

Twitter における市民意見を活用した行政サービスの把握の試み

Analysis of Public Sentiment in Twitter for Local Government Service

関 洋平 *1 猪 圭 *2

Yohei Seki Kei Ino

*1筑波大学 図書館情報メディア系

Faculty of Library, Information and Media Science, University of Tsukuba

*2つくば市役所

Tsukuba City

Active collaboration with the public is the key to improving the administrative services of local government, and social media is an essential tool for understanding public sentiment. In this paper, we propose a method to gauge public sentiment for local government by using Twitter. We conducted an operational test with local government officers and found that our proposal was effective in revealing sentiments that were meaningful for the public, such as avoiding full parking lots, finding potholes in the road, or attracting favorable comments or suggestions to improve festival events.

1. はじめに

ここ数年、シビックテクノロジー（シビックテック）[関 14]*1 と呼ばれる活動が注目を集めており、市民を参画させることで、行政サービスのインフラを改善することが試みられている。一つの代表例として、Jennifer Pahlka は、Code for America*2 と呼ばれる NPO 団体を、2009 年に設立している。この団体では、政府や自治体の行政サービスを効果的で使いやすいものにするために、自治体に情報技術者をフェロー（1 年間ほどの期間限定職員）として派遣して、Web サービスやアプリを開発する仕組みを実現している。もう一つの例は、Open311*3 である。このサービスでは、顧客関係マネジメント（CRM）の技術を導入することで、市役所による道路の補修や、落書きの清掃等について、市民からの要求を追跡するサービスの実現を目指している。日本では、少し異なるが似た試みとして、ちばレポ（ちば市民協働レポート）*4 が注目されている。一方、情報検索やデータマイニングを対象とした研究コミュニティにおいては、都市データを活用した研究が注目されている。ACM SIGIR のチュートリアルとしては、Searching in the City of Knowledge [Veli 13] が 2013 年に開催された。また、ワークショップとして、Information Access in Smart Cities [Albakour 14]、WWW 2015 Workshop on Web Data Science and Smart Cities*5 や Mining Urban Data*6 などが開催されている。

こうした活動において、魅力的なアプリケーションの一つが、ソーシャルメディアから市民の意見を収集し、分析するツール [Musto 15, 栗原 15] を実現することで、行政サービスのコストの削減に繋げることである。本研究では、図 1 に示すように、ソーシャルメディア上の意見を活用することで、市民の本音を反映した行政が実現できると考える。意見分析の研究分野 [Pang 08] では、情報分析や世論調査のための多くの実用的なサービスが開発されており、そのリソースとしては、製品レビュー、映画レビュー、ブログ、マイクロブログなどが

連絡先: 関洋平, 筑波大学図書館情報メディア系, 〒 305-8550
茨城県つくば市春日 1-2, yohei@slis.tsukuba.ac.jp

*1 <http://civictechforum.jp/>*2 <https://www.codeforamerica.org>*3 <http://www.open311.org/>*4 <http://chibarepo.force.com/>*5 <http://www.data2.cn/WebDataScience4SC/>*6 <http://www.insight-ict.eu/mud2>

図 1: 市民の意見分析を通じた行政改善

対象となっている。洗練された意見分析の技術は、通常あるドメインの辞書に依存しているが、ソーシャルメディア上の書き込みにおける話題は多岐のドメインに渡っている。

本研究では、特定の地域を対象として、市民の意見を収集し分析することで、地域の人気スポットや施設を宣伝したり、自治体の行政コストを削減するための枠組みを提案する。この枠組みでは、以下の 2 つの提案を行なう。

1. 特定の地域に住んでいるソーシャルメディア（Twitter）上のユーザのアカウント集合を抽出
2. 1 のアカウント集合から発信される意見を分析する際に、地域のお祭り（イベント）や、行政サービスの内容など、多数のドメインに適合した意見の手がかりを柔軟に拡張

1 点目は、プロフィールに記載されている地域名を利用して Twitter のユーザを抽出した後、フォロワーを使って市民の拡張を行なう。2 点目は、分散表現（埋め込み）に基づき類義語を高精度で抽出する Word2Vec [Mikolov 13] を利用して、ドメインに応じて、種となる肯定表現や否定表現との類似関係を考慮し、意見分析の手がかりを柔軟に拡張する。

本論文の構成は以下の通りである。2 節では、関連研究について紹介する。3 節では、提案手法について記述する。4 節では、自治体との協力に基づく実験と議論を提供する。最後に、5 節で結論を述べる。

