

平成 21 年度 情報工学コース卒業研究報告要旨

間瀬 研究室	氏 名	内 藤 和 宏
卒業研究題目	スポーツイベントにおける Twitter 実況要約の ための重要語抽出手法の検討	

インターネットの普及によって、誰でも簡単に情報を発信できるようになり、様々な事件や出来事に対しての意見や感想を他者と話し合う機会が生まれた。中でもテレビ番組の視聴者がインターネット上で、番組の内容についての感想をリアルタイムに書き込む Web 実況は、近年参加者が増加している。

これまでの Web 実況に関する研究では、テレビ番組に対する視聴者の反応や注目を定量化するために、実況者の書き込みのメタデータの作成などが行われていたが、Web 実況において多くの人が話している会話内容を把握する研究は行われていない。そこで本研究では、実況コメント中に頻出する語に注目して、Web 実況で多くの人が話している内容を抽出することを目的とする。

イベントの Web 実況を行う場合は、マイクロブログ「Twitter」を利用した。実況コメントの抽出は Twitter において特定の話題について語るときに書き込みの末尾に付けられる「ハッシュタグ」を用いた検索によって抽出を行った。

時間の経過による実況の変化を見るために、実況コメント全体を一定の時間間隔で分割した。多くの人が話している内容に含まれる単語は、実況コメント中の出現頻度が高く、またその語を含むコメントは多くの人が話している内容を表すものであると考えた。実況コメント中の語の出現頻度を計算し、次に実況コメントの重要度を決定するために、コメントが含む語の総出現数の和をコメントの重要度として加算していった。

上位に順位付けされたコメントが実際にそのイベント実況において多くの人が話している内容かを評価するための実験を行った。実況の対象にはボクシングを用いた。被験者にまず Twitter での実況コメントを見てもらい実況の様子を把握してもらった。次に、順位付けされたコメントから 10 個を取り出して、順位を伏せて被験者 8 名にそのコメントに対して多くの人が話していると思うものを 5 段階で点数付けしてもらった。そのラウンドにおいて多くの人が話している話の内容であると思うものに対して高い得点を与えるとする。実験を行った 3 つのラウンドの順位付けされた実況コメントに対する点数は図 1 のようになった。

実験の結果から、出現頻度の高い語を多く含む順位の高いコメントが、必ずしもその実況において多くの人が話している内容に合うものではないことが分かった。また順位の高いコメントは、試合内容とは関係のないコメントや実況ではあまり話されていない内容のコメントであることが示された。

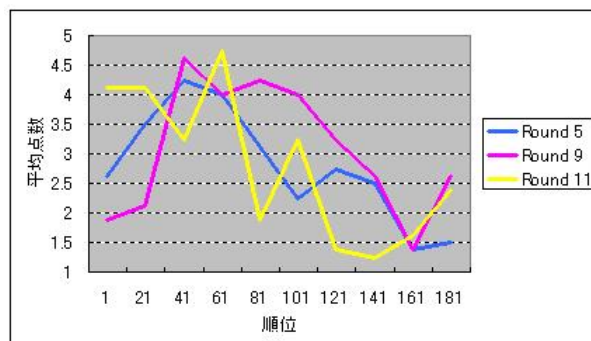


図 1. 順位付けした実況コメントに対する平均点数