

## 論文内容の要旨

博士論文題目 Combining Linguistic Knowledge and Machine Learning for Anaphora Resolution  
(照応解析のための言語学的知識と機械学習手法の融合)

氏名 飯田龍

### (論文内容の要旨)

本論文は、機械学習に基づく照応解析に役立つ文脈的な手がかりを導入する試みについて報告している。従来の照応解析の処理は、人手で作成された規則に基づく手法と照応関係がタグ付与されたコーパスを利用した機械学習に基づく手法に分かれ、それぞれ相補的に研究が進められてきた。規則に基づく手法では、センタリング理論など談話研究に基づいた手がかりを人手で規則に導入する試みがなされており、一方、機械学習に基づく手法では、品詞や文字列の情報など主に表層的な手がかりを素性として導入し、規則ベースの手法と同程度の解析精度を得ている。

先行研究の照応解析の問題は、共参照関係との関係でさまざまな問題設定がなされていた。また、日本語を対象とした照応解析についても厳密な問題設定の議論がないまま、いくつもの解析手法が提案されていた。本論文では、規則ベースの手法で主に導入されてきた言語学的な知見を機械学習ベースの手法で効果的に利用できるモデル化を提案した。3章で日本語を対象とした照応関係・共参照関係の仕様を提案し、分析/評価/学習用のタグ付与コーパス作成について説明している。また、タグ付与作業の際に起こった問題点についても議論している。4章では、先行詞の序列を理論的に説明したセンタリング理論の考えを採用し、この序列を学習ベースのモデルで利用する二つの方法について述べている。一つは、先行詞候補が前文脈から得られた forward-looking center(先行詞候補集合を先行詞らしさの序列に基づき並べたリスト)のどの箇所に含まれているかを示す素性を学習手法で直接利用するやり方である。もう一つは、先行詞候補全体の間の序列を二つの候補の間の関係に分解して考え、二つの候補の間の選好を学習し、候補間で勝ち抜き戦を行うことにより、最終的に最尤の先行詞を決定することで先行詞同定を行うモデルの提案である。日本語ゼロ代名詞の先行詞同定の評価実験を通じて、特に後者が既存手法に比べ大幅に精度向上したことを報告している。5章では、与えられた照応詞の候補が真に照応詞となるか否かを判定する照応性判定の問題に取り組んでいる。既存の手法で個別に利用されてきた先行文脈の情報と照応詞の局所的な文脈情報を効果的に組み合わせることにより、日本語名詞句照応解析において先行詞同定と照応性判定のそれぞれにおいて精度向上に貢献することを示した。6章はで、文内ゼロ照応に統語的なパターンを素性として導入する一手法を提案している。文内ゼロ照応では、ゼロ代名詞と先行詞がどのような構造的な位置関係で出現しているかが解析のための大きな手がかりとなる。この考察に基づき、ゼロ代名詞と先行詞の間の統語的なパターンから有効な素性をマイニングし、ゼロ照応解析に利用する手法を提案した。

氏名	飯田龍
----	-----

(論文審査結果の要旨)

平成18年12月27日に開催した公聴会の結果を参考に平成19年2月14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。

飯田龍は、本博士論文において、自然言語文書の理解にとって重要な課題である照応解析に取り組み、様々な有効な手法の提案と新しい知見を報告した。本論文の貢献は、次のようにまとめることができる。

- 規則に基づく照応解析で利用されているセンタリング理論に基づく知見を機械学習の素性として加え、さらに、照応詞の先行詞候補を2つずつ比較することによって最尤の先行詞を得るトーナメントモデルという新しい手法を提案した。
- 与えられた名詞句が照応詞になっているかどうかを判定する照応性判定問題に対して、先行詞候補に関する情報と照応詞候補の局所的な情報の両者を有効に利用する方法を提案し、従来手法を大きく上回る精度を達成した。
- ゼロ代名詞の先行詞同定において、先行詞が文内にある場合と文外にある場合の解析手法を区別し、特に文内照応の解析のための統語構造に基づく構造的な素性の利用を提案し、精度向上に寄与した。
- 一連の研究において作成した照応関係・共参照関係タグ付きコーパスをNAISTテキストコーパスとして公開した。

日本語の照応解析に関して、多くの有効な解析手法と共通のデータを作成した本研究は、手法としての独創性が高く、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。