

市民意見の収集システムで得られたツイートからの「保育園」「教育」に関する意見抽出

柏野 和佳子¹ 立花 幸子¹ 平本 智弥¹ 関 洋平²

¹ 国立国語研究所 音声言語研究領域 ² 筑波大学 図書館情報メディア系

1. はじめに

自治体の規模によらず市民の意見を収集するための基盤を構築し、市民との協働に基づき自治体による情報発信や施策を展開するための研究を進めている(関2015, Seki2016, 関2017)。そのために、ソーシャルメディアのマイクロブログ(主にTwitter¹を利用)上の行政に関係する意見を対象とした意見分析コーパス(意見ごとにタイプを付与したテキストの集合。関(2013)。)を作成することが本研究の目的である。

Twitterは、マイクロブログの代表と言えるほど普及しており、速報性およびリアルタイム性のある、有用な情報源と考えられている(奥村2012)。Twitterのツイートから要望等の情報を抽出する最近の研究には、山本ほか(2012)、栗原ほか(2015)、川島ほか(2016)などがある。

本発表では、「保育園」や「教育問題」に着目する。それらについて市民がどのようにとらえているかを把握するために、ツイートから必要な情報を抽出し、意見情報のアノテーションの方針を立て、アノテーション属性の分布情報と具体例に基づき今後の課題を明らかにする。

2. 市民意見の収集

次の方法で収集した。

- ① つくば市民データアカウントを、ツイプロ²に掲載されているアカウントとそのフォロー関係を利用して、22,120件収集した。
- ② これらのアカウントから、2015年10月17日～12月27日まで、リアルタイムに発信されたツイートを収集した。
- ③ 上記のツイートのうち、行政サービスの手がかかり語(「保育園」「教育」「防災」「まちづくり」「ごみ」「税」等)を含むツイートを抽出し、さらに要求意図表現(大塚ほか2004, 栗原ほか2015)を含むツイートを抽出した。このうち、本発表の対象として得たツイートは、「保育園」551件と「教育」2,071件とである。

¹ <https://twitter.com/>

² <http://twpro.jp/>

3. 「教育」「保育園」に関するツイートからの意見抽出

3.1 意見を含むツイートの選定

「保育園」551件と「教育」2,071件のツイートを対象に意見を含むものを選定した。その件数は次の表1の通り、それぞれ約6割であった。

表1 意見を含むツイートの選定結果(件)

対象	意見あり	(%)	意見なし	(%)	合計
保育園	338	61.3	213	38.7	551
教育	1,205	58.2	866	41.8	2,071

3.2 意見テキストの抽出

ツイートは最大140文字の短い文章である。そこから意見とみなすテキストを特定し、抽出した。その時、複数の意見が認められる場合は、テキストを分割して抽出した。例えば、次のツイートからは「/」で区切った、下線部分の2か所を意見テキストとして抽出した。「意見あり」とした「保育園338件のツイートを対象に抽出した結果、355件の意見テキストを得た。

tweet³ :

この賃金は酷すぎる。保育園や保育士が足りない以前の問題だ。 / 少子化を建前に対策を疎かにしている厚生労働省の怠慢の罪は重い!!

※/, 下線は筆者らによる。(以降、同様)

→「保育士の賃金」と「厚生労働省の施策」への意見

3.3 意見か否か

抽出対象とする意見は目的によって定義される。例えば、小林ほか(2006)では、対象とする意見を「態度や評価の叙述」「態度や評価の原因・根拠となる事態の叙述」としている。栗原ほか(2015)では、要望を表す「直接要求」「要求意図」に当てはまるテキストに加え、要望の動機になる否定的なテキストを「不満」として定義し、それら3タイプを対象とする意見としている。

本研究では、市民の意見とするもの、意見としないものとして、次のように定めた。

³ 引用表示するツイートは、URL や@+アカウント名、一部の日時、固有名など、個人の特定性の高い情報は一部省略、もしくは伏字で示す。

◆意見とするもの

- a) テーマ(保育園, 教育)に関し, 評価や要望があるものを意見とする。

tweet :

つくばに住んでいる間、今月末まで、〇〇**保育園**というところに子どもを預けていたのだけど、非常に良い**保育園**であった。いろんなイベントがあり、家では体験させづらいことをさせてもらい、いろんな人と関わり、いつの間にか様々な事ができるようになってくれた。

