

実生活 tweet に対する局面の階層的推定法に関する研究
A Study on Hierarchical Estimation Methods for
labeling Multi-Aspects to Real Life Tweets

学生番号：201221606

氏名：山本 修平

Shuhei YAMAMOTO

Twitter は、tweet と呼ばれる短文を投稿する、最も広く普及しているマイクロブログの一つである。ユーザ自らの経験や知識、また日常生活でのイベントなど、身近な「今」が投稿されている Twitter には、他のユーザにとっても新鮮かつ有益な tweet が数多く存在する。例えば、電車の遅延情報は交通機関を利用するユーザに役立ち、スーパーの特売情報はこれから買物に出かけようとしている近郊のユーザを支援できる。これらのような地域性が高く新鮮、かつ他のユーザに有益な tweet を、本研究では「実生活 tweet」と呼ぶ。実生活 tweet は、生活の様々な局面に対応している。例えば、「電車が来ない」という tweet は生活の中の「交通」の局面に対応する。

本研究では、実生活 tweet を抽出するため、人々の生活を 14 の局面に整理し、未知の tweet に対して局面を付与する、階層的推定法を提案する。階層的推定法の第一段階は、大量の tweet から潜在的ディリクレ配分法 (LDA) を用いてトピックを抽出する。第二段階は、少量のラベル付き tweet を用いて、トピックと局面の関連度を算出し、局面毎に閾値を超えた関連度を持つトピックと対応関係を構築する。入力された tweet に対する局面の推定は、tweet に出現する単語をトピックに展開し、第二段階で構築した対応関係に基づいて、複数の局面を tweet に付与する。

提案手法の有効性を評価するために、京都市内で投稿された日本語 tweet を用いて実験を行った結果、未知の tweet に対して適切に局面を推定できることを明らかにした。トピックと局面の対応関係について分析した結果、推定精度が高い局面では、その局面を表現するために必要な単語が集まったトピックが、高い関連度で結びついていることを明らかにした。

研究指導教員：佐藤 哲司

副研究指導教員：池内 淳