

Twitter を用いた音声実況システムにおけるデータ処理方法

三輪 将吾† 小高 知宏† 黒岩 丈介† 白井 治彦‡ 諏訪 いずみ†
 †福井大学工学研究科 ‡福井大学工学部

1. はじめに

近年 Twitter や facebook に代表されるソーシャルメディアの普及により個人がインターネット上に発信する情報量が増加している。ソーシャルメディアでの情報は、ユーザが感じたことを即座に投稿できるためリアルタイム性の高い特徴があり、どこで何が起きているのかを迅速に察知するための手掛かりとして利用することが可能である [1]。

本研究では Twitter でユーザーが投稿する実況 Tweet を用いてプロサッカーの試合の実況を可能にするためのシステムを開発する。ユーザの Tweet 投稿から実況文出力までをリアルタイムで行うことで迅速に試合での出来事を把握することができる。実況文の出力は、テキスト読み上げエンジンを利用して音声による出力を可能にする。音声による出力によって、本システムをラジオの実況中継を聞く様な感覚で利用することが出来る。今回は、リアルタイム性を保持し、違和感のない実況文を音声出力するためのデータの処理方法について検討し実装することを目的とする。

2. システム構成

本研究では Twitter 上でユーザーの投稿する試合関連 Tweet 取得し、それをを用いたサッカーの音声実況を行うためのシステム開発を行う。そのため Twitter からサッカー観戦者が試合時間内に得点や、選手交代などの出来事に関して発言している実況 Tweet をリアルタイムで取得する必要がある。

Twitter 上から必要なデータの取得を行うために Twitter API を利用する。本研究では Twitter Streaming API の status/filter を用いて実況 Tweet 取得する。この API の特徴は、様々な条件をつけて、ストリーミング形式でリアルタイムに Tweet の取得ができるという点である。この Streaming API を PHP 上にて実装し、実況 Tweet の取得を行う。実況 Tweet 取得のために与える条件については 3 章で説明する。

取得した実況 Tweet はテキスト読み上げソフトを用いることで、音声による出力を可能にする。音声出力

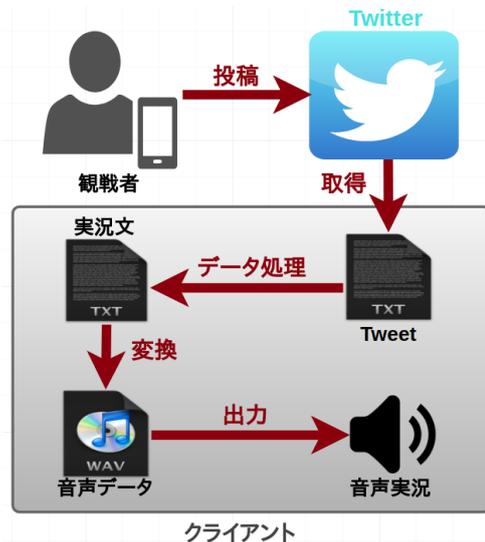


図 1: システム構成

を可能にするために Open JTalk という日本語のテキスト読み上げエンジンを利用する。実況 Tweet 取得の度に Open JTalk に入力し、音声ファイルへと変換しそれを再生する。これをリアルタイムで繰り返し行うことで音声出力を実現する。

本研究で目指す試合観戦者の実況 Tweet 取得から音声出力を行うまでのシステムの一連の流れをについて図 1 に示す。今回は、本システムにおいてリアルタイム性を保持しながら、違和感のない実況文を音声出力を可能にするための実況 Tweet の処理手法について検討し実装を行う。

3 実況 Tweet の取得

試合観戦者の実況 Tweet を取得するための条件として与えたものは次の通りである。

- 1 日本語 Tweet のみ取得
- 2 #チーム名 を含む Tweet の取得
- 3 リプライ、リツイートを排除
- 4 URL を含む Tweet を排除

これらの条件により、無駄な情報を抑えた、観戦者のみの Tweet の取得が可能である。

Data processing method in sound commentary system using Twitter

†Shogo Miwa †Tomohiro Odaka †Josuke Kuroiwa
 ‡Haruhiko Shirai †Izumi Suwa
 †Graduate School of Engineering, University of Fukui
 ‡Faculty of Engineering, University of Fukui

