

# 複合施設における小規模施設に関するツイート発見

安井 豪基<sup>†</sup> 王 元元<sup>†</sup> 細川 侑士<sup>†</sup> 河合由起子<sup>†</sup> 秋山 豊和<sup>†</sup>  
角谷 和俊<sup>††</sup>

<sup>†</sup> 京都産業大学 〒603-8047 京都府京都市北区上賀茂本山

<sup>††</sup> 兵庫県立大学 〒670-0092 兵庫県姫路市新在家本町1丁目1-12

E-mail: †{i1458085,yuanw,i1358103,kawai}@cc.kyoto-su.ac.jp, ††akiyama@cse.kyoto-su.ac.jp,  
††sumiya@shse.u-hyogo.ac.jp

**あらまし** 本研究では、大型ショッピングモールや駅構内のような複合施設における各店舗などの小規模施設に関するツイート発見を行い、関連する施設の Web ページを検出し、そのページ上に該当するツイートならびにツイートの集約情報を提示するシステムの構築を目指す。我々はこれまで、ツイートの発信場所と Web ページの内容（場所名）を抽出することで、ツイートと Web ページのマッピングに基づき Web ページ上に関連するツイートを提示するシステムを構築してきた。そこで、本研究では、さらに時空間分析に基づき複合施設内の店舗ごとに関するツイートを発見する手法およびツイートの情報を集約する手法を提案する。これにより、Web 閲覧ユーザはある複合施設内の各店舗に関する詳細な情報を得ることができる。本論文では、提案手法に基づき施設 Web ページ上にツイートの集約情報を提示するシステムを構築し、各店舗などの小規模施設に関するツイートの発見手法を検証する。

**キーワード** 複合施設, ツイート発見, 集約情報提示

## 1. はじめに

近年、Twitter<sup>(注1)</sup> や Facebook<sup>(注2)</sup> に代表されるソーシャルネットワークサイトを通して、オンライン上でのユーザ間のコミュニケーションを中心としたソーシャルネットワークサービス（以降、SNS と記す）が普及している。また、スマートフォンなどの端末の発達により、場所を問わず任意の時間に情報を発信することができるため、実世界においてある場所に関してリアルタイムに情報の発言を行っている SNS ユーザも多い。それらの発言は、リアルタイムでその場所の状況を把握するために役立つことができる。しかし、Twitter にリアルタイムで随時発信されているツイート数は膨大であり、関心のあるトピックのツイートを取得することが難しい。また、ユーザ間のコミュニケーションにおいても、その SNS 内のコミュニティ（フォローとフォロワー）に限定される場合もあるため、情報の網羅性が低くなってしまふ。例えば、ツイートの内容やハッシュタグでの検索により関連するツイートを取得する手法 [1] では、実空間においてその場所にいない SNS ユーザのツイートも検出する。

また、Oku ら [2] では、位置情報付きツイートをを用いて、緯度と経度を指定し、その場所付近で発信された関連性の高いツイートを取得できる。しかしながら、その場所で発信されたツイートの中で、その場所に関する内容とは無関係のツイートも検出されてしまふ。大型ショッピングモールや駅構内のような複合施設における各店舗などの小規模施設に関するツイートを

選別することが難しい。このように、適切なツイートの取得は手間や時間がかかってしまふ問題点と、網羅性の向上が課題としてあげられる。

我々はこれまで、ツイートの発信場所と Web ページの内容（場所名）を抽出することで、実空間においてツイートを発信しているユーザ（ツイートユーザ）と、発信場所に関連する Web ページを閲覧しているユーザ（Web ユーザ）に対して、SNS と Web といった異種メディア横断型コミュニケーション支援システムを構築してきた [3]。そこで、本研究では、さらに複合施設における小規模施設に関するツイート発見を行い、関連する施設の Web ページを検出し、そのページ上に該当するツイートならびにツイートの集約情報を提示することで、異種メディア横断型コミュニケーション支援システムの拡張を目指す。これにより、複合施設よりその内の小規模施設に関して発信されるツイートの効率的な取得と、ツイートユーザ間だけでなく、Web ユーザとも情報提供、取得ができ、網羅性の向上にも繋がる。本論文は、具体的には、以下の 2 点を実現する。

