

自動運転を見据えた社会実験アンケートにおける自由記述データのテキスト分析

牧野 玲子¹・志賀 孝広²・羽藤 英二³

¹正会員 株式会社豊田中央研究所 社会システム研究領域 (〒480-1192 愛知県長久手町横道 41-1)

E-mail: e4398@mosk.tytlabs.co.jp

²非会員 株式会社豊田中央研究所 社会システム研究領域 (〒480-1192 愛知県長久手町横道 41-1)

E-mail: t-shiga@mosk.tytlabs.co.jp

³正会員 東京大学工学系研究科 社会基盤学専攻 教授 (〒113-8656 東京都文京区本郷 7-3-1)

E-mail: hato@bin.t.u-tokyo.ac.jp

地方都市において過疎化や社会インフラ整備等の課題が顕在化し、地域輸送サービスに対するニーズが多様化・高度化する中、地域特性に応じたモビリティサービスが求められている。本研究では、自動運転を見据えた社会実験（岐阜県中津川市付知町）の後に実施されたアンケート調査から、自由記述に着目し、そのデータをテキスト分析することで住民の意見や期待等を定量的・視覚的に示した。また、自由記述における地域特有の意味付けや文脈を確認することで、少数意見ではあるものの地域にとって重要な課題を抽出することができた。

Key Words: text mining, open-ended questionnaire, resident survey, autonomous driving

1. はじめに

日本では急速な少子化と高齢化が同時に進行し、人口減少社会というかつてない大きな社会経済構造の変動期を迎えている。地方都市においても過疎化や社会インフラ整備等の課題が顕在化し、地域輸送サービスに対するニーズも多様化・高度化する中、地域特性に応じたモビリティサービスが求められている¹⁾。地方都市における将来のモビリティサービスを検討するためには、その地域の特性を理解し、地域特有の課題やニーズを把握することが必要である²⁾。

近年、ChatGPT³⁾に代表されるような大規模言語モデルが急速に発達しており、獲得された知識に基づき多様なタスクの処理が自動化・高度化されつつある。この種の技術では、非常に大規模なテキストデータを用いた事前学習に基づき普遍性のある一般解が提示されるが⁴⁾、タスクの種類によってはそのような大規模データを収集できないことがある。地域課題解決やニーズ探索に対しては、地域に共有されている体験や知識からその課題に特化された特殊解を見出していくことから、依然として地域特有の知識も必要となる。上記の、地域特性に応じたモビリティサービスの設計においては、そのようなアプローチが重要となる。

また、地域理解や住民の意見、意識等の把握には様々な方法があり、アンケート調査が一般的ではあるが、通常のアンケート調査票から多様な意見を抽出するには限界がある。同時に、調査票の設計方針以外の意見を取り漏らしてしまう可能性も大きい。それらの懸念点に対して自由記入欄を設け、回答者に意見を書いてもらうことで、調査票が想定していない住民の意見や意識を共有することが可能となる。また、自由記述内容のテキスト分析により、記述者の意見や感想などを客観的、網羅的に抽出する定量的分析の有用性も示されている⁵⁾。

本研究では、住民アンケート調査における自由記述データに着目し、テキストマイニングを用いた定量的な分析を試みた。さらに地域特有の課題や少数の拾いにくいニーズを把握するため、自由記述原文にさかのぼり、文章全体としての意味付けや文脈の情報を確認しその特徴を捉えた。これらをもとに、地方都市における住民ニーズや地域特性に応じたモビリティサービスの形を提案することを最終目標としている。

2. 社会実験概要

社会実験は、岐阜県中津川市付知町で行われた。付知町は長野県に隣接した中津川市北部に位置し、人口

5,473 人（2020 年 4 月 1 日現在）が居住している⁷⁾。中津川市では、自動運転技術が導入された将来を見据え、さまざまな利用形態を想定した車両を付知地区内で走らせる実験を行い、将来の地方都市におけるモビリティサービスとはどのような形が考えられるのか検討を行っている。特に、地域の特性や住民の生活など様々な場面が想定される中、「付知地区内で人の移動や荷物の輸送を自動運転車両が担うとした時、どのような場面で活用することができるか」に着目し、社会実験を行った⁸⁾。

