

論文内容の要旨

博士論文題目 Discriminative Learning Methods for Interdependent Decision Problems in Natural Language Processing
(自然言語処理における相互依存予測問題に対する識別学習法の研究)

氏名 賀沢秀人

(論文内容の要旨)

自然言語処理の多くの問題では、入出力変数が複数存在し、かつ、それらが相互に強く依存しあうことが多い。本論文では、このような問題を**相互依存予測問題**と呼ぶ。相互依存予測問題においては、出力を予測する際に他の入出力の情報を利用することで、精度良い予測が可能になると考えられる。本論文では、自然言語処理における相互依存予測問題のうち、以下の三つの問題を取り上げ、識別学習の枠組みでの解法を探求している。

文選択問題とは、与えられた文章 (= 文の集合) から、ある基準に基づいて文を選択する問題である。本論文では、文選択問題が分類問題に帰着可能なことを証明したのち、Selection SVM と呼ぶ選択学習法を提案した。人工データと実際の文選択データを用いて、既存の分類学習法および優先度学習法と Selection SVM を実験的に比較した結果についても報告している。

多重トピック文書分類、与えられた文書に該当するトピックを指定されたトピック群のなかから、過不足無く選び出す問題である。一般に、トピックには、共起しやすいものや逆に共起しにくいものなど、様々な相関が存在し、多重トピック文書分類は相互依存予測問題の一種と考えられる。本論文では、トピックの組合せ一つ一つを独立したクラスとして分類学習を行う Maximal Margin Labeling (MML) と呼ぶ学習法を提案している。MML においてはクラス数が膨大になるという問題があるため、それに対処するための近似法および高速な実装法を示している。さらに、Web から収集した文書データを用いて、既存の分類学習法と実験的に比較した結果についても報告している。

配列タグづけ問題は、一般に、隣接するタグの間に強い相関が存在するため、典型的な相互依存予測問題の一つとみなせる。本論文では、最適なタグづけ順を多項式時間で計算するアルゴリズムを説明したのち、実際のタグづけデータを用いた実験により、タグづけ順がタグづけ精度に大きく影響することを示している。最適なタグづけ順を予測する戦略関数を学習する手法を提案し、学習を行わない場合と実験的に比較した結果を報告している。

氏名	賀沢秀人
----	------

(論文審査結果の要旨)

平成18年7月28日に開催した公聴会の結果を参考に平成18年9月4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。

賀沢秀人は、本博士論文において、自然言語処理の中で特に複数の互いに影響しあう出力をもつ問題を相互依存予測問題と定義し、識別学習法に基づく研究を行った。性質の異なる3つの問題を取り上げ、それらを解決する手法を提案し、これまでの研究との比較を行い、実験によって提案手法の利点を示した。問題の定式化、および、それぞれの問題の解決法には高い新規性が認められる。具体的な問題およびそのための提案手法の概要は次の通りである。

1. 文選択問題を分類問題として定式化し、Selection SVM と呼ぶ選択学習法を提案した。人工データと実際の文選択データを用いて、既存の分類学習法および優先度学習法と Selection SVM を実験的に比較し、その優位性を示した。
2. 多重トピック文書分類を、これまでの考え方とは異なり、文書がもつ多重のトピックがそれぞれ正あるいは負の依存関係を有することを指摘し、トピックの組合せ一つ一つを独立したクラスとして分類学習を行う Maximal Margin Labeling (MML) と呼ぶ新しい学習法を提案した。また、MML においてはクラス数が膨大になるという問題に対処するため、近似解法および高速な実装法を提案した。Web から収集した文書データを用いて、既存の分類学習法と実験的に比較し、その優位性を示した。
3. 配列タグづけ問題は、一般に隣接するタグの間に強い相関があり、典型的な相互依存予測問題である。本論文では、最適なタグづけ順を多項式時間で計算するアルゴリズムを提案した。そして、最適なタグづけ順を予測する戦略関数を学習する手法を提案し、従来の一定の順序に従う方法と比較を実験的にを行い、その有用性を示した。

自然言語処理において特に困難な予測問題を相互依存予測問題として定式化するとともに有効な学習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。