

「ソーシャルサイトから情報収集と可視化、話題の自動抽出に関する研究」

研究代表者氏名 ダビド・ラマムジスア（ソフトウェア情報学科、准教授）

＜要旨＞

本研究では、ニュースウェブサイトが提供しているソーシャルデータの解析を行っていた。ソーシャルデータというのはニュース記事に関する内容が読者のフィードバックとコメントなどを想定している。しかしながら、ソーシャルデータにはそれ以外の内容も色々な目的（広告、スパム、サイバーいじめなど）たくさん含めている。コメントのフィルタリング、情報収集、コメント群の可視化、話題の自動抽出の研究が不可欠となっている。自然言語処理、機械学習、データ解析技術などを用いたソーシャルデータの解析フレームワークを提案した。ニュースウェブサイト以外のソーシャルサイトにも提案したフレームワークを再利用と一般化の取り組みをしている。

1 研究の概要

本研究は、ソーシャルサイトから提供されたソーシャルデータの解析を行っている。特にニュース記事に関する読者のコメント群が本研究のソーシャルデータとなっている。本研究の目的は記者や読者のソーシャルデータ解析の支援システムの提案である。どのように面白いコメントや重要なコメントを自動的に抽出できるのの研究に取り組んだ。

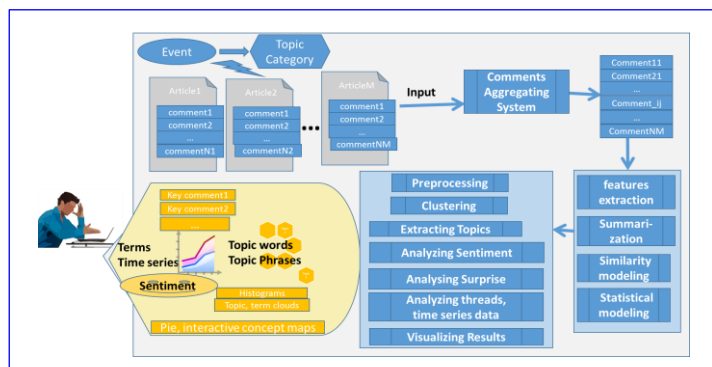


図1：ソーシャルデータ解析のフレームワーク

2 研究の内容

本研究では、ニュースウェブサイトが提供しているソーシャルデータの解析を行っていた。ソーシャルデータというのはニュース記事に関する内容が読者のフィードバックとコメントなどを想定している。しかしながら、ソーシャルデータにはそれ以外の内容も色々な目的（広告、スパム、サイバーいじめなど）たくさん含めている。コメントのフィルタリング、情報収集、コメント群の可視化、話題の自動抽出の研究が不可欠となっている。自然言語処理、機械学習、データ解析技術などを用いたソーシャルデータの解析フレームワーク（図1）を提案した。

3 これまで得られた研究の成果

本研究においては、ニュースウェブサイトが提供しているソーシャルデータの解析を行っていた。ガーディアンニュースウェブサイトで「マレーシア航空 370 便墜落事故」というイベントについての記事群とソーシャルデータを解析し、面白いコメントと記事などの推薦システムの数学モデルを実験した。図2のような読者間とコメントのネットワークが得られた。

4 今後の具体的な展開

提案したフレームワークのモジュールを充実するために様々な手法を検討し、数学モデル、情報システムモデルなどの最新研究を参考し、システムの改善を行う。

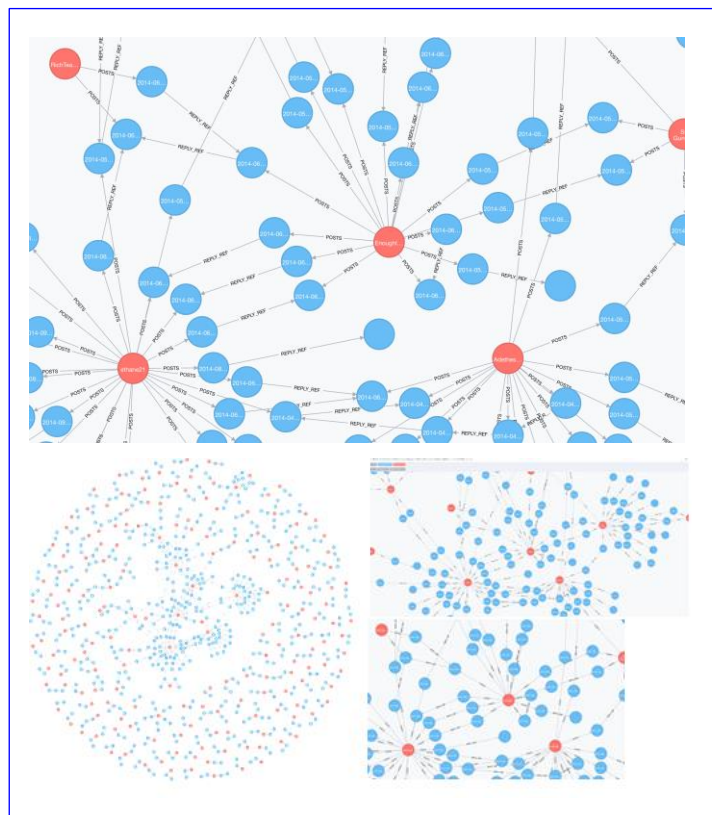


図2：ソーシャルデータ解析のネットワークの例