

## 論文内容の要旨

博士論文題目 A Machine Learning Approach to Natural Language Processing  
(自然言語処理への機械学習によるアプローチ)

氏名 春野 雅彦

(論文内容の要旨)

自然言語処理システムで必要とされる知識の種類と量は膨大である。したがって、実用的システムの開発では多大な人的資源を必要とするだけでなく、プログラムの再利用性等にも問題があった。

本研究では、この問題を解決するため近年利用可能になった言語コーパスを利用して知識を自動的に獲得する機械学習による自然言語処理のアプローチについて報告する。主に、日英対訳コーパスからの教師なし学習 (2,3 章) と、タグ付きコーパスを用いた形態素解析、係受け構文解析の教師付き学習 (4,5 章) を取り扱う。これらは日本語に関する自然言語処理において最も基本的で重要な課題であるためである。

2 章では文の対応が付与されていない対訳コーパスに対して文体応付けを行う。統計的に得られる訳語対と辞書に含まれる訳語対を同時に利用することで文体応付けに必要な情報を得る手法を提案する。実験の結果、様々なジャンル、長さの日英対訳コーパスに対して既存の手法を大きく上回る文体応付け精度が得られた。合わせて、2 章ではフラフィカルインタフェースによって統合的に文体応付けを行う環境 BACCS についても報告する。

3 章では 2 章で文体応付けの行われた対訳コーパスから単語やフレーズの対応 (対訳コロケーション) を取り出す手法を提案する。提案手法はコロケーションの種類を 2 つに分類し、ソートをを用いた高速な学習手法を用いることを特徴とする。株式速報文、日経サイエンスを用いた実験では 8 割を越える精度で分野固有のコロケーションが抽出された。

4 章と 5 章ではそれぞれ形態素解析と係受け構文解析を取り扱う。各課題の表現には効率的な木構造 (階層的な文脈木と決定木) を用いた。両課題において、自然言語に特徴的な頻度の少ない例外的事例に対応するため、誤り駆動の学習法を導入し、複数の木構造を混合して確率予測を行った。実験の結果、既存の統計的手法、人手作成のシステムを大きく上回る解析精度が得られた。更に 5 章では人間による特徴抽出と機械学習の融合が今後の言語工学にとって有望なアプローチであることを具体的に示した。

6 章では複数の辞書、コーパス、自動抽出したコロケーションなどを統合し有機的に利用可能としたシステム AIDA について報告する。AIDA は適応的で柔軟な検索機能の特徴とする。すなわちユーザの入力に意味的、構文的に近い表現をコーパスから抽出する。AIDA は類似度の計算に利用される各属性の重要度もユーザとのインタラクションによって獲得する。

本研究によって機械学習に基づく自然言語処理のアプローチは、殆ど人間の恣意的介入なしに、実用的なレベルの高い精度を達成出来ることが明らかとなった。これらの特徴は、機械学習に基づくアプローチを大規模な実用システムの開発に適用することでより顕著となるであろう。

|    |       |
|----|-------|
| 氏名 | 春野 雅彦 |
|----|-------|

(論文審査結果の要旨)

平成10年7月23日に開催した公聴会の結果を参考に平成10年9月3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。

春野雅彦は、本博士論文において、自然言語処理における様々な問題に対して機械学習手法を適用し顕著な結果を得た。学習の方法は、問題に対する正解例の存在を仮定しない教師なし学習と、問題に対する正解例の存在を仮定する教師付き学習にわけることができる。前者では、日英対訳コーパスの文対応の自動化および単語やフレーズの自動対応付けを行った。また、後者については、日本語の形態素解析および係り受け解析システムのための規則学習を行い、高い精度の解析システムの実現が可能であることを示した。

文対応の自動化については、対訳データから統計的に得られる対訳対と既存の辞書の対訳対を同時に使用することにより、従来手法を上回る精度を達成できることを示した。

対訳コーパスからの単語あるいはフレーズ対応の自動抽出については、固定的な表現、および、非連続的な表現を段階的に抽出する方法を提案し、辞書には記述されていない分野固有の対訳表現を抽出することに成功した。

形態素解析、係り受け解析については、それぞれの問題に適切な学習モデルを提示するとともに、誤り駆動によって複数の学習結果を混合する手法を適用した。その結果、従来統計的手法や人手によって構築されたシステムを上回る精度を達成した。機械学習が実用的な問題に対して有効に働くことを示しており、貴重な成果であると考えられる。

また、辞書、コーパス、および、コーパスから抽出された固有表現を有機的に統合し利用するための環境を構築した。提案者が実現したシステムは、個々のシステムとしても利用可能であるが、それらの具体的な利用形態を示すことでその有用性を確認しており、意義があると考えられる。

機械学習の手法を実用的な自然言語処理に適用した上記の研究は、そのどれもが独創性高く、しかも実用的であることを示しており、自然言語処理および機械学習の分野において極めて高い貢献があると評価する。