

学位論文及び審査結果の要旨

横浜国立大学

氏名	陸 宇傑
学位の種類	博士(工学)
学位記番号	環情博甲第1938号
学位授与年月日	平成29年9月15日
学位授与の根拠	学位規則(昭和28年4月1日文部省令第9号)第4条第1項及び 横浜国立大学学位規則第5条第1項
研究科(学府)・専攻名	環境情報学府 情報メディア環境学専攻
学位論文題目	Social Media Sentiment Analysis: Multilingual Methodology and Monolingual Application (ソーシャルメディアにおける感情分析: 多言語の方法論と単言語の 応用)
論文審査委員	主査 横浜国立大学 教授 森 辰則 横浜国立大学 教授 田村 直良 横浜国立大学 教授 長尾 智晴 横浜国立大学 准教授 富井 尚志 横浜国立大学 講師 白川 真一

論文及び審査結果の要旨

近年、Webを介した利用者の情報流通が盛んになっている。それに伴い、ソーシャルメディアに代表されるCGM(Consumer Generated Media)が広く利用されるようになってきている。これらのWeb上のCGMから、肯定的・否定的といった感情情報を抽出し、収集することが出来れば、利用者がある対象に対する世の中の評価を容易に知ることができ、商品購入等際に意志決定支援が可能になる。また、製品やサービスを提供する企業にとっても商品開発や企業活動などに消費者の声を反映させることが出来る。そのため、文書中から感情情報を抽出し、解析する試みが行われている。

本論文は、ソーシャルメディア、特に、マイクロブログの一つであるTweetのテキストを対象とし、テキストの極性(「肯定的」もしくは「否定的」な意見を述べているのか、あるいは、それらのいずれでもない「中立的」なのか)を自動的に判断する感情分析に関する多言語対応技術と単言語の応用に関する研究をまとめており、序論(第1章)と結論(第7章)を含めて7章から構成されている。第2章では従来研究を紹介するとともに、本論文の位置づけを述べている。第3章と第4章は複数言語のソーシャルメディアを対象として感情分析用の情報を付与した例文集(コーパス)を構築する手法について、第5章は多言語ソーシャルメディアに対応した感情分析手法について、第6章は単言語ソーシャルメディアに対する感情分析結果に基づく「日々のムード」と株価指数との間の相関について具体的な研究を述べている。

第2章においては、テキストに対する感情分析とそれに纏わる技術について関連する研究を述べるとともに、本論文におけるアプローチの仕方を説明し、本論文の貢献について述べている。

第3章では、ソーシャルメディアにおける多言語感情分析を行うための基礎資源となる注釈付きのコーパスを構築するための手法について述べている。既存のコーパスでは扱われていなかったが、感情分析に影響を与える可能性のある要素として、皮肉、反語といった修辞表現に着目し、それらを注釈付けする手法について提案している。さらに、コーパスを作成する際に問題となる複数の注釈者間の判断の「ゆれ」を抑制するために、判断が割れた例文に対して監督者が最終判断を行いピボットデータとして蓄積していく手法を提案しており、その有効性を実験により示している。

第4章では、第3章で構築されたコーパスに対して、基礎的な統計量を分析するとともに、修辞表現とテキストの極性ととの間に相関があることを示している。さらに、ある製品等への一般的な評判を判断するために、ソーシャルメディアのテキストの集合体に対する極性分析を行う必要があるが、その際に、各文の極性判定を集積した値を計算する代わりに、それと高い相関があり、より簡単に抽出可能な手がかりから構成可能な指標を計算する手法を提案し、必ずしもテキスト中の一文一文に対する極性分析を積み重ねていく必要がないことを実験的に明らかにしている。

第5章では、多言語で記されたソーシャルメディアのテキストに対して、機械翻訳手法を用いずに、統一した方法論で感情分析を行う手法について述べている。従来手法では、各言語のテキストを機械翻訳により英語のような分析のための資源が充実している言語に翻訳し、翻訳結果に対して感情分析を行っていたが、機械翻訳のような大規模なシステムを、対象とする言語毎に用意しなければならないという問題点があった。これに対して、提案手法では、深層学習の手法により、各言語の語を同一のベクトル空間に符号化し、符号化された語から構築される文の情報に対して、極性を予測する分類器を作成するという枠組みにより、各言語の単言語コーパスだけを準備すればよいことを示している。

第6章では、感情分析の応用として、Twitter でつぶやかれたある事物に対する記述の感情分析結果に基づく「日々のムード」とそれに関連する株価指数との間の相関について論じている。先行研究では、テキストの感情分析結果だけに依存しているが、本研究では、それに加えて、リツイート数、発信者のフォロワー数といった非テキスト情報を組み合わせることにより新しい「日々のムード」を表す指数を提案しており、その有効性を論じている。

最後の第7章では結論として本研究の成果を総括的に述べ、今後の課題を示している。

以上のように、本論文は、ソーシャルメディアの一つである Tweet のテキストを対象とした感情分析に関する多言語対応技術と単言語の応用について有用な知見をまとめた論文であり、自然言語処理分野の研究に大きく貢献するものである。

本論文の研究内容は、1編の査読付論文誌論文、1編の査読付国際会議論文(正論文相当)、ならびに、4編の大会・研究会論文により公表されており、国内外の学会からも高い評価を得ている。

よって、本論文は博士(工学)の学位論文として十分な価値を有すると審査委員全員が一致して判断した。

平成29年7月12日(水)、総合研究棟 S113 室において、14時30分から16時00分まで博士論文公聴会を行い、その後、同室において審査委員全員出席のもと、審査対象者・陸宇傑氏の最終試験を行った。

始めに審査対象者より博士論文の概要について説明があり、続いて博士論文に関する学術論文の公表状況、大学院講義の単位取得状況などについて説明があった。

これに引き続き、学力試験として、博士論文の内容を中心とする専門分野および情報関連分野における口頭試問を行い、これらの分野の研究に関する深い専門知識と理解力、表現力、および質疑応答における適切な対応能力を同氏が有することを確認した。博士論文に関する公表論文については、査読付論文誌論文、査読付国際会議論文(正論文相当)が各々1編採録されていることから、博士号取得基準を満たしていることを確認した。外国語については、主著の査読付国際会議論文が採択されており、同会議で本人が口頭発表を行っていることをもって学力を確認した。また、博士課程後期修了に必要な単位をすべて修得していることを確認した。

以上により、審査対象者・陸氏は最終試験に合格であると、審査委員全員が一致して判断した。

この審査結果に基づき、環境情報学府情報メディア環境学専攻会議にて審議をし、本論文に博士(工学)の学位論文としての価値があるものとして、環境情報学府教授会に付議することを決定した。その後、平成29年9月4日(月)に開催された環境情報学府教授会において審議並びに無記名投票により、審査対象者・陸宇傑氏に博士(工学)の学位を授与することを決定した。

注 論文及び審査結果の要旨欄に不足が生じる場合には、同欄の様式に準じ裏面又は別紙によること。