

審査の結果の要旨

氏 名 モハンマド アシフ ホッセン カン

本論文は「**Identifying and Summarizing Public Opinion to Real-time Events Using Microblogs as Social Sensors**（ソーシャルセンサとしてのマイクロブログを利用したリアルタイムイベントへの世論の識別と要約）」と題し、代表的なマイクロブログである **Twitter** を対象として、大規模イベント・災害・選挙のようなリアルタイムの社会現象に対しての時々刻々変動していく世論を識別する手法を提案したものであり、全五章から構成されている。

第一章は「**Introduction（序論）**」であり、マイクロブログやソーシャルネットワークの出現により、加工されていない生の世論を解析することが可能になり **Computational Social Science** の分野で注目されていること、一方でマイクロブログによる世間識別の実用化に向けては多数の技術的課題が山積していることを述べた上で、既存手法を整理すると共に、本論文で新たに提案するリアルタイムイベントへの世論の識別手法について概観している。

第二章は「**Classifying Tweets under Limited or Noisy Training Instances** (制限のある訓練事例によるツイートのクラス分類)」と題し、**Bag of Words** 分類器を用いてツイートのクラス分類を行う手法を提案している。**Twitter** 特有の問題として、文字数が 140 に制限されていることから、非常に多数の略語が用いられかつこれらが全く異なる意味で使用される場合が多いことや、前後の文脈をフォロワーが理解しているという前提でのツイートが多く、単純に **Bag of Words** 分類を行っただけでは、複数のクラスに適合してしまう場合が多いことが挙げられる。ここではこれらの問題を回避するため、分類の前処理として、まず発生頻度が高くても特徴抽出の観点からは意味のない連語を **stop words** として取り除く処理、及び曖昧なツイートを取り除く処理を施した上で、**Unigram** と **Bigram** を考慮した **Bag of Words** 分類を行う手法を提案した。また提案手法を用いて、「風邪の罹患」に関連するツイートを抽出する評価実験を行い、**recall**, **precision**, **F-measure** 共、従来手法に比べ性能が改善されることを確認した。更にハッシュタグが付与されていないツイートを、特定の話題を示すハッシュタグ群に分類する評価実験を行い、従来手法に比べ性能が改善されると共に、これが推奨ハッシュタグを提示する方式としても有効であることを示した。

第三章は「**Delineating Real-time Events using Relevant Tweets** (関連ツイートを用いたリアルタイムイベントの記述)」と題し、災害・スポーツイベント・選挙のような特定の社会現象に関連したツイートのみを、教師なしで自動的に選択する手法を提案している。ここでは、**LDA** とそのクラスタ数を決定する既存のアルゴリズムを用いてツイートをクラスタに分類した上で後、議論の焦点と流れを抽出しつつ、類似・重複したツイートを排除するための手法を開発している。また 2012 に行われた米大統領選討論会の時に投稿されたツイートをを用いて議論の焦点の抽出をすすめる実験を行うと共に議論の流れを可視化するソフトウェアを開発した。その結果 81.6%の **precision** と 80%程度の **recall** 率を得ることが出来、提案手法が大規模の実時間イベントの流れを有効に抽出できることを確認した。

第四章は「Summarizing Event-Related Tweets by Jointly Optimizing Multiple Objectives(特定イベントに関連したツイートの複数指標に基づく最適化)」と題し、第三章と同様、特定の社会現象に関連したツイートのみを、教師なしで自動的に選択する手法の提案を行っている。ここでは第三章のように、ツイートをクラスタに分類した後でツイートの抽出を行うのではなく、ツイートのサマリーから得られる内容の最大化、冗長なツイートの最小化、トピックの広がり最大化を同時に行うことを定式化した上で、ラングランジェ緩和法と双対分解を用いてこの問題を解いている。また、評価実験によりこの手法が三章で提案した手法及び既存手法である Tf-Idf 法や Tf-Idf-Cosine 法よりも良好な結果となることを示している。

第五章は「Conclusion & Future Works (結論と今後の課題)」であり、論文の成果と今後の展開をまとめている。

以上これを要するに、本論文は代表的なマイクロブログである Twitter を対象として、大規模イベント・災害・選挙のような社会現象に対して時々刻々変動する世論を識別する手法を、教師ありと教師なしの両方の場合に対して提案したものであり、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）の学位論文として合格と認められる。