

意見属性に対するマルチタスク学習を利用した  
市民意見抽出と複数都市への適用

筑波大学

人間総合科学学術院人間総合科学研究群

情報学学位プログラム

2023年3月

石田 哲也

# 意見属性に対するマルチタスク学習を利用した 市民意見抽出と複数都市への適用

## Citizen Opinion Extraction with Multitask Learning for Opinion Viewpoints and its Application to Multiple Cities

氏名：石田 哲也

Ishida Tetsuya

行政による政策や接客業によるサービスを向上させるためには、都市で暮らす市民の意見を反映させることが重要となる。そこで本研究では、事前学習済みの言語モデルをファインチューニングすることで、Twitter のつぶやきに現れる市民意見を自動抽出する手法を提案する。しかし、つぶやきに現れる市民意見は多様であるため、従来の意見分析研究で用いられている極性や感情の種類といった特定の観点のみの分析では、多様な市民意見を整理できない。また、自治体による政策や接客業のサービスは都市によって異なるため、特定の都市にとらわれない市民意見抽出を行う必要がある。

本研究では、多様な市民意見を整理するため、Martin と White により定式化されたアプレイザル理論に基づく意見タイプや、極性、地域依存性といった複数の属性と各属性のラベルを定義し、言語モデルの分類結果に基づいて、属性ラベルを自動でつぶやきに付与する。そして、つぶやきに付与された複数の属性のラベルを指定することで、多様な市民意見を整理し、条件を満たすつぶやきのみを抽出する。この際、アプレイザル意見タイプを含む、意見に直接関連する複数の意見属性については、同一のモデルを用いてラベルの分類を行うマルチタスク学習を適用することで、分類精度を向上させる。

また、複数の都市の市民意見を抽出するには、全ての都市で機械学習モデルの訓練データを作成する必要があるが、これはコストが大きい。そこで本研究では、既に市民意見の分析を行っており、訓練データを作成済みの都市（ソース都市）のデータと、新たに分析を行う都市（ターゲット都市）の比較的少量のデータを用いた2段階のファインチューニングを適用することで、ターゲット都市における新たな訓練データの作成コストを削減する。さらに、ソース都市のデータを用いてファインチューニングを行ったモデルによる予測の確信度の情報を用いてターゲット都市のつぶやきを選定することで、ターゲット都市におけるデータを効率的に作成する。

実験では、提案手法の評価を行うため、横浜市民と札幌市民による、コロナ禍において対策が必要とされた「保育園」と「飲食店のテイクアウト」に関するつぶやきに対して複数の属性を付与したコーパスを手で作成した。作成したコーパスにおいて、アノテーションの判定者間一致度を表す Fleiss の  $\kappa$  係数が両都市の全属性で 0.6 を超えたことから、各属性が妥当な基準で定義されたことが示された。

各属性のラベル分類実験では、意見属性に対するマルチタスク学習を適用することで、各属性を独立に分類する手法と比較して F 値が向上し、t 検定（有意水準 5%，両側検定）で有意差が認められた。また、都市を横断した各属性のラベル分類では、ソース都市で訓練したモデルによる確信度が上位のターゲット都市のつぶやきを用いた、2段階のファインチューニング手法を適用することで、複数の比較手法と比較して F 値が向上し、t 検定（有意水準 5%，両側検定）による有意差を確認した。

主研究指導教員：関 洋平

副研究指導教員：高久 雅生

# 目次

<b>第1章</b>	<b>はじめに</b>	<b>1</b>
1.1	本研究の目的	1
1.2	市民意見抽出における課題	2
1.2.1	多様な市民意見の整理	2
1.2.2	都市にとらわれない市民意見抽出	4
1.3	本論文の構成	5
<b>第2章</b>	<b>関連研究</b>	<b>6</b>
2.1	ソーシャルメディア上の市民意見分析に関する研究	6
2.2	アプレイザル理論	7
2.3	アノテーションコストの削減に関する研究	8
2.4	事前学習済み言語モデルを利用したマルチタスク学習	9
2.4.1	マルチタスク学習	9
2.4.2	T5	9
2.4.3	BERTを利用したマルチタスク学習	10
2.5	本研究の位置づけ	11
<b>第3章</b>	<b>都市を横断した市民意見抽出手法の提案</b>	<b>12</b>
3.1	つぶやきに付与する複数の属性の定義	12
3.1.1	意見ユニットに付与する意見属性	13
3.1.2	つぶやき全体に付与するつぶやきの属性	14
3.2	複数の属性を組み合わせた市民意見抽出	16
3.3	都市を横断した市民意見抽出	17
<b>第4章</b>	<b>市民意見分析コーパスの構築</b>	<b>20</b>
4.1	つぶやきの収集方法	20
4.1.1	市民アカウントの収集	20

---

4.1.2	市民アカウントによるつぶやきの収集	21
4.2	各属性の人手によるアノテーション	22
4.2.1	アノテーション方針の一致のための訓練	23
4.2.2	残りの全てのつぶやきに対するアノテーション	23
<b>第5章</b>	<b>実験：都市を横断した各属性のラベル分類</b>	<b>25</b>
5.1	実験の方法	25
5.1.1	予備実験：単一都市における各属性のラベル分類	25
5.1.2	都市を横断した各属性のラベル分類	26
5.1.3	各属性の都市を横断したラベル分類精度と単一都市におけるラベル分類精度の比較	27
5.1.4	各属性の分類実験に用いる評価指標	27
5.2	予備実験：単一都市における各属性のラベル分類の結果	28
5.2.1	意見属性のラベル分類の結果	28
5.2.2	つぶやきの属性のラベル分類の結果	28
5.3	都市を横断した各属性のラベル分類の実験結果	30
5.3.1	都市を横断した意見属性のラベル分類	30
5.3.2	都市を横断したつぶやきの属性のラベル分類	31
5.3.3	結果の考察	33
5.4	各属性の都市を横断した分類精度と単一都市における分類精度の比較	35
<b>第6章</b>	<b>実験：都市を横断したつぶやきからの市民意見抽出</b>	<b>38</b>
6.1	都市を横断した市民意見の抽出方法	38
6.2	テストデータの拡張	39
6.3	結果	40
6.3.1	保育園サービス	40
6.3.2	飲食店のテイクアウトサービス	43
6.3.3	自動抽出された市民意見における抽出ミスの割合に関する考察	47
<b>第7章</b>	<b>おわりに</b>	<b>50</b>
7.1	本研究のまとめ	50
7.2	今後の課題	52
<b>謝辞</b>		<b>53</b>

---

参考文献	55
発表論文	61

# 表 目 次

4.1	各属性の判定者間一致度 ( Fleiss' $\kappa$ ) . . . . .	23
4.2	複数都市を対象とした市民意見分析コーパスのデータ数 . . . . .	24
5.1	T5 モデルと BERT モデルを用いた際のマルチタスク学習手法と各属性独立学習手法の <u>意見属性</u> のラベル分類精度の比較 (F 値) . . . . .	29
5.2	T5 モデルと BERT モデルを用いた際の各属性独立学習手法とマルチタスク学習手法の <u>つぶやきの属性</u> のラベル分類精度の比較 (F 値) . . . . .	29
5.3	ターゲット都市の 25%のデータを用いた都市を横断した <u>意見属性</u> のラベル分類精度 (F 値) . . . . .	30
5.4	ターゲット都市の 50%のデータを用いた都市を横断した <u>意見属性</u> のラベル分類精度 (F 値) . . . . .	30
5.5	ターゲット都市の 25%のデータを用いた都市を横断した <u>つぶやきの属性</u> のラベル分類精度 (F 値) . . . . .	32
5.6	ターゲット都市の 50%のデータを用いた都市を横断した <u>つぶやきの属性</u> のラベル分類精度 (F 値) . . . . .	32
5.7	手法毎に得られる教師データ内の各属性のラベルの割合の分布 . . . . .	34
5.8	<u>意見属性</u> の都市を横断した分類精度と単一都市における分類精度の比較 (F 値) . . . . .	35
5.9	<u>つぶやきの属性</u> の都市を横断した分類精度と単一都市における分類精度の比較 (F 値) . . . . .	36
6.1	複数都市を対象とした市民意見分析コーパスの拡張データのデータ数 . . . . .	39
6.2	保育園サービスで自動抽出された市民意見における各属性のモデルの予測ラベルが正しい割合 . . . . .	48
6.3	テイクアウトサービスで自動抽出された市民意見における各属性のモデルの予測ラベルが正しい割合 . . . . .	49

# 目 次

1.1	多様な市民意見の整理のためのフレームワークの概要 . . . . .	3
1.2	都市を横断した市民意見抽出手法の概要 . . . . .	4
3.1	複数の属性を用いた市民意見抽出手法 . . . . .	16
3.2	都市を横断した市民意見抽出 . . . . .	18
6.1	保育園サービスの休園・登園自粛に関する市民意見の出現頻度の時系列順 の推移 . . . . .	40
6.2	飲食店のテイクアウトサービスの市民意見の出現頻度の時系列順の推移 .	44

# 第1章 はじめに

## 1.1 本研究の目的

行政による政策を改善するためには、都市で暮らす市民の意見を反映させることが重要となる。そのため、自治体はパブリックコメントや住民アンケート等の行政への市民参加を促進するための制度を設けている。しかし、このような既存の制度に対しては、「住民の意識・関心が低く、参加者が少ない」、「参加者の偏りや固定化などがある」といった課題が自治体から指摘されており [39, 40]、より多くの市民意見を収集する新たな方法が必要と考える。また、接客業のサービスにおいても、顧客の声をもとにした経営戦略の見直しが行われている [44] ことから分かるように、実際にサービスを利用した市民による意見を取り入れる必要がある。さらに、他の利用者による接客業のサービスの評価は、市民がどのサービスを利用するかを決定する際の判断材料として活用することができる。

このような背景から、本研究では、事前学習済みの言語モデルをファインチューニングすることで、Twitter<sup>1</sup>のつぶやきに現れる市民意見を自動抽出する手法を提案する。Twitter は多くのユーザが利用しているソーシャルメディアであり、ユーザが普段感じたことをつぶやきとして気軽に投稿している。そのため、市民が行政に対して感じている不満や批評といった意見、また、利用した接客業のサービスに関する評価等の多様な意見を抽出することができる。

提案手法の有効性を検証するため、本研究では、新型コロナウイルス感染症の感染拡大によって休園や登園自粛が大きな問題となった「保育園」に関する市民意見と、同じく新型コロナウイルス感染症の感染拡大によって急激に利用者が増加した「飲食店のテイクアウトサービス」に関する市民意見を抽出する。保育園のサービスに関しては、地方自治体のような行政による政策の影響が大きい。そのため、つぶやきに現れる保育園に関する市民意見を分析対象とすることで、本研究で提案する市民意見抽出手法によって行政の政策の改善に有意義な意見を抽出できるかを検証する。また、飲食店のテイク

---

<sup>1</sup><https://twitter.com>



アウトサービスに関しては、保育園サービスの行政的な側面とは対称的に、商業的な側面に関する市民意見の抽出を行うことを目的とする。

また、行政の政策や接客業のサービスが都市によって異なることから、市民の抱える意見も都市によって異なる。そのため、特定の都市のみではなく、複数の都市を対象とした際にも市民意見抽出手法が有効であるかを検証する必要がある。そこで本研究では、政令指定都市である横浜市と札幌市を対象として市民意見抽出を行うことで、複数の都市における手法の有効性を検証する。

## 1.2 市民意見抽出における課題

### 1.2.1 多様な市民意見の整理

従来の意見分析研究では、肯定や否定といった意見の極性 [35, 32, 14, 7] や、喜びや悲しみといった感情の種類 [8, 13, 20, 23] の観点から意見を分析している。しかし、実際に市民が抱えている意見は多様であるため、特定の観点からの分析では不十分である。たとえば、同じ否定意見という条件で市民のつぶやきから保育園に関する意見を抽出しても、保育園に関する政策を公表した行政に対する意見や、コロナ禍で子供を保育園に通わせることに関する不安感情等の様々な意見が抽出されてしまう。また、その意見が都市で暮らす市民ならではの意見、つまり、地方自治体による政策等の地域の特色が関連している意見なのか、全国一般的な意見であるかという情報や、保育士の意見であるか、子育て中の親の意見であるかといった情報を、整理して分析することができない。

このような課題を解決するため、著者の先行研究では、複数の観点からつぶやきを分析し、それらの観点を属性としてつぶやきに付与することで多様な市民意見を整理するフレームワークを提案している [12, 42, 43, 41]。多様な市民意見の整理のためのフレームワークの概要を図 1.1 に示す。このフレームワークでは、つぶやきに現れる意見の極性やアプレイザル理論 [18] に基づく意見タイプといった、意見に直接関連する意見属性と、つぶやきの地域依存性や投稿主の立場といった、つぶやき全体に関する属性を定義する。さらに、これらの属性について、つぶやきに人手でアノテーションを行った市民意見分析コーパスを構築し、教師データとすることで、事前学習済みの言語モデルをファインチューニングし、つぶやきの各属性のラベル分類モデルを準備する。そして、キーワードを含む大量の未知のつぶやきを、訓練済みの各属性のラベル分類モデルに入力することで、全てのつぶやきに複数の属性を付与する。このようにしてつぶやきに付与された

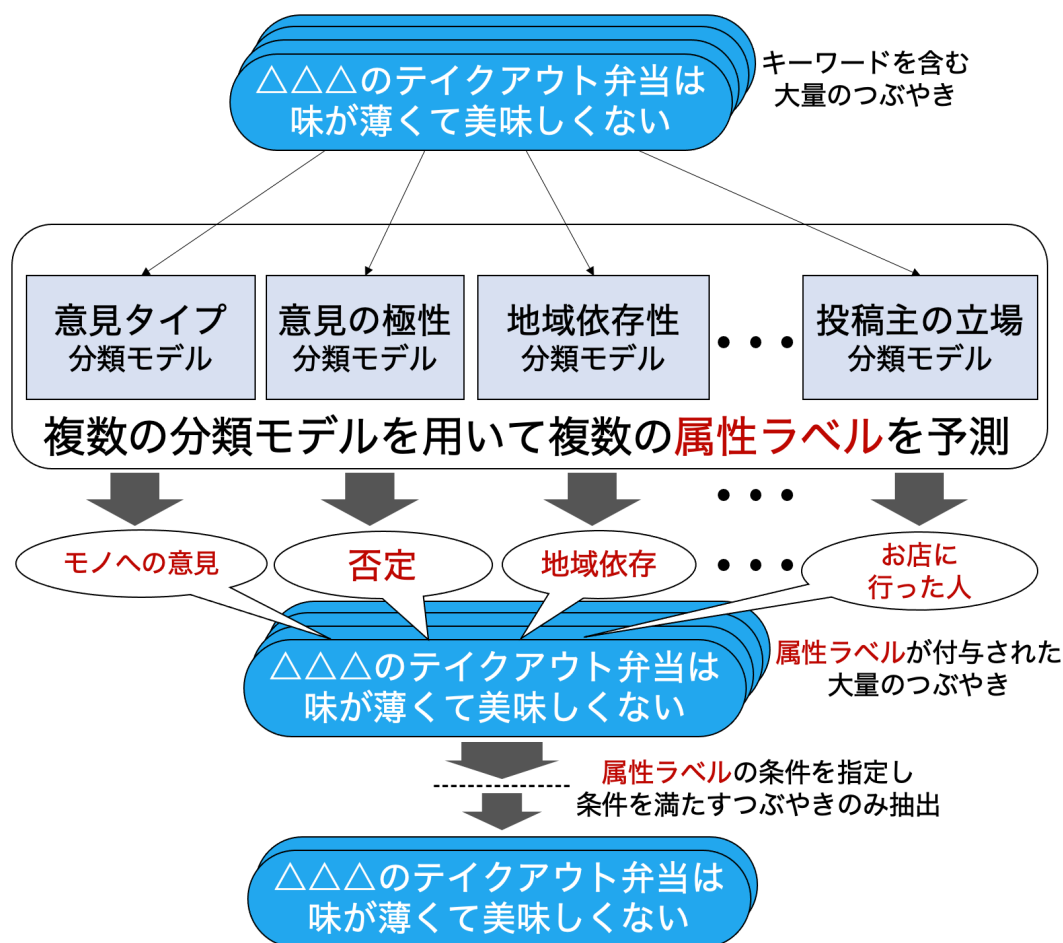


図 1.1: 多様な市民意見の整理のためのフレームワークの概要

複数の属性のラベルを任意の数だけ指定することで、大量のつぶやきの中から、指定された属性の条件を全て満たすつぶやきのみを自動抽出する。

つぶやきの属性ラベルの分類においては、マルチタスク学習手法 [3] を適用することで、分類精度の向上を目指す。マルチタスク学習は、共通のモデルを用いて関連する複数のタスクを同時に学習することで、各タスクを独立して学習するよりも精度向上が実現可能な手法である。本研究で定義する意見属性は、全てつぶやき内の同じ意見に対する属性となっている。そのため、意見属性のラベル分類タスクについては、各タスク間に関連性があり、マルチタスク学習の適用による精度向上が期待できる。意見属性の分類におけるマルチタスク学習手法の有効性については、5章の各属性のラベル分類実験において検証する。

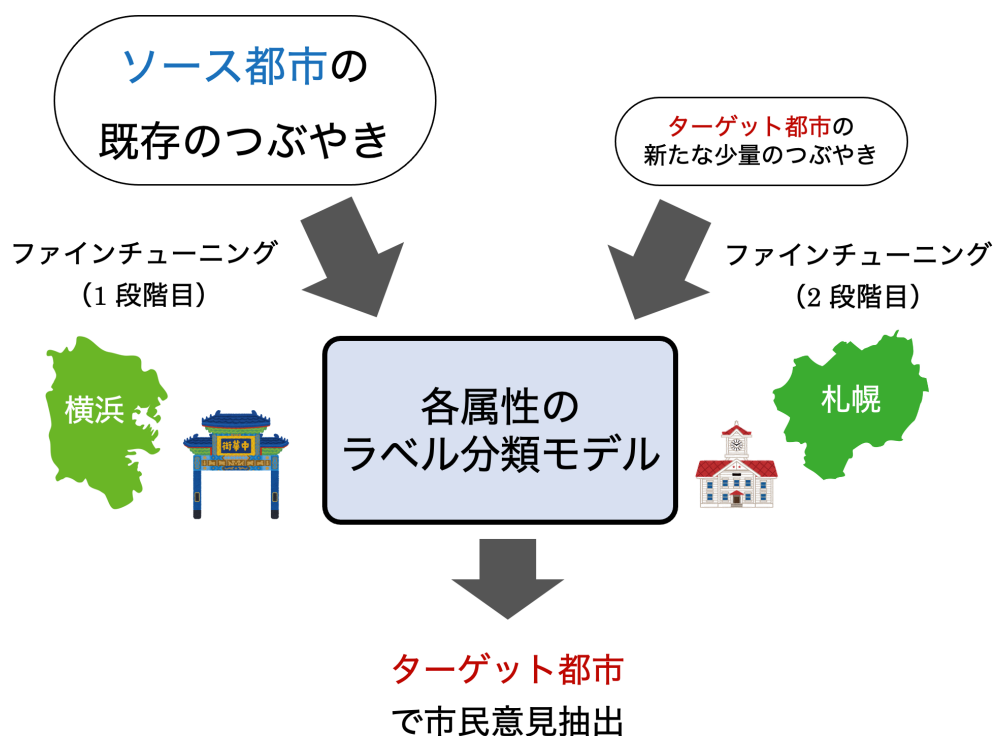


図 1.2: 都市を横断した市民意見抽出手法の概要

### 1.2.2 都市にとらわれない市民意見抽出

地方自治体による政策は、市民の生活している都市によって異なる。また、接客業のサービスについても、利用している都市の店舗によって異なる。そのため、市民の意見も、生活している都市の自治体による政策や普段利用している都市の接客業のサービスによって異なる。このような背景から、つぶやきに現れる市民意見を抽出する際には、特定の都市にとらわれず、複数の都市において市民意見を抽出可能な手法が必要となる。また、都市によって異なる市民意見を、同一のモデルによって全て抽出することは難しく、各都市の市民意見抽出に特化したモデルが必要となる。一方で、これまで分析を行っていた都市とは異なる都市において市民意見を分析するためには、市民意見抽出モデルの訓練に用いるための教師データを新たな都市における分析用に作成する必要があるが、全ての都市において人手で教師データを作成するのはコストが大きく、このような実装方法は現実的ではない。

