

SNS 情報に基づく都市・建築空間像の解析

○菊地 弘祐*¹ 木原 己人*² 遠田 敦*³ 高柳 英明*⁴
木村 謙*⁵ 林田 和人*⁶ 渡辺 仁史*⁷

キーワード: LBSN Twitter kh coder Gephi 駅商圈 ライフスタイル センシング

1. 背景

1.1 都市・建築分野での調査手法

今まである場所の特性や人の行動特性を明らかにするために、さまざまな調査手法が提案されてきた。例えば、PT 調査は広く使われるある場所の利用者が移動経路を明らかにすることで、計画に役立てようとするものである。他にも、都市計画分野ではキャプション評価法¹⁾ 2) や都市のイメージ³⁾、さらにはゲニウス・ロキ⁴⁾ など様々な都市の背景、利用者のオーラル・ヒストリが活用されている。これらの調査手法は一定の成果を収めたものの、量と質のどちらかを限定するという制約がある点が問題であった。前者の例でいえば、量を計測することが可能であるが、利用者ごとの背景まで踏み込んで利用者を分けたいという調査することが出来ない。同様に後者では、調査する人をあらかじめ絞ってしまうため、これらの参加者自体にバイアスがかかっていると考えられる。そのために、新しい手法を提案するべきであると考えられる。

1.2 都市・建築のセンシング手法

都市・建築分野にて質と量を止揚しうるものの1つはセンシングである。

情報化やセンサーノードの小型化により、都市・建築分野においても徐々にセンシングを行い、莫大な情報量から質も担保する手法が提案されてきた。一つは、センサーノードを点在させる方法である。これらの方法は新規にセンサーノードをユビキタスに点在させ、その情報網から量と質を生み出そうとするものであった。ボストンでは、ハーバード大学とベンチャー企業⁵⁾ が共同で気象情報に関するデータをセンシングするノードを点在させ、公開している。また、遠田ら⁶⁾ は住宅にセンサーを点在させ、環境情報の可視化を試みている。この手法の大きな欠点としては、センサー作成に多大なコストを伴うという点である。たとえば、遠田らの例では、一つのセンサーノードを作成するのに1万円ほどかかっている。

これを街全体に点在させるとなると莫大な費用になる。

他の方法としては、既存にある情報のうちGPSなど地理情報の含んだものをセンサー情報としてみなす方法である。MIT SENSEable City Lab⁷⁾ では、携帯電話会社と共同で、発信時間と電波基地局から割り出した場所を蓄積して、一般公開している。また、同様に東日本大震災の後にNTT docomo⁸⁾ がモバイル空間統計という手法を用いて帰宅困難者の行動を可視化した事例がある。これらの事例の致命的なまでに難しい点は、通信会社など情報と場所を検閲できる立場でなければ、既存インフラに流れる情報を閲覧することができないことである。そのため、出来る限りコストをかからないWeb情報のうち公開されているものを利用することが好ましい。

1.3 新しい手法へのアプローチ

数多く存在するWeb情報のうち、我々は地理情報の付与されたソーシャルネットワークを用いることにした。日本全国の情報を網羅し、同じシステムで日本中をセンシングすることがシステム上好ましいと考えられる。SNSであれば、情報がオープンに公開されており、APIによってデータにアクセスすることが容易に可能である。中でもTwitterにはジオタグの付与されたものがあり、これによりユーザー、投稿日時、投稿場所、投稿内容の情報を得ることが出来る。これらの情報は日常生活に関連しており、各ユーザーのアクティビティが記述されている。従って、これらの情報を分析することで、ただ単なるWebの情報とは異なり、場所の特性をその場所に出てくる単語の頻度で抽出出来ると考えられる。

2 目的

地理情報の付与されたソーシャルネットワーキングサービスから取得した情報を分析するためのシステムを構築する。具体的には、取得するシステム、分析する手法、結果の可視化に至るプロセスを構築する。

