中学1年生レベルの文章と比べ、より実用レベルの文章においても、従来の遺伝的アルゴリズムの適用における有効性を損なうことなくより最適な翻訳が行えることを実験結果より述べた。今後は、さらに多くのより実用レベルの文章を用いた評価実験を行い、実用化に向けてより良い翻訳を行うため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) 赤間清:帰納的学習システムLS/1による翻訳の学習, 人工知能学会誌, Vol.2, No.3, pp341-349(1987).
- (3) 佐藤理史:MBT2: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, pp861-871(1991).
- (4) 内山智正, 荒木健治, 宮永喜一, 柄内香次: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).
- (5) 越前谷博, 荒木健治, 桃内佳雄: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, 北海学園大学工学部研究報告, 22, pp.275-283(1994).
- (6) 越前谷博, 荒木健治, 桃内佳雄, 柄内香次:実例からの帰納的学習による機械翻訳手法における遺伝的アルゴリズムの有効性について, 情報処理学会研究報告, NL107-4, pp.75-82(1995).
- (7) 荒木健治, 柄内香次: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).
- (8) 教科書ガイド教育出版版ワンワールド2, 日本教材, 東京(1991).