そこで本研究では、既に市民意見の分析を行っており、教師データを構築済みの都市（以降、ソース都市と呼ぶ）のデータと、これから分析を行おうとしている評価対象の別の都市（以降、ターゲット都市と呼ぶ）の比較的少量のデータを用いて、都市を横断してターゲット都市における市民意見を抽出する手法を提案する。都市を横断した市民意

見抽出手法の概要を図1.2に示す。本手法では、はじめにソース都市のデータを用いることで、つぶやきの各属性のラベル分類モデルをファインチューニングする。続いて、新たに作成したターゲット都市の比較的少量のデータを用いることで、再度モデルのファインチューニングを行う。つまり、複数の都市を横断して、モデルの2段階のファインチューニングを適用する。本手法を用いることで、これまでに市民意見を抽出していた都市とは別のターゲット都市において新たに市民意見を抽出する際に、既存のソース都市のデータを有効活用し、新たな教師データの作成コストを削減して分析を行うことが可能となる。さらに、2段階目のファインチューニングにおいて、ターゲット都市のデータを用いることで、ターゲット都市の市民意見抽出に特化したモデルを構築することができる。

また、ターゲット都市における市民意見抽出に使用する比較的少量の新たなつぶやきの選定についても、ランダムに行うのではなく、効率化する必要があると考える。そこで本研究では、ソース都市のつぶやきを用いて1段階目のファインチューニングを行ったモデルによる、ターゲット都市のつぶやきに対する各属性のラベル分類の予測の確信度の情報を用いることで、効果的にターゲット都市のつぶやきを選定する手法を提案する。ラベル分類モデルの予測の確信度については、高い確信度のつぶやきを対象としてアノテーションを行う手法が有効であるか、もしくは低い確信度のつぶやきを対象とする手法が有効であるかを、5章の都市を横断した各属性のラベル分類実験において検証する。

### 1.3 本論文の構成

本論文の構成を以下に示す。2章では、アプレイザル理論や、ソーシャルメディア上の市民意見分析に関する研究、アノテーションコストの削減に関する研究など、本研究の関連研究について述べ、本研究の位置づけを示す。3章では、提案手法である都市を横断した市民意見抽出手法について述べる。4章では、5章、6章の実験に用いる市民意見分析コーパスの構築のための、つぶやきの収集方法と各属性の人手のアノテーションについて述べる。5章では、各属性のラベル分類実験を通して、提案手法の定量的な有効性について検証する。そして6章では、提案手法を用いて、実際に大量のつぶやきから都市を横断して市民意見を抽出した結果について述べる。最後に7章において、本研究のまとめと今後の課題について述べる。

## 第2章 関連研究

本研究では、アプレイザル理論に基づく意見タイプを含む、複数の属性を定義することで、ソーシャルメディア上のつぶやきに現れる多様な市民意見を整理して抽出する。また、ソース都市で作成済みのアノテーションデータと、ターゲット都市における比較的少量のアノテーションデータを活用することで、新たな教師データの作成のためのアノテーションコストを抑えた上で、ターゲット都市における市民意見を抽出する。さらに、意見属性のラベル分類に対してマルチタスク学習を適用することで、分類精度の向上を目指す。

そこで本章では、はじめに2.1節において、ソーシャルメディア上の市民意見分析に関する研究について述べる。続いて2.2節では、アプレイザル理論とアプレイザル理論を用いた意見分析研究について述べる。さらに、2.3節では、アノテーションコストの削減手法である、能動学習手法と半教師あり学習手法について述べる。2.4節では、マルチタスク学習手法と、本研究で用いる言語モデルである、T5モデルとBERTモデルについて述べる。最後に2.5節で、関連研究を踏まえた本研究の位置づけを示す。

### 2.1 ソーシャルメディア上の市民意見分析に関する研究

行政の政策や接客業のサービスは都市によって異なるため、ソーシャルメディア上に現れる市民意見も、市民が暮らす都市によって異なる。都市別に市民意見を分析するにあたり、全ての都市において人手で教師データを作成するのはコストが高いため、絵文字を用いたルールベース手法でつぶやきの極性を判定する手法 [35] や、つぶやきを用いて事前学習された公開済みの言語モデル [1] の予測結果に基づいてつぶやきの極性を自動推定することで、教師データを作成する手法 [32, 14] が提案されている。これらの手法では、アノテーションコストを削減して市民意見の分析ができる一方で、市民意見の傾向が都市ごとに異なるにも関わらず、都市ごとに教師データを分けること無く、全ての都市について同じ手法で分析を行っている。さらに、ルールベース手法に基づいて全てのつぶやきの極性を分類する場合、人手によるアノテーションとは異なる結果が得

られてしまう [35]. こういった課題を解決するため, 本研究では, ターゲット都市においても, アノテーションコストを抑えた上で比較的少量のアノテーションデータを作成する. さらに, 既に豊富に存在するソース都市の教師データと, ターゲット都市の比較的少量のデータを組み合わせて使用することで, 都市ごとに異なる市民意見に対応する手法を提案する.

また, 上記の研究では, ソーシャルメディア上のつぶやきについて, 肯定や否定といった特定の観点からの分析を行っている. ソーシャルメディア上のつぶやきについて分析している他の多くの研究においても, 極性 [7] や感情の種類 [8, 13, 20, 23] といった特定の観点からつぶやきの分析を行っている. しかし, ソーシャルメディア上の市民意見は多様であるため, 特定の観点からの分析では多様な市民意見の整理が出来ない. そこで本研究では, つぶやきの地域依存性や極性, さらに 2.2 節で述べるアプレイザル理論に基づく意見タイプを含む複数の属性を 3.1 節で定義し, それらの属性を組み合わせることで, つぶやきに現れる多様な市民意見の整理を試みる.

## 2.2 アプレイザル理論

アプレイザル理論 [18] は, 選択体系機能言語学 [10] の立場から提案された意見を体系化する理論である. アプレイザル理論では, テキストに現れる対人メタ機能は仮想的な読者 (putative reader) に対する感情であるという考えに基づき, テキスト内の対人メタ機能を, appraisal, negotiation, involvement の 3 つのシステムから構成されるものとした. また, このうち appraisal については, さらに attitude, engagement, graduation の 3 つの要素から成り立つとした. そして, アプレイザル理論において, テキストに現れる態度 (attitude) は, 「自発的感情の表明 (affect)」, 「人間・組織の振舞や行為を対象とした批評 (judgment)」, 「事物・事象を対象とした評価 (evaluation)」の 3 種類に分類される. つまり, アプレイザル理論における態度の分類体系を用いることで, 市民の抱える意見を事物・事象や人間・組織といった対象に着目して分類することができる.

関 [46] は Yahoo!知恵袋のコミュニティQA 文書の複数のカテゴリについて, アプレイザル理論における態度の分類体系を用いて分析し, 政治カテゴリの文書には人間・組織の振舞や行為を対象とした批評 (judgment) が多く含まれることを示した. Mora ら [21] は, オンラインストアのレビューデータについて, アプレイザル理論における態度の分類体系を用いて分析し, レビューデータには事物・事象を対象とした評価 (evaluation) が多く含まれることを示した. このように, 政治やレビューデータといったカテゴリ毎

に主要な意見は異なる。本研究においても、行政の政策と深い関連がある保育園サービスと、商業的な側面が強い飲食店のテイクアウトサービスという、異なる特徴を持つ2つのサービスに関する市民意見を分析する。そこで、アプレイザル理論の態度の分類体系を用いることで、異なる特徴を持つ複数のサービスに関する、多様な市民意見の整理を試みる。

## 2.3 アノテーションコストの削減に関する研究

アノテーションコストを削減し、新たな教師データを効果的に作成する手法として、能動学習 (active learning) [29, 26] が有効であり、意見分析研究においても用いられている [24, 36, 30]。能動学習手法では、既存のアノテーション済みのデータを教師データとして訓練を行ったモデルを用いて、アノテーションを行っていないデータの中から、新たにアノテーションを行う候補となるデータを選定する。アノテーション候補となるデータの選定の際には、一般的に、現段階でのモデルの予測の確信度が低いデータを選定する [16]。つまり、既存のアノテーションデータのみではモデルの学習が不十分であり、モデルが高い確信度で予測を行えないデータに対して新たにアノテーションを行うことで、モデルの予測性能を向上させている。

一方で、半教師あり学習 (semi-supervised learning) [38, 33] も、限られたアノテーション済みのデータを活用してモデルの性能を向上させることができる手法であり、意見分析研究においても用いられている [11, 31, 22, 4, 27]。半教師あり学習のうち、特に自己訓練 (self-training) [37, 11, 27] や、共訓練 (co-training) [2, 22, 4] の手法では、能動学習手法と同様に、既存のアノテーション済みのデータを教師データとして訓練を行ったモデルを用いて、アノテーションを行っていないデータの予測を行う。しかし、半教師あり学習の場合、能動学習手法とは逆に、モデルの予測の確信度が高いデータに対して予測結果を自動的に付与することで、教師データとして活用する。

能動学習手法では、モデルの予測の確信度が低いデータに対してアノテーションを行うことでモデルの性能が示されている一方で、半教師あり学習手法では、予測の確信度が高いデータを用いることがモデルの性能の向上に繋がることが示されている。本研究においても、ソース都市のデータを教師データとしてファインチューニングを行ったモデルによる予測の確信度を用いて、ターゲット都市のデータから新たにアノテーションを行うデータの選定を行う。そこで実験において、関連研究を参考に、モデルの予測の確信度が高いデータを選定する手法と、モデルの予測の確信度が低いデータを選定する

手法のどちらが有効であるかを検証する。

## 2.4 事前学習済み言語モデルを利用したマルチタスク学習

### 2.4.1 マルチタスク学習

マルチタスク学習 [3] は、共通のモデルを用いて関連する複数のタスクを同時に学習することで、各タスクを同時に学習するよりもモデルの性能を向上させる手法である。マルチタスク学習においては、異なるタスクを同時に学習することで得られた情報が帰納バイアス (inductive bias) [19] として働き、モデルの一般性を向上させることに繋がるため、未知のデータに対する予測性能を向上させることができる。本研究で提案する市民意見抽出手法では、多様な市民意見の整理のために、複数の属性のラベル分類タスクを行う。そこで、マルチタスク学習を用いて関連する属性の学習を同時に行うことで、各属性のラベル分類精度の向上を目指す。

### 2.4.2 T5

T5モデル [25] は、Transformer[34] の構造を基本とした言語モデルであり、意見分析タスクを含む多くの自然言語処理タスクで高い性能を示している。また、大規模なデータを用いて事前学習されたモデルを、様々な下流タスクにファインチューニングすることができる。T5モデルでは、分類タスクや回帰タスク、生成タスク等の全ての自然言語処理タスクの入出力をテキストの形式で統一している。具体的には、モデルの入力は “<prefix>: <input text>” の形式で統一されており、出力形式は “<output text>” の形式で統一されている。入力形式の “<prefix>” の部分には、現在の入力テキスト (“<input text>”) がどのタスクに対応したテキストであるかを表す情報を格納する。例えば, “translate Japanese to English: 私は寿司が嫌いです。” という入力を与えられた場合、モデルは “I don’t like Sushi.” という出力テキストを予測できるよう訓練される。また, “classification: 私は寿司が嫌いです。” という入力を与えられている場合、出力は “私は寿司が嫌いです。” という文章に対応する何らかの分類ラベルとなる。仮にモデルの訓練段階で, “classification” という prefix が与えられた際に肯定や否定と行った極性を判定する分類タスクを行うよう訓練されていた場合、出力は “否定” に対応するラベルとなる。なお, prefix はモデルのユーザが自由に定義することができる。また, 極性の分類と意見タイプの分類等, 複数の分類タスクを共通の T5 モデルを用いて同時に行う場合, “appraisal classification”



と “polarity classification” のように異なる prefix を用意することで、各タスクを別のタスクとしてモデルに学習させることができる。このように、T5 モデルでは、<prefix> を複数種類用意するだけで、共通のモデルを用いて複数の自然言語処理タスクを同時に行うことができるため、マルチタスク学習モデルの構築に適した構造となっている。

著者の先行研究 [42] では、T5 モデルを用いてマルチタスク学習モデルを構築し、意見属性のラベル分類タスクにおいて、各属性を同時に分類するよりも高い精度が実現可能であることを示した。本研究においても、5章で T5 モデルを用いたマルチタスク学習を構築し、各属性のラベル分類タスクにおける有効性を検証する。

### 2.4.3 BERT を利用したマルチタスク学習

BERT モデル [6] は、T5 モデルと同様に Transformer の構造を基本とした言語モデルであり、大規模なデータで事前学習済みのモデルを下流タスクにファインチューニングすることができる。

著者の先行研究 [12] では、BERT を用いたマルチタスク学習手法である、Multi-Task Deep Neural Networks [17] を参考にしてマルチタスク学習モデルを構築し、T5 モデルと同様に、意見属性のラベル分類タスクにおいて、各属性を独立に学習するよりも高い精度が実現可能であることを示した。Multi-Task Deep Neural Networks では、マルチタスク学習を行うにあたり、各タスク間で共有する層と、各タスク固有の層からなるモデルを構築している。この際、各タスクに共通の層が BERT モデルとなっており、BERT モデルから出力された分散表現が、文章分類層や 2 文の類似性判定層などの各タスクに固有の層へと入力される。マルチタスク学習を行う際は、はじめに各タスクのデータを全てバッチに変換し、バッチ変換後の全データをランダムにシャッフルする。こうすることで得られた、全タスクがランダムにバッチ単位で並べられたデータを順に入力することで、モデルの訓練を行う。本研究においても、事前学習済みの BERT モデルをファインチューニングするにあたり、バッチごとにシャッフルされた複数の属性のラベル分類タスクのデータを順にモデルに入力することで、マルチタスク学習を行う。

また、本研究では、5章の各属性のラベル分類実験において、マルチタスク学習手法の有効性の検証に加え、T5 モデルと BERT モデルの性能を比較検証する。

## 2.5 本研究の位置づけ

本研究では、アプレイザル理論に基づく意見タイプを含む複数の属性を定義することで、従来の意見分析研究と比較して詳細な市民意見の整理を行う。また、ソース都市の豊富なアノテーションデータと、ターゲット都市の比較的少量のアノテーションデータを用いた手法を適用することで、アノテーションコストを抑えた上でターゲット都市の市民意見抽出を行う。この際、ソース都市のデータを教師データとしてファインチューニングを行ったモデルによる、予測の確信度の情報を用いて、ターゲット都市のつぶやきから新たにアノテーションを行うべきつぶやきを選定する。つぶやきの選定においては、能動学習手法と半教師あり学習手法を参考に、予測の確信度が高いデータと低いデータのどちらを用いる手法が有効であるかの検証を行う。さらに、各属性のラベル分類には、マルチタスク学習手法を適用する。マルチタスク学習においては、先行研究で用いたBERTモデルとT5モデルの双方を用いることで、本研究の市民意見抽出に対して有効な手法を明らかにする。

## 第3章 都市を横断した市民意見抽出手法の提案

本章では、提案手法の都市を横断した市民意見抽出について述べる。提案手法では、はじめにつぶやきに対して複数の属性を付与した市民意見分析コーパスを人手で作成する。続いて、作成したコーパスを教師データとして事前学習済みの言語モデルをファインチューニングすることで、各属性のラベル分類モデルを訓練する。このようにして訓練を行ったラベル分類モデルに未知のつぶやきを入力することで、全てのつぶやきに対して自動で複数の属性ラベルを付与する。これらの属性ラベルを任意の数だけ指定することで、未知のつぶやきのうち、条件を満たすつぶやきのみを整理して抽出することができる。

また、本研究では、前述の複数の属性を用いた市民意見抽出手法を拡張し、都市を横断して市民意見抽出を行う手法を提案する。都市を横断した市民意見抽出では、ソース都市におけるアノテーション済みのデータと、ターゲット都市における比較的少量のアノテーション済みのデータを用いてモデルの2段階のファインチューニングを行う。本手法によって、ターゲット都市における新たなデータ作成のコストを削減して市民意見抽出を行う。

そこで、はじめに3.1節では、市民意見抽出に用いる複数の属性を定義する。3.2節では、3.1節で定義した複数の属性を用いた市民意見抽出モデルについて述べ、3.3節では3.2節の手法を拡張し、都市を横断して市民意見抽出を行う手法について述べる。

### 3.1 つぶやきに付与する複数の属性の定義

本研究では、アプレイザル理論に基づく意見タイプを含む、意見に直接関連する意見属性を複数定義する。しかし、1つのつぶやきには複数の意見が含まれることが多く、意見属性のラベルを付与する対象として、つぶやき全体という単位は大きすぎる。そこで、意見属性については、1つの意見のみを含む文、または節（以降、1つの意見のみを含む文、もしくは節を意見ユニットと呼ぶ[28].）を対象として属性ラベルを付与する。

### 3.1.1 意見ユニットに付与する意見属性

本研究では、2.2節で述べたアプレイザル理論に基づく意見タイプを定義することで、事物・事象や人間・組織の振舞といった対象に着目した市民意見抽出を行う。しかし、本研究で用いるアプレイザル理論の態度 (attitude) に含まれる意見は、肯定と否定の双方の意見を含むが、アプレイザル意見タイプ属性のみではこれらの極性は判定することができない。そこで、極性の属性を定義することで、肯定的な意見と否定的な意見を区別して分析可能にする。さらに、アプレイザル意見タイプのみでは、肯定や否定といった極性を持たない中立的な立場の意見を網羅することができない。そこで、推測や疑問といった中立的な立場の意見をコミュニケーション意見タイプとして定義することで、本手法で分析可能な意見の網羅性を高める。

