

## TETDM を用いた関連 Tweet 探索の一手法

## A Method for Searching Related Tweets using TETDM

梶並知記<sup>\*1</sup>

Tomoki Kajinami

<sup>\*1</sup>神奈川工科大学情報学部

Faculty of Information Technology, Kanagawa Institute of Technology

This paper reports a method for searching related tweets by a twitter search module “RelaTwi” and other modules. The RelaTwi is constructed on TETDM (Total Environment for Text Data Mining), which searches tweets according to subjective keywords of an input text. This paper shows that the TETDM has potential for supporting search as well as analyzing the input text.

## 1. はじめに

本稿では、TETDM を用いて試作した、Twitter 検索モジュール「RelaTwi」と他のモジュールの連携によって行う、関連ツイートの探索や分析に関する試験的な一手法について報告する。本稿の目的は、入力されたプレーンテキストそのものの分析を主な想定利用法としている TETDM が情報探索と分析が一体化したテキストデータマイニング支援システムとしても利用できる可能性を提示することである。

## 2. TETDM の特徴と本稿の目的

## 2.1 TETDM の特徴

TETDM (Total Environment for Text Data Mining) は、マイニング処理モジュールと可視化インタフェースモジュールから構成される、テキストデータマイニングシステムのプラットフォームである [砂山 11]。ユーザは、自身が行いたいタスクに応じて任意のマイニング処理モジュールと可視化インタフェースモジュールを組み合わせることで、入力データとして与えられるプレーンテキスト（ファイル）を分析できる。TETDM の詳細は TETDM サイトや文献 [砂山 11] に譲り、本稿に關係する特徴をまとめると、以下の 2 点となる。

- 入力データがテキストファイルである。
- 他のマイニング処理モジュールの出力結果を用いたマイニング処理ができる。

## 2.2 本稿の目的

TETDM のプロジェクトでは、多数のユーザの獲得や 1000 以上のモジュールの収集を試みている。本稿の目的は、TETDM へ入力されたテキストに關連するツイートを検索するモジュール「RelaTwi」を構築（と紹介）し、他のモジュールと連携して、TETDM がテキストそのものの分析だけでなく情報探索と分析が一体化したテキストデータマイニング支援システムとしても利用できる可能性を示唆することである。本稿は、学術的に価値のある技術要素を含まず単に TETDM の利用例の 1 つを紹介し、モジュール構築のヒントを読者に提供する類のものである。

連絡先: 梶並知記, 神奈川工科大学情報学部, 〒 243-0292 神奈川県厚木市下荻野 1030, E-mail: kajinami@ic.kanagawa-it.ac.jp, TEL: 046-291-3235

本稿の意義は、TETDM が入力されたテキストの分析だけでなく検索システムとしての応用可能性を示すことにより、TETDM に興味を持つユーザの増加やモジュール作成を試みるユーザの増加に繋がるきっかけの 1 つになることである。

## 3. RelaTwi の構築

## 3.1 RelaTwi の想定ユーザ

RelaTwi を構築するにあたって、RelaTwi を利用するユーザは以下のような意図をもってしていると仮定する。

意図 A あるテキストに關連するツイート（關連ツイート）を探索したい。

意図 B あるテキストに関してよく引用される關連ツイートの発言者を探索したい。

ここで、あるテキストの例は、Web 上のニュース、Blog に掲載されているコラム、商品のレビューなどである。本稿では、ユーザが自身の興味をもったテキストに関して、他人がどのような意見・感想をもっているのか気になり、現在注目を浴びている Twitter を対象に、意見・感想を収集してみたいという状況を想定している。

## 3.2 モジュールに求められる要件

ユーザのもつ、關連ツイートを検索したい意図（3.1 節の意図 A）に対してモジュールに求められる要件は、以下の要件 1 と要件 2 に細分化できる。ユーザのもつ、よく引用される關連ツイートの発言者を探索したい意図（3.1 節の意図 B）に対してモジュールに求められる要件は、要件 3 となる。要件 2 と要件 3 については、それぞれ a と b に分割して記述しているが、a はマイニング処理を、b は可視化処理を想定し、区別している。

