氏名	渡邉太郎（わたなべたろう）
学位の種類	博士（情報学）
学位記番号	情博第111号
学位授与の日付	平成16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	Example-Based Statistical Machine Translation （用例に基づく統計的機械翻訳）
論文調査委員	（主査）教授 奥乃博　教授 松山隆司　助教授 佐藤理史

論文内容の要旨

本論文は，用例に基づく統計的機械翻訳に関する研究をまとめたものである。日英翻訳のように言語構造が大きく異なる言語間の統計的機械翻訳においては，対話コーパスから得られる統計的な対話知識だけでは翻訳性能が出ない問題点に対するアプローチが述べられている。

第1章は序論で，機械翻訳について概観し，従来手法の機械翻訳の問題点を明らかにしている。

第2章では，統計的機械翻訳について詳述し，言語構造が大きく異なる日英翻訳に適用する場合の問題点を整理し，用例に基づく統計的機械翻訳という新たな枠組みを考案している。

用例に基づく統計的機械翻訳では，対訳の対応関係を表現した翻訳モデル，および，単語や句の並び替えの制約を表現した言語モデルを利用し，入力文に対して，両者のモデルの尤度を最大化するようにデコードすることにより翻訳が行われる。第3章から5章までは，本手法の詳細について述べている。

まず第3章では，ビームサーチデコードについて述べている。言語構造が大きく違う日英翻訳において，文頭および文末からの双方探索により，探索エラーは減少するものの，翻訳品質にはほとんど影響しないことを，実験的に示し，用例に基づく統計的機械翻訳において，翻訳モデルと言語モデルの構築，および，デコード技法が重要であることを指摘している。

第4章では，チャンクに基づく翻訳モデルについて述べている。翻訳単位として語の並びであるチャンクを提案し，用例からチャンクを生成し，各チャンク内で翻訳と並び替えを行い，最後にチャンク間の並べ替えを行うという処理方法を構成している。チャンクに基づく翻訳モデルと用例に基づく言語モデルとの共用により，つまり，チャンクという非明示的な制約と用例という明示的な制約とを組み合わせることにより，単語単位の接続確率に基づいた言語モデルを使用した従来手法と比較して，翻訳性能が向上することを，日英翻訳による具体的な実験によって示している。

第5章では，用例検索に基づくデコーダについて述べている。デコーダは，入力文が与えられると，用例検索から得られる対訳を翻訳候補とし，統計的モデルによる目的関数を使用して翻訳候補を修正し，山登り法により最適解を得ている。統計的モデルを使用したビームサーチデコーダと比較して，本手法による翻訳性能の向上を実験的に示している。具体的には，日本語，中国語，韓国語，英語という4つの言語を取り上げ，各組相互間の12通りの翻訳実験を行っており，本手法が多言語翻訳へ適応可能であることを示している。

第6章では，本手法の有効性について述べている。3種類の対話コーパスを用いて，本手法を含めて4種類の翻訳システムの性能比較を行っている。本手法が対話コーパスのサイズが2千程度と小さいときには性能に限界があるものの他の統計的翻訳システムよりも優れており，サイズが大きくなると，人手による翻訳知識を使用したシステムを含めて他のどれよりも優れていることを実験的に示し，本手法の有効性を確認している。

第7章では結論を述べている。

最後に第8章では統計的機械翻訳の今後の展開を述べている。

論文審査の結果の要旨

本論文は，用例に基づく統計的機械翻訳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．日本語と英語という言語構造が大きく異なる言語間の翻訳に統計的機械翻訳を応用するときに従来指摘されていた翻訳モデルの設計問題に対して，単語単位ではなく，単語の並びであるチャンク単位とした翻訳モデルを提案し，実装することにより，従来の単語単位の統計的モデルに基づく統計的機械翻訳と比較して，性能向上を達成した。

2．統計的機械翻訳において従来指摘されていた翻訳候補の膨大な探索空間の問題に対して，用例検索により翻訳候補を抽出し，統計的モデルによる目的関数を使用して翻訳候補を修正し，山登り法により最適解を得るという手法を提案し，実装することにより，従来手法と比較して，簡便で高性能な翻訳システムを実現した。

3．言語構造の大きく異なる多言語間の翻訳システムが実現可能かという問題に対して，167千文からなる対訳コーパスを用いて，日本語，英語，韓国語，中国語の任意の2言語間での翻訳において本手法が有効であることを実証し，統計的機械翻訳が多言語翻訳に適用可能であることを示した。

4．統計的機械翻訳で従来指摘されてきたビーム探索の問題が，探索方向にあるのではなく，翻訳モデルと言語モデルの構築，および，デコード技法が重要であることを実験的に指摘し，それを大規模から小規模までの3つの対話コーパスを用いて，ルールベース翻訳システムや統計的機械翻訳システムを含む3種類のシステムと比較実験を行い，本手法の有効性を実証した。

以上，本論文は，最近注目を浴びている統計的機械翻訳を用いた，言語構造が大きく異なる言語間での翻訳問題に対して，用例に基づく機械翻訳手法を統合することにより，対訳コーパスが大きければルールベース翻訳システムを凌駕する翻訳性能が実現できることを示しており，高度ユーザインタフェースの観点から学術上，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16年2月23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。