- 時空間分析に基づく複合施設内の小規模施設に関するツイートのクラスタリング
- 時間帯ごとに複合施設内の小規模施設に関するツイートの情報集約

提案システムでは、ツイートユーザと Web ユーザのコミュニケーションが可能になることで、ツイートユーザは他のツイートユーザだけでなく、Web ユーザへの問い合わせも可能となり、情報の網羅性の向上が期待できる。また、Web ユーザは、Web ページに関連する複合施設内の小規模施設についての感想や混雑具合といった情報をツイートユーザから取得し、集約した情報を提示することで、リアルタイムに複合施設内の小規模

(注1) : <https://twitter.com/>

(注2) : <http://www.facebook.com/>

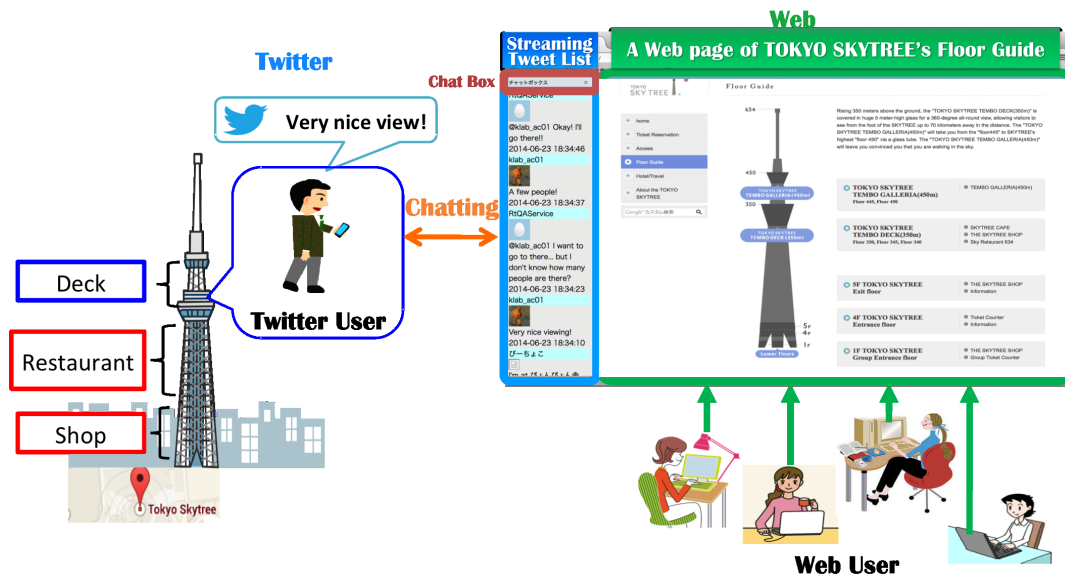


図 1 異種メディア横断型コミュニケーション支援システムの概要図

施設に関する詳細な情報を把握することができる。Web ユーザがメッセージを発信すると、該当 Web ページを閲覧している他の Web ユーザに発信され、ツイートユーザにはメッセージがツイートとして提示される (図 1.)。なお、ツイートユーザは本研究で提供するサービスのアカウントのフォロワーになる必要がある。

本論文では、時空間に基づく複合施設内の小規模施設に関するツイートの発見手法およびツイートの情報集約手法について述べる。また、提案システムを実装し、システム動作について考察する。本論文の構成は以下のとおりである。次章で提案システムの概要を説明し、3章で位置情報付きストリーミングツイートデータの分析手法および時空間に基づくツイートのクラスタリングについて述べる。4章で実装したシステムを検証し、5章で関連研究について述べた後、最後に、6章で本研究のまとめと今後の課題と展開について述べる。

## 2. システム概要

本研究は、場所に関するツイート情報の取得ならびに、ツイートユーザと Web 閲覧ユーザが場所情報に基づきリアルタイムに通信可能なシステムの構築を目指す。

図 1. にシステムの概要を示す。ツイートを発信すると、ツイート閲覧者だけでなく、ツイートの内容と発生場所の位置情報に基づき、関連する Web ページを検出し、それらを閲覧しているユーザのページ上にそのツイートがリアルタイムに提示される。Web 閲覧ユーザにとってはそれら提示されたツイートを閲覧することで、場所に関する現状把握の支援になる。各ツイートに対して返信する場合は、該当するツイートを選択することで、個別に返信できる。ただし、ツイートユーザは本サービスをフォローしている必要がある。

一方、Web 閲覧ユーザが本システムの入力ボックスを用いて情報発信すると、ページを閲覧している他の Web 閲覧ユーザにメッセージが送信される。また、本サービスをフォローして