2. 関連研究

2.1 ソーシャルユーザの居住地推定

ソーシャルメディア上のユーザの居住地を推定するために、先行研究では、投稿内容に着目するアプローチと、ソー

シャルグラフに基づくアプローチが提案されている。前者については、Cheng ら [Cheng 10] が、特定の地域に偏って出現する地域特有の言葉を抽出することで、地域を推定する手法を提案している。このアプローチの欠点は、特に小規模の地域を対象として推定する際に、地域特有の言葉が、他の地域のユーザに使われる可能性が高くなる点にある。本研究では、対象地域を市区町村レベルと考えていることから、ソーシャルグラフに基づくアプローチを採用する。Jurgens [Jurgens 13] は、Twitter ユーザの居住地を、フォロー関係に焦点を当てることで推定するソーシャルグラフに基づくアプローチを提案した。この提案では、ジオタグ付きツイートから、最初の種となる居住地ユーザを推定する。しかし、ジオタグ付きツイートを発信するユーザの数は非常に少ないことが知られている [伊川 15]。我々は、ユーザプロフィールを考慮することによって、種となる居住地ユーザを決定し、フォロー関係を利用して、ユーザを拡張する手法を提案する。

2.2 Twitter 上の意見分析

最近の意見分析の研究は、マイクロブログ、特に Twitter に焦点を当てている [Martinez-Camara 13]。Twitter は、一つの投稿が 140 字以内と短いテキストであることから、出現する単語を見るだけでは、ツイートの極性（肯定・否定・中立）の分類には不十分である。したがって、ツイートの文脈や Twitter 特有の書き方を考慮して、手がかりを拡張する必要がある。Jiang ら [Jiang 11] は、クエリとして与える対象との構文的関係に着目し、ツイートの極性を分類する手法を提案した。しかし、Twitter を対象とした構文解析の精度は必ずしも高くない。Chen ら [Chen 12] は、特定のドメインにおいて、種となる意見表現との共起語を利用して意見表現を拡張した。本研究では、分散表現（埋め込み）に基づき類義語を高精度で抽出する Word2Vec [Mikolov 13] を利用して、ドメインごとに、種となる意見表現と類似する意見表現を、新たな手がかりとして拡張する。拡張する意見表現は、Twitter 特有の表現であるハッシュタグ [Wang 11] や顔文字 [Purver 12] を含む。

3. 市民の意見分析

本研究では、地方自治体の行政サービスにおける利点や改善点を明らかにするために、自治体の市民が発信するツイート集合に現れる意見を分析する手法を提案する。まず、特定の地域の市民によって発信されているツイートを収集する。次に、ツイート集合を肯定、否定、中立のツイート群に分類する。最後に、それぞれのツイート群を話題別にクラスタリングし、タグクラウドインタフェースによって可視化する。以下で、それぞれのステップについて説明する。

3.1 市民ユーザアカウントの抽出

日本の市区町村に住む市民に対応する Twitter のユーザアカウントを、以下の手順に従い抽出する。

1. ユーザプロフィールを使用して、地域名に照合するアカウントを、種アカウントとして抽出する。地域名に照合するアカウントは、Twitter のユーザプロフィール検索サービス *7 を利用して抽出する。
2. 種アカウントをフォロワーを利用して拡張する。拡張にあたっては、フォロワー数の上限値、フレンズ（フォロワー）数の上限値をもうけることで、スパムアカウントを除去する。また、フォローするアカウント集合に含まれる、市

*7 <http://twpro.jp>

表 1: 市民ユーザアカウントの抽出結果

アカウントの抽出方法	アカウント数	F-値
プロフィールのみ	4,548	0.459
フォロワーの拡張 (1 回目)	27,589	0.668
フォロワーの拡張 (2 回目)	31,666	0.677

表 2: 抽出したアカウントの職業属性

職業属性	割合 (%)
学生	50.2
社会人	29.0
企業・団体	10.8
主婦	3.1
bot	1.7
その他	5.2

民アカウント数の下限値をもうけることで、市民となる可能性を向上させる。

予備実験 [長島 16] では、日本のつくば市（市民数約 20 万人）を対象として、フォロワーを 2 回拡張する *8 ことにより市民アカウントを抽出した。また、抽出したアカウントから発信されるツイートのうち、つくば市にある 12 のラーメン店についてのツイートの適合性を、F-値により評価した。結果を表 1 に示す。この結果から、拡張は 2 回で十分収束すると結論づけた。また、市民アカウントから 200 件をランダムに選択し、手で職業属性 [武田 16] を付与した。結果を表 2 に示す。つくば市は学園都市であることから、学生が多く、学生ユーザの割合が少し高くなる。これらの結果から、市民として妥当なアカウント集合が抽出できていると判断した。