3. システム構築方法

今回開発を試みたシステムは大きく分けて情報取得方法、分析方法、可視化手法に分類される(図1)。それぞれの手法や設計思想を以下の項目で説明する。

3.1 情報取得方法

情報取得方法では、Streaming API⁹⁾ という 14% のツイートをリアルタイムに取得できるサービスを利用した。このうち、日本近辺でジオタグを付けて投稿したユーザーを CSV 形式のファイルに書き出した。その際に、ユーザー名ごとのファイルを作成し、ユーザー名、投稿時刻、投稿 ID、緯度、経度、投稿内容を書き出して行った。その後、ジオタグのあるツイートが多い場所を検出するために、GPS 座標を Yahoo! Open Location Platform の地図座標に変換して可視化するプログラムを Java で作成し、東京近辺のジオタグをすべてプロットした(図2)。これにより、駅近辺にジオタグのある Tweet が集中していることを確認し、人を比較的集めやすい駅周辺が場所として適切であると考えた。

3.2 分析方法

分析方針として、駅ごとに利用する人の属性が異なるために、Twitter に出てくる単語やその頻度も異なると考えた。また、この単語と頻度を他の駅との相関係数から駅同士の意味・機能の近さも定量化出来ると考えた。この検証のために、雑誌や Web から単語抽出、Twitter テキストからソート、最後に相関係数の計算というプロセスを考えた。

3.2.1 雑誌のテキスト抽出方法

Twitter というメディアは話題が多岐に渡るため、ノイズが大きく、情報のエントロピーが集約しない可能性が極めて高い。従って、日常生活、つまりライフスタイルに言及したもののみを抜き出し、話題を集約させることを狙った。そのために、ライフスタイルに言及している雑誌をとして Switch、Brutus、Pen を選んだ。この選択の理由として、ある特定なものにフォーカスしていない点が上げられる。そのために、漠然としたライフスタイルの単語を抜き出すのに好ましいと考えた。次に、雑誌を出来る限り高画質でスキャニングし、Adobe Acrobat X Pro を用いて OCR 処理を施した。出力される PDF をテキストに張りつけた。この際に OCR 処理を施した文章は文章の折り返しが改行と認識されるので、計量言語の Ruby を用いて、文章にある改行を削除し、“。”の次に改行させた。