実験では、地区のコミュニティバスとして使用されているワゴン車を自動運転車と想定し（実験時は有人運転）、2022 年 12 月 9 日～15 日の 7 日間、付知地区内にて 3 つの実験を行った（図-1）。実験 1 は、「地区内人・物どこでも輸送」実験といい、「付知地区内において、人と荷物のどちらでも、どこでも輸送する実験」である。住民は出発地と目的地をアプリや電話で予約し、地区内を移動することができる。また、荷物だけの輸送も可能である。実験 2 は、「巡回便」を定期的に走らせる実験である。地区内に巡回ルートを設定し、自動運転を想定した車を巡回させる。実験 3 は、自動運転を想定した車に弁当や食べ物を載せ、地区内の 3 か所で移動販売する実験である（図-2）⁸⁾。これら社会実験期間が終了した後に、地域住民へのアンケート調査を行った。



図-1 実験で使用した自動運転車を想定したワゴン車
(筆者撮影)

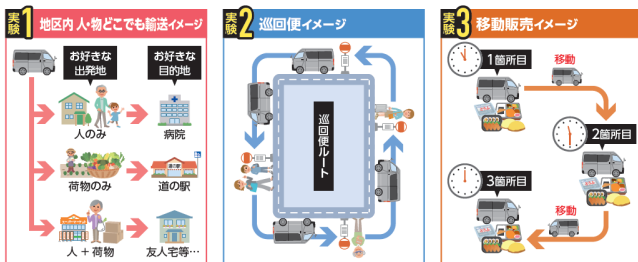


図-2 社会実験の概要（イメージ）
(出典：中津川市次世代交通研究室⁸⁾)

3. 調査概要

(1) 調査対象

社会実験の後、2022 年 1 月時点で付知町に在住している全世帯の 1,650 世帯を対象に、地区ごとの回覧板を通じてアンケート調査を行った。アンケートは、社会実験に参加していない世帯も含めすべての付知町在住世帯を調査対象としたため、アンケート記入の前に社会実験の概要（図-1, 2）や当日の利用状況と利用者の感想紹介等を示し、アンケートへの回答を依頼した。

(2) 調査・分析方法

a) 調査項目

「社会実験に関する住民アンケート調査票」では、世帯調査、社会実験に関する調査、自動運転車両の利用可能性に関する調査、生活満足度・地域愛着度調査の 4 つの項目について質問し調査を行った（表-1）。本稿で分析対象としたものは、「自動運転車両の利用可能性に関する調査」項目にある「問 16 地域の移動・輸送を自動運転車両が担うようになった際、便利になりそうなこと、期待したい暮らしや人生の変化など」の質問に対し、自由記入欄に回答のあった記述内容である。

表-1 調査項目

項目	番号	質問内容
(1) 世帯に関する調査	問1	住所（地区）
	問2	世帯人数
	問3	世帯の方全員の性別、年齢、職業、自動車運転免許有無、通勤・通学先、通勤・通学にかかる交通費
	問4	世帯で保有している自動車の台数
	問5	保有している自動車(保有台数、車種、購入金額、購入時期、今後の予定)
(2) 2021年に実施した社会実験に関する調査	問6	社会実験実施の認知
	問7	実験への参加の有無
	問8	(参加していない方のみ) 参加しなかった理由
	問9	付知町地区に必要なと思うサービス(実験1~3を選択)
	問10	自動運転技術の導入を見据えた“まちづくり”への興味や関心の有無
	問11	自動運転技術の導入を見据えた“まちづくり”で期待すること
(3) 自動運転車両の利用可能性に関する調査	問12	「人の移動」での利用：利用できる時期、移動する人(対象)、目的地、利用時間、利用頻度
	問13	「荷物の輸送」での利用：利用できる時期、輸送する荷物(対象)、目的地、利用時間、利用頻度
	問14	自動運転車両を活用したサービスが導入され、「人の移動」や「荷物の輸送」で利用できるようになった場合の自動車や免許保有への影響
	問15	今回の実験のように、自動運転車両を活用したサービスが地域に導入された場合、利用する金額
	問16	地域の移動・輸送を自動運転車両が担うようになった際、便利になりそうなこと、期待したい暮らしや人生の変化など(自由記述)
(4) 生活満足度、地域愛着度調査	問17	付知町に住むことへの愛着度
	問18	最近の自分の生活の満足度

b) 分析方法

アンケートの自由記入欄に書かれた文章の分析には、KH Coder の Version3 Beta. 03 を使用した⁹⁾。分析の手順は、以下の通りである。はじめに、テキストの欠損値やパソコンで読み込み不可能な表記の削除、誤字脱字の修正を行うことでデータのクリーニングを行った。さらに、「軽トラ」や「スマホ」といった短縮化した表記やカタカナ、ひらがな表記、同義語などを統一し、表記の揺れを置換作業により修正した。分析用データとして整ったところで KH Coder に読み込ませ、形態素分析によりテキストから品詞別に抽出語を抜き出し、データベースとして整える前処理を行った。また、「自動」「運転」のように別々に切り出された語であっても、つなげて「自動運転」とした方が解釈しやすくなる語に関しては、一つの語として強制的に切り出す作業を行った。本稿の分析としては、形態素分析、共起ネットワーク分析、レスポンス分析の3つを行った。

