複合施設におけるツイート分析に基づくタグクラウド生成および可視化

安井 豪基 † 坪井 結香 † 岡山 愛 † 河合由起子 †
王 元元 †† 秋山 豊和 †

† 京都産業大学 〒 603-8047 京都府京都市北区上賀茂本山
†† 山口大学大学院理工学研究科 〒 755-8611 山口県宇部市常盤台 2-16-1
E-mail: †{i1458085,kawai}@cc.kyoto-su.ac.jp, ††{g1244758,g1344270,akiyama}@cse.kyoto-su.ac.jp,
†††y.wang@yamaguchi-u.ac.jp

あらまし 本研究では，ショッピングセンターや高層ビルのような複数の小規模施設が存在する複合施設内で発信されたツイートを分析し，各店舗などの小規模施設に関するツイートの特徴語を用いてタグクラウドの生成を行い，関連する施設の Web ページ上に関連するツイートをマッピングすることで，そのページ上に該当するツイートならびにツイートを集約したタグクラウドを提供するシステムの構築を目指す．我々はこれまで，複合施設を対象にその施設全体に対するツイートを複合施設の場所名に基づき施設の Web ページにそれらのツイートを提示するシステムの構築および各店舗のツイート分類の検証を行ってきた．本論文では，場所や時間帯で変化するツイートの特徴語に注目し，時空間情報に基づくツイート分析による各フロアや小規模施設に関するのツイートのタグクラウド生成・可視化システム，およびタグクラウドとなる特徴語の時間変化に伴う相関分析を行う．

キーワード ツイート発見，タグクラウド，集約情報提示

1. はじめに

近年，スマートフォンの普及により時間や場所を問わずに，Twitter(注1) や Forsquare(注2) の様なソーシャルネットワーキングサービス（以降，SNS と記す）を通して，建造物やイベントなど場所に関連する情報が発信されている．任意の時間や場所で情報を発信することができるため，現地の情報がリアルタイムで発信されることとなり，Web ページの情報と比較すると即時性の高くなり，有益な情報と言える．しかし，そのような情報はリアルタイムで随時発信されており，膨大な情報の中から自分の関心のある話題の情報を取得することは困難である．ユーザの目的に合わせてツイートの内容やハッシュタグで検索を行い関連するツイートを取得する手法 [1] があるが，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取得する可能性がある．[2] では，位置情報付きツイートの緯度経度情報から目的の場所付近で発信されたツイートを収集することで，関連性の高いツイートを取得しているが．都市部などツイートの発信が密集している場所においては，様々な話題が存在するため位置情報だけでは目的のツイートの発見が難しい．このように，目的のツイートを発見するには手間がかかってしまう問題点の解消が課題として上げられる．

我々はこれまで，複合施設などの様々な話題が密集している場所で発信されたツイートに対して，ツイートの内容からクラスタリングを行い，高さ情報を付与し Web ページの内容とマッピングすることで，施設全体の話題だけでなく店舗といった小規模施設ごとに関連するツイートが閲覧可能なシステムの提案および検証を行ってきた．本研究では，場所や時間の流れで変化する特徴語を抽出しタグクラウドの生成を行い，場所に関連する Web ページ上を検出し，その Web ページ上に該当するタグクラウドを提示する．Web ユーザはタグクラウドを選択することで，タグクラウドに関連するツイートを閲覧することができる．これにより，Web ユーザはツイートの話題が様々である場所でも，ユーザの目的の話題のツイートを発見しやすくなりユーザの情報支援につながる．

本論文では，具体的に以下の 2 点を実現する．

- 時空間分析に基づくツイート集合のタグクラウド生成
- ツイートやタグクラウドを用いた可視化システムの構築

提案システムでは，ユーザが閲覧している Web ページに関連するツイートから生成されたタグクラウドが提示され，タグクラウドを選択することでその単語に関連するツイートを提示する．これにより，ユーザはタグクラウドを見ることで，そのページに関連性が高い単語を確認でき，関連のあるツイートを見ることで場所の感想や現状などの情報を知ることができ，情報の網羅性の情報に繋がる．

本論文では，時空間情報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内容によるタグクラウド生成方法について述べる．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 章で位置情報付きツイートデータの分析手法，時空間に基づくツイートのクラスタリング，および，ツイート内容に基づくタグクラウド生成方法について述べる．4 章で提案した手法の検証をし，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．