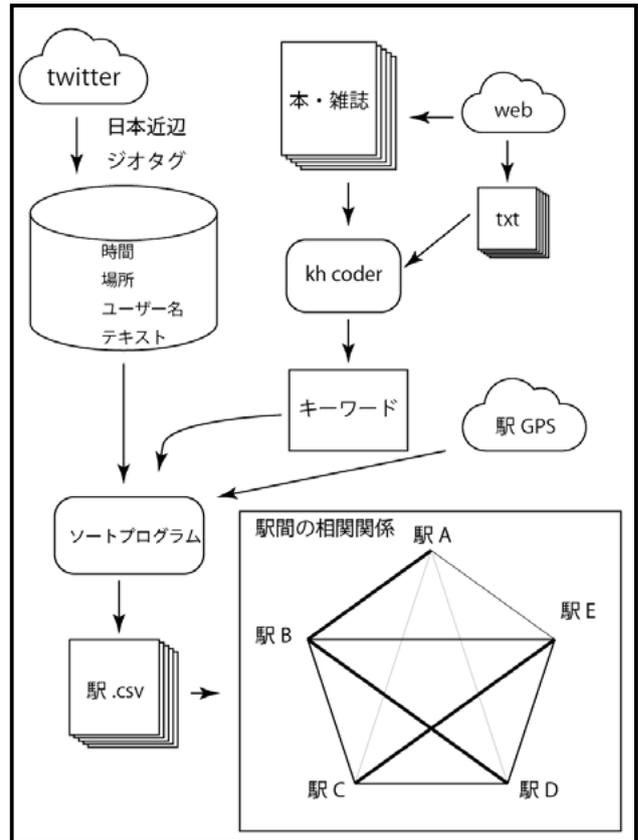


図1 システムの全容

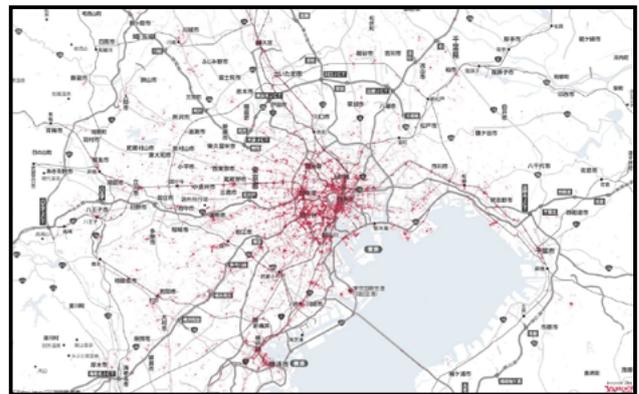


図2 東京周辺にてジオタグをプロットした図

3.2.2 Web のテキスト抽出方法

また Web の単語抽出では、ホームページ、ブログ、facebook の URL を取得し、そこからテキストを抜き出した。まずライフスタイルに関係が深いと思われる単語を上記の雑誌と「下流社会 新たな階層集団の出現」「下流同盟 格差社会とファスト風土」の目次から抜き出した。この際には年齢性別の異なる3名それぞれがライフスタイルに関係が深いと思われる単語を話し合いながら任意に抽出した。ホームページでは、7月に Google にて先述のキーワード集を基に1単語ずつ検索した。その際には Google アカウントからログアウトして、出来

る限り利用者のバイアスの無い状態にした。検索結果から検索結果の1ページ目の中に含まれるHPからライフスタイルに関係性が高い事柄に言及しているものについてURLを抽出した。さらに、Google内のBlog検索を利用し、HPの抽出時と同様の方法でURLを抽出した。同様にfacebookに関してもライフスタイル+FacebookでGoogle検索し、全てのURLを抽出した。これらのURLをPerlのWeb::Scrapeというモジュールを用いてテキストを取得していった。

3.2.3 単語抽出方法

雑誌やWebから採取してきたテキストをkh coderにて単純頻度から名詞を抽出した。そのうち、名詞のカテゴリのうち、名詞、サ変名詞のみを抜き出し、上位4000語を抽出した。表1には抜粋した単語のうち、上位30語づつを示した。これら合計8000語がライフスタイルに関連するキーワードと推定したこの単語集を用いてTwitterのテキストデータを日本全国の駅周辺400m内のキーワードを抽出した。

3.3 Twitterデータのソート

前項にて抽出した単語とTwitterに出てくる単語を対応させた。また、人の集まりやすい駅周辺にジオタグが集中する傾向が見えたため、駅周辺400mが無理なく歩ける商圈として定め、すべての駅のGPS座標をwebページから取得していった。

3.4 駅間の相関関係の計算

駅ごとのジオタグおよびそこに現れる単語とその回数を計算し、単純頻度をJavaにて集計し相関係数を計算してリスト化した。まず、駅圏内のジオタグ数をカウントし、駅名とジオタグ数をCSV形式に書き込んだ。ノード名、ノードID、ラベル名、ジオタグを計算しファイルに書き出した。次に駅圏内(ノード)のデータを計算した。また相関係数リスト(パス)では、始点ノードID、終点ノードID、ラベル名、相関係数を書き出した。これら二つのファイルは事項に用いるアプリケーションにて可視化されるための下準備である。

3.5 データ可視化

可視化の際にGephiを用いた。このアプリケーションはJavaで書かれたオープンソースのネットワーク可視化及び分析が可能なパッケージである。CSVで書かれたノードおよびパスをインポートすることで必要なデータを揃えることが可能である。また、これらのデータをソートし、

