

論文内容の要旨

博士論文題目 Studies on improving two fundamental steps for Chinese natural language processing: word segmentation and spelling check
(中国語の言語解析のための二つの基盤的な処理に関する研究：単語分割とスペルチェック)

氏名 Fei Cheng

(論文内容の要旨)

中国語の文は単語の境界に空白を持たない漢字の列として記述されている，したがって，中国語文の単語分割は中国語の自然言語処理の基本的問題と考えられている．一方，中国語のスペルチェックは，中国語文中で人が犯した綴り誤りを検出し訂正するタスクであり，単語分割の前処理と考えられている．中国語の単語分割とスペルチェックは，このように中国語処理の基本的処理であり，品詞タグ付けや統語解析などの中国語自然言語処理に直接影響を及ぼすため，非常に重要なタスクである．

中国語の単語分割には2つの問題が残されている．一つは，複数の単語分割基準が存在し，それぞれに基づいてタグ付けされたコーパスが存在する現状である．二つ目は，中国語の合成語が極めて生産的であるため，未知語の存在を無視できないことである．我々は，両者の問題が語の内部構造に関係することを前提とし，語の内部構造解析を行うための学習データの構築と，内部構造解析のための機械学習に基づく依存構造解析法の提案を行った．これを利用して合成語の内部構造解析を行うことにより，単語分割の性能を向上させることを示した．さらに，単語分割基準の異なる2つのコーパスを単語分割基準がより整合的なコーパスに変換し，両者を組み合わせたコーパスを学習データとすることによって，より整合性の高い単語分割が行えることを示した．単語の内部構造を利用することによって得られるより大きなコーパスを用いることで，それぞれのコーパスにおいて精度の高い単語分割を行うことができることを示した．

中国語のスペル誤りは，単語の分割情報なしには特定することが難しい，我々は，二つのステージからなるスペルチェック法を提案した．最初のステージでは，複数のシステムによって訂正候補を収集し，次のステージにおいて，精度と再現率のバランスを見ながら最も正しいと思われる訂正候補を選択する．

氏 名	Fei Cheng
-----	-----------

(論文審査結果の要旨)

平成 29 年 12 月 22 日に開催した公聴会の結果を参考に平成 30 年 2 月 13 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、研究活動が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。

Fei Cheng は、本博士論文において、中国語の自然言語処理の基盤的な問題である文の単語分割と綴り誤り訂正の問題に取り組み、その両者に対して新たな手法を提案した。また、それぞれについて、評価実験を行い、その有効性を示した。

中国語は、日本語等と同様に、文の記述の際に明示的な単語分割を行わない言語である。特に、中国語では、単語分割基準が複数存在し、それぞれに基づいてタグ付けされたコーパスが存在すること、および、文が一つの種類の文字（漢字）で記述されるため、綴り誤り等の表記誤りの特定が難しいことが問題となっている。

本論文は、これらの問題に取り組み、それぞれの問題に対する解決策を提案した。本論文の貢献は、次のようにまとめることができる。

1. 中国語の単語分割に対して整合的な分割基準を考えるための有効な方法として、中国語の合成語の内部構造を利用することを提案した。
2. 語の内部構造解析を行うために、中国語の合成語の内部構造を基本的な単語間の係り受け構造としてアノテーションした学習データを構築し、内部構造解析のための機械学習に基づく依存構造解析法の提案を行った。
3. 構築した単語係り受け解析器を利用して合成語の内部構造解析を行い、より細かい粒度でかつ整合性の高い単語分割基準に基づくコーパスを作成することにより、単語分割の性能を向上させることを示した。また、単語分割基準の異なる 2 つのコーパスをより整合的なコーパスに変換して両者を組み合わせたコーパスを学習データとすることによって、より整合性の高い単語分割が行えることを示した。
4. 中国語のスペル誤りの検出と訂正をおこなうための二つのステージからなる手法を提案した。まず最初のステージで複数のシステムによって訂正候補を集め、次のステージにおいて正しい候補を選択することによって、誤りの検出と訂正を行う新たな手法を提案した。

中国語自然言語処理の基本問題である単語分割と綴り誤り検出・訂正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。