3.2 Word2Vec を使用した意見分析

市民が発信するツイート集合に現れる、行政サービスに関連した良い評判や改善点を把握するために、ツイートを、肯定、否定、中立の極性に分類する。ツイート集合は、祭りなどの地域イベントや、道路交通事情など、行政サービスに関連するドメインに応じて絞り込む。極性を分類するための手がかりは、ドメインごとに、以下の手順で獲得する。

1. まず、ツイートを形態素解析器 MeCab*9 にかける。辞書は、Twitter 特有の非定型な表現を抽出することを意識して、mecab-ipadic-neologd*10 を使用する。
2. ツイート中の形態素を Word2Vec*11 を使用して共起関係に基づき類義語をまとめる。単語クラスタのサイズは 10 から 1,000 まで変化させる。
3. 肯定表現と否定表現の種となる表現 *12 と、複数の単語クラスタにおいて頻繁に類義語と見なされる単語を、新たな肯定表現と否定表現として拡張する。この手法により、ドメインごとに、非定型な感情表現を獲得すること

*8 種のアカウントのフォロワーのフォロワーまでを対象として抽出する。

*9 <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

*10 <https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md>

*11 <http://word2vec.googlecode.com/svn/trunk/>

*12 表記揺れも含めて、約 30 の形容詞あるいは強い極性を表す名詞を種となる表現として選定し、否定の種としては、ALAGIN のトラブル表現リスト (<https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-list.html>) を追加した。

ができた。具体的には、地域の祭りをドメインとした場合、肯定表現は、顔文字 (“(*/ω*)”, “\ (^_^) /” 等)、音符の繰り返し (“♪♪”) といった非定型表現や、“みんな”、“最高” といった表現が拡張できた。否定表現については、顔文字 (“(*.*)”), 笑いを表す非定型表現 (“w w w”) や、“人混み”、“ゴミ” 等の表現が拡張できた。

- 最後に、ツイートに含まれる意見表現の種類と出現頻度を考慮して、ツイートを肯定、否定、中立に分類する。

3.3 話題のクラスタリング

本手法では、肯定・否定・中立と分類されたツイート群に対して、形態素を手がかりとしたクラスタリングを適用することで、肯定・否定・中立的な話題を抽出する。クラスタリングは、bayon^{*13} を使用して、repeated bisection 法を適用する。クラスタのラベルは、tf-idf の値が高い名詞・形容詞と、TermExtract^{*14} を使用して抽出されたスコアの高い専門用語を利用して付与し、それぞれのクラスタに含まれるツイート数に基づき、タグクラウドを利用して可視化する。また、収集したツイート集合から、キーワードをクエリとして検索できる支援ツールを、Apache Solr^{*15} を用いて実現した。以上の処理をまとめた全体像を図2に示す。

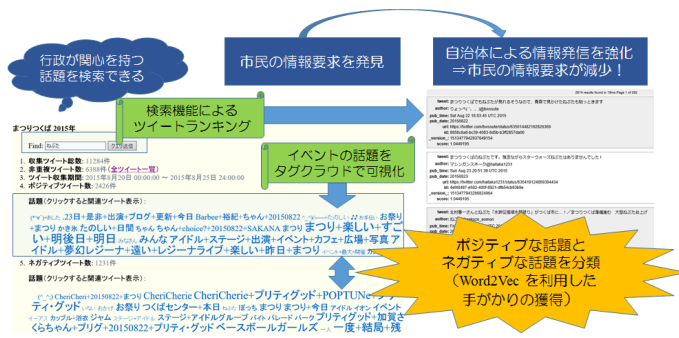


図2: 提案する市民の意見分析ツールの全体像

4. 実験: 市民意見を活用した行政サービスの改善

4.1 実験方法

提案手法の有効性を検証するために、つくば市役所の職員と2014年から2016年にかけて協力し、実験を進めた。市役所の職員の方には、実際に提案システムを使用していただき、行政サービスを向上させるための市民の意見を、Twitterから発見できるか調査した。なお、ツイートは、Twitter streaming API^{*16} を使用して収集した。対象としたタスクは、以下に説明する2つである。

