

街に着目した Twitter メッセージの自動収集と分析システムの提案と試作

松村 飛志[†] 安村 通晃^{††}

Proposal and prototyping of Twitter crawling and analysing system for town information

TAKASHI MATSUMURA[†] and MICHIAKI YASUMURA^{††}

1. 背景、問題点

近年、短いテキストメッセージ「つぶやき」のやりとりを中心とした、マイクロブログと呼ばれる形態の Web サービスが隆盛している。これらの Web サービスの代表例である Twitter では、多くのユーザーが頻繁にメッセージをやりとりしているため、ブログよりもリアルタイム性の高いテキストが常に大量に配信されている。また、こうしたサービスでは「いま自分がどこで何をしているか」について入力することが多いため、実世界の出来事や場所に関する情報が多い。

しかし、Twitter に投稿される膨大なつぶやきは有用で興味深いものと、自分は関心のないものが混じった状態であり、時として必要な情報を見逃してしまうという問題がある。また、実世界の場所・出来事に関するつぶやきを含んでいることから、膨大なつぶやきを分析することによってどの街が盛り上がっているかを推定することができるはずだが、個々のつぶやきがどの場所に言及しているのかわからないという問題がある。

著者自身も、Twitter を利用して 6700 人前後の人々のつぶやきを見ている。主に友人とのチャット的な利用が中心だが、大量のつぶやきを見ているとこうしたメッセージが、外出中などに自分が今いる場所の周辺で何が起きているのかをぼんやりと知る手がかりとなることが出来ることに気がついた。こうした経験から、

Twitter からロケーション情報を抽出・分析することで実世界に関する知見が得られるのではないかと考えた。また、多くの人々のつぶやきを常時チェックすることは実際のところ難しく、自分にとって興味深い情報を見逃してあとから気づくという事が多かった。そのため、自動でこれらの発言から重要と思われる部分を抽出し要約してくれるシステムを開発することにした。

2. 研究目的、解決点

本研究の目的は、膨大なつぶやきから、実際の行動に役に立つ・有用と思われるつぶやきを抽出し提示することである。特に、場所キーワードを導入することで個々のつぶやきが言及している場所を推定し、どこが盛り上がっているのかを知ることが目的である。

これによって、どの場所が盛り上がっているのかがリアルタイムにわかり、またその盛り上がりと言及していると思われるつぶやきの抽出を可能にする。

3. 提案手法

「場所の名前」= 場所キーワードを基準として Twitter のつぶやきを検索し収集する。場所キーワードごとに「勢い」= 単位時間あたりのつぶやき数を計測し記録する。

勢いの長期的な傾向から、現在の勢いが普段に比べて瞬間的に盛り上がっているかどうかを判定できるようにし、盛り上がっていると判定されたらその前後のつぶやきを提示する。

4. システム設計、実装

提案手法を実現するために、

(1) 日常的な会話で使われる場所の呼び名「場所

[†] 慶應義塾大学 政策・メディア研究科

Graduate School of Media and Governance, Keio University

^{††} 慶應義塾大学 情報環境学部

Faculty of Environment and Information Studies, Keio University



図 1 実装したシステムのトップページ
Fig. 1 Toppage of system

- キーワード」を入力し登録できるしくみを用意した。
- (2) そして登録した場所キーワードのリストを元に Twitter の発言をキーワード検索し、収集するクローラを作った。
 - (3) 収集した発言の勢い、Mecab と TFIDF による特徴語の検出、発言しているユーザーを分析するプログラムを用意し、1 時間間隔で定期的に行う。
 - (4) 勢いに基づいて、勢いが早ければ短い間隔で、勢いが遅ければ長い間隔で検索を行い発言を収集するよう調整を行った。

5. 運用

2009 年 8 月 31 日から 2009 年 11 月 13 日まで運用を行った。132 個の場所キーワードに関して Twitter を検索し、1,018,162 発言を収集した。

6. 事例

図 1 が開発したシステムの画面である。運用した結果得られた成果の事例を紹介する。

6.1 イベント、ニュース、事件など

池袋のヤマダ電機がオープンした際には、場所キーワードの表示部分が図 2 のように表示された。システムによって「池袋」というキーワードを含むつぶやきの中で「ヤマダ電機」という単語が特徴的であると検出されている。このほかにも様々な街中でのイベントやニュース、たとえば秋葉原の献血所がリニューアルしたといった情報もキャッチすることができた。

6.2 災害、事故など

災害や事故の際に当事者や付近の人々が Twitter を通じて実況し、情報を伝達することがある。本システ



図 2 登録した場所キーワードの例
Fig. 2 Example of Place Keywords



図 3 盛り上がり検出時のつぶやきの例
Fig. 3 Example of Twitter messages when it fevers

ムでもこうした現象が確認できた。たとえば品川シーサイドの大規模停電の際には「品川」という場所キーワードで「停電」という単語が特徴語として検出され、図 3 のような発言が表示された。このほかにも、台風や地震などの際にも規模や被害状況、経過などを知ることができる。

7. 今後の課題

盛り上がっている場所・時間のつぶやきを提示できるようになったが、時折 spam アカウントや bot などつぶやきが混ざってしまう。これらの有用ではないつぶやきをノイズとして除去する必要がある。

また、現在は収集対象とする場所キーワードを自分で選定し入力しているので偏りがある。Twitter ユーザーから協力者を募り、場所キーワードを登録してもらう事を検討している。

8. おわりに

場所に特化した言葉を手がかりに、膨大なつぶやきの中から、特定の場所に依存したもののみを抽出し、一覧するシステムを提案し、試作した。

参考文献

- 1) Twitter : <http://twitter.com/> (2006).