いるツイートユーザに対してもツイートとして提示できる。ツイートによる返信は、前者の問合せ同様にツイートだけでなく Web ページ上にも提示され、全てのページ閲覧ユーザはこれらの問合せと返信を閲覧できる。

なお、全てのメッセージは WebSocket サーバを経由するため匿名性が保たれる。図 1. では、東京スカイツリー付近にいるユーザがツイートを発信した場合に、そのツイートが東京スカイツリーのページと関連付けられ、Web ブラウザに提示されている。Web 閲覧ユーザは混雑具合や東京スカイツリーの感想など状況に関する問合せができ、一方でツイートユーザは東京スカイツリーの展望時間など、その時その場では確認しにくい情報に関してツイートによる問合せが可能になる。

図 2 に処理の流れを示す。本研究では、Web 閲覧ユーザとツイートユーザとを Web ページとツイートを通してリアルタイム通信可能にするため、リアルタイムに送信されるツイート (以下、ストリーミングツイートする) ならびに Web ユーザがアクセスしている Web ページの URL を取得する。サーバはツイートユーザが発信したストリーミングツイートを取得し、位置情報に基づいて関連するページを取得し、対応付けの管理を行う。取得した関連ページに Web 閲覧ユーザがアクセスすると、対応するツイートを抽出し、ブラウザへ送信および提示する。なお、ツイートユーザは本サーバからメッセージを受信する際には、本サービスのアカウント (注1) のフォロワーとなっている必要がある。Web 閲覧ユーザは提案システムとなるアドオンを用いる必要がある。また、取得した位置情報付きツイートのクラスタリングを行い、ツイートに詳細な店舗名など小規模施設の情報付与する。一方で、Web ユーザが閲覧している Web ページから特徴語を場所名だけでなく店舗名を抽出し、先ほどのツイートに付与した店舗名とマッピングを行い、店舗ごとといった小規模施設に分けてツイートを集約情報にし、Web

(注 1) : <https://Twitter.com/@RtQAService>

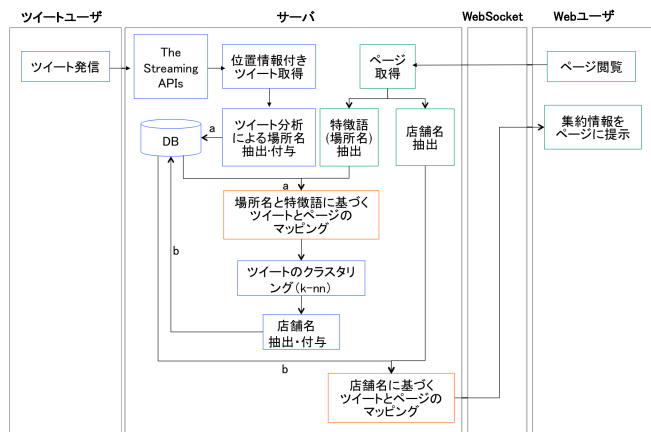


図2 集約情報提示までの流れ

ユーザに提示する。

### 3. 位置情報に基づくツイートと Web ページ間リアルタイムメッセージ通信

Web 閲覧ユーザがアドオンの入力ボックスにメッセージを入力すると、サーバが受信し、同じページを閲覧しているユーザのブラウザへ送信する。また、Web ページに提示されている各ツイートにのみ直接返信も可能である。直接任意のツイートに送信する場合、サーバはブラウザからメッセージを受信し、ツイートしたツイートユーザ（フォロワー）へ送信する。

全ての送受信はサーバを介するため、匿名性が保たれる。

#### 3.1 ストリーミングツイートデータ取得と場所名付与

本論文では、位置情報に基づく問合せを目的としており、ページとツイートを位置情報に基づき関連付ける。そのためまず、指定地域から重複を除いた緯度経度情報を含むストリーミングツイートを The Streaming APIs を用いて取得する。指定地域は、1 度以上異なる南西および北東を指定することで、その 2 点に囲まれた矩形領域のストリーミングツイートを取得できる。

次に、ツイートに場所名付与を行う。取得したストリーミングツイートの緯度経度情報から、Google Place API version 3<sup>(注2)</sup> を用いて、半径 1m の場所名を取得した。評価実験では、取得した場所名は関連する Web ページ取得の際に検索キーワードとして用いられることと、ツイート発信ユーザの移動も考慮し、 $l=5$  とした。また、ツイート内容を形態素解析し、名詞となる単語を取得する。