要件 1 入力として与えられるテキストを特徴付ける主題キーワードの抽出。

要件 2a 主題キーワードを含んだツイート（關連ツイート）の検索。

要件 2b 關連ツイートの提示。

要件 3a 発言者同士のツイート引用・被引用關係の抽出。

要件 3b ツイート引用・被引用關係に基づいた発言者同士の關連の提示。

### 3.3 Relatwi の役割と他のモジュールとの連携

Relatwi の主な役割は、TwitterAPI を用いて主題キーワードを含んだツイートを検索する (3.2 節の要件 2a を満たす) ことと、検索して得られた関連ツイートを提示する (3.2 節の要件 2b を満たす) ための前処理、検索して得られた関連ツイートから発言者同士のツイート引用・被引用関係を抽出する (3.2 節の要件 3a を満たす)、発言者同士の関連を提示する (3.2 節の要件 3b を満たす) ための前処理を行うことである。

TETDM サイト<sup>\*1</sup> からダウンロードできる既存のマイニング処理モジュール「Panoramic」により、入力として与えられるテキストから主題キーワードを抽出することが可能である (抽出手法は文献 [砂山 01] を参照のこと)。したがって、Relatwi には、主題キーワードを抽出するアルゴリズムを含める必要がない。主題キーワードと検索結果である複数のツイートを表形式で提示することとし、既存の可視化インタフェースモジュール「TextDisplayHtml」を利用する。Relatwi は、得られた検索結果に html タグを付与し、それを TextDisplayHtml へ送る。ユーザは、HTML の表示に適した TextDisplayHtml を通して、入力として与えたテキストの主題キーワードと、主要キーワードが含まれる関連ツイートの一覧を眺めることができる。関連ツイートの提示は、TextDisplayHtml と連携することで容易に解決できることになる。

発言者同士のツイート引用・被引用関係を抽出するために、Relatwi は、検索して得られたツイートから発言者の ID を抽出し、同一ツイートに他の発言者の ID が存在する場合、発言者間に関連があると処理を行う。発言者同士の関連はネットワーク構造となるため、発言者同士の関連を提示するために、ネットワーク構造をもつデータの取り扱いに適する可視化インタフェースモジュール「SimpleKeywordMap」[梶並 11] と連携する。SimpleKeywordMap では、関連のある発言者同士がリンクで結ばれる。したがって、多くの発言者から着目されている、すなわち多くの発言者のツイートないに出現する発言者にリンクが集中することになる。ユーザは、可視化されたネットワーク構造を概観することで、よく引用される関連ツイートの発言者を探索することができる。Relatwi は、発言者同士の関連を SimpleKeywordMap が読み込める書式に加工するだけで、可視化処理を行う必要はない。

ユーザに検索結果や発言者同士の関連を提示するまでの大まかな流れをまとめると、以下ようになる。

1. Panoramic が入力されたテキストから主題キーワードを抽出
2. Relatwi が主題キーワードを含む関連ツイートを検索
3. Relatwi が検索結果を TextDisplayHtml と SimpleKeywordMap 向け書式に加工
4. TextDisplayHtml が関連ツイートを、SimpleKeywordMap が発言者同士の関連を提示

### 4. ツイート探索の実行例

Relatwi と他のモジュールの連携による、ツイート探索の例を示す。TETDM に入力したテキストは、2012 年 4 月 16 日の Yahoo ニュース「節電方法教えます...東京都が大阪府・大阪市に」<sup>\*2</sup> であり、以下のような記事内容である。

\*1 <http://tetdm.jp/pukiwiki/index.php?TETDM>  
 \*2 <http://headlines.yahoo.co.jp/hl?a=20120415-00000760-yom-soci>