- アプレイザル意見タイプ

接客業のサービスを改善する際に、提供している商品の改善点を見つける際には事物・事象を対象とした意見が重要となり、店員の接客の改善点を見つける際には人物の振舞を対象とした意見が重要となる。また、コロナ禍における保育士の対応のうち、保護者が満足に感じた対応を分析する際には、人物の振舞を対象とした意見が重要となり、保育園に通うことに対する保護者の意見を知りたい場合、自発的感情が重要となる。そこで、これらの市民意見を整理して抽出するため、2.1節のアプレイザル理論に基づく意見タイプを定義する。ラベルの選択肢は「自発的感情の表明」、「人間・組織の振舞や行為を対象とした批評」、「事物・事象を対象とした評価」、「該当無し」。

- 極性

行政による政策や接客業のサービスについて、改善すべき点を明らかにする際には否定的な市民意見が重要となり、評価されている点を明らかにする際には肯定的な意見が重要となる。そこで、本研究における意見属性の1つとして、従来の意見分析研究においても用いられている極性を判断する。なお、極性が中立となるのは、次に述べるコミュニケーション意見タイプに含まれる意見が表れた場合とする。ラベルの選択肢は「肯定」、「否定」、「中立」、「意見無し」。

- コミュニケーション意見タイプ

アプレイザル理論に含まれる意見は、全て肯定や否定といった極性を持つ意見となっている。そのため、アプレイザル理論に含まれる意見のみでは、認識や言語行

為等を表す、推測や疑問といった中立的な意見については網羅できない。そこで、モダリティおよび言語行為論を参考にして [45]、アプレイザル理論に基づく意見タイプのみでは分析しきれない意見タイプを、コミュニケーション意見タイプとして定義する。これによって、本研究で網羅できる市民意見を拡張する。ラベルの選択肢は「推測」、「提案」、「疑問」、「要求」、「該当無し」。

### 3.1.2 つぶやき全体に付与するつぶやきの属性

市民の抱える意見が生活している都市に特有であるか社会一般的な意見であるかによって、意見を反映させる対象は異なる。そのため、地域依存性の属性を定義することで、都市で暮らす市民特有の意見を整理して抽出可能とする。さらに、単語としてはサービスに関連するものの、つぶやきの意味としてはサービスに関連しない場合がある。このようなつぶやきからの意見抽出を避けるため、サービスとの適合性属性を定義する。また、例えば保育園サービスであれば、意見を抱えている人が保育士であるか、小さい子を持つ親であるかによって意見が異なる場合が想定されるため、投稿主の立場属性を定義し、市民の立場ごとに整理された意見抽出を可能とする。

また、本研究では、保育園サービスと飲食店のテイクアウトサービスの2つのサービスを対象として市民意見抽出を行うが、3.1.1節の意見属性や上記で述べたつぶやき全体に付与する3つの属性は2つのサービスに関するつぶやきに共通して付与する属性となっている。これらの属性に加えて、本研究では各サービスに固有の属性を定義する。各サービス固有の属性は、特定の話題との関連性とする。事前に意見を抽出したいと考えている特定の話題との関連性を判定するモデルを訓練しておき、全てのつぶやきに対して特定の話題との関連性の属性ラベルを付与することで、その話題に関連するつぶやきのみを自動抽出することが可能となる。

- 地域依存性

行政に対する市民意見であっても、その意見が地方自治体への意見であるか、国政政党への意見であるかによって、意見を反映させる対象は異なる。そこで、つぶやきに現れる意見が都市で暮らす市民特有の意見であるか、もしくは社会一般的な意見であるかを区別するために、地域依存性の属性を定義する。本研究では、地域に依存するとは、市や県等の地名や、地域が特定できる飲食店等の施設名を含むもの、もしくはつぶやきの内容が市民の暮らす地域に関連するものであると定義する。ラベルの選択肢は「依存」、「非依存」。

- サービスとの適合性

本研究における市民意見抽出では、はじめに各サービスの分析対象となるつぶやきを、「保育園」や「持ち帰り」といったキーワード検索によって収集する。しかし、保育園サービスにおける「保育園の近くのお店でご飯を食べた」というつぶやきや、飲食店のテイクアウトサービスにおける「仕事を家に持ち帰りした」というつぶやきのように、つぶやき内にサービスに関連する単語が含まれているものの、意味的にはサービスとは関連の無いつぶやきが存在する。そこで、これらの意味的に各サービスと関連のないつぶやきからの市民意見抽出を避けるため、サービスとの適合性属性を定義する。ラベルの選択肢は「適合」、「不適合」。

- 投稿主の立場

保育園サービスに関連する意見であっても、保育士の意見であるか、保育園に子供を通わせる保護者の意見であるかによって主張は異なるが、行政が政策を改善するためには、双方の意見を取り入れた上で改善策を見出す必要がある。そこで、投稿主の立場の属性を定義することで、つぶやきがどのような立場のユーザによって投稿されたものかを分析可能にする。ラベルの選択肢は、保育園サービスが「小さい子を持つ親」、「保育園関係者」、「その他」、飲食店のテイクアウトサービスが「店を利用した人」、「飲食店の従業員」、「その他」。

#### 特定の話題との関連性（保育園サービス）

- 休園・登園自粛との関連性

新型コロナウイルス感染症の感染拡大によって、各自治体から保育園の休園や登園自粛に関する政策が発表され、全国で大きな問題となった。そこで、これらの政策に対する市民意見を抽出するため、つぶやきと保育園の休園や登園自粛との関連性を判断する属性を定義する。ラベルの選択肢は「関連する」、「関連しない」。

- 保育園の定員との関連性

現在日本では、多くの自治体において、保育園の数が足りず、子供を保育園に預けたくても預けられない待機児童問題が発生している。特に横浜市は、国内でも待機児童問題が顕著な都市となっている。そこで、待機児童問題や保育園の合否といった、保育園の定員の話題との関連性を判断する。ラベルの選択肢は「関連する」、「関連しない」。

#### 特定の話題との関連性（飲食店のテイクアウトサービス）

---

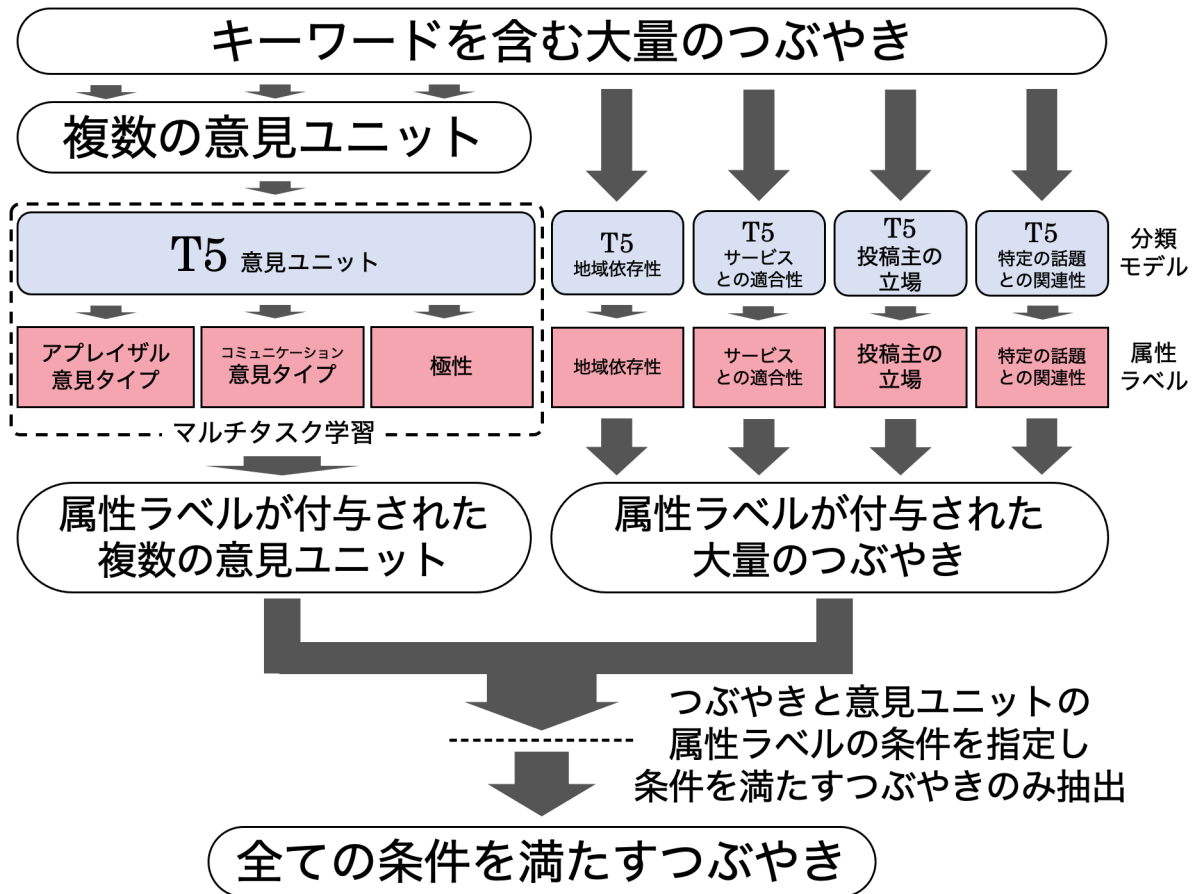


図 3.1: 複数の属性を用いた市民意見抽出手法

- 商品の評価との関連性

飲食店のテイクアウトサービスに関するつぶやきには、実際に購入した商品への評価が多く含まれる。このようなつぶやきは、飲食店の店員が提供している商品を改善するのに役立つことができるだけでなく、市民が飲食店を探す際にも有用なつぶやきとなる。そこで、これらのつぶやきのみを整理して抽出するため、つぶやきに商品の味や量、提供状態といった評価を含むかを判断する属性を定義する。ラベルの選択肢は「関連する」、「関連しない」。

### 3.2 複数の属性を組み合わせた市民意見抽出

複数の属性を組み合わせた市民意見抽出手法を図 3.1 に示す。本手法では、はじめに各サービスに関連するキーワードを含むつぶやきを収集し、つぶやきを 1 つの意見のみが含まれる意見ユニットとなるまで分割する。続いて、意見ユニットとつぶやき全体に

3.1 節で定義した各属性のラベルを人手で付与した市民意見分析コーパスを教師データとして、事前学習済み T5 モデルをファインチューニングすることで、各属性のラベル分類モデルを訓練する（市民意見分析コーパスの構築については、4 章で述べる。）。

意見ユニットに付与するアプレイザル意見タイプ、極性、コミュニケーション意見タイプの 3 つの意見属性は、全て意見ユニット内の同じ意見に対する属性であるため、各属性のラベル分類タスクには関連性がある。また、3.1.1 節で述べたように、極性が中立となる際にはコミュニケーション意見タイプを含み、アプレイザル意見タイプは肯定と否定の意見を含むということからも、各属性のラベル分類タスクには関連性があると考えることができる。そこで、これらの 3 つの意見属性の分類には、2.3 節で述べたマルチタスク学習手法を適用した T5 モデルを用いる。マルチタスク学習は、関連するタスクを共通のモデルを用いて同時に行うことで精度向上が可能なモデルである。そのため、意見属性のラベル分類のように、各タスク間に強い関連性がある場合は、マルチタスク学習手法による精度向上が期待できると考える。一方で、つぶやき全体に付与するつぶやきの属性については、各属性間に関連性は無いため、ラベル分類は独立の T5 モデルを用いて行う。

このようにしてファインチューニングを行った全てのラベル分類モデルに、未知のつぶやきと、つぶやきを分割した意見ユニットを入力し、各属性のラベルを判定する。ラベル分類モデルの予測結果を各属性のラベルとすることで、自動で全てのつぶやき、意見ユニットに対して複数の属性ラベルを付与する。そして、これらの属性のラベルを任意の数だけ指定することで、大量のつぶやきに現れる多様な市民意見を整理し、属性の条件を満たすつぶやきのみを抽出することができる。

### 3.3 都市を横断した市民意見抽出

本節では、3.2 節の複数の属性を組み合わせた市民意見抽出手法を拡張することで、都市を横断して市民意見を抽出する手法を提案する。都市を横断した市民意見抽出では、新たなデータ作成のコストを減らした上で、ターゲット都市における市民意見を抽出することを目的とする。そこで本節では、ソース都市で作成済みのデータを効果的に活用することで、ターゲット都市におけるデータ作成コストを削減可能な手法について述べる。

都市を横断した市民意見抽出手法を図 3.2 に示す。本手法では、はじめに正解ラベルが付与されているソース都市のつぶやきを教師データとして、各属性のラベル分類モデルの 1 段階目のファインチューニングを行う。続いて、ソース都市のつぶやきを用いて



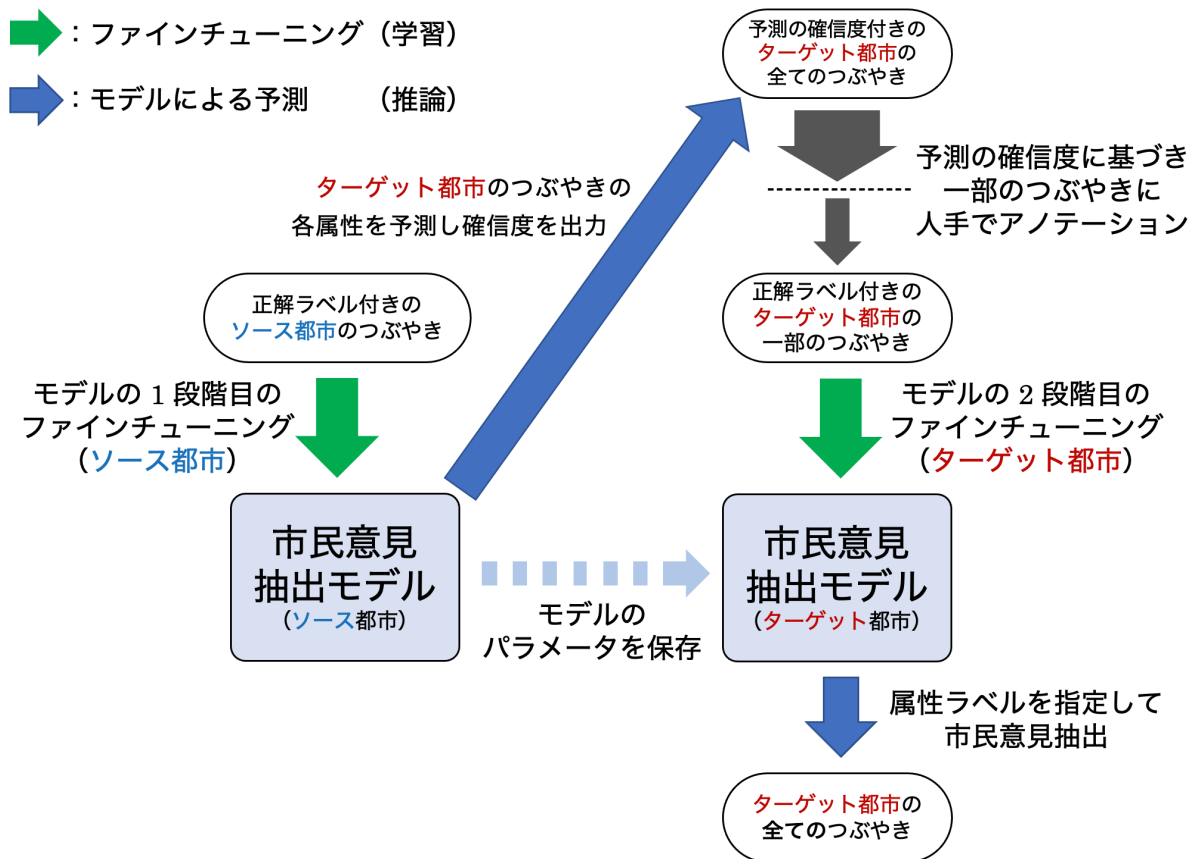


図 3.2: 都市を横断した市民意見抽出

ファインチューニングを行った各属性のラベル分類モデルを用いて、ターゲット都市のつぶやきの各属性のラベルの予測を行う。この際、ラベル分類モデルの予測の確信度についても同時に出力する。これによって、全てのターゲット都市のつぶやきに対して予測の確信度が付与された状態を作り出すことができる。そして、これらのターゲット都市のつぶやきから、予測の確信度の情報に基づき、一部のつぶやきを抽出し、人手によるアノテーションを行うことで、正解ラベルが付与されているターゲット都市の比較的少量のつぶやきを得る。なお、2.3節で述べた通り、本研究では、予測の確信度が高いターゲット都市のつぶやきを用いる手法と、予測の確信度が低いターゲット都市のつぶやきを用いる手法のうち、どちらの手法が有効であるかを、5章の各属性のラベル分類実験において検証する。

このようにして得られた正解ラベル付きのターゲット都市の比較的少量のつぶやきを用いて、ソース都市のつぶやきによって1段階目のファインチューニングを行った各属性のラベル分類モデルの2段階目のファインチューニングを行う。そして、3.2節で述べ

---

た複数の属性を組み合わせた市民意見抽出手法を用いて、ターゲット都市のつぶやきから、市民意見を整理して抽出する。

---

## 第4章 市民意見分析コーパスの構築

本章では、5章、6章の実験で使用する市民意見分析コーパスの構築について述べる。本研究で用いる市民意見分析コーパスは、政令指定都市である横浜市の市民のつぶやき [12, 41, 43, 42] と、同じく政令指定都市である札幌市の市民のつぶやき [42] を用いて構築する。また、コーパスには、新型コロナウイルス感染症の拡大によって話題となった、保育園サービスと飲食店のテイクアウトサービスに関するつぶやきを用いる。4.1節では、市民意見分析コーパスに用いるつぶやきの収集方法について述べる。続いて4.2節では、4.1節で収集したつぶやきに対して各属性の人手によるアノテーションを行った際の方法、結果について述べる。

### 4.1 つぶやきの収集方法

#### 4.1.1 市民アカウントの収集

はじめに、Twitterのプロフィール情報に基づいて、各都市の市民アカウントを収集する [47]。市民アカウントの収集では、まず、ツイプロ<sup>1</sup>の検索APIを用いて、Twitterのプロフィールの所在地欄、もしくはプロフィール文に、「鶴見区」や「神奈川区」といった対象の都市における行政区、もしくは、(鶴見区内の)「朝日町」、(神奈川区内の)「青木町」といった行政区内の町名を含むアカウントを収集する。続いて、収集したアカウントに対して、以下の方法で、各都市の市民アカウントであるかの判定を行う。