以上より、ツイートユーザ id、アイコン画像 URL、緯度、経度、場所名、ツイート内容、単語集合、取得時刻を一定時間管理する。

#### 3.2 ツイートの緯度経度と内容に基づくツイート選別

前節より取得したストリーミングツイートに対して位置情報に基づいた内容判定を行い、ページと関連付ける。ツイートが発信された場所名と関連するかをツイートの内容から判定することで、ツイート発生場所と関係性の低いツイートの除去を行う。

位置情報に基づいたツイート内容判定法は、一定範囲内の一定時間のツイートに多く出現する単語は関連性が高いと考え、場所名に対する特徴語として抽出する。この特徴語を多く含むツイートを場所名に関連するツイートとして選択する。まず、取得したツイート  $t$  の位置情報より、半径  $d$  内に存在する一定時間内のツイート  $n$  個を取得する。次に、下記の式によりツイートの重要度を算出する。まず、ツイート  $t$  に出現する各単語  $i$  のツイートに出現する頻度を抽出し、その平均値を算出する。また、特徴的な単語が出現しても単語数が多い場合は、ツイートの重要度が低下するため、シグモイド関数を用いることで、出現頻度の高い単語には、さらに重要度の重みを増やす方法を取ることにした。

$$\sum_{i=1}^m \left( \frac{\text{単語 } i \text{ が出現するツイート数}}{\text{ツイート総数 } n} \times \frac{1}{1 + e^{-x}} \right) \times \frac{1}{m} \quad (1)$$

$m$  はツイート  $t$  に出現する単語総数である。 $x$  は以下の式で求まる単語  $i$  の DF 値である。

$$x = \frac{\text{単語 } i \text{ が出現するツイート数}}{\text{ツイート総数}} \quad (2)$$

最後に、閾値以上のツイート  $t$  を位置情報に基づいたツイートとする。

#### 3.3 Web ページの場所名抽出

まず、Web 閲覧ユーザの閲覧している Web ページの URL を取得し、その Web ページのスニペットを取得する。次に、スニペットから出現頻度の高い単語を特徴語として抽出する。また、形態素解析よりその特徴語の中から地名を判別し、該当する単語をそのページの場所名とする。尚、複数地名が抽出された場合は全てを場所名とする。

#### 3.4 場所名に基づく Web ページとツイートの対応付け

3.3 節より Web 閲覧ユーザの閲覧している Web ページの場所名が抽出された。また、3.1 節より、ツイートユーザの位置情報付きツイートを The Streaming APIs を用いて取得し、緯度経度から場所名を取得して、さらに、3.2 節では場所に関連するツイートを選別した。ユーザが Web ページを閲覧すると、場所名から関連するツイートを検索し、Web 閲覧ユーザに提示する。ツイートユーザには、緯度経度情報から場所名を抽出し、その場所名と一致する Web ページを対応づける。なお、DB には取得したツイートおよび抽出した場所名を格納する。これらのツイートと Web ページを場所名に基づき、対応付ける。

#### 3.5 リアルタイム双方向通信

リアルタイム問い合わせシステムを構築する上で、Ajax や Comet、WebSocket といった様々な双方向通信手法が存在する。先行研究として、これまで我々は Web サーバと Web ブラウザ間の通信のための双方向通信として Ajax や Comet を用いてきた [4] が、本研究では、より通信ロスの少ない WebSocket<sup>(注5)</sup> を用いる。

(注2) : <https://developers.google.com/place/>

(注5) : <http://gihyo.jp/dev/feature/01/websocket/0001>  
<http://www.atmarkit.co.jp/ait/articles/1111/11/news135.html>

表 1 実験データ

大阪駅 (LUCUA)	
中心緯度経度 (34.703289, 135.496242)	
取得範囲半径 $d = 200\text{m}$	
日時	ツイート数
2015 年 12 月 5 日 (金) 17 時-22 時	122
2015 年 12 月 12 日 (金) 17 時-22 時	99
2015 年 12 月 19 日 (金) 17 時-22 時	133
2015 年 12 月 26 日 (金) 17 時-22 時	138
2015 年 12 月 6 日 (土) 11 時-16 時	151
2015 年 12 月 13 日 (土) 11 時-16 時	147
2015 年 12 月 20 日 (土) 11 時-16 時	172
2015 年 12 月 27 日 (土) 11 時-16 時	129
2015 年 12 月 7 日 (日) 11 時-16 時	188
2015 年 12 月 14 日 (日) 11 時-16 時	148
2015 年 12 月 21 日 (日) 11 時-16 時	160
2015 年 12 月 28 日 (日) 11 時-16 時	134

