

マイクロブログにおけるトピック出現量の時系列変化の抽出に関する 研究

Study on extraction of time series variation for topic popularity in microblogs

学籍番号：201721691

氏名：福山 怜史

Fukuyama Satoshi

近年、Twitterは社会で流行している話題をリアルタイムに反映するメディアとして注目されている。特に、関係するツイート出現量が時間経過によって急上昇するような話題は、他の多くのメディアにおいても注目されている。一方で、Twitterでは統一されたタグによってツイートが分類されていないため、Twitterで特定の話題の時系列的な変化を網羅的に観測することは困難であると考えられる。この問題を解決するため、文書が潜在的に有するトピックを推定する手法であるLatent Dirichlet allocation (LDA)を利用する手法が提案されている。しかし、LDAはツイートのような文書に含まれる単語が少数なケースにおいて学習が困難であることが報告されている。

本研究では、ツイートデータを対象にトピックを学習し、各トピック出現量の時系列的な変化を抽出する手法を提案する。提案手法では、Hashtag Poolingによる擬似文書とBiterm topic modelをそれぞれ利用する。実験では、提案手法で実際のツイートデータからトピックを抽出し、抽出されたトピックの出現量を計算する。Hashtag Poolingを利用する手法では、抽出したトピック出現量が現実世界の事象に対するユーザの興味・関心を反映していることを確認する。Biterm topic modelを利用する手法では、トピックの学習およびトピック出現量を高速に計算する方法を提案し、既存手法と比較して、汎化性能が向上しつつ処理が高速化されていることを検証する。

研究指導教員：佐藤 哲司

副研究指導教員：若林 啓