#### 市民アカウントの判定

市民アカウントの判定においては、はじめに、有名人やbotアカウントからのつぶやきの収集を避けるため、フォロワー数が3,000以上、もしくはフォロワー数が4,000以上のアカウントを除去する。また、前述のツイプロの検索APIを用いた対象の都市の地名による検索のみでは、横浜市中区の市民アカウントの収集を行う際に、「中区」のクエリでアカウントを収集してしまう。しかし、中区は横浜市以外にも、名古屋市や浜松市に

---

<sup>1</sup><https://twpro.jp/>

も存在するため、これでは他都市の市民アカウントを誤って収集してしまう可能性がある。そこで、収集したアカウントの所在地欄を参照し、事前に用意した、各都市の市民アカウントがプロフィール内に記述すると考えられるキーワードとのマッチングを行い、マッチしたアカウントを最終的に市民アカウントと判定する。なお、横浜市と札幌市において、事前に用意したキーワードは以下の通りである。

- 横浜市：よこはま, ヨコハマ, 横浜, yokohama, 横濱, はまっこ, 赤レンガ, 赤煉瓦
- 札幌市：さっぽろ, サッポロ, 札幌, sapporo

この際、所在地欄に「→」や「/」を含む場合、「つくば市→横浜市→新宿区」といったように、これまでに在住してきた都市の遷移を記述している可能性があるため、このような場合は、最後の「→」、もしくは「/」以降のテキストのみを対象としてキーワードのマッチングを行う。また、所在地欄が空欄のアカウント、もしくは、所在地欄とのキーワードマッチングでは市民アカウントと判定されなかったものの、MeCab<sup>2</sup>による形態素解析の結果、所在地欄に地域を表す単語が含まれると判定されたアカウントについては、前述の所在地欄に対するキーワードとのマッチングをプロフィール文に対して行い、マッチしたアカウントを市民アカウントと判定する。

#### 市民アカウントの拡張

また、ツイプロの検索 API を用いて収集した対象都市の市民アカウントをフォローしているアカウントについても、同様に対象都市の市民アカウントである可能性が高い。そこで、アカウントのフォロー関係を用いることで、市民アカウントの拡張を行う。

市民アカウントの拡張では、はじめに、ツイプロの検索 API を用いて収集したアカウントのうち、前述の判定方法によって市民アカウントであると判定されたアカウントのフォロワーのアカウントを全て取得する。そして、取得した市民アカウントのフォロワーに対しても、上記の市民アカウントの判定を行い、条件を満たすアカウントを市民アカウントとすることで、対象都市の市民アカウントの拡張を行う。

### 4.1.2 市民アカウントによるつぶやきの収集

4.1.1 節の方法によって、横浜市民アカウント 82,583 件と札幌市民アカウント 64,790 件を収集し、これらのアカウントのつぶやきを、Twitter の Streaming API を用いて収集

<sup>2</sup><https://taku910.github.io/mecab/>

した。続いて、収集したつぶやきのうち、2020年1月1日から2020年7月11日までの横浜市民アカウントの計28,971,414件のつぶやきと、札幌市民アカウントの計19,286,694件のつぶやきから、「保育園」、「保育士」、「保活」、「待機児童」を含むつぶやきを保育園サービスのつぶやきとして収集し、「持ち帰り」、「テイクアウト」を含むつぶやきを飲食店のテイクアウトサービスのつぶやきとして収集した。この際、リツイートと重複するつぶやきは取り除いた。

市民意見分析コーパスの構築に用いるデータ数は、横浜市、札幌市の両サービスに共通で2,622件のつぶやきと、それらのつぶやきを文単位に区切り直したものとした。つぶやきの文単位への区切り直しには、PythonライブラリであるspaCy<sup>3</sup>を用いた。なお、名詞のみによって構成される文は意見性を含まない場合が多いため、spaCyによる文単位の区切り直しの後に名詞のみの文が抽出された場合、つぶやきの先頭以外の文は1つ前の文に統合し、つぶやきの先頭の文については1つ後の文と統合した。さらに、改行は文の区切りとして処理した。また、ハッシュタグのみで構成されている文についても意見性を含まない場合が多いため、前の文と統合する処理を行った。最後に、閉じ括弧で始まる文については、前の文と統合する処理を行った。

## 4.2 各属性の人手によるアノテーション

4.1節で収集した横浜市民アカウントと札幌市民アカウントのつぶやきに対して、3.1節で定義した各属性のラベルを付与するアノテーション作業を人手で行うことで、市民意見分析コーパスを構築した<sup>4</sup>。横浜市と札幌市の両都市について、アノテーション作業は著者を含む計5名の判定者で行い、全ての属性ラベルのアノテーション結果を多数決により決定した。なお、ラベルの選択肢が3つ以上存在する属性のアノテーションにおいて、3名の選択したラベルが全て異なる場合や、5名中2名ずつが同じラベルを選択し、残りの1名が異なるラベルを選択した場合は、多数決で1つのラベルを決定することができない。このような場合は、判定者間で意見を交換し、最終的に全ての結果を多数決で決定した。

---

<sup>3</sup><https://spacy.io/>

<sup>4</sup>筑波大学図書館情報メディア系倫理審査承認（通知番号 第20-28号）

表 4.1: 各属性の判定者間一致度 ( Fleiss'  $\kappa$  )

サービス	属性	横浜市		札幌市	
		チーム A	チーム B	チーム C	チーム D
保育園	アプレイザル意見タイプ	0.643	0.633	0.687	0.720
	極性	0.673	0.693	0.769	0.753
	コミュニケーション意見タイプ	0.768	0.788	0.827	0.776
	地域依存性	0.856	0.888	0.747	0.829
	サービスとの適合性	0.721	0.707	0.783	0.750
	投稿主の立場	0.841	0.830	0.791	0.854
	休園・登園自粛との関連性	0.847	0.846	0.850	0.885
	定員との関連性	0.879	0.865	0.909	0.902
テイクアウト	アプレイザル意見タイプ	0.704	0.771	0.772	0.767
	極性	0.778	0.766	0.828	0.805
	コミュニケーション意見タイプ	0.818	0.819	0.874	0.808
	地域依存性	0.905	0.887	0.916	0.798
	サービスとの適合性	0.913	0.835	0.891	0.909
	投稿主の立場	0.885	0.785	0.914	0.856
	商品の評価との関連性	0.828	0.803	0.900	0.834

#### 4.2.1 アノテーション方針の一致のための訓練

アノテーション作業においては、まずはじめに、全判定者のアノテーション方針を一致させるための訓練を行った。訓練では、保育園サービスと飲食店のテイクアウトサービスの各 250 文と、それらの文を含むつぶやきを対象に、5 名全員で各属性のラベルの判定を行った。Fleiss の  $\kappa$  係数 [9] を用いたアノテーションの判定者間一致度が全属性で 0.6 (Substantial Agreement[15]) 以上となり、全判定者のアノテーション方針が一致した時点で訓練を終了した。なお、アノテーションを行う際に、1 文に複数の意見が含まれると判断した場合は、1 つの意見のみを含む意見ユニットとなるまで文を分割した。意見ユニットへの文の分割作業については、全判定者で意見を交換し、過半数の判定者間で意見が一致した場合のみ行った。

#### 4.2.2 残りの全てのつぶやきに対するアノテーション

全判定者のアノテーション方針が一致したことが確認できた後に、残りの全てのつぶやきに対するアノテーション作業を行った。著者は全てのつぶやき、文に対するアノテ

ションを行い、両都市において、残りの4名の判定者を2名ずつに分けることで各3名ずつのチームとし、各チームで全体の半数のつぶやき、文のアノテーションを実施した。なお、4.2.1節の訓練時と同様に、過半数の判定者が、1文に複数の意見が含まれると判断した文においては、1つの意見のみが含まれる意見ユニットとなるまで文を分割した。構築した市民意見分析コーパスの、Fleissの $\kappa$ 係数[9]を用いた各チームのアノテーションの判定者間一致度を表4.1に示す。表4.1より、横浜市、札幌市の全属性において、Fleissの $\kappa$ 係数が0.6 (Substantial Agreement[15]) 以上となっており、判定者によって属性のラベル判定に大きな差異は生まれなかったことが示された。

アノテーション作業によって得られた市民意見分析コーパスの、各都市におけるつぶやきと意見ユニットのデータ数を表4.2に示す。

表 4.2: 複数都市を対象とした市民意見分析コーパスのデータ数

都市	サービス	つぶやき	意見ユニット
横浜市	保育園	2,622	7,916
	テイクアウト	2,622	6,671
札幌市	保育園	2,622	8,113
	テイクアウト	2,622	7,620

# 第5章 実験：都市を横断した各属性のラベル分類

本章では、4章で構築した横浜市民アカウントと札幌市民アカウントのつぶやきからなる市民意見分析コーパスを用いて、3章で提案した手法による都市を横断した各属性のラベル分類を行うことで、手法の有効性を検証する。

また、都市を横断した実験に先立ち、3.2節で述べたマルチタスク学習手法の有効性について検証し、さらに、BERTモデルとT5モデルの性能の比較を行う。

## 5.1 実験の方法

### 5.1.1 予備実験：単一都市における各属性のラベル分類

著者の先行研究 [12] では、BERTモデルを用いて横浜市民アカウントのつぶやきを対象とした各属性のラベル分類を行う場合、マルチタスク学習手法を適用し、3.1.1で定義した3つの意見属性を同時に学習することでラベル分類精度が向上することを示した。また、著者の先行研究 [42] では、T5モデルを用いて意見属性の分類を行う際にも、マルチタスク学習手法が有効であることを示した。そこで、都市を横断した各属性のラベル分類実験に先立ち、5.2節の予備実験において、単一都市におけるマルチタスク学習手法の有効性の検証を行い、さらに、BERTモデルとT5モデルの性能の比較を行う。意見属性のラベル分類においては、共通のT5モデルを用いて各属性を同時にマルチタスク学習する手法を提案手法とし、各属性を独立のT5モデルを用いてラベル分類する手法と、BERTモデルを用いたマルチタスク学習手法、BERTモデルを用いて各属性を独立にラベル分類する手法を比較手法とする。つぶやきの属性のラベル分類では、各属性に関連性は無いため、各属性を独立のT5モデルでラベル分類する手法を提案手法とし、T5モデルを用いたマルチタスク学習手法と、BERTモデルを用いた各属性独立のラベル分類手法、BERTモデルを用いたマルチタスク学習手法を比較手法とする。



なお、本研究で用いる BERT モデルと T5 モデルは、Python の Hugging Face transformers ライブラリで公開されている事前学習済みモデル<sup>1</sup>とし、各属性のラベル分類では、これらの事前学習済みモデルのファインチューニングを行う。

### 5.1.2 都市を横断した各属性のラベル分類

5.3 節の都市を横断した各属性のラベル分類実験では、はじめにソース都市の市民意見分析コーパスの全データを用いて、各属性のラベル分類モデルの1段階目のファインチューニングを行い、モデルのパラメータを保存する。続いて、ターゲット都市のデータを25%、50%と一部用いて、パラメータを保存済みのラベル分類モデルの2段階目のファインチューニングを行う。なお、ターゲット都市から一部使用するデータの選定においては、1段階目のファインチューニング後のモデルの予測の確信度に基づき、以下の手法について比較を行う。

- ターゲット都市の25%のデータを用いる場合
  - (1) モデルの予測の確信度が上位25%のデータを選定する手法;
  - (2) モデルの予測の確信度が下位25%のデータを選定する手法;
  - (3) ランダムに25%のデータを選択する手法。
- ターゲット都市の50%のデータを用いる場合
  - (1) モデルの予測の確信度が上位50%のデータを選定する手法;
  - (2) モデルの予測の確信度が下位50%のデータを選定する手法;
  - (3) モデルの予測の確信度が中間の50%のデータを選定する手法;
  - (4) モデルの予測の確信度上位25%と下位25%のデータを選定する手法;
  - (5) ランダムに50%のデータを選択する手法。

モデルの予測の確信度に基づくターゲット都市におけるデータの選択では、はじめにコーパスに含まれるソース都市のデータを全て使って各属性のラベル分類モデルをファインチューニングし、コーパスに含まれるターゲット都市の全データの各属性のラベルを推定する。そして、ターゲット都市のデータから、モデルの予測の確信度<sup>2</sup>の情報を用いることで、2段階目のファインチューニングに用いるためのデータを選定する。

5.3.3 節では、都市を横断した各属性のラベル分類実験に関する考察を行う。5.3.3 節の考察では、1段階目のファインチューニング後のモデルの予測の確信度が上位のつづ

---

<sup>1</sup>BERT モデル : <https://huggingface.co/cl-tohoku/bert-base-japanese-v2>

T5 モデル : <https://huggingface.co/sonoisa/t5-base-japanese>

<sup>2</sup>T5 モデルの予測の際に使用しているビームサーチのスコア (Hugging Face transformers ライブラリ中では、パラメタ sequences\_scores と定義) のうち、予測ラベルに対応するスコアを確信度とする

やきを用いる手法と、確信度が下位のつぶやきを用いる手法の双方において、ターゲット都市において作成された教師データの各属性のラベルの分布を算出する。算出した各属性のラベルの分布をもとに、手法ごとに得られる教師データの特徴を分析する。

### 5.1.3 各属性の都市を横断したラベル分類精度と単一都市におけるラベル分類精度の比較

本章の最後に、5.4節において、都市を横断した各属性のラベル分類精度と、都市を横断せずに単一都市のデータのみを用いた際の各属性のラベル分類精度を比較する。単一都市のデータのみを用いた各属性のラベル分類では、ターゲット都市の100%のデータを用いているのに対し、都市を横断して各属性のラベル分類を行う場合は、ソース都市のデータは100%用いているものの、ターゲット都市のデータは一部しか用いていない。そのため、これらの手法の各属性のラベル分類精度を比較することで、既存の都市とは異なるターゲット都市で新たに市民意見の分析を行う際に、データ作成コストを抑えた上で、各属性のラベル分類の精度差をどの程度まで抑えられるかを検証する。都市を横断した各属性のラベル分類には、5.3節の都市を横断したラベル分類実験において最も精度が高かった手法を用い、単一都市のデータを用いて各属性のラベル分類を行う手法には、5.2節の単一都市における各属性のラベル分類実験で最も精度が高かった手法を用いる。

### 5.1.4 各属性の分類実験に用いる評価指標

本実験の評価指標は全てF値とし、5分割交差検証を用いることで、コーパス内の全てのデータに対する各属性のラベル分類精度を算出する。5分割交差検証における各検証では、全体の64%のデータを訓練データ、16%のデータを検証データ、20%のデータをテストデータとして用いる。なお、テストデータについては、全ての実験で同じデータを用いる。ターゲット都市のデータを25%、50%と一部用いる手法においては、モデルの2段階目のファインチューニングに使用する5分割交差検証における訓練データ、検証データの割合のみ変化させる。ファインチューニングのエポック数は、5分割交差検証の各検証において、2,3,4,5エポックのうち、検証データのF値が最も高いものを用いる。

## 5.2 予備実験：単一都市における各属性のラベル分類の結果

### 5.2.1 意見属性のラベル分類の結果

単一都市のデータを用いた意見属性のラベル分類精度を表 5.1 に示す。表 5.1 より、意見属性のラベル分類では、全属性において、提案手法である T5 モデルを用いたマルチタスク学習手法が最も高い精度となっている。さらに、提案手法と比較した際に、全ての手法において、1 対の対応のある両側  $t$  検定で有意差が有り（有意水準 5%）、その効果量（Pearson's  $r$ ）が大きい（large [5]）ことを確認した<sup>3</sup>。また、BERT モデルを用いた場合に注目しても、全ての属性において、各属性独立学習手法と比較してマルチタスク学習手法が高い精度となっている。これらの実験結果から、意見属性のラベル分類には、各属性間の関連性に注目し、同一のモデルを用いたマルチタスク学習手法が有効であると分かる。また、T5 モデルと BERT モデルでは、全属性において T5 モデルのラベル分類精度が高いことから、意見属性のラベル分類における T5 モデルの有効性が示された。これらの結果を踏まえ、5.3 節の都市を横断した実験における意見属性のラベル分類には、T5 モデルを用いたマルチタスク学習手法を使用する。

### 5.2.2 つぶやきの属性のラベル分類の結果

単一都市のデータを用いたつぶやきの属性のラベル分類精度を表 5.2 に示す。表 5.2 より、つぶやきの属性では、横浜市、札幌市の全 18 属性中、13 属性で T5 を用いた各属性独立学習手法が最高精度となっている。また、横浜市の保育園サービスの地域依存性、定員との関連性、札幌市の保育園サービスの地域依存性の 3 属性でのみ、T5 を用いたマルチタスク学習手法が各属性独立の T5 を上回る精度となっているが、F 値の差は 0.02 未満と小さい。この結果から、つぶやきの属性では、各属性独立の T5 モデルを用いる手法が有効であると分かる。また、T5 モデルを用いた各属性独立学習手法と BERT モデルを用いた各属性独立学習手法を比較すると、札幌市のテイクアウトサービスの地域依存性を除く全属性で T5 モデルが高精度となっており、札幌市のテイクアウトサービスの地域依存性においても F 値の差は 0.001 と極めて小さい。よって、つぶやきの属性においても、BERT モデルと比較した際に T5 モデルが有効であると分かる。これらの結果を踏まえ、5.3 節の都市を横断した実験におけるつぶやきの属性のラベル分類には、各属性独立の T5 モデルを使用する。

---

<sup>3</sup>表 5.1, 表 5.2 の  $r$  (効果量) は Pearson's  $r$  を示しており、 $0.5 \leq r$  で大きい (large),  $0.3 \leq r < 0.5$  で中程度 (middle),  $0.1 \leq r < 0.3$  で小さい (small) とされている [5]。以降の表でも効果量は同様に Pearson's  $r$  を用いる。

---

表 5.1: T5 モデルと BERT モデルを用いた際のマルチタスク学習手法と各属性独立学習手法の意見属性のラベル分類精度の比較 (F 値)

都市	サービス	属性	T5		BERT	
			マルチタスク (提案手法)	各属性独立	マルチタスク	各属性独立
横浜市	保育園	アプレイザル意見タイプ	<b>0.701</b>	0.636	0.633	0.563
		極性	<b>0.849</b>	0.792	0.824	0.725
		コミュニケーション意見タイプ	<b>0.764</b>	0.739	0.695	0.672
	テイクアウト	アプレイザル意見タイプ	<b>0.771</b>	0.698	0.720	0.621
		極性	<b>0.869</b>	0.807	0.838	0.744
		コミュニケーション意見タイプ	<b>0.764</b>	0.760	0.712	0.697
札幌市	保育園	アプレイザル意見タイプ	<b>0.705</b>	0.651	0.665	0.590
		極性	<b>0.864</b>	0.774	0.844	0.721
		コミュニケーション意見タイプ	<b>0.764</b>	0.754	0.700	0.680
	テイクアウト	アプレイザル意見タイプ	<b>0.770</b>	0.709	0.714	0.646
		極性	<b>0.877</b>	0.805	0.855	0.759
		コミュニケーション意見タイプ	<b>0.793</b>	0.786	0.699	0.689
p 値			-	*	*	*
r (効果量)			-	0.866	0.916	0.980