表1 抽出した単語の単純頻度

	名詞		サ変名詞				
映画	7267	素材	1369	デザイン	4033	完成	1125
自分	6291	アルバム	1354	監督	3690	附	1110
世界	5962	情報	1310	生活	2092	制作	1058
作品	5783	子供	1293	関係	1966	一緒	1054
写真	4284	ラーメン	1284	存在	1955	参加	1036
音楽	4073	パン	1282	料理	1906	展示	995
時代	4072	アイテム	1262	ト	1883	恋愛	986
人気	2615	感じ	1258	仕事	1817	泊	950
年代	2361	映像	1257	公開	1655	開発	916
人間	2177	歴史	1235	使用	1650	研究	915
女性	2169	作家	1231	開催	1598	新作	909
人生	1885	子ども	1197	発売	1502	変化	883
ブランド	1831	魅力	1187	代表	1477	利用	882
生まれ	1820	イン	1179	意味	1456	製作	845
モデル	1744	会社	1150	話	1448	体験	842
現代	1671	スタイル	1149	オープン	1416	企画	822
社会	1627	カフェ	1143	表現	1406	機能	812
シャツ	1570	テーマ	1134	紹介	1403	持	801
中心	1541	ライブ	1128	注目	1391	デビュー	780
場所	1515	シリーズ	1096	イメージ	1322	演出	753
空間	1495	ワイン	1082	登場	1319	実現	752
商品	1480	特徴	1076	限定	1317	受賞	742
一つ	1478	徒歩	1071	活動	1307	演奏	734
自身	1467	デザイナー	1060	建築	1271	経験	731
世紀	1461	大人	1049	販売	1262	広告	693
文化	1458	ハウス	1043	発表	1199	パス	692
言葉	1444	ショップ	1039	出演	1164	編集	688
ライフスタイル	1432	テレビ	1037	旅	1163	予約	668
人々	1411	イベント	1028	展開	1160	誕生	666
スト	1398	技術	1019	予定	1128	特集	644

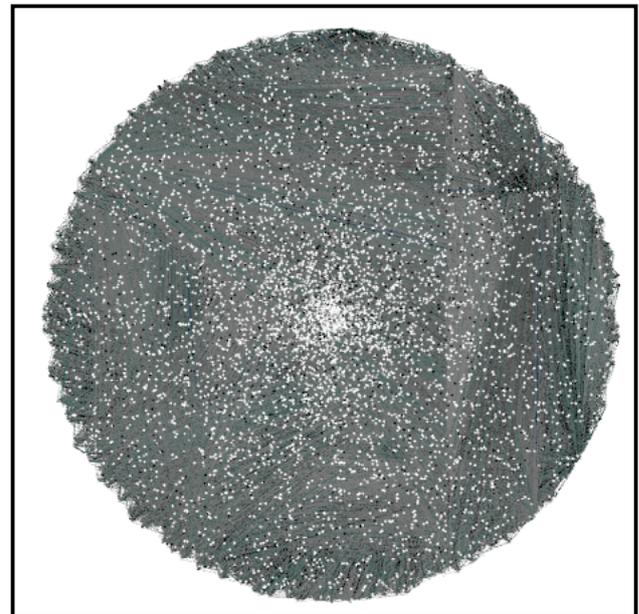


図3 Gephiによって可視化された駅間の相関関係値に応じて可視化(図3)することができる。

4. 結果

我々の提案したプロセスは一通り動く事を確認した。まず、情報取得部分では、24時間365日体勢でTwitterから情報を得ている。また、分析においても一通り駅間の相関関係を計算することに成功した。最後に、Gephiを用いた可視化もとりあえず成功している。

しかし、問題は依然として残っている。まず、キーワードを決める際に雑誌とWebから情報を得て単語を抽出したが、この手法の妥当性に問題がある。雑誌の選定やWebの選定などを複数人と行っていたりするものの、まだど

の単語がライフスタイルにふさわしいかを決めるには至っていない。さらに、ライフスタイルに関連する単語は恐らく時間とともに変化していくものであるが、こういう自動的にテキストを取得し、単語を羅列は出来ない。また、今は等しく扱っているものの、Tweet 数の少ない駅や Twitter 利用者の少ない駅の扱いを決めなければならない。

また、データ数が莫大なため、メモリ容量による分析サイズの限界が生じてしまった。さらに可視化では、9000 に近いデータを可視化するとほぼ意味がなくなってしまう。そのために、メモリ消費量を考えても駅の関連の意味を考えてもスケラブルに可視化をしていく必要があると思われる。

