

複数編集距離を用いた機械翻訳の自動評価*

5 Y-7

秋葉 泰弘† 今村 賢司† 隅田 英一郎†
ATR 音声言語通信研究所†

1 はじめに

近年、機械翻訳システムによる機械訳を自動的に評価する手法の研究が盛んになって来た [5, 8, 2, 3]. そのような自動評価法の 1 つに、単一の編集距離に基づき機械訳に ABCD [6] の様なランク付与を行う研究がある [8, 2]. しかし、全てのランクを高い精度でランク付けする編集距離を設計する事は難しい [3].

以下本稿では、複数の編集距離を用いてランク付け規則を自動学習し、学習された規則で自動的にランク付けを行う手法を提案する. また、提案手法を実験的に評価した結果を報告する.

2 提案手法

提案手法は、図 1 に示すように、学習フェイズと評価フェイズからなる. 学習フェイズでは、予めランク付けされた各機械訳を複数の編集距離を用いてベクトル表現し、これらベクトルを訓練事例としてランク付け規則を決定木形式に自動学習する. 評価フェイズでは、ランク付けを行う機械訳を学習フェイズと同様にベクトル表現し、学習した決定木を用いて自動的にランク付けする.

各機械訳から生成されるベクトルの各要素は、編集距離の各バリエーション (後述) により計算された、該機械訳と正解の人手訳との最小編集距離 [8] である. 例えば図 1 の場合、機械訳と各人手訳について計算される 3 つの編集距離の最小値が、最小編集距離となる.

本稿で利用する編集距離の基本形 (ED1) は、形態素の正規形と品詞の組を編集単位とし、挿入、削除、置換 [8] を編集操作とする. 編集距離は、図 2 に示すように、DP [5] で計算される. 他の編集バリエーションと該基本形との違いは、以下の点を考慮するか否かで全部で 16 個の編集距離 (ED1 ~ ED16) が定義される.

1. 編集操作に交換 [5] をも利用し、入換え可能な語順に対処する.
2. 編集単位を内容語に限定し、機能語の抜けや入換え可能性に対処する.

S(k): 翻訳対象原文 (k は文番号)
T(k): S(k) の機械翻訳出力
Rank(k): T(k) のランク
T(k, h): S(k) を人手で翻訳した結果の複数正解例 (h は正解番号)
EDv: 後述の編集距離のバリエーション (v はバリエーション番号)
Cv(k): 各 T(k, h) と T(k) との編集距離 EDv の最小値
U: ランク未知の機械翻訳出力
Cv(U): U に対応する複数正解各々 U との編集距離 EDv の最小値

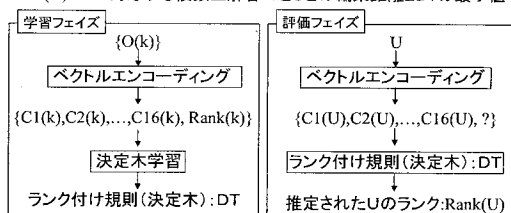


図 1: 提案手法の概要

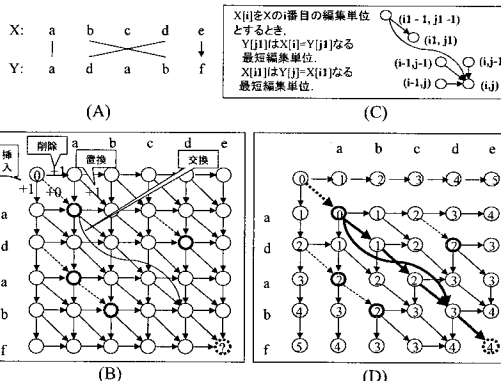


図 2: 編集距離: ED1 [8], ED9 (交換が可能な ED1) [5]

3. 内容語について、正規形の代わりに意味コードを利用し、異表記語や同義語を一致させる. 図 3 の「来る」と「訪れる」の意味コード 283 のように、共通意味コードがあれば一致したと見なす.
4. 複数正解の 2 つ以上に現れた形態素をキーワード¹と考え、編集単位をキーワードに限定し、翻訳すべき重要語の有無をチェックする. 図 3 の人手訳 2 に表れる「彼」はキーワードでない.

3 実験評価

提案手法を評価するために、旅行予約会話文 [7] 中の約 340 個の英文を ATR で研究開発中の英日機械翻訳

¹ここではキーワードは内容語に限定されず、機能語も含み得る.

*Automatic Evaluation of Machine Translation systems by using Multiple Edit Distances.

Yasuhiro Akiba† Kenji Imamura† Eiichiro Sumita†

† ATR Spoken Language Translation Research Laboratories, 2-2-2 Hikaridai Seika-cho Souraku-gun Kyoto 619-0288, Japan