- b) テーマ(保育園, 教育)に関する疑問や感想も意見とする。

◆意見としないもの

- c) 発信元が一市民でないものは意見としない。団体, 企業等が発信する案内や広告など。
 d) テーマ(教育, 保育園)が主対象でない, 出来事, 事実の列挙は意見としない。

tweet :

保育園にて素敵なおじさんに会釈される。あれ、どこかでお目にかかったことがある。他学部の先生か、学会で一緒した人か...?といろいろと記憶を掘っていくうちに、ときどき行く銭湯の番台に座っている方とお顔が一致。すっきりした。

- e) 個人的な応答, メッセージは意見としない。

tweet :

まず、Z 頑張ってくれてありがとう。それと、教育局でも頼りない先輩達を手伝ってくれてありがとう。来年は**教育局**の女の子たちをサポートしてあげてください。相談とか全力で乗るからね、良かったら頼ってください。来年もよろしくね、クラ代頑張れ！#いいねした人に一言

- f) リツイートのみのもものや, web 上の記事タイトル, 記事内容等を引用しているだけのものは意見としない。ただし, 引用前後に意見らしきものがあるば, それは抽出する。

4. 意見の対象とタイプの付与

次に、「保育園」の意見テキスト 355 件を対象に, 意見の「対象」と「タイプ」とを付与した。「対象」は, できるだけ同じ話題のものを統一できるように定めた。「タイプ」は, 評価 (ポジ), 評価 (ネガ), 肯定・賛同, 否定・非難, 疑問, 不安, 指摘(中立), 要望, 提案, 不満, 以上 10 種である。

付与した結果について, 表 2 にタイプ別件数を, 表 3 に意見の対象別 (3 件以上の上位 15 種) のタイプ内訳件数を示す。表 3 のようにまとめることにより, どのような問題について, どのような意見が出ているのか把握が可能になることがわかる。

表 2 意見のタイプ別件数 (件)

タイプ	件数
評価ポジ	38
評価ネガ	34
肯定・賛同	26
否定・非難	47
疑問	50
不安	3
指摘(中立)	16
要望	5
提案	60
不満	76
合計	355

表 3 意見の対象別 (上位) のタイプ内訳件数 (件)

対象	タイプ	件数	対象	タイプ	件数			
保育園不足	評価(ポジ)	2	保育園の管理	評価(ポジ)	4			
	評価(ネガ)	19		否定・非難	1			
	肯定・賛同	6		疑問	1			
	否定・非難	15		不安	1			
	疑問	21		指摘(中立)	2			
	指摘(中立)	9		提案	1			
	要望	1		不満	2			
	提案	37		計	12			
	不満	36		保育園建設 反対	評価(ポジ)	1		
	計	146			肯定・賛同	5		
保育園の利用	評価(ポジ)	13	否定・非難		2			
	評価(ネガ)	6	疑問		1			
	肯定・賛同	3	不満		1			
	否定・非難	7	計		10			
	疑問	9	保育士の給料		肯定・賛同	1		
	指摘(中立)	2			否定・非難	2		
	提案	7			疑問	3		
	不満	12			提案	1		
	計	59		計	7			
	保育園の騒音	評価(ポジ)		1	保育料	評価(ネガ)	1	
評価(ネガ)		3		肯定・賛同		1		
肯定・賛同		5		指摘(中立)		2		
否定・非難		8		提案		1		
疑問		10		不満		1		
提案		2	計	6				
不満		1	公園内に 保育園設置	肯定・賛同		1		
計		30		否定・非難		3		
保育園の設置		評価(ポジ)		3		提案	1	
		否定・非難		2		計	5	
	疑問	2		少子化	評価(ネガ)	1		
	要望	2			提案	1		
	提案	3			不満	3		
	不満	4			計	5		
	計	16			学童保育	不満	3	
	入園申請	評価(ネガ)				1	計	3
		否定・非難	2			家庭保育手当	評価(ポジ)	2
		不安	2				提案	1
指摘(中立)		1	計				3	
要望		1	入園準備				評価(ポジ)	1
提案		1		不満			2	
不満		6		計			3	
計		14		保育園開園 断念			肯定・賛同	2
合計								提案
							計	3

