

Twitterの実況ツイートを利用したタイムライン上の ネタバレ情報検知*

齊藤 令†
電気通信大学大学院

寺田 実‡
電気通信大学

1 はじめに

SNSは他ユーザとの交流や今話題の情報を入手するために日夜利用されている。しかし、興味のある情報だからといって必ずしも今必要な情報とは限らない。例えば、楽しみにしていたコンテンツの結果を知ってしまう(ネタバレ)が挙げられる。ネタバレ情報は結果やそれに至る過程を知ってしまうとそのコンテンツの面白さを減衰させてしまうことがあり、これからそのコンテンツを楽しむユーザにとって大きなストレスとなりうる。

また、SNSを利用するユーザの増加により、ネタバレ情報が目に入りやすい環境になっている。特に、スポーツ等の試合経過・結果は、多くの人にとってその速報性が重要であるので、公式サイトやニュースサイトでも積極的に発信される。そして、ユーザもその経過や結果から様々な感想や考察等を発信されるので、積極的にネタバレ情報が発信されている。

そこで本研究では、積極的に発信されるスポーツ等のネタバレを除去することを目的とする。ユーザが今放送している番組についてツイートする「実況」というものに積極的に発信されるネタバレの手がかりがないかと考えた。Twitterでは、番組に関連するハッシュタグをツイートに付加して「実況」を行うのが一般的である。このようなツイート(以下「実況ツイート」)を利用してネタバレ判別分類器を作成し、ネタバレなツイートの推定を試み、提案手法の有効性を検証する。

2 関連研究

ネタバレを遮断することを目的とした研究は過去にも幾つか行われている。

中村の研究[1]では、ブラウザベースでフィルタリングを行いネタバレ情報を遮断する研究である。ネタバレ

レ情報の抽出は、オブジェクト名リストと正規表現辞書を利用している。オブジェクト名リストには、主に固有名詞が入っておりチーム名及び選手名(正式名称やニックネームなど)から構成されている。正規表現辞書には、スポーツ等の試合に関する汎用性の高い単語や英数字、オブジェクト名リストのオブジェクトをマッチさせる正規表現が格納されている。このことからネタバレ情報の検出はルールベースによるもので比較的シンプルなものとなっている。

実況ツイートを利用した研究も幾つか行われている。Jeonらの研究[2]では、Twitter上でネタバレ情報を検出するにあたって重要な4つの要素「頻出人名」「頻出動詞」「URLつきツイート」「ツイートの時制」を提唱し、これらの要素を元にSVMで機械学習したモデルを使い、既存の機械学習モデルと比べた研究と比較した研究である。本研究では別の要素を追加し、分類器の向上を目指す。

内容が鮮明にわかる良いツイートを使って試合を要約[3]や、ユーザの属性(応援しているチーム)に応じた試合の要約[4]などといったTwitterの実況ツイートに関する研究も行われている。本研究では、このような実況ツイートは試合の内容が明らかにわかってしまうため遮断対象である。

Samuelらの研究[5]では、単語内に同じ文字が連続しているものを利用して投稿者の強い感情を表す単語を抽出する研究である。実況ツイートでは、このような投稿はよく見るので、連続した文字を検知して分類器のパラメータとして加える。

3 提案手法

ネタバレかどうかを分類するためのSVMによる分類器を作成し、分類器を用いてネタバレの推定を行う。

3.1 データセット

2015年7月に開催されたプロ野球オールスターゲーム第2戦のTwitter上の実況ツイートをおよそ5000件

*Spoiler detection on the timeline using the Live Commentary Tweets on Twitter

†Saitou Ryou, Graduate School of University of Electro-Communications

‡Terada Minoru, University of Electro-Communications

