

Twitter を用いた地域別食生活調査

鈴木 貴文[†] 堀 幸雄[‡] 今井 慈郎[‡]

香川大学 大学院 工学研究科 信頼性情報システム工学専攻[†]

香川大学 工学部[‡]

1. はじめに

現在，インターネット上には様々なコミュニケーションのためのサービスが存在している．その中でも人気が高いサービスの一つに Twitter がある．Twitter はマイクロブログの一種であり，140 文字以下の短い文章をつぶやくことで通常のブログや SNS よりもリアルタイム性の高いサービスを実現している．Twitter では「今なにをしているか？」が発言されるため，ユーザの生活に密接した情報が多いことが特徴である．ユーザが今どこにいるか，何を食べたかなども頻繁につぶやかれ，ユーザは気軽につぶやきを投稿できる．

しかし，Twitter に投稿されるつぶやきはユーザ個人の行動や嗜好はわかっても全体としての流れは把握しづらいという問題がある．この問題は現在の流れでつぶやかれている単語を表示する「buzzter」[1]や，県別で人気のあるユーザが表示される「まちつい」[2]などのウェブサービスによってフォローされているが，これらのサービスには地域毎の特性を調査するような機能はない．地域毎の特性がわかれば，各地域へ出店しようとする中小企業や個人事業の支援となるはずである．

Zhiyuan Cheng[3]らは位置情報や地域を推測できるようなつぶやきの内容からあるユーザのフォロワーのうち半分は100マイル以内の距離に住んでいるという研究結果を出した．しかし，位置情報を利用し，ユーザの生活スタイルを推測することはできていない．

本研究では，位置情報を登録したユーザがつぶやいた内容から何の食品に興味を持っているかを明らかにし，地域間の食文化の違いを発見することを目的とする．

2. システム概要

Twitter にはユーザの嗜好や生活スタイルに密

Analysis of Regional Food Life on Twitter

[†] Takafumi Suzuki, Graduate School of engineering, Kagawa University

[‡] Yukio Hori, Yoshiro Imai, Faculty of Engineering, Kagawa University

接したつぶやきが投稿される傾向にある．

Twitter にはユーザ登録時点で自分の住んでいる場所を登録できる機能があり，この情報を元に各県のログを作り，県別住人が興味を持っているものを推測する．ジャンルを食品に限定することによって結果を得やすくする．食料品は多くの人が毎日消費しているものであり，他のものよりも調査がしやすいはずである．

一般公開されているつぶやきのうち，日本語で書かれたものを収集し，ログとしてサーバに蓄積する．つぶやきに含まれる食品の名前を調べることによってそのユーザが何を食べたか，何を食べたがっているのかを読み取る．各県の住民の好みを推測し，結果を出力する．図1は本システムのフローチャートである．結果はhtmlで出力する．

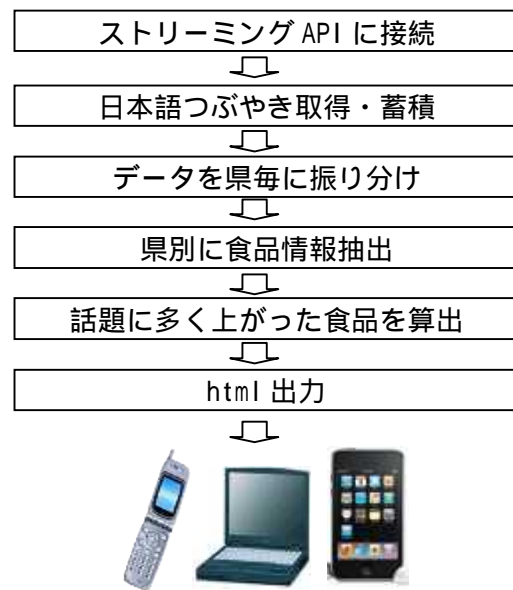


図1. システムフローチャート

3. システム設計

本システムの構成を図2に示す．

まず Twitter に投稿されたつぶやきを取得し，サーバに蓄積する．取得するつぶやき情報は public_timeline であり，これは全体に公開されているつぶやきとなっている．Twitter が提供す