4 結果

(1) 分析対象

中津川市付知町在住の全世帯 1,650 世帯に配布したアンケート調査のうち、714 世帯から回答があり、回答率は 43.8%であった(表-2)。そのうち、自由記述のあった世帯は 225 世帯(13.6%)であり、これを本稿の分析対象とした。

表-2 アンケート調査の配布・回答数

	アンケート配布	回答世帯	自由記述のあった世帯
世帯数	1,650	714	225
全世帯に占める割合	-	43.8%	13.6%

(2) 頻出語の抽出

自由記入欄に書かれた 225 世帯の自由記述を対象に形態素分析した結果、自由記述の総抽出語数は 6,925 語であり、今回、分析に使用した総語数は 3,151 語であった。語の出現回数は表-3 のようであった。自由記述でよく使われた語は、「思う」「車」「買い物」「移動」「免許」「生活」「返納」「利用」「人」「運転」「自動運転」であった。「病院」と「通院」は別の語として出現回数をカウントしているが、「病院・通院」としてカウントし直すと、頻出上位の語となる。アンケートの質問項目で使われている「車」「移動」といった語や、「思う」といった動詞を除くと、「買い物」「免許」「生活」「返納」「人」「病院・通院」といった語が自由記述でよく使われた語であり、頻出語から地域の特徴を捉えることができる。

表-3 よく使われた語の出現回数

出現順位	抽出語	出現回数	出現順位	抽出語	出現回数
1	思う	80	23	行ける	16
2	車	41		時間	
3	買い物	39		自由	
4	移動	38	27	病院	15
5	免許	36		地域	
6	生活	35	30	通学	14
7	返納	34		良い	
9	利用		29	31	外出
10	人	28	32	機会	12
12	自動運転			事故	
14	高齢者	27	35	自分	11
16	便利	22		高齢	
18	期待		19	38	助かる
20	増える	付知町			
20	行く	18	44	減る	9
20	中津川市			現在	
20	車両	17	44	少ない	9
20	必要			町	
20	今	17	44	特に	9
20	子ども			老人	
20	出来る	17	44	バス	9

(3) 自由記述の要約

抽出語の出現頻度と抽出語同士の関連性を要約する目的において、共起ネットワーク分析を行った。共起ネットワークとは、出現パターンの似通った語、すなわち、抽出語間の共起性の強さをネットワーク図で示したもので、円の大きさは語の頻度の多さを示し、円をつなぐ線は関連性の強さを示している。今回の分析では、関連性の強さを示す共起関係の算出には Jaccard 係数¹⁾を使用し、頻出語の最小出現回数を 9 回、描画する共起関係の絞り込みを描画数 60 と設定し分析を行った。その結果、9 つのサブグラフが抽出された(図-3)。それぞれのサブグラフに示された語とその文脈を確認し、下記のように自由記述を要約することができた。

次に、アンケートの自由記述全体を見た時の中心的な語(媒介中心性の高い語)、すなわち、ネットワークのハブとなる語を確認すると、「自動運転」「高齢者」「人」であった(図-4)。これらは、9 つに要約された自由記述をつなぐ語であり自由記述全体から見て重要な語であると言える。