(注1)：https://twitter.com/
(注2)：https://ja.foursquare.com/

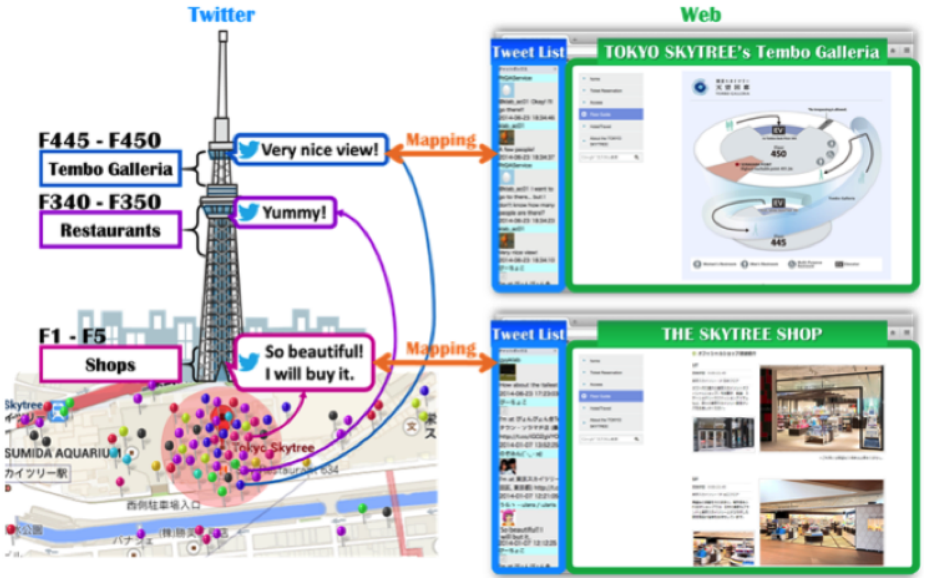

図 1　異種メディア横断型コミュニケーション支援システムの概要図

2. システム概要

本研究は，場所に関連するツイート情報の取得ならびに，場所や時間帯ごとの特徴語からタグクラウドを生成しユーザが目的に応じてツイートを閲覧できる可視化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，ツイートの内容と発信場所の位置情報に基づき，発信場所に関連する内容のツイートを抽出する．抽出されたツイートを話題ごとにクラスタリングし，その内容に関連する Web ページを検出し，Web 上に関連するツイートを提示する．Web ユーザ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．また，ツイートの内容に基づく特徴語をタグクラウドとして提示し，タグクラウドを選択することでその特徴語に関連するツイートだけを閲覧することができる．これにより，タグクラウドよりユーザの関心を持ったツイートだけをすぐ閲覧することができる．

システムの例を挙げると，東京スカイツリー付近にいる Twitter 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，タグクラウドに「待ち時間」「混雑」などが提示され Web ユーザは東京スカイツリーの混雑具合などの現状を知ることができる．

図 2 に処理の流れを示す．本研究では，Web ユーザが閲覧している Web ページに関連するツイートをリアルタイムで更新するため，リアルタイムに発信されているツイートならびに Web ユーザがアクセスしている Web ページの URL を取得する．サーバは発信されたツイートを取得し，ツイート内容と位置情報に基づきクラスタリングを行い関連性のある Web ページとの対応付け，およびにツイート分析に基づくタグクラウド生成および管理を行う．取得した関連ページに Web ユーザがアクセスすると，対応するツイートおよびにタグクラウドをブラウザへ送信および提示する．なお，通信を行うために Web ユーザは提案システムのアドオンを用いる必要がある．

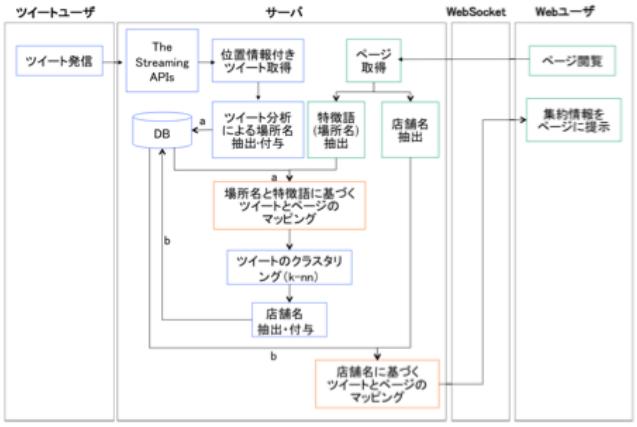

図 2　集約情報提示までの処理流れ

3. 時空間情報に基づくツイート分析およびタグクラウド生成

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 Twitter Developers の Streaming API 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 vervion 3(注1) を用いて，半径 lm の場所名を取得した．評価実験では，

(注1)：https://developers.google.com/place/

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，名詞と形容詞の単語を取得する．

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う．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．

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

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．

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

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 Streaming API 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 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．

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

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．

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

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，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．

3.6 タグクラウド生成のための特徴語抽出

タグクラウドを生成するために，前節のクラスタリングされたツイートを用いてクラスごとの単語に重要度の重みを算出し特徴語抽出を行う．TF-IDF に基づき，各クラスを 1 つのドキュメントとして各クラスごとに下記の式より，単語に重要度を付与する．

$$\text{TF}=\frac{\text{単語 } i \text{ の出現回数}}{\text{すべての単語の出現回数}}$$
$$\text{DF}=\frac{\text{単語 } i \text{ が出現したクラス数}}{\text{総クラス数}}$$

これにより，重みが付与された単語を用いてツイート内容に関連するタグクラウド生成を行う．同じ場所・時間帯で発信されたツイートに含まれる単語を取得し，特徴語の重みをフォントの大きさとしタグクラウドを生成する．

4. 実装および検証

本研究では，ツイートやタグクラウドを用いた可視化システムの構築を目的としている．本章では，タグクラウド生成における時空間情報に基づくツイート分類の検証を行う．なお，2015 年 7 月 13 日から 2015 年 12 月 17 日まで Twitter Developers の Streaming API で日本全国の位置情報付きツイートを取得した．

表 1　実験データ 1

大阪駅（LUCUA）	
中心緯度経度（34.703289，135.496242） 取得範囲半径 d = 200m	
時間帯 (2015 年 10 月)	ツイート数
6 時-9 時 (早朝)	847
9 時-12 時 (午前)	1007
12 時-15 時 (お昼)	1209
15 時-18 時 (午後)	1496
18 時-21 時 (夕方)	1734
21 時-24 時 (夜)	1073