る API は 1 時間あたりの取得限度が決まっているため十分ではない．しかし Twitter には可能な限り漏れなくタイムラインを取得したい場合にストリーミング API が用意されている．ストリーミング API には 3 つのレベルがあり，Twitter 社と契約無しで利用できる sample レベルを用いる．このレベルでは全てのタイムラインの 5% が取得できる．取得したデータはデータ容量削減のため，地域データを含む「ユーザ情報」と 1 時間毎に蓄積された「つぶやき情報」に分けて記録する．

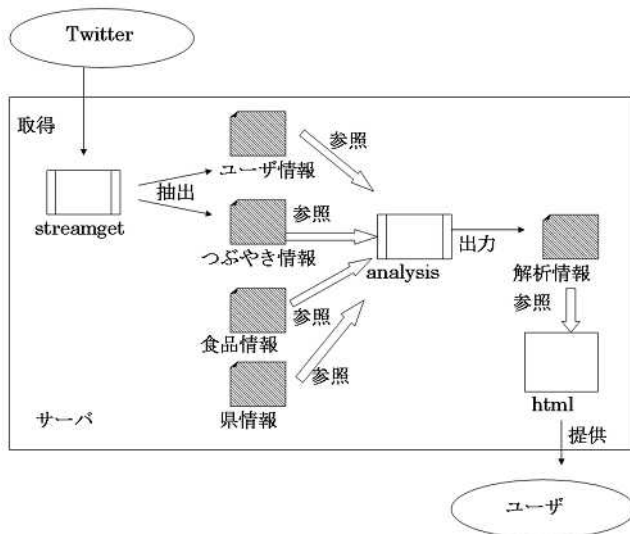


図 2．本システムの構成

次に取得したデータを解析する．先に県毎の情報と食品目の情報を用意しておく．食品情報は日本でよく食べられているラーメンやカレーといった食品を 290 種を手動で登録した．県毎に振り分けられたデータから，パターンマッチによってつぶやきから食品情報抽出し，各項目の回数を計算する．



図 3．出力結果

食品情報の抽出と計算が終われば結果を html に出し，一般ユーザが閲覧できるようにする．

グラフを用いることで食品の興味に占める割合が一目でわかり，品目の内容がわかる．食品ごとに何がつぶやかれているかを見ることで本当にその食品に興味を抱いているかを判断できる．県別にどの食品が食べられたかだけでなく，食品別にどの県民に人気があるかの結果も出力する．図 3 はこの例である．

4．ログ収集実験

作成したプログラムを使用し，実際にログ情報の蓄積を行った．その解析を行い，つぶやきのなかにもどの程度の食品の情報が含まれるかを調査した．1 週間分の情報量では，12 月 2 週目に取得した 7 日間の発言数は 967487 件で，食品情報はそのうち 1496 件だった．これは 0.155% の割合である．12 月の 1 カ月の発言数は 3608374 件，そのうち食品情報を含む発言は 5285 件であった．これは全体の 0.147% である．

この実験では，取得したつぶやきにフィルターをかけているため，元のデータとの差を比較する．日本語表記のみの場合を all，日本語表記かつ住所を登録している場合を filter とした．図 4 はその件数別のグラフである．

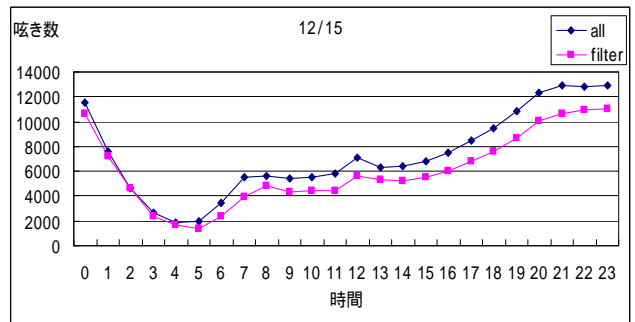


図 4．取得件数別グラフ

5．おわりに

本稿では，Twitter を用いた地域の食生活調査手法とそのシステム設計を提案した．つぶやきの収集を行い，その実験結果を述べた．今後は実際に公開し，一般の食生活調査との比較をする．またつぶやきに含まれる食品の割合の検出向上を目指す．

参考文献

[1] <http://buzztter.com/ja>
 [2] <http://machi.userlocal.jp/>
 [3] Zhiyuan Cheng, James Caverlee, Kyumin Lee, "You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users", http://students.cse.tamu.edu/zcheng/papers/cheng_cikm10.pdf