

論文内容の要旨

博士論文題目 Automatic Generation of Syntactically Well-formed and Semantically Appropriate Paraphrases
(統語的・意味的に適格な言い換えの自動生成)

氏名 藤田 篤

(論文内容の要旨)

言い換えとは、ある言語表現を、意味や内容を保持したまま同一言語の別の表現に変換する処理である。言い換え技術は、機械翻訳、文書要約、質問応答などの多くの自然言語処理アプリケーションに共通のミドルウェアとして近年注目を集めている。本論文では、日本語文を対象とした語彙・構文的言い換えに焦点を当て、統語的・意味的に適格な言い換えを自動生成するための課題に取り組む。

言い換への出力は入力と同じ自然言語表現である。このため、言語的に適格な表現を生成する必要がある。言い換えは同一言語内の変換であるため文中の一部を別の等価な表現に置き換えることで生成できる。しかし、文の一部を機械的に置き換えると、言い換えた箇所と周辺の表現とのつながりが不自然になったり、言い換え前の意味が変わったりする可能性がある。言い換えに用いる変換規則には、その規則を適用できるあらゆる条件を記述することが現実的には不可能だからである。

この問題に対し、我々は文の一部を書き換えて言い換への候補を生成する処理の後処理として、候補生成時に生じる変換誤りを言語学的制約に基づいて排除する機構を提案する。この枠組みでまず、様々な変換誤りが上で例示したような言い換への種類に依らず生じることを示す。変換誤りのうち、形態的誤りは従来の自然言語生成の分野で多くの知見がある。このため我々は統語・意味的な誤りの検出に焦点を当てる。とくに、シソーラスや格フレーム辞書などの既存の言語資源では正しく検出できない、「不適格な動詞格構造」の検出に取り組む。本論文では、コーパスから収集できる大規模な共起用例(正例)における単語の分布に基づく手法と、人手で収集できる小規模な誤りの例(負例)を用いる手法を組み合わせる誤り検出モデルを構築する。また、負例を人手で収集する際の効率化について検討する。

言い換への自動生成においては、いかにして多様な言い換を生成するかも重要な課題である。近年、言い換と見なせるテキスト対から言い換への関係にある表現対を抽出するという研究が盛んである。しかし、このアプローチからは建設的な知見は得られていないし十分な規模のリソースも蓄えられていない。そこで我々は、多くの種類の言い換で必要となる語彙意味的知識を同定するという目的で、動詞の意味表現について検討する。具体的には、語彙概念構造(LCS)が持つ意味レベルの抽象的な情報、および統語構造と意味構造を結び付ける情報に着目する。本論文では、LCSに基づく言い換生成モデルを構築し、意味的構造変換、表層表現生成におけるLCSの振る舞いについて調査する。また、言い換への生成実験を通じて、現在利用可能なLCS辞書/体系の問題点を整理し、既存の表層的なアプローチに対する利点、LCSの言い換への応用可能性を示す。

氏名	藤田 篤
----	------

(論文審査結果の要旨)

平成16年12月27日に開催した公聴会の結果を参考に平成17年2月17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。

現在盛んに研究が行われている言い換え技術において、生成される文が文法的あるいは意味的に適格であることを保証することが重要な問題の一つである。藤田 篤は、本博士論文において、統語的・意味的に適格な文生成を可能にするために次のようないくつかの観点からの手法を提案した。提案された手法は、特定の言い換え技術のみを対象にしておらず、汎用性の高い方法である。

1. 言い換え候補生成の後処理として、変換時に生じる文法的誤りを排除する機構を提案した。それを汎用的な言い換えシステムの中で実現する方法を示した。
2. 統語・意味的な誤りの中でも慣用的な用法に違反する誤りは、荒い意味分類を記述したシソーラスやそれに基づく意味制約を記述した格フレーム辞書などの情報では十分に検出することができない。本論文では、大規模なコーパスから収集可能な具体的用例を正例とし、言い換え出力から生じた誤り事例を負例として用いるそれぞれの学習モデルと、それらを混合した誤り検出モデルを提案した。提案手法は、少数の負例によっても有意な誤り検出性能の向上を達成ことができ、実用的に優れていることを示した。
3. 上記手法において、人手でチェックしなければならない負例の収集に対して、選択的な収集法を適用し、効率よい事例の収集が可能であることと、その効果を示した。
4. 語彙概念構造という語の意味の一般的な表現法に基づき、意味レベルの等価性を保証しつつ、多様な言い換えを系統的に行うことが可能な方法を示した。特に、機能動詞構文をもつ事例の系統的な言い換えを実装した。このような言い換えの生成実験を通じて、現在のLCS体系の問題点を整理と従来の言い換え手法に比した利点を示し、LCSの言い換え技術への応用可能性を明らかにした。

このように、言い換えによって生成される文の統語的・意味的な適格性を保証する種々の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。