\*提案手法に対して 1 対の対応のある両側 t 検定で有意差有り (有意水準 5%)

表 5.2: T5 モデルと BERT モデルを用いた際の各属性独立学習手法とマルチタスク学習手法のつづやきの属性のラベル分類精度の比較 (F 値)

都市	サービス	属性	T5		BERT	
			各属性独立 (提案手法)	マルチタスク	各属性独立	マルチタスク
横浜市	保育園	地域依存性	0.925	<b>0.927</b>	0.915	0.920
		サービスとの適合性	<b>0.909</b>	<b>0.909</b>	0.903	0.905
		投稿主の立場	<b>0.709</b>	0.675	0.681	0.648
		休園・登園自粛との関連性	<b>0.854</b>	0.827	0.851	0.847
		定員との関連性	0.850	<b>0.869</b>	0.841	0.839
	テイクアウト	地域依存性	<b>0.926</b>	0.916	0.915	0.915
		サービスとの適合性	<b>0.987</b>	<b>0.987</b>	0.978	0.976
		投稿主の立場	<b>0.885</b>	0.878	0.850	0.847
		商品の評価との関連性	<b>0.855</b>	<b>0.855</b>	0.823	0.804
札幌市	保育園	地域依存性	0.884	<b>0.903</b>	0.876	0.894
		サービスとの適合性	<b>0.862</b>	0.847	0.844	0.844
		投稿主の立場	<b>0.648</b>	0.608	0.614	0.587
		休園・登園自粛との関連性	<b>0.795</b>	0.780	0.781	0.788
		定員との関連性	0.728	0.634	0.704	<b>0.751</b>
	テイクアウト	地域依存性	0.906	0.899	<b>0.907</b>	0.898
		サービスと関連性	<b>0.985</b>	0.984	0.983	0.984
		投稿主の立場	<b>0.888</b>	0.885	0.844	0.858
		商品の評価との関連性	<b>0.853</b>	0.851	0.795	0.809
p 値			-		*	*
r (効果量)			-	0.433	0.770	0.627

\*提案手法に対して 1 対の対応のある両側 t 検定で有意差有り (有意水準 5%)

表 5.3: ターゲット都市の 25%のデータを用いた都市を横断した意見属性のラベル分類精度 (F 値)

ターゲット都市	サービス	属性	予測の確信度に基づくターゲット都市のつぶやきの選択手法		
			上位 25%	下位 25%	ランダム 25%
横浜市	保育園	アプレイザル意見タイプ	<b>0.648</b>	0.609	0.635
		極性	<b>0.816</b>	0.772	0.793
		コミュニケーション意見タイプ	<b>0.759</b>	0.719	0.734
	テイクアウト	アプレイザル意見タイプ	<b>0.733</b>	0.626	0.722
		極性	<b>0.856</b>	0.762	0.816
		コミュニケーション意見タイプ	<b>0.795</b>	0.628	0.757
札幌市	保育園	アプレイザル意見タイプ	<b>0.665</b>	0.609	0.661
		極性	<b>0.812</b>	0.762	0.795
		コミュニケーション意見タイプ	<b>0.772</b>	0.706	0.736
	テイクアウト	アプレイザル意見タイプ	<b>0.739</b>	0.648	0.706
		極性	<b>0.852</b>	0.750	0.810
		コミュニケーション意見タイプ	<b>0.793</b>	0.554	0.748
p 値			-	*	*
r (効果量)			-	0.848	0.901

\*上位 25%手法に対して 1 対の対応のある両側 t 検定で有意差有り (有意水準 5%)

表 5.4: ターゲット都市の 50%のデータを用いた都市を横断した意見属性のラベル分類精度 (F 値)

ターゲット都市	サービス	属性	予測の確信度に基づくターゲット都市のつぶやきの選択手法				
			上位 50%	下位 50%	中間 50%	上位 25% +下位 25%	ランダム 50%
横浜市	保育園	アプレイザル意見タイプ	<b>0.688</b>	0.643	0.669	0.659	0.665
		極性	<b>0.843</b>	0.794	0.814	0.819	0.819
		コミュニケーション意見タイプ	<b>0.780</b>	0.700	0.749	0.773	0.759
	テイクアウト	アプレイザル意見タイプ	<b>0.766</b>	0.642	0.713	0.750	0.731
		極性	<b>0.876</b>	0.782	0.811	0.858	0.841
		コミュニケーション意見タイプ	0.786	0.660	0.719	<b>0.800</b>	0.783
札幌市	保育園	アプレイザル意見タイプ	<b>0.691</b>	0.649	0.683	0.678	0.687
		極性	<b>0.846</b>	0.783	0.809	0.820	0.826
		コミュニケーション意見タイプ	<b>0.789</b>	0.698	0.741	0.771	0.755
	テイクアウト	アプレイザル意見タイプ	<b>0.768</b>	0.667	0.719	0.739	0.736
		極性	<b>0.884</b>	0.778	0.832	0.863	0.833
		コミュニケーション意見タイプ	<b>0.803</b>	0.636	0.676	0.801	0.769
p 値			-	*	*	*	*
r (効果量)			-	0.929	0.859	0.796	0.896

\*上位 50%手法に対して 1 対の対応のある両側 t 検定で有意差有り (有意水準 5%)

## 5.3 都市を横断した各属性のラベル分類の実験結果

### 5.3.1 都市を横断した意見属性のラベル分類

都市を横断した意見属性のラベル分類の精度を表 5.3, 表 5.4 に示す。

- 表 5.3 より, ターゲット都市の 25%のデータを用いて都市を横断した意見属性のラベル分類を行う場合, 全属性において, 1段階目のファインチューニングを行ったモデルによる予測の確信度が高いターゲット都市のデータを選定し, 2段階目のファインチューニングを行う手法が最も高い精度となっている。

- 表 5.4 より，ターゲット都市の 50% のデータを用いる場合，横浜市のテイクアウトサービスのコミュニケーション意見タイプを除き，1 段階目のファインチューニングを行ったモデルによる予測の確信度が高いターゲット都市のデータを選定する手法が最高精度となっており，さらに，ターゲット都市の 25% のデータを用いる場合と比較して高精度となっている。

以上の結果より，都市を横断して意見属性のラベル分類を行う場合，1 段階目のファインチューニングを行ったモデルによる予測の確信度が高いターゲット都市のデータを選定し，2 段階目のファインチューニングを行う手法が有効であると分かる。

### 5.3.2 都市を横断したつぶやきの属性のラベル分類

都市を横断したつぶやきの属性のラベル分類精度を表 5.5，表 5.6 に示す。

- 表 5.5 より，ターゲット都市の 25% のデータを用いて都市を横断したつぶやきの属性のラベル分類を行う場合，札幌市の保育園サービスの地域依存性を除く全属性で，1 段階目のファインチューニングを行ったモデルによる予測の確信度が高いターゲット都市のデータを選定し，2 段階目のファインチューニングを行う手法が最も高い精度となっている。また，札幌市の保育園サービスの地域依存性においても，確信度上位のデータを選定する手法とランダムにデータを選定する手法の精度差は，0.005 と極めて小さい。
- 表 5.6 より，ターゲット都市の 50% のデータを用いる場合，全 18 属性中 12 属性で，1 段階目のファインチューニングを行ったモデルによる予測の確信度が上位のデータを選定して 2 段階目のファインチューニングを行う手法が最高精度となっている。この結果から，つぶやきの属性においても，確信度上位のデータを用いる手法が有効であると分かる。しかし，表 5.6 より，つぶやきの属性では，確信度上位 25% と下位 25% のデータを合わせた手法が 4 属性で最も高い精度となっている<sup>4</sup>。表 5.5 では，確信度が下位 25% のデータを選定する手法の精度が低いことから，確信度が下位 25% のデータがモデルの分類に有効とはいえない。さらに，つぶやきの属性では，表 5.5 の確信度上位 25% のデータを選定する手法と，表 5.6 の確信度上位 50% のデータを選定する手法の精度差が意見属性と比較して小さい。そのため，つぶやきの属性では，確信度上位 25% のデータに，モデルの学習に有効なデータが多く含まれていると考えられる。

<sup>4</sup>確信度上位 50% 手法と，確信度上位 25%+下位 25% 手法では，1 対の対応のある両側  $t$  検定で有意差が確認できた（有意水準 5%）ものの，効果量  $r$  は中程度（middle[5]）である。

表 5.5: ターゲット都市の 25% のデータを用いた都市を横断したつづやきの属性のラベル分類精度 (F 値)

ターゲット都市	サービス	属性	予測の確信度に基づくターゲット都市のつづやきの選択手法		
			上位 25%	下位 25%	ランダム 25%*
横浜市	保育園	地域依存性	<b>0.926</b>	0.913	0.924
		サービスとの適合性	<b>0.914</b>	0.909	0.913
		投稿主の立場	<b>0.697</b>	0.578	0.658
		休園・登園自粛との関連性	<b>0.845</b>	0.774	0.837
		定員との関連性	<b>0.887</b>	0.043	0.862
	テイクアウト	地域依存性	<b>0.919</b>	0.901	0.904
		サービスとの適合性	<b>0.988</b>	0.984	0.983
		投稿主の立場	<b>0.876</b>	0.853	0.874
		商品の評価との関連性	<b>0.858</b>	0.806	0.850
札幌市	保育園	地域依存性	0.886	0.864	<b>0.891</b>
		サービスとの適合性	<b>0.854</b>	0.838	0.850
		投稿主の立場	<b>0.672</b>	0.630	0.623
		休園・登園自粛との関連性	<b>0.815</b>	0.782	0.790
		定員との関連性	<b>0.813</b>	0.657	0.750
	テイクアウト	地域依存性	<b>0.907</b>	0.874	0.893
		サービスとの適合性	<b>0.986</b>	0.976	0.985
		投稿主の立場	<b>0.899</b>	0.872	0.892
		商品の評価との関連性	<b>0.844</b>	0.800	0.836
p 値			-		*
r (効果量)			-	0.412	0.643

\*上位 25%手法に対して 1 対の対応のある両側 t 検定で有意差有り (有意水準 5%)

表 5.6: ターゲット都市の 50% のデータを用いた都市を横断したつづやきの属性のラベル分類精度 (F 値)

ターゲット都市	サービス	属性	予測の確信度に基づくターゲット都市のつづやきの選択手法				
			上位 50%	下位 50%*	中間 50%*	上位 25% + 下位 25%*	ランダム 50%*
横浜市	保育園	地域依存性	<b>0.930</b>	0.922	0.924	0.926	0.927
		サービスとの適合性	0.914	0.913	<b>0.915</b>	0.914	<b>0.915</b>
		投稿主の立場	<b>0.715</b>	0.589	0.663	0.701	0.702
		休園・登園自粛との関連性	0.844	0.778	0.821	0.836	<b>0.855</b>
		定員との関連性	0.883	0.655	0.811	<b>0.884</b>	0.845
	テイクアウト	地域依存性	<b>0.924</b>	0.898	0.910	0.917	0.917
		サービスとの適合性	<b>0.989</b>	0.985	0.983	0.987	0.984
		投稿主の立場	<b>0.886</b>	0.870	0.877	0.878	0.877
		商品の評価との関連性	0.853	0.846	0.851	<b>0.856</b>	0.850
札幌市	保育園	地域依存性	<b>0.893</b>	0.882	0.890	0.887	0.888
		サービスとの適合性	<b>0.862</b>	0.847	0.855	0.855	0.861
		投稿主の立場	<b>0.674</b>	0.577	0.638	0.667	0.660
		休園・登園自粛との関連性	0.819	0.808	0.811	<b>0.829</b>	0.811
		定員との関連性	0.778	0.709	0.717	<b>0.789</b>	0.777
	テイクアウト	地域依存性	<b>0.914</b>	0.882	0.904	0.910	0.901
		サービスとの適合性	<b>0.986</b>	0.969	0.980	0.985	0.985
		投稿主の立場	<b>0.904</b>	0.881	0.892	0.898	0.898
		商品の評価との関連性	<b>0.860</b>	0.831	0.853	0.851	0.847
p 値			-	*	*	*	*
r (効果量)			-	0.615	0.658	0.458	0.595

\*上位 50%手法に対して 1 対の対応のある両側 t 検定で有意差有り (有意水準 5%)

### 5.3.3 結果の考察

5.3.2節の都市を横断した各属性のラベル分類実験において、1段階目のファインチューニングを行ったモデルによる予測の確信度が、ターゲット都市のデータの選定に与える影響を分析するため、手法ごとに選定されたデータの各属性のラベルの分布を算出した。結果を表5.7に示す。表5.7の「全体」の列は、市民意見分析コーパスの全データの各属性のラベルの分布を表しており、「確信度上位」と「確信度下位」の列は、それぞれソース都市のデータを用いて1段階目のファインチューニングを行ったモデルによる確信度が上位50%、下位50%のターゲット都市のデータを選定した際の各属性のラベルの分布を表している。表5.7から、確信度が上位のデータを選定することで、各属性のラベルの分布を均衡に近づけられることが分かる。たとえば、表5.7の横浜市のアプレイザル意見タイプに着目すると、市民意見分析コーパス全体においては、「該当無し」ラベルの分布が保育園サービスで56.4%、テイクアウトサービスで73.0%と過半数を占めている。一方で、確信度上位50%に着目すると、「該当無し」ラベルの割合が減り、他のラベルが占める割合が高くなっている。他の多くの属性においても、確信度上位50%のデータを選定することで、同様にラベルの分布が均衡に近づいていることが分かる。モデルの予測の確信度が高いデータは、モデルにとって予測を行いやすいデータであり、特徴的なデータであると考えられることができる。そのため、ターゲット都市のデータのうち、ラベルの種類に関わらず、モデルの予測の確信度が上位のデータ、つまり、特徴的なデータのみを選定することで、作成された教師データのラベルの分布を均衡に近づけることができ、分類精度が向上したと考える。



表 5.7: 手法毎に得られる教師データ内の各属性のラベルの割合の分布

都市	付与する対象	サービス	属性	ラベル	各ラベルの割合						
					保育園			テイクアウト			
					全体	確信度上位	確信度下位	全体	確信度上位	確信度下位	
横浜市	意見ユニット	-	アプレイザル意見タイプ	自発的な感情の表明	18.2%	25.1%	11.2%	7.0%	13.4%	0.5%	
				人間・組織の振舞や行為を対象とした批評	13.7%	15.3%	12.2%	5.4%	9.6%	1.3%	
				事物・事象を対象とした評価	11.8%	11.8%	11.7%	14.6%	27.3%	1.9%	
			極性	該当無し	56.4%	47.8%	65.0%	73.0%	49.7%	96.3%	
				肯定	14.9%	8.6%	21.2%	19.5%	14.7%	24.3%	
				否定	29.4%	29.7%	29.1%	8.5%	14.7%	2.3%	
				中立	17.0%	23.9%	10.2%	17.0%	31.5%	2.5%	
				該当無し	38.6%	37.7%	39.5%	54.9%	39.1%	70.8%	
				推測	5.3%	8.3%	2.4%	1.8%	3.5%	0.0%	
			コミュニケーション意見タイプ	提案	1.4%	2.7%	0.1%	1.3%	2.5%	0.0%	
				疑問	9.4%	15.0%	3.8%	3.9%	7.8%	0.0%	
				要求	8.2%	10.5%	6.0%	12.9%	25.6%	0.2%	
				該当無し	75.7%	63.5%	87.8%	80.2%	60.6%	99.7%	
				依存	76.9%	68.1%	85.8%	52.8%	41.3%	64.3%	
				非依存	23.1%	31.9%	14.2%	47.2%	58.7%	35.7%	
	つぶやき全体	-	地域依存性	依存	80.2%	66.6%	93.7%	94.7%	89.4%	99.9%	
				不適合	19.8%	33.4%	6.3%	5.3%	10.6%	0.1%	
				不適合	19.8%	33.4%	6.3%	5.3%	10.6%	0.1%	
			投稿主の立場	保育園サービス	小さい子を持つ親	51.9%	37.0%	66.9%	-	-	-
				保育園関係者	9.6%	17.1%	2.1%	-	-	-	
				その他	38.5%	45.9%	31.1%	-	-	-	
		テイクアウトサービス	飲食店を利用した人	-	-	-	31.7%	27.3%	36.0%		
		飲食店の従業員	-	-	-	33.1%	19.4%	46.9%			
		その他	-	-	-	35.2%	53.4%	17.1%			
		保育園	休園・登園自粛との関連性	関連する	19.3%	18.2%	20.4%	-	-	-	
				関連しない	80.7%	81.8%	79.6%	-	-	-	
				関連する	9.6%	19.2%	0.1%	-	-	-	
定員との関連性	関連する		90.4%	80.8%	99.9%	-	-	-			
	関連しない		9.4%	19.2%	0.1%	-	-	-			
	関連する		90.4%	80.8%	99.9%	-	-	-			
テイクアウト	商品の評価との関連性	関連する	-	-	-	19.0%	30.5%	7.5%			
		関連しない	-	-	-	81.0%	69.5%	92.5%			
札幌市	意見ユニット	-	アプレイザル意見タイプ	自発的な感情の表明	15.6%	18.2%	13.0%	5.4%	8.6%	2.2%	
				人間・組織の振舞や行為を対象とした批評	17.3%	25.2%	9.3%	7.2%	9.3%	5.1%	
				事物・事象を対象とした評価	23.5%	32.3%	14.8%	23.6%	33.0%	14.3%	
			極性	該当無し	43.6%	24.3%	62.9%	63.8%	49.1%	78.5%	
				肯定	20.3%	18.9%	21.7%	25.1%	21.7%	28.6%	
				否定	36.0%	42.9%	29.2%	11.0%	19.8%	2.3%	
				中立	12.8%	15.7%	10.0%	15.9%	29.9%	1.8%	
				該当無し	30.8%	22.6%	39.1%	47.9%	28.5%	67.3%	
				推測	5.8%	10.3%	1.3%	2.6%	5.1%	0.1%	
			コミュニケーション意見タイプ	提案	2.4%	4.5%	0.2%	2.5%	4.9%	0.2%	
				疑問	10.1%	14.3%	5.8%	3.2%	6.2%	0.2%	
				要求	8.0%	12.2%	3.9%	13.9%	19.6%	8.1%	
				該当無し	73.7%	58.7%	88.7%	77.8%	64.1%	91.4%	
				依存	66.2%	45.3%	87.3%	48.3%	53.0%	43.5%	
				非依存	33.8%	54.7%	12.7%	51.7%	47.0%	56.5%	
	つぶやき全体	-	地域依存性	依存	66.2%	44.5%	88.0%	93.3%	86.9%	99.7%	
				不適合	33.8%	55.5%	12.0%	6.7%	13.1%	0.3%	
				不適合	33.8%	55.5%	12.0%	6.7%	13.1%	0.3%	
			投稿主の立場	保育園サービス	小さい子を持つ親	54.5%	36.5%	72.5%	-	-	-
				保育園関係者	7.8%	13.6%	2.1%	-	-	-	
				その他	37.7%	49.9%	25.4%	-	-	-	
		テイクアウトサービス	飲食店を利用した人	-	-	-	34.8%	31.5%	38.2%		
		飲食店の従業員	-	-	-	33.7%	15.2%	52.2%			
		その他	-	-	-	31.5%	53.3%	9.6%			
		保育園	休園・登園自粛との関連性	関連する	20.1%	35.8%	4.3%	-	-	-	
				関連しない	79.9%	64.2%	95.7%	-	-	-	
				関連する	5.7%	10.7%	0.7%	-	-	-	
			定員との関連性	関連する	94.3%	89.3%	99.3%	-	-	-	
				関連しない	5.7%	10.7%	0.7%	-	-	-	
				関連する	94.3%	89.3%	99.3%	-	-	-	
テイクアウト	商品の評価との関連性	関連する	-	-	-	24.0%	22.3%	25.7%			
		関連しない	-	-	-	76.0%	77.7%	74.3%			