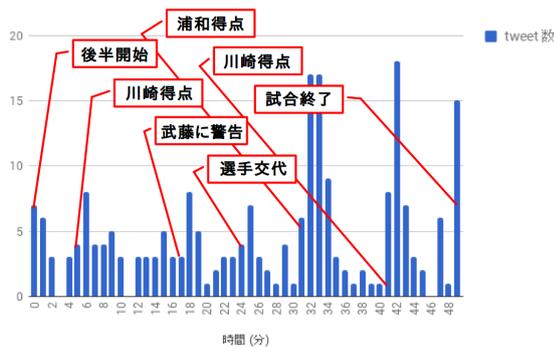


図 2: Tweet 数の変位

実際にこの条件で実況 Tweet の取得実験を行い、得られたデータの特徴について調査した。対象とした試合の前半の時間帯に 100 件、後半の時間帯に 230 件の Tweet を取得した。図 2 は対象とした試合(後半)の時間に取得した Tweet 数の 1 分ごとの変位と、試合での出来事の関係を表すグラフをである。

4. 実況 Tweet の処理

音声出力では一つの実況文の読み上げに一定の時間を必要とする。今回対象とした試合ではそのまま出力しても、リアルタイム性はある程度確保できる結果となった。しかしデータ量の多い試合で問題なくリアルタイム性を保持して行うためには、一定時間ごとの出力に制限を設ける必要がある。読み上げに掛かる時間を考慮して 10 秒に一つの実況 Tweet について出力を行うよう設定を行う。

図 2 から、試合における特定のイベント発生時に実況 Tweet が増加することが分かる。これらのポイントでは発生イベントに対して言及する Tweet が多数見られ、その内容が酷似していた。このことから一定時間において取得した Tweet の内容は似た内容の Tweet が集中すると考えられる。そこで 10 秒間に取得できた全ての Tweet から内容を判断し、重要 Tweet のみ出力する手法を検討する。そのために Text Summarization API を用いる。この API は入力された文章ごとの意味を読み取り、特徴的な文章を抽出することができる。そのため、複数の Tweet から重要であると考えられるものを実況文として出力することができ、試合状況をより正確に把握することが出来ると考えられる。図 3 に得点時に取得した実況 Tweet の例を示す。Text Summarization API を用いて図 3 の Tweet から 1 つの Tweet を抽出すると「武藤のゴールで一点返した!」という Tweet が要約文として出力された。この結果から、「武藤」が「ゴール」したという重要な内容が含まれた実況 Tweet を抽出できることが分かった。そこで新たに注目度の高い試合を対象に、データの

```

20:33:42: オフサイドっぽかったけど、武藤ゴール！ギリギリだった？(^^)/
20:33:39: 武藤サ——(°▽°)——!!
20:33:35: きたあ——！むと——！
20:33:31: ラファもそろそろ不調から脱出してほしいな
20:33:28: 武藤のゴールで1点返した！
20:33:07: 武藤雄樹ゴール
20:33:07: 武藤さめた！！！！オフサイドじゃない！
20:33:04: 武藤ゴール！！アウエーゴール1点目
20:33:02: 武藤サ——(°▽°)——!!
20:33:01: ナイス武藤！！

```

図 3: 得点時の実況 Tweet 例

取得を行う。前述した処理を行い音声出力を行うことでリアルタイム性を保持しながら、違和感のない正確な音声実況が可能であるかについて検証を行う。その結果については当日示す。

5. 考察

サッカーの試合観戦者の実況 Tweet から試合の音声実況を行う本システムにおいて、リアルタイム性を保持したより正確な実況を可能にするためのデータ処理方法について検討を行った。音声実況ではテキストによる実況と異なり、一定時間おける出力が制限される。そのため、今回のように一定時間で適したデータ量の実況文の出力はリアルタイム性確保のために有効であると考えられる。また複数の Tweet から一つの実況文を出力することで、より注目度の高い出来事について把握が可能であると考えられる。個人が投稿するソーシャルメディアでの情報はリアルタイム性が高い反面、誤った情報が拡散される可能性があるため要約文により正しい情報のみを提供することが可能なのではないかと考えられる。

本研究ではサッカーのようなスポーツの試合のだけでなく、災害時等の現場の状況把握を行うシステムとしての応用も視野に入れており、そのための手法についても今後検討したい。

6. まとめ

サッカーの試合観戦者による実況 Tweet から試合実況を行うシステムの提案を行った。実況 Tweet を取得することで、その試合での出来事を捉えることが可能であった。実況文はテキスト読み上げエンジンにより音声出力を行うことで様々な場面での利用が期待できる。今回はそのために必要と考えられるデータの処理方法について検討を行った。今後はその他のイベント実況への適用も検討していく。

参考文献

- [1] 齊藤 令, 寺田 実. 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. 第 78 回全国大会講演論文集 2016.1 (2016): 539-540.