

## Twitter からの鉄道運行情報の検索 Retrieval of Railway Operation Information from Twitter

河邊 拓也†      山田 剛一†      絹川 博之†  
Takuya Kawabe   Koichi Yamada   Hiroshi Kinukawa

### 1. はじめに

鉄道の運行では、しばしば事故や災害等により列車に遅延や運休が起こる。その時、鉄道会社は Web 等で運行情報を発表する。しかし、鉄道会社が発表する運行情報は発表されるのが遅く、また、些細な遅延などであればそもそも発表がされないことがある。そのため、実際の運行と鉄道会社が発表する運行情報に違いが生まれ、鉄道利用者は目的の列車に乗ることができない等の問題が発生する場合がある。この問題を解決するため、Twitter に着目する。

Twitter は、マイクロブログ、ミニブログとも言われ、140文字以内の短文を介して、コミュニケーションを取るサービスである。Twitter は、現在の状況を即座に投稿しやすい、情報を広く共有しやすいという特徴上、鉄道の運行情報についての投稿も多く存在する。

本研究では、Twitter を利用することで、鉄道会社が発表していない情報を含めた現在の鉄道運行情報をユーザに提供することを目的としている。Twitter から運行情報に関する tweet を検索し、得られた tweet群から運行情報を抽出して表示するシステムの作成を目指す。

本論文では、Twitter からの鉄道運行情報の検索について述べる。

### 2. 運行情報の項目

運行情報をユーザにわかりやすく提供するための項目として以下を対象とする。

- (1) 路線名  
事故や災害等により列車ダイヤの乱れが発生した路線の名称である。
- (2) 運行状況  
「運転見合わせ」や「遅延」、「運休」等の運行に関する現在の状況である。
- (3) 原因  
列車ダイヤに乱れをもたらした詳しい原因である。

運行状況や原因を表すのに駅名が含まれる場合があるが、本研究は運行情報を路線単位で扱うため、この2つの項目に含まれる駅名については扱わないものとする。また、(1) 路線名 (2) 運行状況 は運行情報として最低限必要な項目である。

### 3. 駅の略称の特徴

路線名を特定するために駅名の抽出を必要とすることがある。Twitter では駅を正式名称だけでなく略称で表現することがしばしばあり、駅名を抽出する際には略称の存在を考慮しなければならない。駅の略称で多くみられる特徴として以下のものが挙げられる。

- (1) 正式名称の先頭2文字から作られる

- (2) 正式名称の1文字目と3文字目から作られる
- (3) 正式名称の末尾2文字から作られる
- (4) 正式名称が 東, 西, 南, 北, 上, 中, 下, 新, 本 のいずれかから始まる時、正式名称の先頭2文字から作られる

駅の正式名称の字数により、みられる特徴に違いがあり、3文字では(1)、4文字では(2)(3)(4)、5文字では(2)、となっている。

### 4. 運行情報表示システム

#### 4.1 システムの概要

本システムは、Twitter から鉄道の運行情報を得るために、Twitter検索、対象外ツイート除去、運行情報の抽出、路線単位での運行情報作成、表示、を行う。運行情報表示システムの流れを図1に示す。

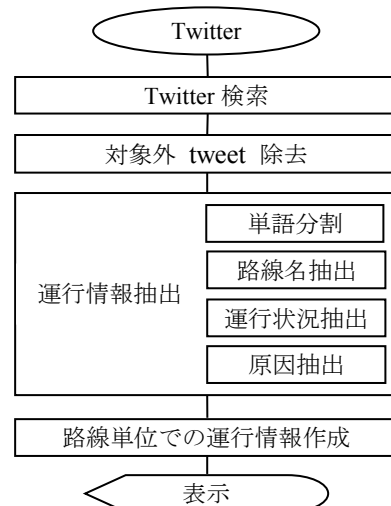


図1. 運行情報表示システムの流れ

#### 4.2 Twitter検索

Twitter の検索を利用し、鉄道の運行情報に関する tweet を取得する。運行情報に関連する語句で OR検索することで、鉄道の運行情報に関する tweet を含む検索結果を得る。

#### 4.3 対象外 tweet の除去

本研究では現在の運行情報を抽出、表示することを目的としている。そのため、投稿から時間の経っている tweet、当日の内容でない tweet は対象外とし、Twitter検索で得た tweet群からこのような対象外 tweet を除去する処理を行う。

- (1) 投稿から時間の経っている tweet の除去  
時間の経っている tweet を投稿から3時間以上経過した tweet と定義する。tweet の投稿時刻と現在の時刻を比べ、投稿から3時間以上経過している tweet は対象外と判断し、除去する。
- (2) 当日の内容でない tweet の除去