表 5.8: 意見属性の都市を横断した分類精度と単一都市における分類精度の比較 (F 値)

ターゲット都市	サービス	属性	都市を横断して分類	単一都市で分類
横浜市	保育園	アプレイザル意見タイプ	0.688	<b>0.701</b>
		極性	0.843	<b>0.849</b>
		コミュニケーション意見タイプ	<b>0.780</b>	0.764
	テイクアウト	アプレイザル意見タイプ	0.766	<b>0.771</b>
		極性	<b>0.876</b>	0.869
		コミュニケーション意見タイプ	<b>0.786</b>	0.764
札幌市	保育園	アプレイザル意見タイプ	0.691	<b>0.705</b>
		極性	0.846	<b>0.864</b>
		コミュニケーション意見タイプ	<b>0.789</b>	0.764
	テイクアウト	アプレイザル意見タイプ	0.768	<b>0.770</b>
		極性	<b>0.884</b>	0.877
		コミュニケーション意見タイプ	<b>0.803</b>	0.793

## 5.4 各属性の都市を横断した分類精度と単一都市における分類精度の比較

本節では、都市を横断した各属性のラベル分類精度と、都市を横断せずに単一都市のデータを用いた際の各属性のラベル分類精度を比較する。

5.3 節の結果を踏まえ、都市を横断した意見属性のラベル分類は、ターゲット都市の確信度上位 50% のデータを選定して T5 を用いたマルチタスク学習を行う手法の精度を都市横断の精度とする。つぶやきの属性については、ターゲット都市の確信度上位 50% のデータを選定して各属性独立の T5 モデルを用いる手法の精度を都市横断のラベル分類精度とする。また、5.2 節の結果を踏まえ、単一都市におけるラベル分類では、意見属性は各属性を T5 を用いたマルチタスク学習でラベル分類した際の精度を単一都市の精度とし、つぶやきの属性は各属性独立の T5 モデルを用いる手法の精度を単一都市の精度とする。結果を表 5.8、表 5.9 に示す。

- 表 5.8 より、意見属性の分類においては、都市を横断したラベル分類精度が全 12 属性中 6 属性で単一都市におけるラベル分類精度よりも高精度となっており、ターゲット都市の 50% のデータを効果的に選定することで、ターゲット都市のデータのみを 100% 用いる手法と同等の精度が得られることが分かる。特にテイクアウトサービスにおいては、横浜市と札幌市の両都市で、都市を横断した極性とコミュニケーション意見タイプのラベル分類精度が単一都市における精度を上回っており、アプレイザル意見タイプにおいても、横浜市で 0.005、札幌市で 0.002 と精度差は

表 5.9: つぶやきの属性の都市を横断した分類精度と単一都市における分類精度の比較 (F 値)

ターゲット都市	サービス	属性	都市を横断して分類	単一都市で分類
横浜市	保育園	地域依存性	<b>0.930</b>	0.925
		サービスとの適合性	<b>0.914</b>	0.909
		投稿主の立場	<b>0.715</b>	0.709
		休園・登園自粛との関連性	0.844	<b>0.854</b>
		定員との関連性	<b>0.883</b>	0.850
	テイクアウト	地域依存性	0.924	<b>0.926</b>
		サービスとの適合性	<b>0.989</b>	0.987
		投稿主の立場	<b>0.886</b>	0.885
		商品の評価との関連性	0.853	<b>0.855</b>
札幌市	保育園	地域依存性	<b>0.893</b>	0.884
		サービスとの適合性	<b>0.862</b>	<b>0.862</b>
		投稿主の立場	<b>0.674</b>	0.648
		休園・登園自粛との関連性	<b>0.819</b>	0.795
		定員との関連性	<b>0.778</b>	0.728
	テイクアウト	地域依存性	<b>0.914</b>	0.906
		サービスとの適合性	<b>0.986</b>	0.985
		投稿主の立場	<b>0.904</b>	0.888
		商品の評価との関連性	<b>0.860</b>	0.853
<i>p</i> 値			-	*
<i>r</i> (効果量)			-	0.569

\*都市を横断した分類手法に対して1対の対応のある両側 *t* 検定で有意差有り (有意水準 5%)

極めて小さい。この結果から、テイクアウトサービスは都市を横断したラベル分類が行いやすいと分かる。原因として、保育園サービスに関する市民意見は、各都市の行政によるサービスの影響が大きく、都市ごとに差が現れやすい一方で、テイクアウトサービスに関しては、都市によって店名等の差は存在するものの、つぶやきに現れる市民意見の表現については、都市ごとに大きな差が無いためであると考えられる。

- 表 5.9 より、つぶやきの属性の分類においては、都市を横断したラベル分類精度が全 18 属性中 15 属性で単一都市におけるラベル分類以上の精度となっており、ターゲット都市の 50% のデータを効果的に選定することで、ターゲット都市のデータのみを 100% 用いる手法以上の精度が実現可能であると分かる。つぶやきの属性においては、地域依存性のラベル分類には地名や店名が手がかりとなり、休園・登園自粛との関連性のラベル分類には「休み」、「休園」といったキーワードが手がかりとなるように、各属性のラベル分類のための手がかりが意見属性と比較して明確である。そのため、モデルの予測の確信度が上位のデータ、つまり、特徴的なデータを選定する手法が意見属性と比較してさらに効果的であると考えられる。

本節で検証を行った，都市を横断した各属性のラベル分類では，全ての属性において，ターゲット都市のデータは50%のみを用いている．しかし，表5.8のアプレイザル意見タイプ属性のように，都市を横断した分類精度が単一都市における分類精度と比較して低くなってしまっている属性がある．このような一部の属性の分類精度をさらに向上させ，より高精度な市民意見抽出を行うために，一部の属性のみ，ターゲット都市のデータを50%よりも多く使用する手法の有効性が期待できる．一部の属性のみ，重点的にターゲット都市のデータ量を増やすことで，ターゲット都市の全体のアノテーションコストは大幅に上げることなく，効率的に分類精度を向上させることができると考える．

---

## 第6章 実験：都市を横断したつぶやきからの市民意見抽出

本章では、3章の提案手法に基づき、実際につぶやきから都市を横断して市民意見を抽出 [12, 41, 43] し、横浜市民と札幌市民の市民意見を比較、分析する。

### 6.1 都市を横断した市民意見の抽出方法

本章の分析では、はじめに、6.2節で述べるように、4章で作成した市民意見分析コーパスに含まれない新たなつぶやきを収集することで、テストデータの拡張を行う。これによって、コーパスに含まれない未知のつぶやきも対象とした市民意見の分析が可能となる。

6.3節では、6.2節で拡張したテストデータを対象として、都市を横断した各属性のラベルの予測を行う。次に、モデルの予測結果に基づき、市民意見の出現頻度の10日間ごとの時系列順の推移を分析する。なお、出現頻度の時系列順の分析に使用するつぶやきは、保育園サービスではモデルが休園・登園自粛と関連すると予測したつぶやきとし、テイクアウトサービスでは全てつぶやきとする。そして、時系列の条件に加えて、各属性のラベルの条件を複数指定することで、全ての条件を満たすつぶやきのみを自動抽出する。また、この際に、アプレイザル意見タイプのラベルの条件を変化させることで、抽出される市民意見にどのような差異が現れるかを検証する。都市を横断した各属性のラベル分類では、5章の結果を踏まえ、意見属性のラベル分類には、ソース都市のデータを用いて1段階目のファインチューニングを行ったモデルによる確信度が上位50%のターゲット都市のデータを選定したT5のマルチタスク学習モデルを使用し、つぶやきの属性のラベル分類には、ソース都市のデータを用いて1段階目のファインチューニングを行ったモデルによる確信度が上位50%のターゲット都市のデータを選定した各属性独立のT5モデルを用いる。

## 6.2 テストデータの拡張

本章では、4章で作成した市民意見分析コーパスに加えて、4.1節と同様の方法で2020年1月1日から2020年7月11日までのつぶやきを収集することで、分析対象となるテストデータを拡張する。2020年1月1日から2020年7月11日までの全てのつぶやきの件数は、横浜市の保育園サービスで9,535件、テイクアウトサービスで21,014件、札幌市の保育園サービスで4,623件、テイクアウトサービスで14,980件となっている。4章で作成した市民意見分析コーパスについては、判定者の就業可能時間数に合わせて、両都市の両サービスにおいて2,622件のつぶやきを全体から抽出した。そのため、追加のテストデータとして、市民意見分析コーパスに含まれるつぶやきと重複しないよう、横浜市の保育園サービスで6,913件のつぶやき、テイクアウトサービスで18,392件のつぶやきを新たに抽出し、札幌市の保育園サービスで2,001件のつぶやき、テイクアウトサービスで12,358件のつぶやきを新たに抽出した。これらの新たに抽出したつぶやきから得られた文数は、横浜市の保育園サービスで22,256文、テイクアウトサービスで54,772文、札幌市の保育園サービスで5,986文、テイクアウトサービスで36,145文となった。また、市民意見分析コーパスに含まれるつぶやきについても、5.3節の都市を横断した各属性のラベル分類実験における5分割交差検証の各テストデータを、引き続きテストデータとして使用する。拡張後のデータのデータ数を表6.1に示す。

本章では、市民意見分析コーパス内のデータと、表6.1における新たに収集したデータを合わせたデータを全てのテストデータとし、これらのテストデータに対する、都市を横断した各属性のラベル分類モデルによる予測結果を用することで、市民意見分析を行う。

表 6.1: 複数都市を対象とした市民意見分析コーパスの拡張データのデータ数

都市	サービス	コーパス内のデータ		新たに収集したデータ		合計	
		つぶやき	意見ユニット	つぶやき	意見ユニット	つぶやき	意見ユニット
横浜市	保育園	2,622	7,916	6,913	22,256	9,535	30,172
	テイクアウト	2,622	6,671	18,392	54,772	21,014	61,443
札幌市	保育園	2,622	8,113	2,001	5,986	4,623	14,099
	テイクアウト	2,622	7,620	12,358	36,145	14,980	43,765

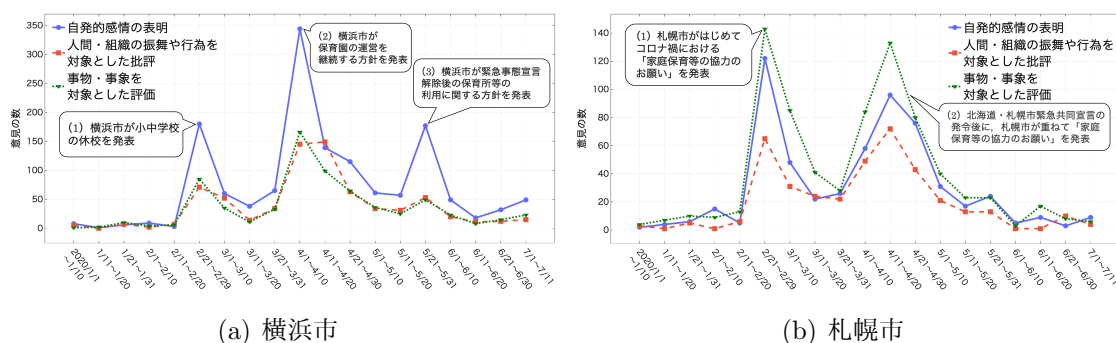


図 6.1: 保育園サービスの休園・登園自粛に関する市民意見の出現頻度の時系列順の推移

## 6.3 結果

### 6.3.1 保育園サービス

保育園サービスのつぶやきのうち、都市を横断したラベル分類モデルによって休園・登園自粛と関連すると推定されたつぶやきに現れる市民意見の出現頻度の時系列順の推移を図 6.1 に示す。

#### 横浜市の市民意見の抽出結果

図 6.1(a) より、横浜市民のつぶやきに現れる休園・登園自粛に関する意見は、3つの期間において急増していると分かる。これらの期間は、(1) 横浜市が小中学校の休校を発表した2月下旬、(2) 横浜市が保育園の運営を継続する方針を発表した4月上旬、(3) 横浜市が緊急事態宣言解除後の保育所等の利用に関する方針を発表した5月下旬、であり、都市を横断して抽出した市民意見の出現頻度の時系列順の推移が、行政による政策と強く関係していることが分かる。

そこで、市民意見の出現数が最大となっている(2)の横浜市が保育園の運営を継続する方針を発表した4月上旬の期間において、地域依存性：依存，サービスとの適合性：関連する，投稿主の立場：小さい子を持つ親，休園・登園自粛との関連性：関連する，アプレイザル意見タイプ：自発的感情の表明，という条件を全て満たすつぶやきのみを自動抽出した。抽出された横浜市民の市民意見の例は以下の通りである<sup>1</sup>（なお、上記の属性の条件を指定することで抽出された市民意見は全部で259件あり、以下には抽出された市民意見のうち、典型的な例を記している。）。

- いっそのこと休園にしてほしい。でも横浜市は原則開園って。。。

<sup>1</sup>本論文に記載している全てのつぶやきは表現を一部改変している。

保育士の皆さんもほんと辛いだろうに。

- 超個人的な望みを言わせてもらおうと、いっそ保育園を休園にしてくれたら私も仕事を休めるんですがね…（妊婦なのに都内まで電車通勤）

上記のような条件を指定することで、横浜市が保育園の運営を継続する方針を発表した4月上旬に、小さい子を持つ横浜市民は保育園の休園や登園自粛に関してどういった感情を抱いているのか、という情報を含むつぶやきのみを、大量のつぶやきから整理して抽出することができる。実際の抽出結果から、横浜市民はこの時期に保育園を開園させるという横浜市の方針に対して否定的であり、不満感情を抱いていると分かる。こういった横浜市の政策に対する意見は横浜市で暮らす市民ならではの意見となっており、政策の改善のために有用な意見である。このように、都市を横断した市民意見抽出手法を用いることで、ターゲット都市である横浜市における新たなデータ作成コストを半減させた上でも、横浜市民ならではの意見を抽出可能であると分かる。

一方で、子供を保育園に通わせることに対して保護者が不安を感じているコロナ禍において、保育士や行政はどのような振舞をすれば良いのかという情報も、保護者の不安を軽減するために重要な情報となる。そこで今度は、上記の市民意見抽出と同様の4月上旬の期間において、地域依存性：依存、サービスとの適合性：関連する、投稿主の立場：小さい子を持つ親、極性：肯定、アプレイザル意見タイプ：人間・組織の振舞や行為を対象とした批評、というように、アプレイザル意見タイプの条件を変化させ、肯定的な意見という条件を付け加えた上で市民意見を抽出した。なお、今回は保育園の生活全般に関しての保護者の意見を抽出するため、休園・登園自粛との関連性属性は指定しない。抽出された横浜市民の市民意見の例は以下の通りである（なお、上記の属性の条件を指定することで抽出された市民意見は全部で102件あり、以下には抽出された市民意見のうち、典型的な例を記している。）。

- 保育園ではマスクも消毒も徹底しているし、保護者にもマスクと消毒が要求される。送迎は玄関先で基本は保護者を中まで入れないし、外部講師を呼ぶこと、行事、保護者会も全部中止。出来ることは全部やってくれている。
- 良かった、横浜市だと保育園を自粛で休んだとしても保育料は返金になるんだね。ありがとうございます。

肯定的な人間・組織の振舞や行為を対象とした批評、という条件を指定することで、小さい子を持つ保護者が肯定的に感じている保育士（人間）や行政（組織）の振舞のみを



整理して抽出することができる。上記の1つ目の例である保育士の振舞に対する肯定的な批評は、コロナ禍という未曾有の状況において、保育士がどのように行動するべきかというベストプラクティスのような情報となっている。そのため、他の多くの保育士が参考にすることで、コロナ禍における保護者の不安を取り除くことに繋げることができる。また、上記の2つ目の例である行政に対する批評については、横浜市が今後も続けていくべき政策を明らかにすることに繋がるだけでなく、他の自治体が登園自粛に関連する政策を検討する際に参考とすることができる。このように、アプレイザル意見タイプの条件を変化させることで、目的に応じた市民意見を整理して抽出することができる。

### 札幌市の市民意見の抽出結果

図6.1(b)より、札幌市民のつぶやきに現れる休園・登園自粛に関する意見は、2つの期間において急増していると分かる。これらの期間は、(1)札幌市がはじめてコロナ禍における「家庭保育等の協力のお願ひ」を発表した2月下旬、(2)北海道・札幌市緊急共同宣言の発令後に、札幌市が重ねて「家庭保育等の協力のお願ひ」を発表した4月中旬、であり、札幌市をターゲット都市とした際にも、都市を横断して抽出した市民意見の出現頻度の時系列順の推移が、行政による政策と強く関係していることが分かる。さらに、都市によって行政による政策の内容や政策の発表時期が異なるため、市民意見の出現頻度の推移も都市によって異なるが、提案手法による市民意見抽出によって、都市ごとの市民意見の増加のタイミングの差を捉えられていると分かる。また、図6.1から、札幌市民が抱える休園・登園自粛に関する市民意見では、横浜市民が抱える自発的な感情表明とは異なり、事物・事象を対象とした評価が高い割合を占めている。そこで、(2)の北海道・札幌市緊急共同宣言の発令後に、札幌市が重ねて「家庭保育等の協力のお願ひ」を発表した4月中旬において、前述の横浜市における抽出条件から、意見タイプ属性のラベルを変化させて市民意見を抽出する。抽出する意見の属性ラベルの条件を、地域依存性：依存、サービスとの適合性：関連する、投稿主の立場：小さい子を持つ親、休園・登園自粛との関連性：関連する、アプレイザル意見タイプ：事物・事象を対象とした評価、とした際に、抽出された札幌市民の市民意見の例は以下の通りである（なお、上記の属性の条件を指定することで抽出された市民意見は全部で91件あり、以下には抽出された市民意見のうち、典型的な例を記している。）。