サブグラフ①: 自動運転により買い物、通院・病院、通学等が自由に行けることへの期待

サブグラフ②: 人と会うといった外出機会の創出

サブグラフ③: 付知町や中津川市内の近隣地区への移動

サブグラフ④: 自動運転車両による高齢者の事故の減少

サブグラフ⑤: 現在と老後の両方に期待

サブグラフ⑥: 生活に必要な免許と免許返納への懸念

サブグラフ⑦: 現在、特に問題を感じていないこと

サブグラフ⑧: 自動運転車の利用で便利になること

サブグラフ⑨: 老人の利用やバスの本数が少ないこと

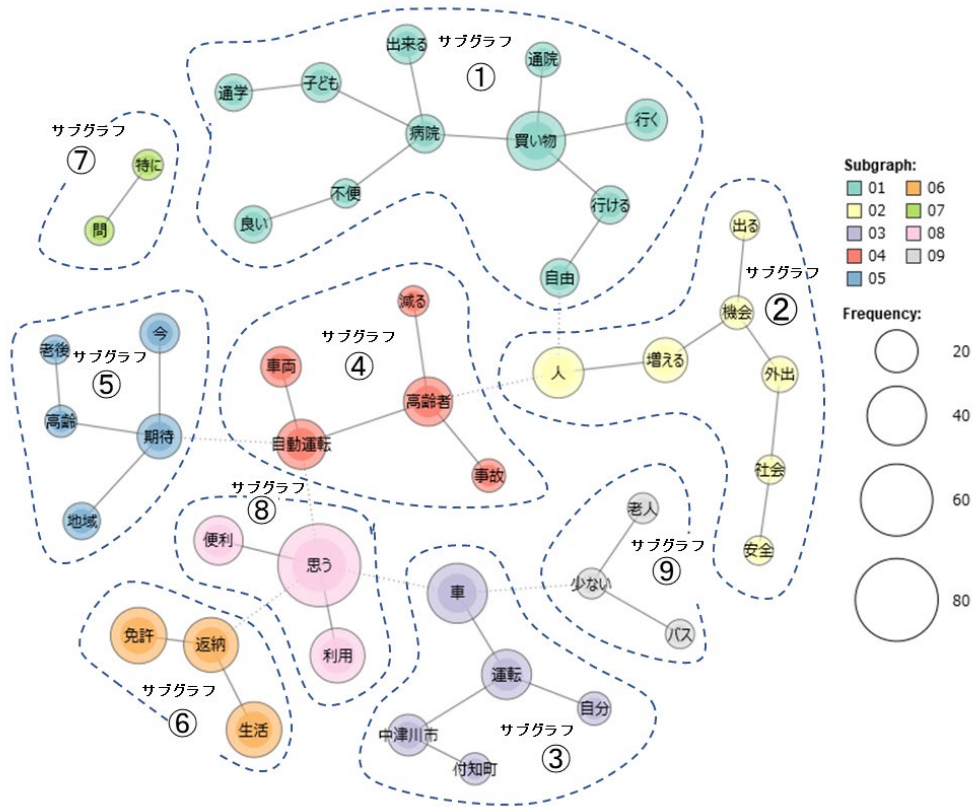


図-3 自由記入欄に記述された語の共起ネットワーク

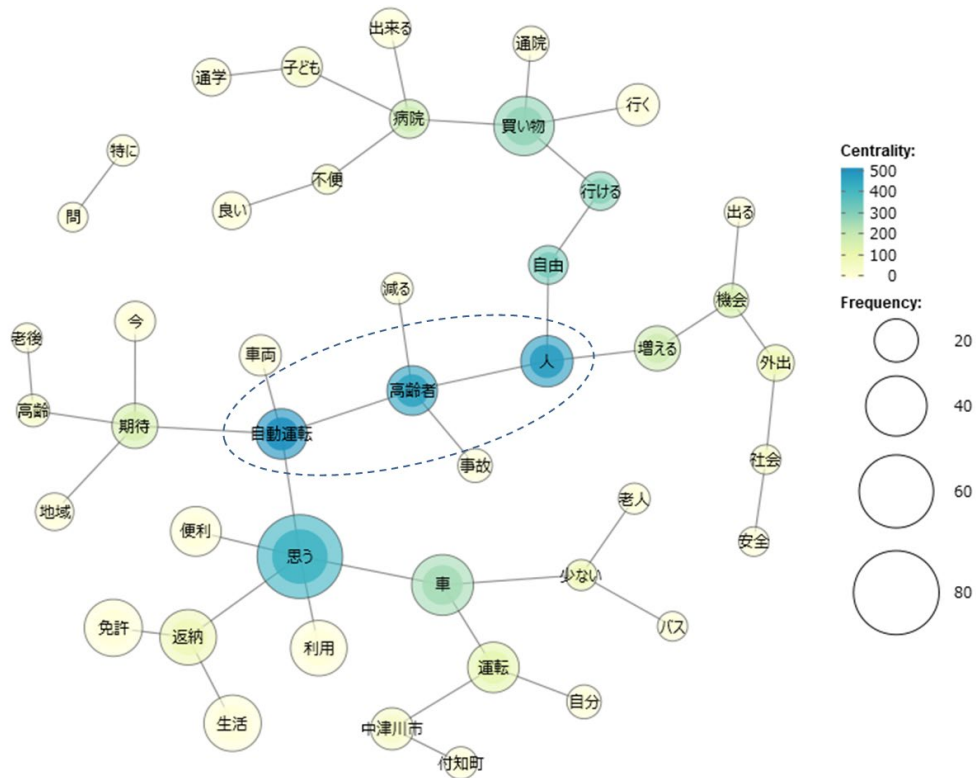


図-4 自由記述全体の中で中心となる語

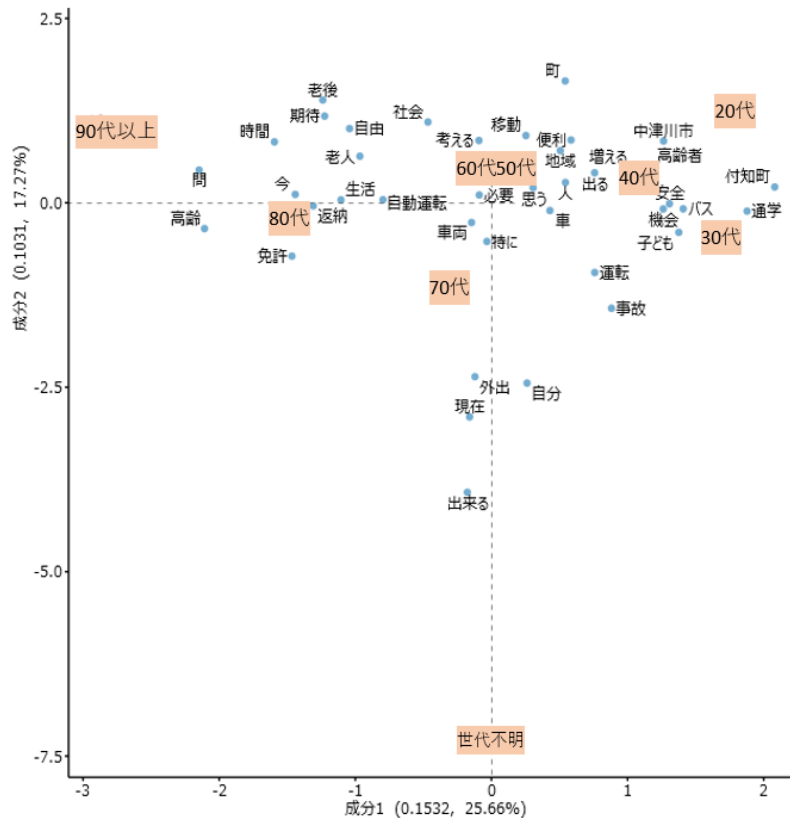


図-5 世代による回答傾向

注) 図中の点(・)は単語を示し、四角の枠内は外部変数(世代)を表す。成分1,2のカッコ内の数値は(固有値, 寄与率)であり、固有値は各軸(成分)が含んでいる情報の大きさを示す指標、寄与率はその軸(成分)によって多次元空間における語の分布の何%が表現されているかを表している。