表 2　実験データ 2

東京ディズニーリゾート	
中心緯度経度（35.6290692，139.8829573） 取得範囲半径 d = 800m	
時間帯 (2015 年 8 月)	ツイート数
6 時-9 時 (早朝)	965
9 時-12 時 (午前)	1745
12 時-15 時 (お昼)	1538
15 時-18 時 (午後)	1518
18 時-21 時 (夕方)	1498
21 時-24 時 (夜)	1144

表 3　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	ブックス，ライフスタイルグッズ
1F〜8F	レディス，メンズファッション
B1F	スイーツ，フード，コスメ

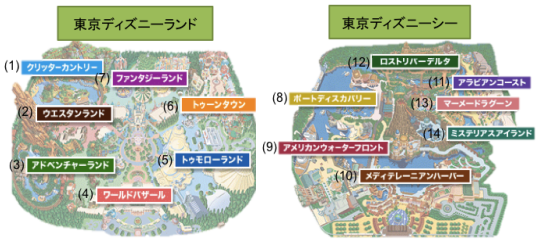

図 3　東京ディズニーリゾートの 14 つの小規模施設のカテゴリ

提案手法では，施設を中心から半径 dm 内で発信されたツイートを対象として，ツイートをカテゴリごとにクラスタリングし，カテゴリごとの重要度の高い単語をタグクラウドとし抽出する．よって本実験では，カテゴリごとの各時間帯のツイートに出現する単語の重要度を算出し，各時間帯の単語のランキングをスピアマンの順位相関係数により出現する単語を比較し，時間推移で出現する単語の変化を検証をする．

検証対象は，大型ショッピングモール「LUCUA」および「東京ディズニーリゾート」を中心とした各々半径 200m 内, 半径 800m 内で発信したツイートとし，期間は 2015 年 10 月および 2015 年 8 月の一ヶ月にて発信されたツイートとした (表 1, 表 2)．また,「LUCUA」内の各階ならびに「東京ディズニーリゾート」の小規模施設をクラスとした (表 3，図 3)．

教師データのクラスは，20 代の大学生 13 人による主観的評価に基づき，ツイート内容が各クラス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クラス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．

この教師データを用いてツイートを各クラスに分類し，各クラスで出現する特徴語の重要度のランキングを算出した．

4.1　時間変化による特徴語抽出の検証

本実験では，各フロアにおいて時間推移におけるツイートの特徴語の変化を検証する．時間帯ごとのランキングの上位 30 件の特徴語をスピアマンの順位相関係数を用いて比較する．比較する時間帯は，早朝と午前 (6 時-9 時と 9 時-12 時)，午前とお昼 (9 時-12 時と 12 時-15 時)，お昼と午後 (12 時-15 時と 15 時-18 時)，午後と夕方 (15 時-18 時と 18 時-21 時)，夕方と夜 (18 時-21 時と 21 時-24 時) で行う．

実験データ 1(LUCUA) のスピアマンの順位相関係数の結果を図 4 に示す．1F は 6 時-9 時と 9 時-12 時のような同じ午前中と夕方夜間での相関が約 0.4 となった，また，9 時-12 時と 12 時-15 時のような午前中と正午や 12 時-15 時と 15 時-18 時正午と夕方の相関が低くなった．9F，10F は正午以降の相関はやや高いが，同じ午前中でも相関が低く，午前中ではなく正午以降にフロアに関連するツイートが増えることが確認できた．

全体でみると，夕方の相関が 0.4 に収束しているが，それ以外はばらつきがあり，全体的な相関は低いため，フロアごとに時間帯によってツイートの話題が変化し，重要度の高い単語においては同じ単語がほとんど出現しないことが確認できた．出現する単語を確認すると，全体的にフロア，時間帯ごとに出現する単語が大きく異なる場合が多かった．相関の低い時間帯では，フロアに関連する内容ではない単語の出現が多く見られた．例えば，レストランフロアにおいてお昼 (12 時-15 時)，夕方 (18 時-21 時) では食べ物に関する単語が多く出現しているのに対し，6 時-9 時では食べ物に関連する単語が出現しなかったため，相関が低くなったと考えられる．これはフロアの店舗がまだ開店前のため，関連するツイートがなかったためである．各フロアで抽出された単語は，そのフロアに関連する単語が多く含まれることが確認できる．また，同じフロアでも時間帯が変化することで特徴語と変化することが確認できた．

また，東京ディズニーランドにおける結果を図 5，図 6 に示す．相関はエリア (1) や (9) が約 0.4 程度と低い相関がみられ，それ以外では，0.2 以下とほとんど相関がみられなかった．このことから，LUCUA と同様に時間帯で出現している特徴語が変化していることが確認できた．

次に，異なる時間帯での順位相関係数を検証した．具体的には，早朝とお昼 (6 時-9 時と 12 時-15 時)，早朝と夕方 (6 時-9 時と 18 時-21 時)，早朝と夜 (6 時-9 時と 21 時-24 時)，お昼と夕方 (12 時-15 時と 18 時-21 時)，お昼と夜 (12 時-15 時と 21 時-24 時) の相関を比較した．LUCUA における結果を図 7 に示す．結果を見ると，9F，10F における午前との相関が低く，それ以外は 0.3 程度となった．これにより，時間帯，フロアごとに出現する単語が大きく変化することが確認できた．