- 絶対に必要な人は別だけど、やはり保育園は休園とか自粛にはならないのか。色々あるけど、先生も親もどっちも精神的に限界な気がする。。。 #休園 #札幌 #北海道

- やはり恐れてたことが起こってきた。友人の医療関係者の子供の保育園が預けるのを断ってきたらしい。旦那が各々の事情でいないと母親は働けなくなるよね。本人たちは濃厚接触者ではないのに、現実で差別が起きてきてる。 #札幌市 #保育園 #登園拒否

事物・事象を対象とした評価の市民意見を抽出することで、コロナ禍にも関わらず保育園が休園や登園自粛にならないという事象に対する混乱を表す意見や、保育園で発生した医療関係者への差別という事象に対する意見を抽出できる。事物・事象を対象とした評価、という条件を用いると、上記の例のように、市民は具体的にどういったことに対して意見を述べているのかを自動抽出することができる。さらに、否定的な意見という条件を加えることで、市民が否定的な感情を抱いている何かしらの対象をつぶやきから整理して抽出可能となっている。これらの抽出結果を用いることで、行政が解決すべき具体的な課題を明らかにすることができる。また、上記の例に含まれている「#札幌」、 「#札幌市」からも分かる通り、これらの意見は札幌市民ならではの意見となっている。つまり、提案手法を用いることで、札幌市をターゲット都市とした場合においても、新たな教師データの作成コストを半減させた上で、札幌市民ならではの意見を抽出することができる。と分かる。

なお、本節では、多くの属性の条件を指定することで、詳細に整理された市民意見を自動抽出した。一方で、行政は、多様な市民意見を分断することなく幅広く集める必要もあると考える。この場合、投稿主の立場属性の条件を指定せず、小さな子を持つ親の意見と保育園関係者の意見を合わせて収集する手法や、極性の条件を指定せず、肯定的な意見も否定的な意見も分断することなく収集する手法のように、属性の条件を指定しすぎないことで、幅広い市民意見を自動抽出することも可能となっている。

### 6.3.2 飲食店のテイクアウトサービス

飲食店のテイクアウトサービスのつぶやきについて、都市を横断したラベル分類モデルによって推定された市民意見の出現頻度の時系列順の推移を図 6.2 に示す。

#### 横浜市の市民意見の抽出結果

図 6.2(a) より、横浜市の飲食店のテイクアウトサービスに関する市民意見は、4月1日から4月10日の期間に急増し始めていると分かる。この期間は、神奈川県を対象とした第1回緊急事態宣言が発令された4月7日を含む期間であり、この期間に横浜市の多

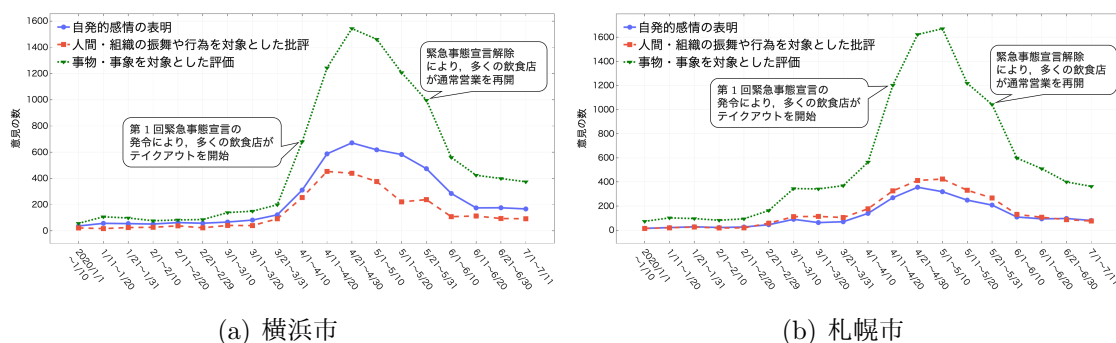


図 6.2: 飲食店のテイクアウトサービスの市民意見の出現頻度の時系列順の推移

くの飲食店がテイクアウトサービスを開始したと考えられる。また、4月1日から4月10日の期間に急増し始めた市民意見の出現数は、4月の下旬に最大となり、緊急事態宣言が解除された5月25日以降は急激に現象していくことが分かる。このように、提案手法を用いた市民意見の出現頻度の時系列順の推移は、社会の状況と関連していることが分かる。また、飲食店のテイクアウトサービスに関する市民意見では、事物・事象を対象とした評価意見が多くなっている。これは、飲食店のテイクアウトというサービスの性質上、実際にテイクアウトサービスを利用した客や、商品の宣伝のつぶやきを投稿する立場の飲食店の店員による、テイクアウト商品を対象とした評価に関する内容のつぶやきが多くなるためであると考えられる。そこで、横浜市民の意見数が最大となっている4月下旬の全てのつぶやきから、地域依存性：依存，サービスとの適合性：関連する，投稿主の立場：飲食店を利用した人，商品の評価との関連性：関連する，極性：肯定，アプレイザル意見タイプ：事物・事象を対象とした評価，という属性ラベルの条件を満たすつぶやきのみを自動抽出した。抽出された横浜市民の市民意見の例は以下の通りである（なお，上記の属性の条件を指定することで抽出された市民意見は全部で464件あり，以下には抽出された市民意見のうち，典型的な例を記している。）。

- #六角橋商店街の #餃子M で2種類の餃子をテイクアウト。きのこ餃子はとろとろで噛むと濃厚な具が流れてくる。何個でも食べたい。テキサス餃子はカレー味の餡にコーンとチーズブレンド。タレは不要で何個でも食べたい。 #横浜飯
- 地球の中華そばさんの3回目のテイクアウトで超純水採麵天国屋のリスペクト麺を購入。スープと麺を味わうために、ネギだけを乗せました。スープが旨すぎて、麺も好き！最近の家での楽しみになっていました。また買います！ #地球の中華そば #テイクアウト

上記の属性ラベルの条件を指定することで、横浜市民が購入したテイクアウト商品の中でも、良い評価を得ている商品に関する市民意見のみを整理して自動抽出することができる。これらの意見は、飲食店のテイクアウトサービスを利用したことがなく、どの店舗に行けば美味しい商品が食べられるか分からない横浜市民や、横浜市を訪れた観光客が、どこの店に行くかを判断する際に有用な情報となる。さらに、上記に示したつぶやきの例に含まれる「餃子 M」、「地球の中華そば」は、横浜市にのみ存在する飲食店である。つまり、飲食店のテイクアウトサービスにおいても、札幌市民のつぶやきと比較的少量の横浜市民のつぶやきを用いる提案手法によって、ターゲット都市である横浜市ならではの市民意見が抽出できることが分かる。

また、コロナ禍においては、多くの飲食店が初めてテイクアウトサービスの提供を開始した。そのため、飲食店の従業員は、こういった接客をすれば利用者が満足に感じるのかといったベストプラクティスの情報を得ることができれば、慣れていないコロナ禍における接客を改善することができる。そこで今回は、サービスとの適合性：関連する、投稿主の立場：飲食店を利用した人、商品の評価との関連性：関連する、極性：肯定、アプレイザル意見タイプ：人間・組織の振舞や行為を対象とした批評、というように、アプレイザル意見タイプの条件を変化させてつぶやきを自動抽出した。なお、今回は地域に関わらず飲食店のテイクアウトサービスに関するベストプラクティスを抽出するため、地域依存性の属性は指定しない。抽出された横浜市民の市民意見の例は以下の通りである（なお、上記の属性の条件を指定することで抽出された市民意見は全部で 56 件あり、以下には抽出された市民意見のうち、典型的な例を記している。）。

- 鶏恵がテイクアウトやってるってつぶやきが流れてきたから、特製つくねハンバーグ丼を買ってみた。オーダーをしてから調理してくれるからできたてです。サラダがセットでマスクまでついてきて 540 円はお得ですよ！ #横須賀 #鶏恵
- 家の近く中華料理店ではじめてお持ち帰りしたけど、麺と具を分けてくれた 🍜 麺分けて炒めてくれてるので、うちまで持ち帰っても中華街に負けないくらいの美味しさ

アプレイザル意見タイプのラベルを人間・組織の振舞や行為を対象とした批評とし、極性のラベルを肯定と指定することで、上記の 1 つ目の例における「オーダーをしてから調理してくれる」や、2 つ目の例における「麺と具を分けてくれた」といった、店員の振舞に対する肯定的な表現を抽出することができる。これらの利用者による意見は、飲

食店の従業員が参考にすることで、これまでに行ったことのないテイクアウトサービスの改善に繋げることができる。

### 札幌市の市民意見の抽出結果

図 6.2(b) より、札幌市の飲食店のテイクアウトサービスに関する市民意見は、4月11日～4月20日の期間に急増し始めている。この期間は、第1回緊急事態宣言の対象範囲が拡大され、北海道が緊急事態宣言の対象地域となった4月16日を含む期間であり、この期間に札幌市の多くの飲食店がテイクアウトサービスを開始したと考えられる。前述したように、横浜市において飲食店のテイクアウトサービスに関する市民意見が急増し始めたのは、第1回緊急事態宣言が発令された4月7日を含む期間となっている。このように、提案手法による都市を横断した市民意見抽出では、横浜市で市民意見が急増し始めた4月7日を含む期間と、札幌市で市民意見が急増し始めた4月16日を含む期間とといったように、数日単位の都市間の差異を反映した市民意見の出現頻度の時系列順の推移を分析することができる。なお、緊急事態宣言が解除された時期は、横浜市と札幌市に共通で5月25日となっており、この期間を過ぎると市民意見の数が急激に減少していくことは両都市で共通している。

札幌市においても、市民意見の数が最大となっている5月上旬のつぶやきの中から、横浜市と同様に、地域依存性：依存、サービスとの適合性：関連する、投稿主の立場：飲食店を利用した人、商品の評価との関連性：関連する、極性：肯定、アプレイザル意見タイプ：事物・事象を対象とした評価、という属性ラベルの条件を全て満たすつぶやきを自動抽出した（なお、上記の属性の条件を指定することで抽出された市民意見は全部で392件あり、以下には抽出された市民意見のうち、典型的な例を記している。）。

- ひろちゃんのザンギは、塩ザンギが超ウまいお弁当屋さん。我が家で定番のテイクアウトです！札幌にはいくつが店舗があって、事前に電話すると揚げたてが買えますよー #札幌テイクアウト #StayHome
- 昨日と同じく、スープカレー 34 でテイクアウトしたハンバーグカレーを食べました。やっぱ34のハンバーグはビーフ100%で美味しい！ #スープカレー #札幌カレー部 #スープカレー 34

このような属性ラベルを指定することで、札幌市においても、テイクアウト商品に関する大量のつぶやきの中から、商品の良い評価の情報を含むつぶやきのみを整理して抽出することができる。また、「#札幌テイクアウト」、「#札幌カレー部」といった表現が

ら、上記のつぶやきは札幌市ならではの内容であることが分かる。さらに、「ザンギ」と「スープカレー」が札幌市の名産グルメであることから、都市を横断した市民意見抽出によって、札幌市における教師データ作成コストを半減させた上で、札幌市特有の市民意見を抽出可能であると分かる。

### 6.3.3 自動抽出された市民意見における抽出ミスの割合に関する考察

6.3.1 節、6.3.2 節では、複数の属性ラベルの条件を指定することで、つぶやきから市民意見を整理して自動抽出した。結果として、行政の政策に反映可能な市民意見や、接客業のサービスの改善に有用な市民意見、利用する接客業のサービスを選択する際に参考となる意見が自動抽出可能になったことが分かった。一方で、抽出された市民意見のうち、有用な意見だけではなく、抽出ミスの意見がどの程度存在するかの分析も重要となる。そこで本節では、抽出された市民意見のうち、4章で作成した市民意見分析コーパスに含まれるつぶやき、つまり、人手による正解ラベルが付与されているつぶやきに注目することで、抽出された市民意見における抽出ミスの割合に関する考察を行う。

#### 6.3.3.1 保育園サービス

保育園サービスにおける抽出ミスの意見の割合について、6.3.1 節ではじめに行った横浜市の市民意見抽出を例に考察する。6.3.1 節では、はじめに、横浜市が保育園の運営を継続する方針を発表した4月上旬の期間において、地域依存性：依存、サービスとの適合性：関連する、投稿主の立場：小さい子を持つ親、休園・登園自粛との関連性：関連する、アプレイザル意見タイプ：自発的感情の表明、という条件を全て満たすつぶやきのみを自動抽出した。抽出された市民意見は全部で259件あり、そのうち、4章で作成した市民意見分析コーパスに含まれる意見は、80件存在した。これらの80件の意見のうち、モデルによる予測結果が全ての属性において人手によるアノテーションと同一の意見、つまり、予測が完全に正しい意見は、41件であった。このことから、複数の属性を組み合わせた上でも、人手によるアノテーションと比較して51.25%の意見は正しく抽出できていると分かる。

しかし、モデルによる各属性の予測ラベルが、コーパスに付与されているラベルと異なる場合においても、全ての意見が行政の政策の改善に有用で無いとは限らない。この点について考察するため、モデルの予測ラベルがコーパスのラベルと異なる意見を確認する。はじめに、自動抽出された市民意見において、モデルの予測ラベルがコーパスの

表 6.2: 保育園サービスで自動抽出された市民意見における各属性のモデルの予測ラベルが正しい割合

属性	地域依存性	サービスとの適合性	投稿主の立場	休園・登園自粛との関連性	アプレイザル意見タイプ
予測が正しい割合	98.75%	98.75%	75.00%	87.50%	81.25%

ラベルと等しい割合を属性ごとに算出した。結果を表 6.2 に示す。表 6.2 より、投稿主の立場とアプレイザル意見タイプの予測の精度が低いことが、モデルの予測が全属性で正しい意見の割合の低下の原因となっていることが分かる。モデルが投稿主の立場の予測を誤っている例として、「学校は休校になるのに、保育園は休園にならないのはナゾ...(絶対に乳幼児の方がやばいと思う...)」というつぶやきを、小さな子を持つ親によるつぶやきであると予測していた。このつぶやきを確認するだけでは、投稿主に小さな子がいるかの判断はできない。しかし、このような意見も、保育園が休園にならないことに対する市民の不安感情という観点では、行政にとって有用な意見であることが分かる。また、モデルがアプレイザル意見タイプの予測を誤っていた例として、「緊急事態宣言で小学校や幼稚園は休みなのに、保育園は感染防止対策の協力の要請?! これまでと変わらない...」という意見を、市民の自発的感情として抽出していた。「これまでと変わらない」という意見は事象に対する評価表現であるため、人手によるアノテーションでは、正解ラベルは事物・事象を対象とした評価となっている。しかし、このような意見も、保育園が休みにならないことに対する市民の感情であり、行政にとっては有用な意見であると考えられる。本研究で作成した市民意見分析コーパスでは、各属性に対応するラベルは 1 つに定めており、提案手法においても、全てのデータに対してモデルは各属性につき 1 つのラベルを判定している。しかし、上記の例のように、最も正確なアプレイザル意見タイプのラベルは事物・事象への評価であるものの、感情表明として抽出されたとしてもノイズにはならない意見も存在する。条件を指定した複数の属性のうち、モデルの予測が 1 属性でも誤っていた意見は 48.75%となっていたが、上記で示した例のように、モデルの予測は誤っているものの行政の政策の改善に有用な意見を除くと、著者の確認した範囲では、抽出ミスは 18.75%となっていた。つまり、抽出された市民意見のうち、80%以上の意見は行政の政策の改善に有用な意見であると分かる。

### 6.3.3.2 飲食店のテイクアウトサービス

飲食店のテイクアウトサービスにおいても、保育園サービスと同様に、6.3.2 節ではじめに行った横浜市の市民意見抽出を例に、抽出ミスの意見の割合を考察する。6.3.2 節で

表 6.3: テイクアウトサービスで自動抽出された市民意見における各属性のモデルの予測ラベルが正しい割合

属性	地域依存性	サービスとの適合性	投稿主の立場	商品の評価との関連性	極性	アプレイザル意見タイプ
予測が正しい割合	89.13%	100.00%	91.30%	97.83%	100.00%	97.83%

は、はじめに、4月下旬において、地域依存性：依存，サービスとの適合性：関連する，投稿主の立場：飲食店を利用した人，商品の評価との関連性：関連する，極性：肯定，アプレイザル意見タイプ：事物・事象を対象とした評価，という属性ラベルの条件を満たすつぶやきのみを自動抽出した．抽出された市民意見は全部で464件あり，そのうち，4章で作成した市民意見分析コーパスに含まれる意見は，46件存在した．このうち，モデルによる予測結果が全ての属性において人手によるアノテーションと同一の意見は，36件であった．このことから，飲食店のテイクアウトサービスでは，人手によるアノテーションと比較して78.26%の意見が正しく抽出できていると分かる．

飲食店のテイクアウトサービスにおいても，モデルの予測ラベルがコーパスのラベルと等しい割合を属性ごとに算出した．結果を表6.3に示す．表6.3より，テイクアウトサービスにおいては，地域依存性と投稿主の立場の予測精度が他の属性と比較して低くなっている．モデルが投稿主の立場の予測を誤っている例としては，「肉ねぎ炒め！金家は定食もウマし 🍡 #テイクアウト OK #タンタンメン金家 #横浜グルメ」というつぶやきを，お店を利用した人によるつぶやきと予測していた例がある．「#テイクアウト OK」という表現は，主に飲食店の公式アカウントが用いる表現であり，このつぶやきは実際には飲食店の公式アカウントのつぶやきとなっている．しかし，市民が飲食店を選ぶ際には，具体的な商品に対する肯定的な評価意見としてこのようなつぶやきも参考とすることができる．地域依存性の予測を誤っている例としては，「南関あげいなり」のように，地名を含む商品名に関するつぶやきを，地域依存性有りと予測してしまっている例があった．モデルの予測を1属性でも誤っていた意見は28.84%であったが，前述の地域依存性の誤りのような明らかな抽出ミスの意見は，著者が確認した範囲では13.04%となっていた．つまり，飲食店のテイクアウトサービスにおいては，自動抽出された市民意見のうち，85%以上の意見を有効活用できると分かる．



