

マイクロブログにおけるリツイート行動の要因分析 ～世界陸上と甲子園を題材に～

小林 竜也

要旨

近年、知名度が高まりユーザー数が増大してきているマイクロブログ Twitter を題材としてデータを扱う研究が盛んに行われている。Twitter (ツイッター) とは、ツイッター社によって提供されている無料情報サービスである。mixi や facebook 等の SNS とは違い、ミニブログ、マイクロブログ、と言ったジャンルに分類される。ツイートと呼ばれる 140 文字の制限がある短文を投稿する手軽かつ簡単なリアルタイムでツイートが流れていくブログであるため、利用者が増大している。最近のテレビ番組では、ニュース番組やバラエティ番組で、該当番組に対するツイートを公開したり、それを話題としてトークをするような形の利用もされている。NHK では 1 日のつぶやきビッグデータと称し、前日に比べて発言回数の多い単語をピックアップしている。本研究は、関連研究として、Nasir Naveed らの研究 Bad News Travel Fast: A Content-based Analysis of Interestingness on Twitter を参考としている。この論文では英語の Tweet のコンテンツに着目し、ReTweet される際に関わる要因の分析を行っている。本研究でもこれに倣い、日本語で書かれたスポーツ大会に関するツイートを対象に、ReTweet の要因分析を行う。その際、関連研究同様、コンテンツのみに着目したため投稿日時やフォロワー数などを無視している。分析手法として、回帰分析、決定木、回帰木、傾向スコアを用いた。ReTweet されるか否か、ReTweet の回数について、各要素がどれ程の働きをするのかを分析した。調べた要素は頻出語、URL、ユーザーネーム、顔文字、感情語、陸上関連用語、野球関連用語である。各要素は出現したか否かのバイナリで調べている。対象としたデータは世界陸上モスクワ 2013、また第 95 回夏の甲子園大会に関する Tweet を利用した。分析手法によって多少異なる結果が得られたが、共通して、専門用語や感情語よりも、頻出語の方が正負、特に正の働きをすることが多い、ということが分かった。