また，東京ディズニーランドにおける結果を図 8 と図 9 に示す．エリア (1) や (9) では低い相関がみられ，それ以外ではほとんど相関がみられなかった．また，エリア (2) の早朝とお昼の相関が高くなっているが，これはツイート数が少なく話題がほぼ一致していたために相関が高くなった．これにより，時間帯，エリアごとに出現する単語が変化していることが確認できた．

次に，LUCUA で出現した重要度の高い単語群を表 4 に示す．下線は，各フロアにおいて異なる時間帯で複数回出てきた単語を示し，太字は，各フロアに関連する単語を示している．午前

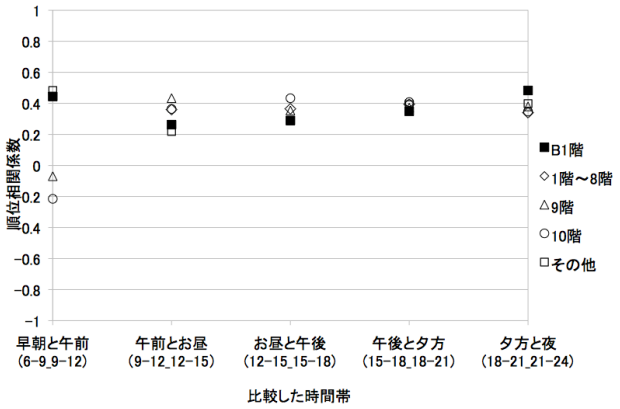

図 4　時間推移による時間帯ごとの spearman の順位相関係数 (LUCUA)

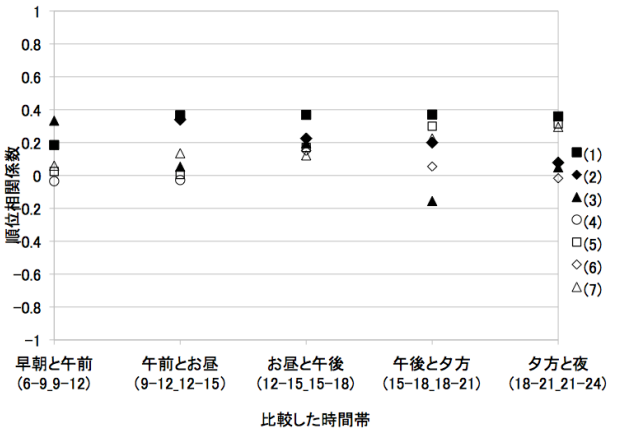

図 5　時間推移による時間帯ごとの spearman の順位相関係数 (TDL(1)～(7))

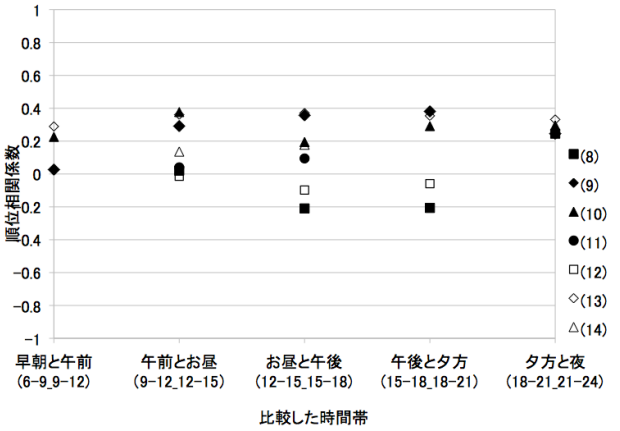

図 6　時間推移による時間帯ごとの spearman の順位相関係数 (TDL(8)～(14))

中，正午，午後で出現する単語が大きく変わったことが確認できる．またレディス，メンズファッション (1～8 階) では，「いちご」や「マシュマロ」など重要度の高い単語として食べ物に関するものが出現している．これは各フロアにあるカフェやフードコートで発信されたツイートであると考えられる．

また，東京ディズニーリゾートで抽出された重要度の高い単語群を表 5 に示す．LUCUA よりも，重要度の高い単語として

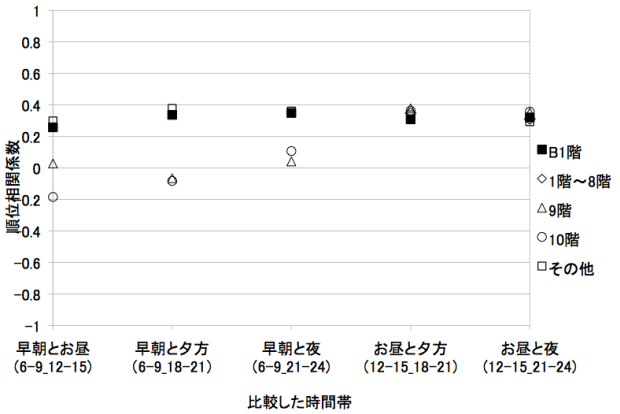

図 7　異なる時間帯ごとの spearman の順位相関係数 (LUCUA)

