

論文内容の要旨

博士論文題目 医療文書の自動点字翻訳における精度向上法

氏名 菅野 亜紀

(論文内容の要旨)

点字は視覚障害者のための触読表音文字であり、主要な情報獲得手段の一つである。医療現場では、近年の個別化医療の進展に伴い、患者個人に対する医療情報の提供が必要になりつつあるが、点字での情報提供への取り組みは進んでいない。従来開発されてきた自動点字翻訳プログラムは、点訳ボランティアの支援という立場で開発されており、しかも、医療現場での使用は考慮されていない。そこで、本研究では、点字翻訳の専門家と同等の高い点訳精度を実現する方法の解明を目的とする。その際、特に医療機関での使用を可能にすることを目指した。

最初に、自動点字翻訳プログラムの評価用コーパスとして、新聞記事を基にした通常文書評価用コーパスと特定機能病院の患者向けの文書を含む医療文書を基にした医療文書評価用コーパスを作成した。同時に評価指標として、分かち書きの精度(F1)、漢字の読みや記号の変換の正解率に加えて、これらの積を点訳精度と定義した。次に、日本点字表記法2001年版に記載された点字表記規則を分析した。加えて、我々の自動点字翻訳プログラム eBraille の点訳精度と他の自動点字翻訳プログラムのそれとを比較した。更に、プログラムの辞書へ医療用語や東洋医学用語を追加した場合の点訳精度、Support Vector Machine に基づく統計的学習モデルを導入した分かち書きについて評価実験を行い、点訳精度の向上に有効な手法の解明に取り組んだ。

日本点字表記法の分析の結果、点字表記規則は、文字種、品詞、活用形、出現形、読み、音韻変化、モーラ(拍)の数を指標として作成する必要があることが明らかとなった。これらを指標とした点字表記規則を作成し eBraille の点訳エンジンに実装した結果、eBraille の点訳精度は他の自動点字翻訳プログラムよりも有意に高かった。また、点訳精度の向上には点訳対象の文書に適した辞書の語彙構成にすることが有効であることが示された。更に、形態素解析結果とモーラの数とを学習素性とした統計的学習モデルと点訳エンジンの分かち書き出力を学習素性に追加した統計的学習モデルにルールベースの分類器を組み合わせた手法が、医療文書の分かち書きへの適用に有用であることが示された。

氏名	菅野 亜紀
----	-------

(論文審査結果の要旨)

平成 22 年 7 月 30 日に開催した公聴会の結果を参考に平成 22 年 9 月 1 日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

菅野 亜紀は、本博士論文において、医療文書の点字翻訳を行うシステムの開発を通じて、以下のようなシステム構築と点訳手法を提案した。また、その有用性について様々な視点から評価を行った。本学論文は以下の点で評価することができる。

1. これまでに開発されてきた医療文書点訳エンジン eBraille0.81 の性能の評価を行い、エラー解析を行うことによって、新たな点訳規則を実装し、これまで開発されてきた他の点訳システムに比べて優れた性能をもつシステムを実装した。
2. 上記システムを実際の医療現場 (大学病院) の外来案内と入院案内の点訳に利用し、評価を行った。このシステムに対して、兵庫県より「ユニバーサルデザイン賞」を受賞している。
3. 点訳の分かち書き精度だけでなく、漢字や記号の読みの精度を含む総合的な評価尺度を提案した。
4. 統計的機械学習に基づく分かち書き手法と規則に基づく eBraille の分かち書き法を統合するためのいくつかの方法を提案した。まず最初に、eBraille の分かち書き結果の情報を統計学習システムの素性に与えることにより、eBraille のもつ分かち書き規則を統計学習システムが間接的に学習できることを示した。また、規則の基づく手法と統計学習に基づく手法に特徴的な違いがあることを示し、両者の出力を適切に選択する分類器を導入することにより、そのいずれよりも高い精度を達成することを示した。

医療文書の点訳システムの開発と自動点訳手法を提案した本研究は、独創性が高く、しかも実用的であり、自然言語処理の分野において高い貢献があると評価する。よって、本論文は、博士 (工学) の学位論文として価値あるものと認める。