5. まとめ

ジオタグに紐付けされた SNS の情報を取得、分析、可視化を試みた。まず、Twitter を用いて、twitter の情報を書き出した。次に、ライフスタイルに関連する雑誌と Web を選定し、ライフスタイルに関連する単語を抽出した。さらに、ジオタグの多さは駅周辺に多いことを確認した上で、駅周辺の単語とその単純頻度を計算した。これらからすべての駅間において、相関係数を計算した。可視化では、Graphi が用いられた。ライフスタイルに関連する単語を抜き出す際に恣意性が見られ、またデータ数の多さからメモリ容量の限界が発生するという点が見られたが、一通りこれらのプロセスを経ることが出来た。今後は相関係数及び単語頻度からある駅の SNS 上の意味や、他の駅間との近接性や類似性を定量化していきたい。

本研究は、科学研究費補助金挑戦的研究 (11538100) 「ソーシャルネットワークワーキングサービスに投稿された記事に基づく都市・建築空間像の解析」の助成を受けたものである。

既往研究

- 1) 古賀誉章, 高明彦, 宗方淳, 小島隆矢, 平手小太郎, 安岡正人: キャプション評価法による市民参加型景観調査 都市景観の認知と評価の構造に関する研究 その 1, 日本建築学会計画系論文集, No. 517, pp. 79-84, 1999
- 2) 小島隆矢, 古賀誉章, 宗方淳, 平手小太郎: 多変量解析を用いたキャプション評価法データの分析 都市景観の認知と評価の構造に関する研究 その 2, 日本建築学会計画系論文集, No. 560, pp. 51-58, 2002
- 3) Lynch, K. (1960) The Image of the City (Harvard-MIT

joint center for urban studies series), Massachusetts: The MIT Press

- 4) Hayden, D. (1995) The Power of Place: urban landscape as public history, Massachusetts: The MIT Press
- 5) Murty, R. N., Mainland, G., Rose, I., Chowdhury, A. R., Gosain, A., Bers J. and Welsh M. (2008) CitySense: An Urban-Scale Wireless Sensor Network and Testbed, 2008 IEEE Conference on Technologies for Homeland Security, pp. 583-588
- 6) 遠田敦, 菊地弘祐, 中川純: Arduino・Pachube・SketchUp と連携した建築モニタリングシステムの基礎的研究, 日本建築学会技術報告集 16(33), 791-794, 2010
- 7) Colonna, M., Lovisolo, P., Parata, D. and Ratti, C. (2011) Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome, 2011 IEEE Transactions on Intelligent Transportation Systems, pp. 141-151
- 8) Wireless Wire News: 「モバイル空間統計」で可視化された、東日本大震災時の帰宅困難者 (2011.5.6), 参照日 2012.10.11: http://wirelesswire.jp/News_in_Japan/201105251956.html
- 9) Twitter Developers: Public Stream (2012-10-08 08:24), 参照日 2012.10.11: <https://dev.twitter.com/docs/streaming-apis/streams/public>
- 10) Yahoo! Japan Developer Network: Yahoo! スタティックマップ API (n.d.) 参照日 2012.10.11: <http://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/static.html>
- 11) 樋口 耕一: KH Coder Index Page (n.d.) 参照日 2012.10.11: <http://khc.sourceforge.net/>
- 12) MeCab: Yet Another Part-of-Speech and Morphological Analyzer (n.d.) 参照日 2012.10.11: <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>
- 13) CodePlus: 駅データ (2012.10.10) 参照日 2012.10.11: <https://www.ekidata.jp/>

-
- * 1 早稲田大学大学院建築学専攻 博士課程
 - * 2 滋賀県立大学大学院環境科学専攻 修士課程
 - * 3 東京理科大学建築学科 助教 博士 (建築学)
 - * 4 滋賀県立大学環境科学部 准教授 博士 (工学)
 - * 5 A&A Co. Inc. 研究員 博士 (建築学)
 - * 6 早稲田大学建築学科 准教授 博士 (工学)
 - * 7 早稲田大学建築学科 教授 博士 (工学)