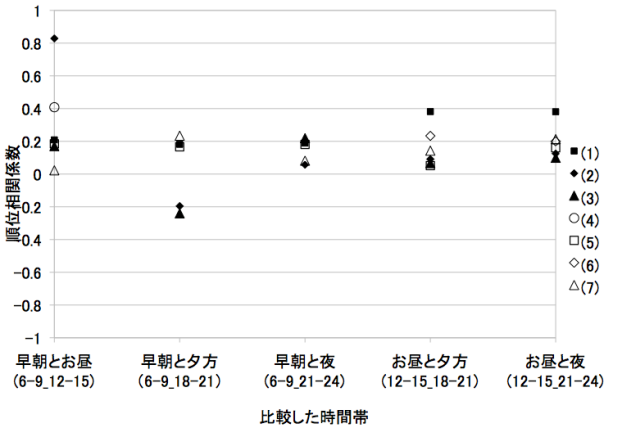

図 8　異なる時間帯ごとの spearman の順位相関係数 (TDL(1)～(7))

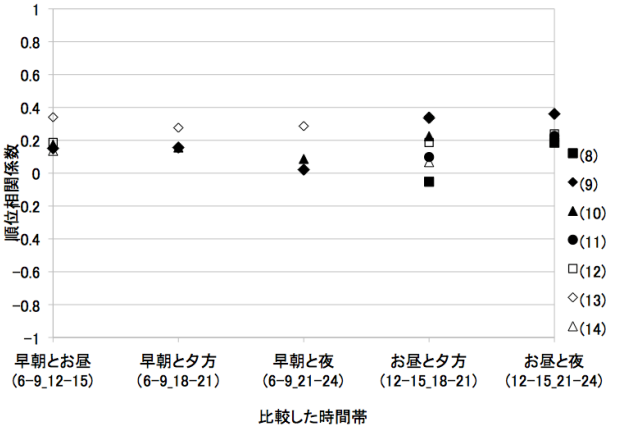

図 9　異なる時間帯ごとの spearman の順位相関係数 (TDL(8)～(14))

関連性の高い単語が多く出現していることが確認できた．一方，各エリアにおいて異なる時間帯で複数回出現している単語は，LUCUA よりと比べて多かった．また，(3) アドベンチャーランドのエリアではお昼と午後 (12 時-15 時と 15 時-18 時) においては，「クリスタル」，「パレス」といったレストランの名前が出現しており，時間帯に合わせて特徴語的な単語が抽出されていることが確認できた．

全体でみるとおおむね，LUCUA と同様に時間帯の変化に伴いエリアごとに特徴語を抽出できた．

表 4　各フロアのツイートから抽出された重要度の高い単語群 (上位 15 件)

フロア＼時間帯	6 時〜9 時 (早朝)	9 時〜12 時 (午前)	12 時〜15 時 (お昼)	15 時〜18 時 (午後)	18 時〜21 時 (夕方)	21 時〜24 時 (夜)
コスメ, フード, スイーツ (B1 階)	**ブルズ**, ロック, こと, オープン, **スタバ**, **スターバックスコーヒー**, ステータス, テラス, 来週, 男子, 笑笑, 系, アントレ, コート, セントラル	ミナス, ルミナリエ, ルー, **屋台**, 金, 毎度, 靴, 靴磨き, なし, ぼ, 上, 手, 時期, 洋服, 活躍	朝, ハウス, 蘭, イメージ, スタッフ, チャンス, ファミマ, 七夕, 今月, 天の川, 存在, 平日, 強い, 新月, 男性	いいね, 一同, 下, 夢, 夫婦, 本屋, 茶目だ, 週間, アロイ, カオマンガイ, 火, ひととき, **アールグレイ**, **エッグ**, **タルト**	丸, **珈琲**, 但馬, **ごちそう**, **パラダイス**, 安心, 福, スーパー, キャンプ, シャベル, スプーン, 参上,**オモニ**, **カウンセリング**, スペース	観光, 福, つら, もと, **ハイボール**, マーケット, テンション,**オモニ**, 専門, クラフト, ホビー, 売場, 東急, 際, **お好み焼き**
レディス, メンズ ファッション (1〜8 階)	—	デザート, **ハロウィン**, 甘い, 食後, かんてん, ほんとだ,**イヤリング**, 今頃, 水上, でんしゃ, **雑貨**, **セレクション**, 期間, 現在, **デザイン**	あかり, 週, 豚まん, やつ, ゼミ, 兄さん, 姉さん, 学期, 学校, 応援, 授業, 新しい, 東, 薩摩, えびす	**いちご**, サリー, ジュー, **スムージー**, **バナナ**, あかね, **ユニクロ**, チャップリン, 心臓, みなさま, **ガールズ**, 客, 無事だ, こいつ, **ネイビー**	ジム, ゆるい, ダーク, **マシュマロ**, **スープ**, 中毒, みせ, 代官, 山, 念, 美, 位置, 刀, 意味合い, 新郎	みーちゃん, 次, とゆう,**イヤリング**, **ガールズ**, **コレクション**, 催事, 先ほど, **入荷**, 数, **限定**, す, みなさま, れい, タイプ
ブックス, ライフ スタイルグッズ (9 階)	外, 時間, マルチメディア, ヨドバシカメラ, 利用, 梅田, 大阪	からあげ, 空, クラシカル, 二, 相手, 真心, く, ちゃー, **紀伊國屋**, マルシェ, デザイン, 優しい, 幸せだ,うめ, なまけもの	**紀伊國屋**, **フロッピー**, いそ, げい, 寄り道, うち, しと, なま **イヤホン**, ショー, ジム, 入院, 手続き, エボルタ, 周年	トンカチ, ペンチ, 不器用だ, 完成, タッセル, ファー 上品だ, 主張, 季節, 春先, いっか, ショート ブレッド, 本店, 眼	電球, うるさい, かわいそうだ, やけど, **スピーカー**, 客, 放送, 普段, 子,からあげ, 勝者, 有終の美, かいもの, かや, にんにく	上, 飲食 ,うめ, **ケーブル**, コーデ, タリーズ, 満足, 私服, 居心地, **レンズ**, たけ, ヘルプ, 寒い, 彼, 眠い
レストラン (10 階)	**キッチン**, ルクア, 北, 大阪	エンゲージリング, 個, 展, 庭園, 目黒, 美術, いっぱいだ, スタッフ, 帰り道, 昨夜, 楽しみだ, 誕生, さ, ダイヤモンド, チェーン	**腹ごしらえ**, **スンドゥブ**, **山芋**,**韓国**,**たこやき**, 樹, 流,リフレッシュ, たけ, 釜, **カオマンガイ**, **キッチン**, **ツリー**, **マンゴ**, **料理**	ボルガ, 池,リフレッシュ, **腹ごしらえ**, おしい, アクセサリー, パール, 仕方ない, 具合, 右下, 運命,**たこやき**, **ねぎま**, まあまあだ, 懇親	**回転**, **寿司**, カプリチョーザ, ずみさん,**たこ焼き**, アタリ, **あまい**, ほとり, あつ, みん, アズナス, 上司, 口コミ, 専務, 社長	いす, むら, ジャンプ, 練習, 薩摩, **クッキング**, スタジオ, **パスタ**, **ラザニア**, 休日, 体験, 教室, **料理**,**韓**, 刺