(4) 属性による回答傾向

属性ごとの回答傾向や特徴を捉えるため、コレスポネンデンス分析を行った。この分析は、語間の関係性を散布図と呼ばれるグラフで視覚的に把握する方法である。互いに関連性の強い語同士は、原点から見て同一方向に布置される性質があり、原点からの距離と方向によって、回答傾向が近いかどうか、特徴的な語は何であるか等を可視化することができる¹²⁾。本稿では、世代、性別ごとの回答傾向について分析した。

a) 世代による回答傾向

回答傾向が似通っていたのは、世代の近い「20代」「30代」「40代」で、原点から右側の同一方向に布置していた(図-5)。20代~40代の近くに記載されている特徴的な語は、「子ども」「高齢者」「通学」「安全」「バス」「利用」といった子どもや高齢者の交通手段や移動に関する語であった。それに加え「付知町」「中津川市」といった地名に関する語も記載されていた。

また、「50代」「60代」はともに原点に近い位置にあり、回答傾向が似通っていると言える。特徴的な語として「地域」「移動」「便利」「必要」とあり、移動の

必要性と便利さを記述していた。「80代」「90代以上」は原点から左側の同一方向に位置し、回答が近い傾向にあると言える。「高齢」「老後」「老人」といった年齢的なことや、「免許」「返納」「今」「時間」「問」といった特徴的な語が記載されていた。