# 第7章 おわりに

## 7.1 本研究のまとめ

行政による政策を改善するためには、都市で暮らす市民の意見を反映させることが重要となる。さらに、接客業のサービスを改善するためにも、実際にサービスを利用した市民の意見を取り入れる必要がある。そのため、本研究では、事前学習済みの言語モデルをファインチューニングすることで、Twitter のつぶやきから市民意見を自動抽出することを目的とする。Twitter は多くのユーザが利用しており、日々感じたことを気軽に投稿しているため、Twitter を用いることで様々な市民意見を抽出することができる。しかし、Twitter に現れる市民意見は多様であるため、従来の意見分析研究で用いられている極性や感情の種類といった特定の観点のみでは、多様な市民意見を整理できない。そこで本研究では、アプレイザル理論に基づく意見タイプを含む複数の属性を定義し、これらの属性を組み合わせることで多様な市民意見を整理する。さらに、地方自治体による政策や接客業の店舗のサービスは都市によって異なるため、市民意見も都市によって異なる。しかし、全ての都市で市民意見抽出モデルの訓練に用いる教師データを作成するのはコストが大きい。そこで本研究では、既に教師データを作成済みのソース都市のデータと、比較的少量のターゲット都市のデータを用いてターゲット都市における市民意見を抽出する。これによって、ターゲット都市における新たなデータ作成コストを抑えることができる。以上の本研究の目的と、市民意見抽出における課題について1章で述べた。

2章では、関連研究であるアプレイザル理論と、ソーシャルメディア上の市民意見分析に関する研究、アノテーションコストの削減に関する研究、マルチタスク学習と本研究で用いる言語モデルについて論じ、本研究の位置づけを述べた。

3章では、提案手法の都市を横断した市民意見抽出手法について述べた。提案手法では、はじめに、アプレイザル理論に基づく意見タイプ、極性、コミュニケーション意見タイプの3つの意見属性を含む、本手法で用いる複数の属性を定義した。続いて、複数の属性を組み合わせた市民意見抽出モデルについて述べた。市民意見抽出モデルでは、

3つの意見属性のラベル分類タスクの関連性に着目し、マルチタスク学習手法を適用することで、意見属性のラベル分類精度の向上を目指す。提案手法の最後に、市民意見抽出モデルを拡張し、都市を横断して市民意見を抽出する手法について述べた。都市を横断した市民意見抽出では、ソース都市で1段階目のファインチューニングを行ったモデルによる予測の確信度を用いて、ターゲット都市の比較的少量のデータを選定し、人手によるアノテーションを行う。そして、正解ラベル付きのターゲット都市のデータを用いて2段階目のファインチューニングを行い、ターゲット都市における市民意見を抽出する。

4章では、実験に用いる市民意見分析コーパスについて述べた、市民意見分析コーパスには、横浜市民アカウントと札幌市民アカウントによる保育園と飲食店のテイクアウトの2つのサービスに関するつぶやきを含み、これらのつぶやきに対して人手による各属性のラベルのアノテーションを行うことでコーパスを作成した。アノテーションの判定者間一致度を表す Fleiss の  $\kappa$  係数は、全ての属性において 0.6 (Substantial Agreement) 以上となり、判定者によって属性ラベルの判定に大きな差異が生まれないと示された。

5章の実験では、提案手法の有効性を評価するため、都市を横断した各属性のラベル分類を行った。実験では、都市を横断した各属性のラベル分類に先立ち、単一都市のデータを用いて、BERT モデルと比較した際の T5 モデルの有効性と、意見属性のラベル分類におけるマルチタスク学習の有効性を確認した。都市を横断したラベル分類では、ソース都市のデータを用いて1段階目のファインチューニングを行ったモデルによる予測の確信度が上位のターゲット都市のデータを用いて2段階目のファインチューニングを行う手法が有効であることを示した。また、都市を横断したラベル分類手法を用いることで、ターゲット都市におけるデータ作成コストを半減した上で、ターゲット都市のデータを 100%用いて単一都市でラベル分類した際の精度と同等かそれ以上の精度が実現可能であることを示した。

6章の実験では、はじめに分析対象のテストデータを拡張し、実際につぶやきから都市を横断して市民意見を抽出した。実験では、提案手法を用いて抽出した市民意見の出現頻度の時系列順の推移が行政による政策や社会の状況と強く関連していることを示した。また、複数の属性を用いた市民意見抽出モデルを用いることで、保育園に関する自治体の政策に対する市民の不満感情を含むつぶやきや、テイクアウト商品に対して高い評価をしているつぶやきを大量のつぶやきから整理して抽出できることを示した。さらに、都市を横断した市民意見抽出によって、ターゲット都市におけるデータ作成コストを抑えた上で、ターゲット都市ならではの市民意見が抽出可能であることを示した。

## 7.2 今後の課題

各都市の自治体が抱える行政課題は多様であり、特定のサービスに関して行政の政策がうまく機能している都市もあれば、市民から理解を得られていない都市もある。このような場合、特定の都市における行政課題を解決するためには、他の多くの都市における市民意見を分析し、市民からの理解を得るための政策を考案する必要がある。本研究では、政令指定都市である横浜市と札幌市に着目し市民意見の分析を行ったが、今後はさらに多くの都市を対象として市民意見を抽出し、特定の都市でサービスの改善を行うには他のどの都市の政策を参考とすべきか、また、それはどういった政策であるのか、といった情報を市民のつぶやきから取得できるフレームワークの検討を進めていきたい。

また、各自治体が抱える行政課題や接客業が、どのようなサービスに関連するものであるかといった情報も多様である。本研究では、提案手法の評価を行うため、行政の政策と強い関連のある保育園サービスと、商業的な側面が強い飲食店のテイクアウトサービスを例として市民意見の分析を行った。今後は、対象となるサービスを拡張し、より多くのサービスに関する市民意見を分析していきたいと考えている。

# 謝辞

はじめに、本研究を行うにあたって、いつも熱心な指導をいただいた図書館情報メディア系関洋平准教授に深く感謝の意を表します。関先生には、筑波大学学群生の頃から約3年間、研究指導をしていただきました。この3年間、学術論文の執筆や学外での研究発表など、研究室配属以前までの自分からは想像もできないほど刺激的で貴重な体験をたくさんさせていただきました。関先生のもとでの充実した3年間の研究活動を通して、人間として大きく成長できたと強く感じています。ここに改めて感謝の意を表します。

副指導教員を務めていただいた筑波大学図書館情報メディア系高久雅生准教授には、博士前期課程の期間のみならず、学群生の頃より研究に関する貴重なアドバイスを数多くいただきました。副指導教員を快く引き受けてくださり、いつも優しく、そして丁寧に指導をいただいた高久雅生准教授に、深く感謝の意を表します。

また、国立情報学研究所の神門典子教授、国立国語研究所の柏野和佳子准教授、一橋大学の櫻惇志准教授には、論文執筆や研究会発表の際に、共著者としてあらゆる側面で研究を支えていただきました。各分野の最前線でご活躍されている研究者である皆様と共に研究を行えたことを、心より誇りに思います。ここに深く感謝いたします。

研究室の先輩である小邦将輝さんには、修了後も研究に関する助言や励ましの言葉を数多くいただきました。小邦さんはいつでも自分の目標であり、小邦さんのおかげで、最上級生となった現在でも、研究室配属当時の初心を忘れることなく研究を行うことができました。ここに感謝の意を表します。

さらに、研究室の先輩である劉依泓さんには、昨年度までの在学期間中に面倒を見ていただいただけでなく、修了後も励ましの言葉をたくさんいただきました。とても頼りになる先輩がいたおかげで、安心して研究を行うことができました。ありがとうございました。

研究室メンバーの魏徴さんと許一帆さんとは、2年間、同期として研究活動をさせていただきました。心優しく明るい性格のお2人が同期にいたおかげで、とても充実した2年間になりました。ありがとうございました。

研究室の後輩の妹尾考さん、西村千恵子さん、儀忠躍さん、三原江理子さん、河南直

希さん、榊原智仁さん、佐々木謙人さん、宋亮玄さん、米丸周吾さんには、先輩として至らない自分を支えていただきました。ありがとうございました。

また、コンテンツ工学研究室の稲福和史さん、大峠和基さん、熊田大雅さんとは、7D棟 140号室の合同研究室で共に研究活動をさせていただきました。皆様のおかげで合同研究室が活気ある部屋になり、いつも楽しく研究を行うことができました。ありがとうございました。

融合知能デザイン研究室の金承彦さん、上保研究室の船越大輝さん、物理ベースコンピュータグラフィックス研究室の松浦一輝さんとは、学群1年次の頃から1番近い距離で学生生活を過ごし、特に大学院入学後は、分野は違えど、切磋琢磨し合って研究活動に励める良い同期として多くの時間を共にしました。皆様のおかげで学生生活が充実し、研究活動にもいつも楽しく励むことができました。ありがとうございました。

本研究の一部は、科学研究費補助金基盤研究B（課題番号19H04420）、挑戦的研究（萌芽）（課題番号22K19822）、2022年度国立情報学研究所公募型共同研究（採択番号22S0103）の助成を受けて遂行されました。

## 参考文献

- [1] Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa Anke, and Leonardo Neves. TweetEval: Unified benchmark and comparative evaluation for tweet classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 1644–1650, Online, November 2020. Association for Computational Linguistics.
- [2] Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory, COLT' 98*, p. 92–100, New York, NY, USA, 1998. Association for Computing Machinery.
- [3] Rich Caruana. Multitask learning. *Machine Learning*, Vol. 28, pp. 41–75, 1997.
- [4] Jing Chen, Jun Feng, Xia Sun, and Yang Liu. Co-training semi-supervised deep learning for sentiment classification of mooc forum posts. *Symmetry*, Vol. 12, No. 1, 2020.
- [5] Jacob Cohen. A power primer. *Psychological Bulletin*, Vol. 112, No. 1, pp. 155–159, 1992.
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, 2019.
- [7] Dimitar Dimitrov, Erdal Baran, Pavlos Fafalios, Ran Yu, Xiaofei Zhu, Matthäus Zloch, and Stefan Dietze. TweetsCOV19 - A Knowledge Base of Semantically Annotated Tweets about the COVID-19 Pandemic. In *Proceedings of the 29th ACM*

*International Conference on Information & Knowledge Management* October 2020, pp. 2991–2998, 2020.

- [8] Luca Dini and André Bittar. Emotion analysis on Twitter: The hidden challenge. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, pp. 3953–3958, 2016.
- [9] Joseph L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, Vol. 76, No. 5, pp. 378–382, 1997.
- [10] M. A. K. Halliday. *An Introduction to Functional Grammar*. Edward Arnold, 1985.
- [11] Yulan He and Deyu Zhou. Self-training from labeled features for sentiment analysis. *Information Processing & Management*, Vol. 47, No. 4, pp. 606–616, 2011.
- [12] Tetsuya Ishida, Yohei Seki, Wakako Kashino, and Noriko Kando. Extracting citizen feedback from social media by appraisal opinion type viewpoint. *Journal of Natural Language Processing*, Vol. 29, No. 2, pp. 416–442, 2022.
- [13] Mohammed Jabreel and Antonio Moreno. A deep learning-based approach for multi-label emotion classification in tweets. *Applied Sciences*, Vol. 9, No. 6, p. 1123, 2019.
- [14] Zhijing Jin, Zeyu Peng, Tejas Vaidhya, Bernhard Schoelkopf, and Rada Mihalcea. Mining the cause of political decision-making from social media: A case study of COVID-19 policies across the US states. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pp. 288–301, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [15] J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, Vol. 33, No. 1, pp. 159–174, 1977.
- [16] David D. Lewis and William A. Gale. A sequential algorithm for training text classifiers. In Bruce W. Croft and C. J. van Rijsbergen, editors, *SIGIR '94*, pp. 3–12, London, 1994. Springer London.
- [17] Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. Multi-task deep neural networks for natural language understanding. In *Proceedings of the 57th*

- Annual Meeting of the Association for Computational Linguistics*, pp. 4487–4496, 2019.
- [18] James Robert Martin and Peter Robert Rupert White. *The Language of Evaluation: Appraisal in English*. Palgrave Macmillan, 2005.
- [19] T. M. Mitchell. The need for biases in learning generalizations. Technical Report CBM-TR-117, Rutgers University, New Brunswick, NJ, 1980.
- [20] Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. SemEval-2018 Task 1: Affect in Tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pp. 1–17, New Orleans, Louisiana, June 2018.
- [21] Natalia Mora and Julia Lavid-Lopez. Building an annotated dataset of app store reviews with appraisal features in english and spanish. *Proceedings of the Second Workshop on Computational Modeling of People’s Opinions, Personality, and Emotions in Social Media*, pp. 16–24, 2018.
- [22] Minlong Peng, Qi Zhang, Yu-gang Jiang, and Xuanjing Huang. Cross-domain sentiment classification with target domain specific information. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2505–2513, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [23] Matthew Purver and Stuart Battersby. Experimenting with Distant Supervision for Emotion Classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 482–491, 2012.
- [24] Stefan Rübiger, Mishal Kazmi, Yücel Saygın, Peter Schüller, and Myra Spiliopoulou. SteM at SemEval-2016 task 4: Applying active learning to improve sentiment classification. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pp. 64–70, San Diego, California, June 2016. Association for Computational Linguistics.
- [25] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learn-



- ing with a unified text-to-text transformer. *Journal of Machine Learning Research*, Vol. 21, No. 140, pp. 1–67, 2020.
- [26] Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B. Gupta, Xiaojiang Chen, and Xin Wang. A survey of deep active learning. *ACM Comput. Surv.*, Vol. 54, No. 9, oct 2021.
- [27] Salim Sazzed and Sampath Jayarathna. Ssentia: A self-supervised sentiment analyzer for classification from unlabeled data. *Machine Learning with Applications*, Vol. 4, p. 100026, 2021.
- [28] Yohei Seki, Lun Wei Ku, Le Sun, Hshin Hshi Chen, and Noriko Kando. Overview of Multilingual Opinion Analysis Task at NTCIR-8 - A Step Toward Cross Lingual Opinion Analysis. In *Proceedings of the Eighth NTCIR Workshop Meeting*, pp. 209–220, 2010.
- [29] Burr Settles. Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2010.
- [30] Heereen Shim, Dietwig Lowet, Stijn Luca, and Bart Vanrumste. Lets: A label-efficient training scheme for aspect-based sentiment analysis by using a pre-trained language model. *IEEE Access*, Vol. 9, pp. 115563–115578, 2021.
- [31] Nadia Felix F. Da Silva, Luiz F. S. Coletta, and Eduardo R. Hruschka. A survey and comparative study of tweet sentiment analysis via semi-supervised learning. *ACM Comput. Surv.*, Vol. 49, No. 1, jun 2016.
- [32] Christopher Stelzmüller, Sebastian Tanzer, and Markus Schedl. Cross-city analysis of location-based sentiment in user-generated text. In *Companion Proceedings of the Web Conference 2021*, WWW '21, p. 339–346, New York, NY, USA, 2021. Association for Computing Machinery.
- [33] Jesper E. van Engelen and Holger H. Hoos. A survey on semi-supervised learning. *Machine Learning*, Vol. 109, pp. 373–440, 2020.
- [34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In

*Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 6000–6010. Curran Associates, Inc., 2017.

- [35] Soroush Vosoughi, Helen Zhou, and Deb Roy. Enhanced Twitter sentiment classification using contextual information. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 16–24, Lisboa, Portugal, September 2015. Association for Computational Linguistics.
- [36] Fangzhao Wu, Yongfeng Huang, and Jun Yan. Active sentiment domain adaptation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1701–1711, Vancouver, Canada, July 2017. Association for Computational Linguistics.
- [37] David Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics, ACL '95*, p. 189–196, USA, 1995. Association for Computational Linguistics.
- [38] Xiaojin Zhu. Semi-supervised learning literature survey. Computer Sciences TR 1530, University of Wisconsin-Madison, 07 2008.
- [39] 一般財団法人地方自治研究機構. 市区町村における住民参加方策に関する調査研究 part1. [http://www.rilg.or.jp/htdocs/img/004/pdf/h24/h24\\_14\\_01.pdf](http://www.rilg.or.jp/htdocs/img/004/pdf/h24/h24_14_01.pdf).
- [40] 一般財団法人地方自治研究機構. 市区町村における住民参加方策に関する調査研究 part3. [http://www.rilg.or.jp/htdocs/img/004/pdf/h24/h24\\_14\\_03.pdf](http://www.rilg.or.jp/htdocs/img/004/pdf/h24/h24_14_03.pdf).
- [41] 石田哲也, 関洋平. 市民意見分析のための複数の属性の定式化と検証. ARG 第16回 Web インテリジェンスとインタラクション研究会, pp. 25–30. ARG Web インテリジェンスとインタラクション研究会, 2020.
- [42] 石田哲也, 関洋平, 櫻惇志, 柏野和佳子, 神門典子. 都市を横断した市民意見抽出に関する課題と手法についての検証. 情報処理学会研究報告, 第 2022-IFAT-148/2022-DBS-175 巻, pp. 1–6, 2022. 32 号.

- [43] 石田哲也, 関洋平, 柏野和佳子, 神門典子. 複数の属性の関連性に着目したソーシャルメディアからの市民意見抽出. 第13回データ工学と情報マネジメントに関するフォーラム, DEIM2021-D13-5. 日本データベース学会, 2021.
- [44] 青木保彦, 三田昌弘, 安藤紫. シックスシグマ. ダイヤモンド社, 1998.
- [45] 大塚裕子, 乾孝司, 奥村学. 意見分析エンジン - 計算言語学と社会学の接点 -. コロナ社, 2007.
- [46] 関洋平. コミュニティQAにおける意見分析のためのアノテーションに関する一検討. 自然言語処理, Vol. 21, No. 2, pp. 271–299, 2014.
- [47] 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-B4-3. 日本データベース学会, 2016.

# 発表論文

## 査読付学術雑誌論文

- (1) Tetsuya Ishida, Yohei Seki, Wakako Kashino, and Noriko Kando. Extracting Citizen Feedback from Social Media by Appraisal Opinion Type Viewpoint. *Journal of Natural Language Processing*. 2022, Vol. 29, No. 2, p. 416-442.
- (2) (投稿中) 石田哲也, 関洋平, 櫻惇志, 柏野和佳子, 神門典子. 都市を横断した市民意見抽出の評価. *自然言語処理*. 2023, vol. 30, No. 2.

## 国内会議論文

- (1) 石田哲也, 関洋平. 市民意見分析のための複数の属性の定式化と検証. ARG 第16回 Web インテリジェンスとインタラクション研究会, pp. 25-30. ARG Web インテリジェンスとインタラクション研究会, 2020. (優秀研究賞)
- (2) 石田哲也, 関洋平, 柏野和佳子, 神門典子. 複数の属性の関連性に着目したソーシャルメディアからの市民意見抽出. 第13回データ工学と情報マネジメントに関するフォーラム, DEIM2021-D13-5. 日本データベース学会, 2021.
- (3) 石田哲也, 関洋平, 櫻惇志, 柏野和佳子, 神門典子. 都市を横断した市民意見抽出に関する課題と手法についての検証. 情報処理学会研究報告, 第 2022-IFAT-148/2022-DBS-175 巻, pp. 1-6, 2022. 32 号. (学生奨励賞)