最初のタスク [つくば市役所 市長公室 15] は、地域の祭りイベント (まつりつくば^{*17}) に関して自治体が提供する行政サービスについて、鍵となる話題を発見できるか検証する。このイベントは、1年に1回、2日間続けて開催され、40万人以上の市民や観光客が参加する。2012年から2015年にかけてイベント開催期間に発信されたツイート数を表3に示す。内容が完全に重複するリツイートについては、前処理として削除した。

表3: まつりつくばについて収集したツイート数

年	収集ツイート数	非重複ツイート数
2012	4,876	4,204
2013	6,204	4,359
2014	7,114	4,803
2015	11,284	6,388

表4: 2014年から2015年の否定的コメント数の減少

Terms	否定的コメント数	
	2015	2014
駐車	15	44
交通規制	7	12

2番目のタスクは、つくば市の道路トラブルに関するコメントを収集することを目標とする。道路トラブルに関して市民が発信したツイートは、2015年の12月27日から、2016年の2月19日にかけて収集した。重複ツイートは、最初のタスクと同様に削除した。

4.2 結果と考察

最初のタスク [つくば市役所 市長公室 15] については、2014年に市役所職員が使用した結果、駐車場が満車であることや見つかからないこと、あるいは交通規制に関する否定的な話題が見つかった。コメントの具体例を以下に示す。

- まつりつくばの駐車場の位置がいまいちつかめない……市役所どこだ
- まつりつくばってどこに車停めれるんだろう。駅周りってあんまり駐車場がないイメージで行けない。
- いつもの駐車場に停めて東京行くはずがまつりつくばのせいで駐車場がいっぱい時間ロス。
- つくば行くのだが まつりつくばで交通規制ないか不安だ・ω・
- まつりつくばの交通規制におどらされる (---;)

この結果に基づき、2015年のイベントでは、市役所職員が公式のTwitterアカウントを使用して、駐車場の状況や交通規制のスケジュールについて周知することを行った。イベント終了後に、“駐車”あるいは“交通規制”を含む否定的なコメントの数を数えて比較したところ、2014年から2015年にかけて減少したことがわかった。結果を表4に示す。

この結果から、提案手法は、行政サービスに対して否定的なコメントを含む話題を発見することに有効であることがわかった。さらに、自治体職員がソーシャルメディアを通じて駐車場の状況や交通規制スケジュールについてアナウンスすることで、市民が満足し否定的なコメントをしなくなることがわかった。

2番目のタスクについては、歩道の補修や、道路陥没、冠水した道路についてのコメントが見つかった。発見したコメントの具体例を、以下に示す。

- 完全に人の通りの多さに対する歩道の整備のなさが原因。道路作った奴に文句言え
- 道路(水没)の脇が激しく陥没してたり罌だらけ

*13 https://code.google.com/p/bayon/wiki/Tutorial_ja
 *14 <http://gensend.ll.it.u-tokyo.ac.jp/termextract.html>
 *15 <http://lucene.apache.org/solr/>
 *16 <https://dev.twitter.com/streaming/overview>
 *17 <http://www.matsuri-tsukuba.com/>

- んー、この道路の構造よろしく無いなあ、住居側から雨水を排水する設備って役所に申請して金出せば可能なんだろう？雪や大雨のたびに冠水するぞコレ

これらの否定的なコメントは、*Open311* のサービス^{*18} やちばレポ (ちば市民協働レポート)^{*19} で公開されているものと内容的に共通しているものが多い。この結果から、提案手法は、ソーシャルメディアから自治体へのフィードバックコメントの発見に役立たせることができる可能性があると言える。

5. おわりに

本研究では、市民の意見を発見するための手法を提案し、発見した意見が行政サービスの向上に有効であることを、市役所の職員との実験を通して明らかにした。今後は、多様な都市をカバーできるように提案手法を拡張し、都市間の行政サービスを比較するためのシステムを実装する予定である。