ツイートユーザから Web 閲覧ユーザへの配信では、ツイートユーザの位置情報付きツイートを The Streaming APIs を用いて取得し、そのツイートの緯度経度情報から場所名の付与を行い、DB にツイートと場所名を格納する。3.3 節より、Web 閲覧ユーザが閲覧している Web ページの場所名から、場所名とマッチするツイートを DB に問い合わせ、該当するツイート情報を Web 閲覧ユーザに提示する。

Web 閲覧ユーザからツイートユーザへの配信では、3.4 節より、Web 閲覧ユーザは Web ページ上に提示されたツイートに対してサーバを経由し、ツイートとしてメッセージを送信することが可能である。

### 3.6 各小規模施設へのクラスタリング

k 近傍法を用いて複合施設内におけるツイートを小規模施設ごとに分類する。

最近傍法とは、判別対象のデータが、どの学習データに一番類似しているかで判別する手法である。データ同士の類似度は、ユークリッド距離を用いる。つまり、ユークリッド距離の値が低いほど類似度が高いということになる。ツイートの内容の名詞と形容詞を形態素解析により取り出し、3.2 節の式 (2) より単語ごとの DF 値を求める。全ての単語の DF 値を各ツイートに当てはめるため、全てのツイートに出現する単語数が  $n$  種類の場合、各ツイートのベクトルは  $n$  次元空間で表される。ツイートのベクトルから以下の式により、ツイートの類似度の算出を行う。

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

上記の式より、判別対象のデータとそれぞれの学習データの類似度を求め、類似度が高い順に学習データのクラスを抜き出し、最も多いクラスが判別対象のデータのクラスとなる。

$$\text{識別クラス} = \begin{cases} j \text{ where } \{c_j\} = \max\{c_1, \dots, c_k\} \\ \text{reject where } \{c_i, \dots, c_j\} = \max\{c_1, \dots, c_k\} \end{cases}$$

判別対象のデータとの類似度が高い順に学習データのクラスを  $k$  個取得し、その内、最も多いクラスがクラス  $j$  の場合、判別対象のデータのクラスは、クラス  $j$  と識別される。ただし、

表 2 大阪駅 (LUCUA) の各階層のカテゴリ

階層	カテゴリ
10F	レストラン
9F	ボックス, ライフスタイルグッズ
8F	ボディケア, コスメ
7F	メンズファッション
6F	アクセサリ
5F	レディースファッション
4F	シューズ, レッグウェア
3F	ファッション雑貨
2F	レディース, メンズファッション
1F	レディース, シーズングッズ
B1F	スイーツ, フード, コスメ
関係なし	その他

最も多いクラスが複数存在する場合、識別不可となる。

これにより、ツイートを各小規模施設に分類を行う。

## 4. 実装および検証

本研究では、複合施設内の小規模施設に関するツイート発見ならびに集約情報の提示を目的としている。本章では、実装による評価としてプロトタイプを構築し、複合施設内の小規模施設に関する Web ページに集約情報を提示するため、ページに提示されるツイートの評価実験を行う。本プロトタイプでは、サーバは、Apache httpd 2.4, java, php5.5 を用いた。クライアント側は、Javascript を用い、Firefox 23.0.1 上で動作確認した。なお、2014 年 7 月 30 日から 2014 年 12 月 31 日まで The Streaming APIs で日本全国のツイートデータを取得した。今回の検証では、上記のツイートデータから金曜日、土曜日、日曜日のツイートデータを使用した。

提案手法では、ツイートの発生位置の半径  $d\text{m}$  内のツイートを対象として、クラスタリングされたツイートの内容がある複合施設内の店舗など小規模施設ごとに基づいているかを判定する。対象は、大阪駅周辺の大規模ショッピングモール「LUCUA」を中心とした半径  $d=200\text{m}$  内で発信したツイートとし、期間は 2014 年 12 月 5 日から 28 日までの約一月間における金、土、日曜日の週末とした (表 1)。また、「LUCUA」内の各階を小規模施設とした (表 2)。

正解データの決定は、20 代の大学生 13 人による主観的評価に基づき、ツイート内容が各カテゴリに対しての関連性の評価を行い決定した。被験者が各ツイートの各カテゴリに対して関係性を評価する。評価は、少し関係している:1, そこそこ関係している:2, かなり関係している:3, 関係なし:0 とした。平均値が最大のカテゴリをそのツイートのカテゴリとした。なお、店舗名が含まれているだけでなく、店舗に関する感想や店舗に関する問合せも関係するツイートとして評価してもらった。

