

論文内容の要旨

博士論文題目 Hierarchical Word Sequences Structures for Language Modeling
(階層的語序列構造に基づく言語モデリング)

氏名 吳 暁一

(論文内容の要旨)

言語モデリングは自然言語処理の基礎研究であり幅広く多くの技術に応用されている。言語モデルの最も重要な役割は自然言語の文の確率を推定することである。文の確率を推定するため、ほとんどの言語モデリングに関する先行研究は n-gram に基づく手法を用いてる。しかし、n-gram 言語モデルの仮説はロバスト性が不足しており、データスパースネスの問題が深刻である。たとえ大規模コーパスおよび MKN のような高性能なスムージング手法を用いても、3-gram に基づいたものであっても、言語モデルの性能に大きな悪影響を与えている。本論文では、認知文法 (Cognitive Grammar) 構造の基本概念に基づき、階層的語序列 (Hierarchical Word Sequence, HWS) 構造を提案する。この構造に基づき、色々な仮説を用い、特定の言語知識に依存せず、完全に教師なしで語序列を再配列することが可能である。更に、この HWS 構造を構築するため、「仮説指向」、「NST 指向」、「ニューラルネットワーク指向」の三つの異なる手法を提案する。n-gram 言語モデルが文の左側から右側へ確率を推定するのに比べて、HWS はより柔軟的な戦略を用いる。評価として再配列した語序列と伝統的な n-gram 語序列を直接実験と間接実験で比較した結果、提案モデルの方はより高い性能が得られ、データスパースネス問題を軽減すると同時に、機械翻訳など自然言語処理応用にもさらなる精度をもたらすことがわかった。これにより、HWS が n-gram 言語モデルの代用になることを示すことができた。

氏名	呉 暁一
----	------

(論文審査結果の要旨)

平成 28 年 10 月 31 日に開催した公聴会の結果を参考に平成 29 年 5 月 29 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。

呉 暁一は、本博士論文において、従来の n-gram に基づく言語モデルを凌ぐ階層的な言語モデルを提案した。本博士論文の貢献は次のようにまとめることができる。

1. ラネカーの認知文法で提案された構文スキーマに基づく階層構造を大規模の言語データから自動的に構築する手法を提案し、従来の n-gram に基づく言語モデルに比べて、データに対する被覆率やパープレクシティにおいて優れていることを示した。
2. 階層的な言語モデルとして、三つの異なるモデルを提案し、その性質と性能について評価を行った。すべてのモデルは未解析の大規模な言語データの存在だけを仮定しており、自動的にモデルを学習することができる。また、従来手法との比較実験を行い、提案手法の優位性を示した。
3. 提案モデルの応用面での性能を評価するため、統計的機械翻訳の言語モデルとして利用した実験を行い、n-gram に基づく手法に比べてわずかではあるがその優位性を示した。

階層構造を持つ複数の新しい言語モデルを提案し、その優位性を示した本研究は、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。