

審査の結果の要旨

氏名 任 勇

本論文は「Minimally Supervised Approaches to Emotion Classification Using Unlabeled Data (ラベルなしデータを用いた情緒分類への弱教師ありアプローチ)」と題し、英文5章から構成されている。本研究は、テキスト内容からその書き手の情緒を推定する情緒分類(Emotion Classification)タスクにおいて、ラベル無しデータを利用した学習方法を提案することによって、十分な量のラベル付きデータが入手困難な場合であっても高精度な分類器の構築を可能としている。

第一章は「Introduction (序章)」である。情緒分類がCGM(Consumer Generated Media)解析のための技術として大きな注目を集めている一方、中国語やヒンディー語など、ラベル付きデータを大量に入手することが容易でない言語においては、精度の高い処理を実現することが困難であることが、研究動機として論じられている。これに加えて、本研究における学術的な貢献点と、本論文の構成がまとめられている。

第二章は「Problem Definition and Related Work (問題定義と関連研究)」と題し、情緒分類というタスクの定式化を行い、関連する研究を概観している。

第三章は「Sentiment Polarity Classification Using Graph-based SSL (グラフに基づく半教師有り学習を用いた感情極性分類)」と題し、グラフに基づく半教師有り学習アルゴリズムを基礎として、少数のラベル付きデータと大量のラベル無しデータから、高精度な極性分類器を学習する新しい手法を提案している。ラベル伝播等のグラフに基づく半教師有り学習アルゴリズムは、ラベル無しデータを活用するための方法として従来から広く知られている。しかし、データ群に潜在するグラフ構造をどのように発見するか、アルゴリズムをどのように初期化するかという点に関しては、従来研究ではアドホックで発見的な手法が用いられるに止まっており、汎用性に問題が残されていた。これに対して本章では、グラフノード間の類似度(コサイン類似度やJaccard係数など)に基づいてエッジの取捨選択を行い、グラフ構造を自動的に決定する手法を提案している。さらに、ノードの次数や有名度(Page Rankアルゴリズムを用いて計算される)を利用することによって、グラフに基づく半教師有り学習アルゴリズムの初期化を行う手法を提案している。実験では、3種類の大規模な中国語データを用いて手法の有効性を検証しており、精度の大幅な改善が実現可能であることが示されている。

第四章は「Horrible Stories Identification Using Word Embedding (単語埋め込みを用いた怪談の同定)」と題し、読み手に「恐ろしい」という感情を喚起させるようなテキストの同定というタスクを題材として、少数のラベル付きデータからでも効果的に特徴量を抽出可能とする方法を提案している。単語という離散データを、数十～数百次

元から成る連続値ベクトルに変換する新しい手法を考案し、この技術を応用することによって、表層が異なる単語間の意味的な近接性を捉えることができるような特徴量を、少数のラベル付きデータからでも抽出することを可能にしている。中国語のCGMデータ (hint fiction) を用いた実験を行い、提案手法の有効性を実証している。

第五章は「Conclusion (結言)」であり、本研究の成果と今後の研究方向性について総括を行っている。

以上これを要するに、本論文は、英語などの一部の言語を除き多くの言語において大規模な言語資源を入手することが実際上困難となっている現状において情緒分類処理の精度を向上させるため、ラベル無しデータを利用した学習方法を提案することにより、十分な量のラベル付きデータが入手困難な場合であっても高精度な分類器の構築を可能とし、大規模な実験によってその実用性と有効性を明らかにしており、電子情報学上貢献するところが少なくない。

よって本論文は博士 (情報理工学) の学位請求論文として合格と認められる。