上記のツイートとカテゴリのデータセットを 5 分割し交差検証を行い、ツイートを施設の各階層にクラスタリングする精度の検証を行った。小規模施設に分類されたツイートの精度を求めるため、評価値と分類の差異で検証を行った。まず、ツイートの各クラスの被験者の評価結果の平均値を正規化する ( $p_k$ )。次に、k 近傍法により推薦された各クラスの数を正規化する

表3 実験結果

	金曜日	土曜日	日曜日
二乗平均誤差	0.103	0.097	0.089

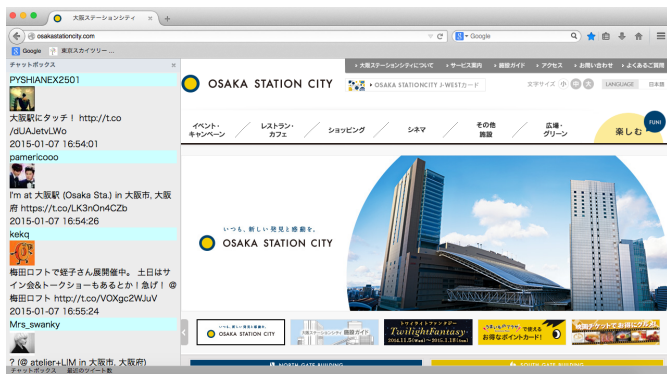


図3 Web閲覧インタフェース

( $q_k$ ). この2つの二乗平均誤差を算出し、全てのツイートの平均で評価を行った。なお、 $c$ はクラス数とする。

$$\text{二乗平均誤差} = \sqrt{\frac{\sum_{k=1}^c (p_k - q_k)^2}{c}}$$

本実験の結果を表3に示す。全ての曜日において、0.09から0.10と概ね良好な結果となり、約9割が正しく分類できることを確認した。手法によるツイート識別を確認すると、「伝説のすた丼 チャーハン 並盛りでも満足なこの量。出来立てで美味しかったし、650円で大満足！」のツイートが「レストラン」に分類できており、「美味しい」のような形容詞で識別できたことが確認できた。ただし、「美味しい」と「おいしい」が別の単語として扱われたため、識別ができない場合があったため、文字揺れの対応が必要である。

また、二乗平均誤差の値が大きくなったツイートを確認すると、ツイート内に店舗名が出現した場合、誤った識別になることが多かった。これは、今回検証した土曜日のツイート総数が1日あたり約600件程度で、その内、ツイート内容に特定の店舗名を含めているツイートが多く、同じ店舗に関して発信しているツイートがほぼなかったためと考えられる。日曜日では、ツイートが「(店舗名) 梅田店」のような内容が多く存在した。「梅田店」は様々なカテゴリのツイートで出現する単語だが、そのツイートの多くがチェーン店の集中している「スイーツ、フード、コスメ」のカテゴリで出現しているため、店舗名が含まれている「梅田店」が含まれるツイートのカテゴリは「スイーツ、フード、コスメ」と誤った識別をした。改善案として、学習データを増加させて、Webページやフロアマップなどから小規模施設の店舗名を取得しておき、識別の重みを考慮する。また、今回は学習データの時間を考慮していないので、時間帯を考慮することがあげられる。

## 5. 関連研究

近年、Twitterをテキストマイニングの対象とした研究は活発に行われており、Twitterに投稿されたツイートを分析することでイベントの検出や位置情報の取得を試みた研究も数多く