謝辞

本研究の一部は、筑波大学研究基盤支援プログラム (B タイプ)、科学研究費補助金基盤研究 B (課題番号 25280110) の助成を受けて遂行された。

参考文献

- [Albakour 14] Albakour, M. D., Macdonald, C., Ounis, I., Clarke, C. L. A., and Bicer, V.: Information Access in Smart Cities (i-ASC), in *Advances in Information Retrieval, Lecture Notes in Computer Science*, Vol. 8416, pp. 810–814 (2014)
- [Chen 12] Chen, L., Wang, W., Nagarajan, M., Wang, S., and Sheth, A. P.: Extracting Diverse Snetiment Expressions with Target-dependent Polarity from Twitter, in *Proc. of the 6th International AAAI Conf. On Weblogs and Social Media*, Dublin, Ireland (2012)
- [Cheng 10] Cheng, Z., Caverlee, J., and Lee, K.: You Are Where You Tweet: A Content-Based Approach to Geolocating Twitter Users, in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pp. 759–768, Toronto, Canada (2010)
- [Jiang 11] Jiang, L., Yu, M., Zhou, M., Liu, X., and Zhao, T.: Target-dependent Twitter Sentiment Classification, in *Proc. of the 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, pp. 151–160, Portland, Oregon (2011)
- [Jurgens 13] Jurgens, D.: That’s What Friends Are For: Inferring Location in Online Social Media Platforms Based on Social Relationships, in *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM 2013)*, pp. 273–282, Boston, USA (2013)
- [Martinez-Camara 13] Martinez-Camara, E., Martin-Valdivia, M. T., Urena-Lopez, L. A., and Montejo-Raez, A. R.: Sentiment Analysis in Twitter, *Natural Language Engineering*, Vol. 11, No. 2, pp. 1–28 (2013)
- [Mikolov 13] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient Estimation of Word Representations in Vector Space, in *Proc. of Workshop at ICLR* (2013)
- [Musto 15] Musto, C., Semeraro, G., Gemmis, de M., and Lops, P.: Developing Smart Cities Services through Semantic Analysis of Social Streams, in *Proceedings of the 24th International Conference on World Wide Web (WWW’15 Companion)*, pp. 1401–1406, Florence, Italy (2015)
- [Pang 08] Pang, B. and Lee, L.: *Opinion Mining and Sentiment Analysis*, Now Publishers Inc (2008)
- [Purver 12] Purver, M. and Battersby, S.: Experimenting with Distant Supervision for Emotion Classification, in *Proc. of the 13th European Chapter of the Association for Computational Linguistics (EACL 2012)*, pp. 482–491, Avignon, France (2012)
- [Veli 13] Veli, B. and Vanessa, L.: Searching in the City of Knowledge, in *Proc. of the 36th Int’l ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 1123–1123, Dublin, Ireland (2013)
- [Wang 11] Wang, X., Wei, F., Liu, X., Zhou, M., and Zhang, M.: Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach, in *Proc. of the 20th ACM Conf. on Information and Knowledge Management (CIKM 2011)*, pp. 1031–1040, Glasgow, Scotland, UK (2011)
- [つくば市役所 市長公室 15] つくば市役所 市長公室 シティプロモーション室：つくば市と筑波大学共同研究-地域活性化への新たな SNS 活用報告 Twitter のつぶやきを行政サービスの向上につなげています。(2015), https://www.city.tsukuba.ibaraki.jp/dbps_data/_material/_files/000/000/018/912/No173.pdf
- [伊川 15] 伊川 洋平：ソーシャルメディア位置情報分析を行なう前の Tips, ARG 第 7 回 Web インテリジェンスとインタラクティブ研究会特別セッション「不揃いなデータ達の分析を行なう前の Tips」(2015), <http://www.slideshare.net/YoheiIkawa/tips-55961281>
- [関 14] 関 治之：シビックテックとオープンデータ, 庄司昌彦 (編), 智場 (オープンデータ特集号), 第 119 巻, pp. 101–114, 国際大学 GLOCOM (2014)
- [栗原 15] 栗原 理聡, 佐々木 彬, 松田 耕史, 岡崎 直観, 乾 健太郎：Twitter を利用した地域毎の要望抽出, 人工知能学会 第 29 回 全国大会, pp. 1H3–3 (2015)
- [長島 16] 長島 里奈, 関 洋平, 猪 圭：地域ユーザに着目した口コミツイート収集手法の提案, 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM2016), pp. B4–3 (2016)
- [武田 16] 武田 直人, 関 洋平：Twitter の投稿場所を考慮したユーザ属性推定, 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM2016), pp. A6–1 (2016)

*18 <http://www.open311.org/>

*19 <http://chibarepo.force.com/>