

平成 31 年 2 月 18 日

豊橋技術科学大学長 殿

情報・知能工学 専攻
学位審査委員会
委員長 梅村 恭司

論文審査及び最終試験の結果報告

このことについて、博士学位論文審査を実施し、下記の結果を得ましたので報告いたします。

学位申請者	Abu Nowshed Chy		学籍番号	第 145309 号
申請学位	博士（工学）	専攻名	大学院工学研究科博士後期課程 情報・知能工学専攻	
博士学位論文名	Exploiting Temporal and Semantic Information for Microblog Retrieval through Query Expansion and Reranking Approaches (クエリの拡張と再順位付けアプローチによるマイクロブログ検索のための時間的および意味的情報の活用)			
論文審査の期間	平成 31 年 1 月 17 日 ~ 平成 31 年 2 月 18 日			
公開審査会の日	平成 31 年 2 月 7 日	最終試験の実施日	平成31年2月7日	
論文審査の結果*	合格		最終試験の結果*	合格
審査委員会(学位規程第6条)				
学位申請者にかかる博士学位論文について、論文審査、公開審査会及び最終試験を行い、別紙論文内容の要旨及び審査結果の要旨のとおり確認したので、学位審査委員会に報告します。				
委員長	梅村 恭司			
委員	秋葉 友良		青野 雅樹	

※論文審査の結果及び最終試験の結果は「合格」又は「不合格」の評語で記入すること。

論文内容の要旨

本論文は、ツイートに代表されるマイクロブログに対して、高性能な検索手法を提案したものである。ツイートは、政治家の政策を語る場であったり、アーティストのコンサート告知であったり、多種多様な用途で、世界中で使用されている簡便な SNS のひとつである。しかし、長文の多い通常のブログやニュース記事と異なり、一度の「つぶやき」が短文で、ハッシュタグや URL、省略語が多いこと、また即時性を有することなど、独自のテキスト文化を形作っている。ツイートに含まれる有用で貴重な情報を抽出するためには、多くの挑戦的な問題を含有している。こういった問題に対処できる技術を幾つかの側面で提案し、実証実験したのが本論文である。

論文の構成は第 1 章で、対象とする問題の全体像を述べ、第 2 章で問題解決のために提案する技術を理解するためのテキストマイニング、情報検索、深層学習等の背景知識をまとめている。第 3 章では、世界最先端の関連技術を述べている。第 4 章では、マイクロブログ検索のための新しい提案である各種の特徴量、特に即時性を捉える特徴量、セマンティックな言語モデル、ならびにそれらを用いた検索後のリランキング（再順位付け）手法を提案している。第 5 章では、検索意図に相当するトピックに着目した検索質問の拡張方法として、ユーザの検索意図と検索質問の間のセマンティックギャップを埋めるフレームワークを示している。特に従来の擬似適応フィードバック手法と異なり、深層学習を導入し、候補ツイートを時間的関連度という新尺度のもとで選択する技術を述べている。第 6 章では、本論文を総括し、今後の展望について述べている。

審査結果の要旨

SNS の中でもマイクロブログとしての Twitter は、世界で 3 億人以上のアクティブなユーザが利用し、政治的なメッセージから、様々な出来事に関するつぶやきまで、多くのユーザが存在している。各ツイートは、主として短い文章からなるが、そのメッセージ性は、世界のトレンドを把握する重要な情報源である。この背景から、米国政府機関が支援する TREC（国際的なテキスト検索コンテスト）では、2011 年以降 Twitter を利用したコンテストが開催されている。

本論文では、TREC で提供されたデータを使って、Twitter に代表されるマイクロブログに対する検索問題に焦点をあてている。マイクロブログ検索では、通常の文書検索と大きく異なる幾つかの挑戦的な課題を内包している。マイクロブログが非常に短い文章であること、マイクロブログ特有の俗語、省略語、ハッシュタグ、URL などが多いこと、同じ対象でも多数の別表現（語彙ミスマッチ）が存在すること、リアルタイム性が強く特定の注目されるトピックで多数の投稿が集中することなど独特の文化があり、これらを考慮した検索手法が必須となる。

本論文では、検索技術の中でも重要な 2 つの技法に焦点をあてている、一つは、検索結果のリランキング（再順位付け）手法であり、もう一つは検索質問拡張技術である。

第 4 章では、マイクロブログの再順位付け技術に焦点をあて、20 以上の独自の特徴量を定義している。特にリアルタイム性に着目し、「最新を表現するスコア」と「時系列データとしてのバースト性」という新特徴量を導入し、これらの特徴量を用いた再順位付け手法で、TREC で発表された従来の検索手法よりも優れた検索精度を達成している。

第 5 章では、検索質問拡張に関する新技术を述べている。具体的には、検索で常に問題となるユーザの検索意図と、実際の検索時にクエリとして入力される語句との時間的・意味的なギャップを埋めるため、「時間的関連度」、「セマンティック関連度」等の新特徴量を導入している。また、検索時におけるトピックに着目し、一定時間内に集中して投稿されるツイートを選択するために、C-LSTM と呼ばれる深層学習手法を適用し、効果的な検索質問拡張に成功している。結果として、TREC データにおいて、世界最高性能の検索精度を達成している。

上述のように、マイクロブログを対象とした情報検索において世界最先端技術を開拓した点と、当該分野の技術発展に貢献した点は高く評価できる。以上により、本論文は博士（工学）の学位論文に相当するものと判定した。