図 10　システムのインタフェース (全ての時間帯を選択)

図 11　システムのインタフェース (6 時-9 時の時間帯を選択)

4.2　特徴語を用いたタグクラウド生成

システムのインタフェースを図 10，図 11 に示す．重要度の高い特徴語をタグクラウドとして表示し，その場所に合った特徴語を提示する．図 10 では，ユーザが東京ディズニーランドの Web ページを閲覧している場合であり，右下に東京ディズニーランドに関連する全てのツイートが表示され，左下に全ての時間帯におけるタグクラウドが表示されている．図 11 では，時間帯を 7-11 時を選択した場合であり，タグクラウドが 6-9 時における特徴語が表示される．また，タグクラウドの「開園」を選択することで右下に「開園」に関連するツイートのみが表示される．これにより，ユーザは関心のある特徴語に関連するツイートをすぐに閲覧することができる．

5.　関 連 研 究

近年，Twitter をテキストマイニングの対象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

表 5　東京ディズニーリゾートの各場所のツイートから抽出された重要度の高い単語群 (上位 15 件)

時間 / 場所	6 時〜9 時 (早朝)	9 時〜12 時 (午前)	12 時〜15 時 (お昼)	15 時〜18 時 (午後)	18 時〜21 時 (夕方)	21 時〜24 時 (夜)
(1) クリッターカントリー	ファス, カウンター, **大量だ**, ひと, 三, マイ, シャツ, 二, 多い, 一, **スティッチ**, 人	閉店, 信者, 先週, 前後, 単語, 広, 溺死, 自分, カーペット, ジャスミン, フライング, どん, 周り, カフェ, バナナ	最近, 沿い, カリブソ, かっこいい, 限定, バイユー, 川, 探検, **ヤバイ**, 場所, ちょ, なん, 手, セバスチャン, 子	傘, 邪魔, 京成, 新だ, いけ, さい, 兄ちゃん, 後ろ, 鑑賞, 上, 兄弟, 晴, 秀夫, 静止, **スプラッシュマウンテン**	甥, コスト, 丈, 上, 主張, 激しい, 美, 谷, 首, シンドバッド, スタンバイ, ブック, 切り札, 早計, 海底	えび, おやつ, スタンバイ, ピーナッツ, 危険だ, 好調だ, 批判, **マウンテン**, たで, カスト, ファン, ほう, チェック, 木, 猪
(2) ウエスタンランド	**ビッグだ**,**サンダー**, **マウンテン**,千葉,浦安	**ビッグだ**, **ひどい**,**サンダー**, ファスト, パス,**マウンテン**, の,千葉,浦安	**ビッグだ**,**サンダー**,**マウンテン**	**ビッグだ**, ラスト,**サンダー**, 前, **アトラクション**	嫁, **ヤバイ**, ます, 土産, 屋, いま, **子供**, **サンダー**,**ビッグだ**, 分	草,**ビッグだ**,**サンダー**, **マウンテン**,浦安,千葉
(3) アドベンチャーランド	**アロハ**,**プレゼンツ**, **魅惑**,**クルーズ**, **ジャングル**,**カリブ**, **マイ**,**スティッチ**	**１５**,**カリブ**,**海賊**, きょう, **スタンバイ**, **クルーズ**, 階段, スプラッシュ	バカだ, 親, 故, 英, 里, て, **ほうき**, **クリスタル**,**パレス**, 掃除, **水たまり**, カウンター, タートルトーク, 版, 長男	カップ, スーベニア, 遠出, **クリスタル**,**パレス**, **プレゼンツ**,**魅惑**, らん, ネズミ, 宇宙, **海賊**,**アロハ**, 初, **カリブ**,**スティッチ**	**アドベンチャー**, クリスタルスカル, ジョーンズ,宮,魔, **プレゼンツ**,**魅惑**,**アロハ**, **マイ**,インディ,**クルーズ**, **ジャングル**,**スティッチ**	ラッピング, 相談, ライン, **ジャンル**,**海賊**, タグ,**カリブ**, 黒い, **アドベンチャー**,インディ, クリスタルスカル, ジョーンズ,宮,魔, 筋肉
(4) ワールドバザール	**パレード**,**ルート**, **城**,千葉,浦安	パートナー, **像**, 待機, 通り,**ルート**, **パレード**, 裏, ない,**城**, 抽選,千葉,浦安	放水,**ルート**, **パレード**, 今回, **城**, 夏, **楽しい**, 千葉,浦安	**ルート**,**パレード**, **城**,浦安,千葉	—	—