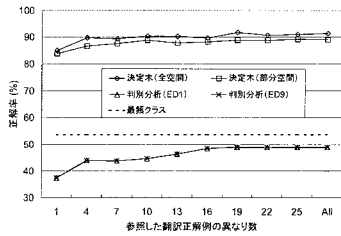


図 4: クローズドテストでの実験結果

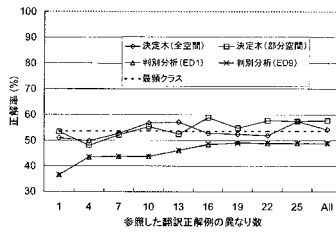


図 5: オープンテストでの実験結果

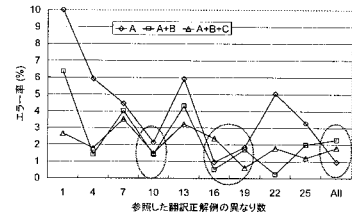


図 6: 翻訳率の推定エラー率

機械訳 今度観光のスペインに来るカップルです。		機械訳の形態素列 今度 普通名詞 (156) 観光 サ変名詞 (892) の 格助詞 () スペイン 普通名詞 (709) に 格助詞 () 来る 本動詞 (283 312) カップル 普通名詞 (530) です 判定詞 () 。 記号 ()	
複数正解[人手訳]: 人手訳1: スペインへ観光に来るカップルです。 人手訳2: 彼らは観光しにスペインに来るカップルです。 人手訳3: 観光でスペインを訪れるカップルです。		人手訳1の形態素列 スペイン 普通名詞 (709) へ 格助詞 () 観光 サ変名詞 (892) に 格助詞 () 来る 本動詞 (283 312) カップル 普通名詞 (530) です 判定詞 () 。 記号 ()	
人手訳2の形態素列 彼 代名詞 () ら 人称接尾辞 () は 係助詞 () 観光 サ変名詞 (892) する 補助動詞 () に 格助詞 () スペイン 普通名詞 (709) に 格助詞 () 来る 本動詞 (283 312) カップル 普通名詞 (530) です 判定詞 () 。 記号 ()		人手訳3の形態素列 観光 サ変名詞 (892) で 格助詞 () スペイン 普通名詞 (709) に 格助詞 () 訪れる 本動詞 (283 786) カップル 普通名詞 (530) です 判定詞 () 。 記号 ()	

図 3: 機械訳と複数正解 (人手訳)

システム TDMT-EJ[6] に翻訳させ、出力された機械訳を、提案手法により自動ランク付けを試みた。比較のために、従来法[2]で採用された判別分析を用いて自動ランク付けも行った。

機械訳へのランク付け正解は、日本語を母国語とする、英語が堪能な3名が行ったランク付けの多数決ランクとした。翻訳の複数正解例は、1原文当り高々26事例で、その内訳は、25人の翻訳家による人手訳、及び前出の旅行予約会話文[7]中の対訳である。機械訳の日本語は日英機械翻訳システム TDMT-JE[6]の形態素解析部を用いて解析した。意味コードは TDMT で採用している角川新類語辞典²[1]を用いた。決定木は、C4.5[4]で学習した。ランク付けを行うランク判別関数(決定木と判別分析の判別関数)の正解率は、10-cross validation で計算した。

図4, 図5は、順にクローズドテスト及びオープンテストの結果で、参照する翻訳正解例の異なり数を変化させた場合に、各手法によるランク付け正解率がどの様に変化したかを示す。決定木(全空間)及び決定木(部分空間)が提案手法の結果を示す。決定木(全空間)は、2節で定義した全ての編集距離を使った場合の結果であり、決定木(部分空間)は一部の編集距離のみを使った場合の結果である。判別分析(ED1)及び判別分

析(ED9)が従来法の結果で、判別分析(ED1)はED1で判別分析した場合の結果で、判別分析(ED1)はED9で判別分析した場合の結果である。提案手法は、従来の単一の編集距離を用いた方法より、判別性能が大幅にアップした。

図6は、決定木(全空間)によるオープンテストの Confusion matrix から算出した翻訳率(或ランク以上の機械訳の割合[6])と、人手で算出した翻訳率との誤差を示す。提案手法による翻訳率の推定誤差は、ベストケースで2~3%に収まった。

4 おわりに

本稿では、ランク付け済み機械訳を複数の編集距離を用いてベクトル表現し、これらベクトルを訓練事例としてランク付け規則を決定木形式に自動学習し、学習した決定木を用いてランク付けを行う自動評価法を提案した。提案手法を旅行予約会話文[7]の翻訳結果を用いて評価したところ、提案手法は、単一の編集距離を用いた従来法より、判別性能が大幅にアップした。また、翻訳率[6]の推定誤差は、2~3%であった。

参考文献

- [1] 大野晋, 浜西正人: “類語新辞典,” 角川書店 (1981).
- [2] 安田志志 他: “対訳コーパスを用いた表層的類似度に基づく翻訳能力自動評価法,” 信学技法, NLC2000-63, pp.97-102 (2000).
- [3] Akiba, Y. et al.: “Using Multiple Edit Distances to Automatically Rank Machine Translation Output,” Proc. MT Summit VIII, to appear (2001).
- [4] Quinlan, J.R.: “C4.5: Programs for Machine Learning,” San Mateo: Morgan Kaufmann (1993).
- [5] Su, K. et al.: “A New Quantitative Quality Measure for Machine Translation Systems,” Proc. COLING'92, pp. 433-439 (1992).
- [6] Sumita, E. et al.: “Solutions to Problems Inherent in Spoken-language Translation: The ATR-MATRIX Approach,” Proc. MT Summit VII, pp.229-235 (1999).
- [7] Takezawa, T.: “Building a Bilingual Travel Conversation Database for Speech Translation Research. In Proceedings of the 2nd International Workshop on East-Asian Language Resources and Evaluation - Oriental COCOSDA Workshop'99 pp.17-20 (1999).
- [8] Thompson, H.S.: “Automatic Evaluation of Translation Quality: Outline of Methodology and Report on Pilot Experiment,” Proc.the Evaluator's Forum pp.215-223 (1991).

²著者らは、類語新辞典の意味コードの利用を許可して頂いた、角川書店に感謝する。