ある。

Arakawaら[5]は位置情報ツイートから位置依存性の高い文字列を抽出する手法を述べている。位置情報ツイートから得たエリアを100キロ四方のグリッドに分割し、それぞれのグリッド内のツイート含有率を計算し、ツイート含有率がある閾値を超えたエリアを最終的に1キロ四方のグリッドまで走査することにより、1つのキーワードに対して複数の位置依存性を抽出することができる。この研究では、位置情報とツイートのコンテンツを対応付けている。本研究でも、位置情報とツイートのコンテンツ内容を関連付けているが、こちらは、特定の場所や建造物を中心とした位置の重要性の高い文字列の抽出を行っている。また、YamamotoらのTwitterに投稿された実生活情報から有用性の高いものを抽出し局面に応じた記事をユーザに提示するシステム[6],[7]やツイートから地震や台風などのイベントの検出を試みた研究として榊らの研究[8]がある。Twitterのタイムラインを監視しておくことでリアルタイムでイベントの検出を行い、高い精度を得られた。これらの研究では、コンテンツベースでの抽出を行っているが、本研究では、コンテンツ内容と緯度経度情報の関連付けを行っている。また、Yamaguchiら[9]は、位置情報が既知であるユーザのツイートを用いてローカルイベントの検出を行い、検出されたローカルイベントに関する発信を行った。位置情報が未知であるユーザの位置情報を推定する。Nicholsらの研究[10]は、ツイートのコンテンツ内容の変化に注目しており、更新の量の急増などでイベント内の重要な瞬間の識別を行う。Ribeiroらの研究[11]では、ツイート内容を識別し緯度経度から区域でのイベント発生を検出する。本研究では、ツイートの内容に着目し、特定の単語の出現頻度が高くなれば、イベントが発生したと見なしている。

位置情報付き画像ツイートを用いてイベント検出を試みた研究として、Nakajiら[12]はあらかじめ特定のキーワードや期間を設け、位置情報付きツイートを収集し、解析することで画像付きのイベント検出を試みた。Kanekoら[13]はイベントのキーワードをシステムにより自動的に抽出することで多くのイベントを抽出することで未知のイベントのキーワードを得られるようにし、キーワードを用いて収集した画像を解析することで、ユーザが知らないイベントでも画像により視覚的にとらえることができるようにした。これらの研究は、それぞれ位置情報ベースとコンテンツベースで別々に取り扱っているが、本研究では、この2つを同時に取り扱う。Takemuraら[14]は、Twitterユーザを、広く一般のユーザが興味を示す情報を発信するのかが、一部のユーザのみが興味を示す情報を発信するのかの範囲を示すため、対象局所性と定義される指標を用いた分類を行う手法を提案している。本研究では、Twitterの位置情報と内容に基づいて発信されたツイートが発信された場所に関連しているかを判別する。

オンライン上でのユーザ間のコミュニケーションを行う研究として、質問応答サイトの回答を対象にした研究として、Yahoo!知恵袋を対象にして知恵袋の質問回答情報をクラスタリングし、クラスごとに機械学習を行って最も質問に適した回答となりうる可能性が高い回答を判定する手法を述べた[15]や、教師つ

き負例と教師なし正例からなる学習コーパスからの SVM 学習器を作成し、不適切な回答の発見を半自動化するシステムの作成を行った [16] がある。また、ある質問に対して一つ以上の回答の組（以下、QA コンテンツと記す）は急激に増えている。QA コンテンツは質問に詳しい専門家がベストアンサーを決めているわけではなく、閲覧ユーザの投票で決定したり、質問者自らが決定するため、質問に対する回答が不十分な場合がある。そこで、高田ら [17] は Web 情報を用いてコンテンツを補完することで、QA コンテンツの利用者が回答の信憑性を確認したり、補足的な情報を得ることができる手法を提案している。本研究では、Web ページに関連するツイートの集約情報を Web ユーザに提示するシステムの構築を目標とする。

## 6. ま と め

本研究では、大型ショッピングモールや駅構内のような複合施設における各店舗などの小規模施設に関するツイートの発見を行い、ツイートに店舗名といった詳細な情報の付与し、そのツイートと関連する施設の Web ページ上に該当するツイートならびにツイートの集約情報を提示するシステムの構築を目指した。そのため、時空間情報に基づき複合施設内の店舗ごとに関連するツイートに分類する手法の提案および検証を行った。実験の結果より、本手法でのツイート分類は約 9 割が正しく分類され有効性を確認できた。

今後の課題として、他の手法においての分類精度の比較を行う。今回は学習データとテストデータが共に少なかったため、データ数の増加に伴う処理時間も考慮に入れる必要がある。また、今回はフロアごとの分類を行ったため、店舗ごとのより小規模な分類を行う予定である。次に、分類されたツイートの情報を集約し、関連する Web ページを閲覧しているユーザに情報を提示するシステムの構築を行う予定である。

## 【謝 辞】

本研究の一部は、総務省戦略的情報通信研究開発推進事業 (SCOPE) および JSPS 科研費 26280042 の助成を受けたものである。ここに記して謝意を表す。

