

氏名	あきばやすひろ 秋葉泰弘
学位(専攻分野)	博士(情報学)
学位記番号	情博第138号
学位授与の日付	平成17年3月23日
学位授与の要件	学位規則第4条第1項該当
研究科・専攻	情報学研究科知能情報学専攻
学位論文題目	Automatic Evaluation Methods for Machine Translation Systems (機械翻訳システムの自動評価手法)

論文調査委員 (主査) 教授 奥乃 博 教授 石田 亨 教授 河原達也

### 論 文 内 容 の 要 旨

本論文は、機械翻訳の自動評価に関する研究をまとめたものである。音声翻訳システムや機械翻訳システムの構築や運用時に必要不可欠な翻訳文の評価を自動化する技術の確立に対するアプローチが述べられている。

第1章は序論で、機械翻訳の自動評価について論じ、3つの課題を提示している。第1の課題は、テストセットに対して複数の翻訳システムの翻訳品質をオフラインで相対評価するベンチマーク評価法である。第2の課題は、同一原文に対して複数の翻訳システムから得られる翻訳文同士の翻訳品質をオンラインで相対評価して最良の訳文を選択する最良訳選択法である。第3の課題は、ユーザが満足する翻訳品質を満たさない翻訳文をオンラインで排除する低品質翻訳文排除法である。

第2章では、翻訳システムの研究動向、人手で行なわれている翻訳品質の主観評価、および自動評価の従来研究について概観し、第1章で提示した3つの課題について具体的な問題点を明らかにしている。

第3章から5章までは、3つの課題に対する本研究の詳細について述べている。複数の翻訳システムを用いて、大規模なテストデータにより、本研究の手法の有効性を実験的に示している。

まず第3章では、ベンチマーク評価法について述べている。全く異なるタイプの複数の翻訳システムの翻訳品質を、オフラインでテストセットによる相対評価を行う。予め準備した参照訳(正解訳)と評価対象の翻訳文との編集距離による評点を求める際、複数の編集距離を用いることにより、会話に現れるような短文中の些細な誤りに対して寛容な評価を可能にしている。この結果、複数翻訳システムの翻訳品質の相対評価において、人手による主観評価に基づくものと自動化によるものとの一致率が、従来よりも向上し、ほぼ人手と同等の相対評価が達成できることを実験的に示している。

第4章では、最良訳選択法について述べている。目的言語の統計的言語モデルと統計的翻訳モデルのスコアに基づき、翻訳文同士をオンラインで相対評価する。本課題では、予め参照訳が準備できないので、対訳コーパスから学習を行って作成した統計的言語モデルと統計的翻訳モデルを使用する。さらに、各モデルの統計的な振れによる影響を抑制するために、多重検定を使用する。この結果、従来技術の問題点であったスコアの統計的振れのために難しかった最良訳選択の精度が向上することを実験的に示している。

第5章では、低品質翻訳文排除法について述べている。翻訳文を構成する各単語の事後確率に基づきオンラインで各訳語の信頼度を推定する。複数の翻訳システムから得られるNベストリストを併合し、事後確率を求める。事後確率を求める際、統計的翻訳モデルから得られる各訳語の信頼度を用いることにより、網羅的に翻訳候補を生成しない用例翻訳システムに対しても、低品質翻訳文排除が高精度で行えることを実験的に示している。

第6章は、結論であり、本研究のまとめと今後の展望を述べている。

### 論 文 審 査 の 結 果 の 要 旨

本論文は、機械翻訳の自動評価に関する研究をまとめたものであり、得られた主な成果は次の通りである。

1. 予め準備した参照訳（正訳）に基づいた評価対象の翻訳文への評点付与のために新たな編集距離を考案したこと。  
この編集距離は、従来の編集距離とその拡張の組合せであり、従来法ではうまく評点付与ができなかった会話に現れるような短文中の些細な誤りに対して寛容な評価を可能にしている。この結果、複数翻訳システムの翻訳品質の自動相対評価において、人手による主観評価に基づくものとはほぼ同等の相対評価が達成できることを実験的に示した。
2. 目的言語の統計的言語モデルと統計的翻訳モデルのスコアに基づき、翻訳文同士をオンラインで相対評価する技術を開発したこと。複数の翻訳システムから得られる訳の中からの最良訳選択では、予め参照訳が準備できないので、対訳コーパスから学習を行って作成した統計的言語モデルと統計的翻訳モデルを使用している。従来スコアの統計的な振れのために最良訳が求まらないという問題に対して、両者の統計的モデルを用いた検定を適用し、最良訳選択の高性能化を達成した。
3. 翻訳文を構成する各単語の事後確率に基づきオンラインで各訳語の信頼度を推定する技術を開発したこと。複数の翻訳システムからの N ベスト翻訳結果を併合し、その事後確率を求める際、統計的翻訳モデルから得られる各訳語の信頼度を用いる。この結果、単純な N ベスト併合では網羅的に翻訳候補を生成しない用例翻訳システムでは不良翻訳を排除できなかった問題を提案手法により解決し、高精度で不良品質翻訳文を削除できることを実証した。

以上本論文は、音声翻訳システムや機械翻訳システムの構築や運用時に必要不可欠な、翻訳の評価を自動化する技術確立し、その効果を具体的に示しており、学術上、実際上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成17年1月20日実施した論文内容とそれに関連した試問の結果合格と認めた。