(5) トゥモローランド	**インク**,**モンスターズ**, **ライド**,**コンティニュー**, **スター**,**エンカウンター**, あん, **スティッチ**, 入口, 千葉, 浦安	ホール,**アストロ**, **イヤー**,**プラスター**, ばあい,**ライト**, **エンカウンター**, 昼食, 空,**トゥモロー**, **スペース**, たん	ほど, クラッシュ, **ダースベーダー**, みこ, 体力, 卒業, 海外, **トゥモロー**, 美女, 美男, 高い,**インク**, テラス, **モンスターズ**,**ライド**	まま,**トゥモロー**,**イヤー**, 点, **インク**,**モンスターズ**,**ライド**, **エンカウンター**,**コンティニュー**, **シアターオーリンズ**,**スター**, **スペース**,**アストロ**, **プラスター**,**ライト**	**トゥモロー**, 飯,**アストロ**, **イヤー**,**プラスター**, 江ノ島, 行き先,**コンティニュー**, から, オシャレだ, ゴリラ, バターン, ヒャダイン, 初回, 勤務	**スペース**,**アストロ**,**イヤー**, **プラスター**, 保育, 売り物, 失格, 立場,**インク**, **コンティニュー**, **モンスターズ**,**ライド**, **マウンテン**,**ライト**,**トゥモロー**
(6) トゥーンタウン	—	本日, 年,国,夢,ディズニー, ディズニーランド,東京	ほしい, 胃,**ガジェット**, **カートゥーン**, スピン, **ロジャーラビット**, 家, **ミニー**,千葉,浦安	**ガジェット**,娘,浦安,千葉	リクエスト,**ガジェット**, ノリ, 風, **たのしい**, 国, 一, 雨,夢, いい, ディズニー, ディズニーランド, 東京,千葉,浦安	まどか, 即決,**ガジェット**, いこう, 連絡, **悪い**, 娘,国,夢, 明日, ディズニー, ディズニーランド, 東京,浦安,千葉
(7) ファンタジーランド	**ハニー**,**ハント**,**プーさん**, **ホーンテッドマンション**, **スモール**,**ワールド**, 分, 千葉, 浦安	短い, **面白い**, あほ, **ものすごい**, ウォッチング, 人間, 金子, 一也, 意志, 横濱, 炸裂, 藤田, 野, 魂,**ピノキオ**	**冒険**,**スモール**,**ピノキオ**, 旅行,**クイーン**, **ハート**, **ケット**, **バン**,**ワールド**, **ホール**, シンデレラ, **ハニー**,**ハント**,**プーさん**	**ハニー**,**ハント**, おでこ, 今度,**プーさん**, あご, かくし, 仲良し, 青春, **ファンタジー**, **ホーンテッドマンション**, **スモール**,**ワールド**	体勢, バザール, ジョー,**冒険**, **ワールド**,**ピノキオ**, ものすごい, 家族, **ホーンテッドマンション**, **ハニー**,**ハント**, 前, 私	ハチミツ, おかげ,**女王**, 門, 雪, ゲーム、 ランニング, 下位, 割, ごう,**ハニー**, **ハント**, ループ, 充, ふう
(8) ポートディスカバリー		ジェリーフィッシュ, ジャンピン, プール, ワール,**ストームライダー**, インディジョーンズ, タワーオブテラー, レイジングスピリッツ, **アクアトピア**, アドベンチャー,千葉,浦安	**ストーム**, 娘,スコール, ヤバい, スタンバイ, 並,**ストームライダー**, 雨,千葉,浦安	**ストームライダー**, 出港, 波, 海中, 潜水, 艦, 間, 泡, 時間,浦安,千葉	適当だ,**ストームライダー**, 休憩, ところ,千葉,浦安	**キャプテン**, **デイビス**, 人生,**ストームライダー**, ラスト, アトラクション
(9) アメリカンフロント	**タワー**,**テラー**,**マニア**, **ストーリー**, 年, ディズニーシー, 千葉, 浦安, 東京	**ビッグバンドビート**, ごはん, 優雅だ, うまい, 腹, 山野, 組, おかげ, 無事だ, 爆睡, **タワー**,**テラー**, ご飯, 朝,マニア	アイドル, まーさん, 童貞, 代表, 安藤, 弥, 立花, 龍, **タワー**,**テラー**, もろい, 泊, 流, 焼き魚, 遠征	階,**タワー**,**テラー**, 冬季, 物足りない, おばちゃん, パーカー, 撮影, ネックレス, 今週, **ビッグバンドビート**, 青春, きょう, バージョン, パワー	さむい, つう, ちび, ぐう **ビッグバンドビート**, 合致, やだ,**タワー**,**テラー**, 安定,**ストーリー**, **タワーオブテラー**, テンション	もち, 今月, ワン, 不安定だ, 前日, 片手, 開始,**タワー**, **テラー**, 天気, 一昨日, 活動, 個, 2, 怖い
(10) メディテレーニアンハーバー	ものすごい, 休日, しい, 変更, 期間, ず, ランド, の, 人, 夏, 日, ディズニーシー, 東京	奇跡, こいつ, 場面, 実行, 有言, 大詰め, 会見, 旦那, 男性, 留守, 発表, 自由だ, 舞台, 陣	イベント, 活, キール, グラン, ブッフェ, ロワイヤル, ゆめ,グランデ,シェラトン, トーキョー, 海, ワイン, 満喫,グランデ, シェラトン, 最後だ	おいしい, もったいない, ブルース, 滝, リング, うずうず, カクテル, かな, ど, パーク, 中, グランデ,シェラトン, トーキョー, プール	あやめ, ぐみ, 乾杯, 学生, 生活, 撮影, ダブル, 中島, 人工, 山田, 耳, ブッフェ, 充実, 本気だ, デート	クリスマス, アンバー, サーダ, チャージ, 思ひ, 白黒, イディオム, 満喫, 式, エリック, ワンピ, 来年, 好きだ, かっこよい, ハピネス