## 文 献

- [1] Shingo Tajima, Taketoshi Ushiyama: A Method for Composing Ad-hoc Following Networks on Twitter for Sharing Information among Event Participants, *International Journal of ADADA*, Vol. 17, No. 4, pp. 199-124, 2014.
- [2] Kenta Oku, Koki Ueno and Fumio Hattori: Mapping Geotagged Tweets to Tourist Spots for Recommender Systems, In Proc. of 2014 IIAI 3rd International Conference on Advanced Applied Informatics (IIAI 2014), pp.789-794, 2014.
- [3] Yuanyuan Wang, Gouki Yasui, Yuji Hosokawa, Yukiko Kawai, Toyokazu Akiyama and Kazutoshi Sumiya: TWin-Chat: A Twitter and Web User Interactive Chat System, In Proc. of the 23rd ACM International Conference on Information and Knowledge Management (CIKM 2014), pp. 2045-2047, 2014.
- [4] 松井 優也, 河合 由起子: 人と情報の検索および相互作用を目指したソーシャルサーチシステムの研究開発, *日本ソフトウェア科学会コンピュータソフトウェア (ソフトウェア論文)*, Vol. 28, No. 4, pp. 196-205, 2011.
- [5] Yutaka Arakawa, Shigeaki Tagashira and Akira Fukuda:

- Relationship Analysis between User's Contexts and Real InputWords through Twitter, *IEEE Globecom 2010 Workshop on Ubiquitous Computing and Networks(UbiCoNet 2010)*, pp.1813-1817, 2010.
- [6] Shuhei Yamamoto and Tetsuji Satoh: Two Phase Extraction Method for Multi-label Classification of Real Life Tweets, In Proc. of the 15th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2013), pp. 16-25, 2013.
  - [7] Shuhei Yamamoto and Tetsuji Satoh: Two Phase Extraction Method for Extracting Real Life Tweets using LDA, In Proc. of the 15th Asia-Pacific Web Conference (APWeb 2013), *Lecture Notes in Computer Science 7808*, pp. 340-347, 2013.
  - [8] Takeshi Sakaki, Makoto Okazaki and Yutaka Matsuo: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, In Proc. of the International World Wide Web Conference (WWW 2010), pp. 851-860, 2010.
  - [9] Yuto Yamaguchi, Toshiyuki Amagasa, Hiroyuki Kitagawa and Yohei Ikawa: Online User Location Inference Exploiting Spatiotemporal Correlations in Social Streams, In Proc. of the 23rd ACM International Conference on Information and Knowledge Management (CIKM 2014), pp. 1139-1148, 2014.
  - [10] Jeffrey Nichols, Jalal Mahmud and Clemens Drews: Summarizing Sporting Events Using Twitter. In Proc. of the 2012 ACM International Conference on Intelligent User Interfaces (IUI 2012), pp. 189-198, 2012.
  - [11] S. S. Ribeiro, C. A. Davis, D. R. R. Oliveira, W. Meira, T. S. Goncalves and G. L. Pappa: Traffic Observatory: A System to Detect and Locate Traffic Events and Conditions Using Twitter. In Proc. of the 5th ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN 2012), pp. 5-11, 2012.
  - [12] Yusuke Nakaji and Keiji Yanai: Visualization of Real World Events with Geotagged Tweet Photos, In Proc. of IEEE ICME Workshop on Social Media Computing (SMC 2012), pp. 272-277, 2012.
  - [13] Takamu Kaneko and Keiji Yanai: Visual Event Mining from Geo-tweet Photos, *IEEE ICME Workshop on Social Multimedia Research (SMMR 2013)*, pp. 1-6, 2013.
  - [14] Hikaru Takemura and Keishi Tajima: Tweet Classification Based on Their Lifetime Duration, In Proc. of the 21st ACM International Conference on Information and Knowledge Management (CIKM 2012), pp. 2367-2370, 2012.
  - [15] 西原 陽子, 松村 真宏, 谷内田 正彦: QA サイトにおける質問に適した回答の判定, *言語処理学会 NLP 若手の会第 2 回シンポジウム*, 2007.
  - [16] Daisuke Kobayashi and Naohiro Matsumura: Automatic Gender Estimation of Bloggers' Gender, In Proc. of International Conference on Weblogs and Social Media (ICWSM 2007), pp. 279-280, 2007.
  - [17] 高田 夏希, 山本 裕輔, 小山 聡, 田中 克己: 質問応答コンテンツに対する Web による回答補完, 第 1 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2009), C4-6, 2009.