4.1 「対象」の付与と課題

一覧性を高めるためには, できるだけ対象を絞る必要がある。一方で, 絞りすぎると, 意見の焦点が異なる集合になり, かえって一覧性を損なう場合も出てくる。

例えば, 表 3 の「保育園の騒音」「保育園の設置」「保

育園建設反対」「公園内に保育園設置」「保育園開園断念」という対象は、すべて「保育園の設置」一つにまとめることが可能である。しかし、それぞれに「騒音」「反対運動」「公園内」「断念」という別々の焦点がある。焦点で区別しないと、例えば、意見の肯定、否定(あるいは評価のポジ、ネガ)の向きが、設置推進なのか反対なのかの区別がわかりにくくなる。よって、異なる焦点がある場合は、一つの話題とせず、その焦点別のまとまりでとらえる方が一覧性が高く、わかりやすいと判断した。

しかしながら、表3に記載しなかった2件以下の「対象」は、保育ママ制度、病児保育、財政問題、市役所の対応、退職世代の保育園支援など、28種にもなってしまう。これらは、もう少しまとめる余地があるかもしれない。3件以上の15種も含めて、「対象」の設定は再検討の必要性があると考えている。

さらに、一つの意見が複数の対象をもつ場合の扱いについても今後の課題である(今回は一意に定めた)。

4.2「タイプ」の付与と課題

4.2.1 評価(ポジ)、評価(ネガ)、肯定・賛同、否定・非難について

広くとらえれば、「肯定・賛同」は「評価(ポジ)」に、「否定・非難」は「評価(ネガ)」にまとめることも考えられるが、区別した。まず、「評価(ポジ)」「評価(ネガ)」は、評価表現等により価値判断を行っているものと認められるものに対して付与した。一方、「肯定・賛同」「否定・非難」は、価値判断なし(あるいは不明)で賛同か非難かをしているものと認められるものに対して付与した。次に、例を示す。

「保育園不足」の tweet :

評価(ポジ): つくば市も認可保育園がどんどん増えて改善の努力を感じる

評価(ネガ): [認可保育園の増設は]潜在的な需要が多すぎて焼け石に水なんだろうな。

肯定・賛同: [「保育園も増やせない」との引用に対し]これ、本当にそうだよね。

否定・非難: [「保育園落ちた、日本死ね!」に対し]そもそも保育園そのものがなかったら文句いわねーんじゃないか、こいつらw

※[]は筆者らによる補足。(以降、同様)

しかしながら、価値判断の有無の認定にはゆれが生じやすく、今後の作業による評価表現の定義や、付与結果の見直しはまだ必要だと考えている。

また、次の例のように、「面白い」という同じ評価表現が素直な評価である場合と、皮肉のような場合がある。何に対しての評価であるかをアノテーションすることも検討したい。

「保育園の設置」の tweet :

評価(ポジ): 新宿駅に渋谷区の認可保育園が入るのなかなか面白い

「保育園の利用」の tweet :

評価(ポジ): 意識高い人が意識高い保育園に預けて園から弁当がどうのと意識高いこと言われて大変すぎる! っていうてるのおもしろえな

4.2.2 疑問、不安、指摘(中立)について

価値判断や、肯定、否定といった態度がみえず、また、直接的な要望もみえず、疑問や感想を述べているような意見に対し、疑問を帯びているものを「疑問」、不安を帯びているものを「不安」とし、それ以外を広く「指摘(中立)」と付与した。次に、例を示す。

「保育園の管理」の tweet :

疑問: 保育園は子どもが来なくても保護者に連絡しないのか。

不安: しかし、給食がやはり食べられるか不安...

指摘(中立): うちの保育園は予防接種にけっこう厳しい。[その是非については触れていない]

なお、「疑問」は「否定・非難」や「不満」との区別がゆれが生じやすい。

4.2.3 要望、提案、不満について

要望を直接的に述べているものに「要望」を、要望を提案の形で示しているものに「提案」を、栗原ら(2015)と同様に要望の動機を否定的に示しているものに「不満」を付与した。先の表2のとおり、「要望」はあまりなく、「提案」「不満」が多数あった。ツイート上に「提案」「不満」という形で要望を述べるものが多いことがわかる。次に、例を示す。

「保育園の設置」の tweet :

要望: お散歩や運動を公園に頼ってる小規模保育園がいくつもあることも知って欲しい。

提案: 駅前の18歳以下は入れない施設が保育園になればいいのに。[駅前の有効活用を要求]

不満: 車1台しか通れない上に急な坂の中間にある保育園に、子供らを送っていくのが無理(T_T) [車のアクセスのよい場所での設置を要求]