(11) アラビアンコースト		**カーペット**,**ジャスミン**, **フライング**,**フード**, お腹, **シンドバッド**,**ブック**, ブックバナナ,**コート**, **ストーリー**,千葉,浦安	仮眠,**ブック**, カレー,**コート**, 種,**シンドバッド**,**フード**, **ブック**,**カーペット**, **フライング**, **ジャスミン**,**ストーリー**	**カーペット**,**フライング**, **ジャスミン**,浦安,千葉	**フード**,**コート**,千葉,浦安	**シンドバッド**,**ブック**, **ストーリー**,浦安,千葉
(12) ロストリバーデルタ	**レイジングスピリッツ**, 千葉,浦安	**レイジングスピリッツ**, **インディ**,**クリスタルスカル**, **ジョーンズ**,**宮**,**魔**, **アドベンチャー**	**レイジングスピリッツ**, クリア, 回転, コースター, 度,**クリスタルスカル**, **ジョーンズ**,**宮**,**魔**, **アドベンチャー**,**インディ**, 千葉,浦安	ループ,度, ふう, ファスト, パス, うまい,**アドベンチャー**, **インディ**,**クリスタルスカル**,**ジョーンズ**, **宮**,**魔**,**レイジングスピリッツ**	**レイジングスピリッツ**, 千葉,浦安	**レイジングスピリッツ**, 浦安,千葉
(13) マーメイドラグーン	一般, 成功, そだ, ろ, **マーメイド**, **ラグーン**, サプライズ, マニア, 楽しみだ, ストーリー, 早い, 時間, の, ディズニーシー, 東京	**トリトン**, いいね, 郡, ブロードウェイ, ミュージック, トリプル, ハピネス, パスポート, 以内, 無料,マジック,ランプ, 久々だ, ふた, 満足	けん, たい, レイパン, 水浴び, 都会, 成功, 足先, オイラ, 帽子, サングラス, 我, マジック,ランプ, ブロードウェイ,ミュージック	一行, 次男, 津田, アルコール, 嫌いだ, 後押し, レイン, ロス, 化粧, 整理, 集合, 1, アリス, ティー, パーティー	先生, 完成, 室井, 吉, 客, 悪化, 無視, 鈴木, ウォーシップガンナー, オウガ, ソフト, タクティクス, パパ, **リトルマーメイド**, 三國	スローモーション, 展開, エルフ, ナイフ, ブルック, マック, 一般, 大丈夫だ, 需要, GM, もっか, 今頃, 必死だ, **カリブソ**, **キッチン**
(14) ミステリアスアイランド	**アース**,**センター**,千葉,浦安	**アース**,**センター**,千葉,浦安	加速, 運行, 並, **アース**, 中止,**センター**, すごい, ところ, あと, ファスト, 分,パス, 千葉,浦安	**アース**,**センター**,浦安,千葉	劇,**アース**,**センター**, ファスト,パス,千葉,浦安	**アース**,**センター**, ラスト,浦安,千葉

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，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，クラスタ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，複数施設内で発信されたツイートを分析し，各店舗などの小規模施設に関するツイート発見ならびに関連するツイートの特徴語を用いてタグクラウドの生成を行い関連する施設の Web ページ上に関連するツイートをマッピングすることで，そのページ上に該当するツイートならびにツイートを集約したタグクラウドを提供するシステムの構築を目指した．そのため，時空間情報に基づくツイート分析による各店舗のツイートのタグクラウド生成の検証を行った．実験の結果，場所や時間帯ごとにツイートの特徴語も変化することが確認できた．

今後の課題として，Web ページからフロア情報の自動抽出手法の検討，デモシステムの公開発表が挙げられる．

[謝　　辞]

本研究の一部は，総務省 SCOPE（ICT イノベーション創出型研究開発），JSPS 科研費基盤研究 (B) (26280042) および基盤研究 (C) (15K00162) の助成を受けて実施された．ここに記して謝意を表す．

文　　献

[1] Shingo Tajima, Taketoshi Ushiama: 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: TWinChat: 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 Classication 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 he 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.