†東京電機大学大学院 未来科学研究科  
Graduate School of Science and Technology for Future Life,  
Tokyo Denki University

正規表現を用い、テキスト中に日付が明記されている tweet を探し、その日付を抽出する。抽出した日付が当日の日付と異なる場合、その tweet は対象外と判断し、除去する。

#### 4.4 運行情報の抽出

##### (1) 形態素解析による単語分割

tweet を形態素解析し、単語分割を行う。形態素解析には lucene-gosen [1] を使用する。形態素解析した結果から名詞のみを抽出し、リストに格納する。

##### (2) 路線名の抽出

路線名を抽出するために駅情報データベース [2] を利用する。駅情報データベースを参照し、tweet 中の名詞が路線名であるかを調べる。しかし、たとえば tweet 中から「新宿線」と路線名を抽出したとき、その路線は「都営新宿線」なのか「西武新宿線」なのかかわからない。このようなときは、それぞれの所属駅が tweet 中に含まれるかを調べる。このとき、駅の略称の特徴を利用し、駅の正式名称だけでなく略称についても調べる。そして、tweet 中に出現した所属駅の数により、どちらの路線を示しているのかを特定する。

##### (3) 運行状況の抽出

運行状況に関連する語句のテーブルを事前に作成しておく。テーブルを参照し、tweet のテキスト中にその語句が含まれているかを調べる。含まれているのであれば、該当する語句を運行状況リストに格納する。

##### (4) 原因の抽出

原因として想定される語句のテーブルを事前に作成しておく。テーブルを参照し、tweet のテキスト中にその語句が含まれているかを調べる。含まれているのであれば、該当する語句を原因リストに格納する。

#### 4.5 路線単位での運行情報作成

運行情報は路線単位で扱う。そのため、tweet 単位で抽出した運行情報を路線単位にまとめる必要がある。同一路線名が抽出された複数の tweet から運行状況リスト、原因リストを取得し、それぞれのリストを統合することにより、路線単位の運行情報を作成する。

### 5. 評価実験と考察

#### 5.1 実験概要

本システムの有効性を確認するため評価実験を行う。実験は、全国の路線を対象とし、リアルタイムに検索してきた tweet を用いる。307件の tweet を対象に、人手で抽出した 236件の運行情報と本システムの結果を比べ、評価を行う。評価には以下の指標を用いる。

$$\text{精度} = \frac{\text{正しく抽出した運行情報の数}}{\text{抽出した運行情報の数}}$$

$$\text{再現率} = \frac{\text{正しく抽出した運行情報の数}}{\text{全ての運行情報の数}}$$

$$\text{F値} = \frac{2(\text{精度} \times \text{再現率})}{\text{精度} + \text{再現率}}$$

#### 5.2 実験結果

実験の結果を表 1、表 2 に示す。

表 1. 運行情報の抽出結果

全ての運行情報の数	236件
正しく抽出した運行情報の数	198件
誤って抽出した運行情報の数	63件
抽出した運行情報の数	261件

表 2. 抽出結果の再現率と精度

精度	75.9%
再現率	83.9%
F値	79.7%

誤って抽出した運行情報は、さらに 2種類に分類できる。誤って抽出した運行情報の分類を表 3 に示す。

表 3. 誤って抽出した運行情報の分類

元の情報が運行情報であるもの	21件
元の情報が運行情報でないもの	42件

#### 5.3 考察

(1) 表 3 より、元の情報が運行情報でないものから誤った運行情報を抽出しているものが多いことわかる。その原因として、対象外 tweet が除去できていないものがある点が挙げられる。これは、tweet 中のテキストの時制の解析を行うことで解決が可能であると考えられる。これにより、精度を上げることができる。

(2) 元が運行情報であるものから誤った運行情報を抽出しているものを正しい運行情報が抽出できるようにすることで、精度と再現率をともに上げることができる。元が運行情報であるものから誤った運行情報を抽出する原因として、運行情報が複数含まれる tweet の存在があげられる。これは、tweet のテキストを文ごとに分割し、運行情報を抽出することで解決が可能であると考えられる。

### 6. おわりに

#### 6.1 成果のまとめ

本研究では、Twitter から tweet を検索し、得られた tweet 群から運行情報を抽出するシステムを提案し、実装した。実験の結果、F値が 79.7% という値を得た。

#### 6.2 今後の課題

今後の課題としては、運行情報抽出率の改善、運行情報の表示法の提案と実装が挙げられる。

#### 謝辞

本研究に際して、lucene-gosen の開発者の方々、駅データを作成されました「駅データ.jp」の方々に深く感謝いたします。

#### 参考文献

- [1] lucene-gosen : <http://code.google.com/p/lucene-gosen/>  
 [2] 駅データ.jp : <http://www.ekidata.jp/>