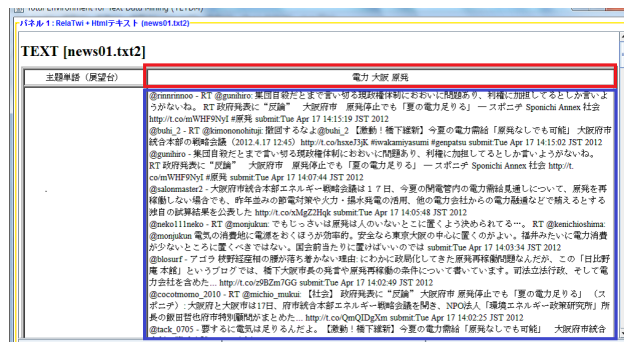


図 1: Relatwi による関連ツイート検索結果の例

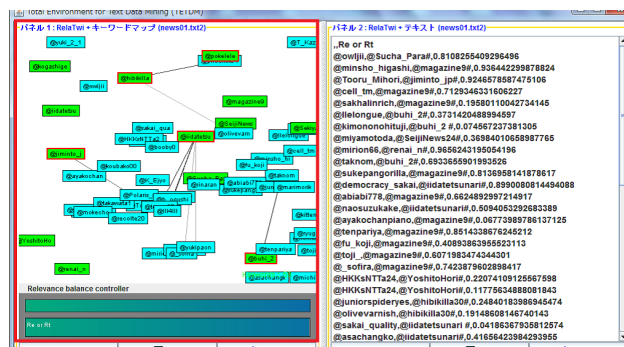


図 2: SimpleKeywordMap との連携例

記事内容 (一部省略)

原子力発電所が軒並み止まり、夏場に電力需給が逼迫(ひっぱく)する恐れが出てきたことを受け、東京電力福島第一原発事故後に深刻な電力不足を経験した東京都が、大阪府と大阪市に節電方法を伝授することになった。

(中略)

また、都は震災後、東京電力に頼らない電源を増やすため、天然ガス発電所やコージェネレーション(電力と熱の同時供給)施設の普及に取り組んでいる。環境局長はこうした取り組みを大阪側に紹介する。

図 1 は、Relatwi による関連ツイート検索の結果を、TextDisplayHtml と連携して表示したものである。図中、赤い実線の矩形で囲った部分に、入力されたテキストの主題キーワードを、青い実線の矩形で囲った部分に関連ツイートが表示されている。主題キーワードは「電力」「大阪」「原発」である。

図 2 は、ツイート引用・被引用関係に基づいた発言者同士の関連を表示したものである<sup>\*3</sup>。図中、赤い実線の矩形で囲った部分が SimpleKeywordMap のインタフェースである。インタフェース内の矩形が、それぞれ発言者 (@ID をラベルとしている) である (SimpleKeywordMap の右側のウィンドウには、関連ツイートから抽出した発言者同士の関連を SimpleKeywordMap 用の書式で表示している<sup>\*4</sup>)。

\*3 図で示している関連は、仮のものである。口頭発表では実データのデモをお見せする予定である  
 \*4 仮のデータである。

## 5. おわりに

本稿では, TETDM を用いて試作した, Twitter 検索モジュール RelaTwi と他のモジュールの連携によって行う, 関連ツイート探索や分析に関する試験的な一手法について報告した. 本稿では, ある Web 上のニュース, Blog に掲載されているコラム, 商品のレビューなどに興味をもち, 自身が興味をもったそれらに対する他者の意見・感想を Twitter を用いて収集したいと考えるユーザを想定し, RelaTwi を構築した. ユーザは, 自身が興味を持ったテキストそのものをクエリとしてシステムに入力することができる. RelaTwi は, マイニング処理モジュール Panoramic と連携することで, クエリとして与えられたテキストの主題キーワードが含まれているツイートを関連ツイートとして検索し, 可視化インタフェースモジュール TextDisplayHtml を通してユーザに提示する. また, 可視化インタフェースモジュールの 1 つである SimpleKeywordMap との連携により, よく着目されている関連ツイートの発言者を探索したいユーザの支援も可能とした.

実行例を示し, 本稿は, 入力されたプレーンテキストそのものの分析を主な想定利用法としている TETDM が情報探索と分析が一体化したテキストデータマイニング支援システムとしても利用できる可能性を提示した.

今後の課題としては, TETDM ユーザのさらなる獲得に向けて, 具体的な利用状況を想定したモジュールの開発と公開, 既存モジュールと連携して特定タスクの遂行を支援する利用例の公開を継続していく必要があると考える.

## 参考文献

- [梶並 11] 梶並知記: TETDM を利用したシンプルキーワードマップの構築, マイニングツールの統合と活用 & 第 6 回情報編纂研究会資料, pp.37-41 (2011).
- [砂山 01] 砂山渡, 谷内田正彦: 文章の特徴を表すキーワードを発見して重要文を抽出する展望台システム, 電子情報通信学会論文誌 D-I, Vol.J84-D-I, No.2, pp.146-154 (2001).
- [砂山 11] 砂山渡, 高間康史, Danushka Bollegala, 西原陽子, 徳永秀和, 串間宗夫, 松下光範: Total Environment for Text Data Mining, 人工知能学会論文誌, Vol.26, No.4, pp.483-493 (2011).