5 意見の対象とタイプの付与以外の課題

5.1 「意見なし」ツイートからの情報抽出

具体的な評価や要望を含まず、日々の保育園での出来事や、保育園児をもつ親の日常の出来事つぶやきは意見としなかった(3.3 節 d))。しかしながら、それらは 63 件あり、そこから市民が保育園を利用するうえでの気づきや生活ぶりをとらえることができる点で、抽出すべき情報があるとも思われた。今後検討したい。例えば、次のようなものである。

tweet :

遠足の水遊び中に、よその保育園の大きい男の子が小さい男の子の頭をつかんで何度も水に沈める場面を見かけて、他人の僕が水に飛び込もうとした瞬間に小さい子が脱出してたけど、すぐ飛び込まない親も馬鹿だし、そういう子に育てた親はクソだ。殺人現場を見なくてよかった

→この保育園の引率体制に問題のあることがわかる？

tweet :

下の娘は保育園で最も月齢が小さいためか、よく別クラスの保育士さんやお迎えの母さん方にプニプニされている。日々の癒やしらしい。赤子のほっぺと太ももは尊いな！

→この保育園の人間関係が良好であることがわかる？

tweet :

娘が4日連続の高熱で入院になるかもしれない、上司からも会社休まないでの圧力電話があって、先輩からあなたが休むから私が怒られたとメールがあって、旦那は出張で、明日は保育園の夏祭り実行委員会。いやもうまじ無理。

→子どもの発熱時の苦勞ぶりがわかる。

5.2 ツイートの形式への対応

引用のみのものも意見とはしなかった(3.3 節 f))。しかし、リツイートや引用という行為自体が意見の表明とみられることもできる。それらの扱いを検討したい。RT や引用のパターンは多用であり、引用か、発信者の意見であるか、そもそも判断に迷うものも実はあった。その区別もあわせて見直す必要がある。

また、# (ハッシュタグ)、顔文字、絵文字、ネット上の特有表現などの整理、扱いの指針を検討したい。

5.3 ツイートの特異な内容への対応

あまりに非現実的な提案や、妄想、誹謗中傷のようなものの扱い方も考えたい。

6. おわりに

意見分析コーパスの作成に際し、計算機によって市民意見ありと判断したツイートのうち、まずは「保育園」と「教育」のツイートから、人手によって意見を抽出した。さらに、「保育園」の意見には対象とタイプとを付与した。今回付与した意見のタイプは、評価(ポジ)、評価(ネガ)、肯定・賛同、否定・非難、疑問、不安、指摘(中立)、要望、提案、不満の10種である。

意見の抽出、分類の目的は、自治体の行政の対象になる事柄や話題に関し、市民の意見や意識の把握を容易にすることである。今回の「保育園」に関する意見の抽出・分類の試みにより、その実現可能性をさぐり、そのための課題を明らかにした。

【謝辞】本研究は、JSPS 科研費 16H02913 の助成を得ています。

【参考文献】

Seki, Y. (2016) "Use of Twitter for Analysis of Public Sentiment for Improvement of Local Government Service", *Proc. of the 2nd IEEE Int'l Conf. on Smart Computing (SMARTCOMP 2016)*, St. Louis, MO, USA, pp.325-327.

大塚裕子, 内山将夫, 井佐原均(2004)「自由回答アンケートにおける要求意図判定基準」『自然言語処理』11(2), pp.21-66.

奥村学(2012)「マイクロブログマイニングの現在」『第3回集合知シンポジウム』、『電子情報通信学会技術研究報告 言語理解とコミュニケーション』111(427), pp. 19-24.

川島崇秀, 佐藤哲司, 神門典子(2016)「Twitter からの消費者ニーズの抽出手法に関する提案」『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(DEIM 2016) 論文集』B5-1(5頁).

栗原理聡, 佐々木彬, 松田耕史, 岡崎直観, 乾健太郎(2015)「Twitterを利用した地域毎の要望抽出」『第29回人工知能学会全国大会』1H3-3, pp.1-4.

小林のぞみ, 乾健太郎, 松本裕治(2006)「意見情報の抽出/構造化のタスク仕様に関する考察」『情報処理学会研究報告』NL(171), pp.111-118.

関洋平(2013)「意見分析コーパスの現状と課題」『情報処理学会論文誌データベース』6(4), pp.85-103.

関洋平(2015)「行政サービスのオープンデータ」『ITSCJ Newsletter』106, pp.12-14.

関洋平(2017)「自治体における市民意見を利用した新たな気づき」『スマートシティ「都留市モデル」シンポジウム』

山本修平, 佐藤哲司(2012)「Twitter からの実生活情報の抽出法の提案」『第4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 2012)』F3-4(8頁).