「70代」は、他の世代とは原点からの方向が違っており、回答傾向の近い世代はいなかった。70代の近くには「特に」「現在」「自分」「外出」といった語が布置されており、それら特徴的な語からは、現在、ご自分で外出できている様子が伺える。しかし、前述の80代以上の免許返納や高齢化に対する記述と比較してみると、70代と80代の間には明らかな回答傾向の違いが見られる。

b) 性別による回答傾向

性別による回答傾向を見ると、「男性」はほぼ原点に位置し、「女性」は原点より右上に位置することから、男女に回答傾向の違いが見られた(図-6)。

「男性」の特徴的な語として、「車」「運転」「利用」「便利」「買い物」「通院」があり、買い物や病院への移動や運転の便利さなどを記述していた。日常生活での具体的な移動や移動手段に対して自動運転への期待が高

といったことに対する意見や期待は視覚化できていなかった。形態素解析による語の出現回数は「出会い」3回、「コミュニケーション」と「接する」各2回、「関わり」と「つながり」各1回であり、個々の語で見た場合に共起ネットワーク分析では視覚化されにくい。たとえ地域にとって意味のある文脈に含まれる語であったとしても、定量的なテキスト分析では表出させることが困難であることがわかった。

今回、自由記述の文脈を確認することで「人とのつながり」に対する期待を確認することができた。その他にも重要な地域特有の意見が拾い上げられていない可能性はあるものの、地域を理解するために文脈を確認することの重要性は示すことができたと考える。

この地域の特性を鑑みると、ほとんどの世帯が自動車を保有し、主な移動手段は自動車である。『付知町において、「人がつながる」とはどういうことか』という点が、今後の重要な論点の一つであると考えられる。

表-4 「人」が含まれる記述文

	記述文
1	人の移動はどこでも自由な時間に好きな場所へ行ける。
2	高齢者の自動車事故の減少、人とのコミュニケーションの増加
3	人との関わりが活力になることもある。
4	日常的に必要な活動が可能になり、人と出会う機会も増えると思う。
5	田舎でも同じことが実現できれば利用する人は増える。
6	現在自動車を持っていない人が受けている外出の制限や移動の労力を解消
7	買い物に困っている人がいなくなればいい。
8	足の悪い人がスムーズに乗り降りしやすいこと
9	移住してくれる人が増える事が期待できます。
10	付知町の人は解答する人がかなり少なくなってしまう事が予想
11	仕方なく運転している人も居ると思う。
12	免許のない人が離れた地区の人と出会う。
13	人とのつながり、免許がなくても移動手段の確保
14	若い人が増え、老人もイキイキ暮らせる町になると良い。
15	自動運転車両は若い人向けではないのですか。
16	人それぞれ、いろいろな生活形態があるのですべてに合わせるのは無理
17	人との接する機会があり、楽しみが増える。
18	友人宅や教室まで本当に予定に合わせて人が移動できる様になれば豊かに
19	人に頼らず行きたい所に自由に行ける。
20	自由に人と関われる事が望ましい。
21	外に出る機会が増え、人とのコミュニケーション増える。
22	人の移動、荷物の輸送を業としている方もあり、むずかしい。
23	車を運転できない人、できなくなってしまった人や老人等には便利になる。
24	介助する人も同乗し、通院・リハビリ・買い物ができる。
25	足の弱い人も気楽に出かけるように。

5 結論

本研究では、将来のモビリティサービスの提案に向けた社会実験（中津川市付知地区）の後に実施されたアンケート調査の結果を分析した。アンケートの自由記入欄から得られたテキストデータに着目し、形態素解析の適用や共起ネットワーク分析・コレスポネンシ分析により、語の関係性を定量的・視覚的に分析・把握した。さ

らに、自由記述原文にさかのぼり、文章全体としての意味付けや文脈の情報を確認することで、少数意見ではあるものの地域にとって重要な課題を抽出することができた。

地域には歴史や文化があり、大規模テキストデータを対象とした分析だけでは捉えることができない様々な地域課題やニーズが潜んでいる。それらを取りこぼさず、拾い上げていくよう努め、地域特性に応じた将来のモビリティサービスの形を検討していく。

謝辞：本研究は、中津川市と東京大学との共同研究によるアンケート調査データに基づき、東京大学 情報理工学系研究科 次世代知能科学研究センターと豊田中央研究所による社会連携研究部門「モビリティ知能社会デザイン」のプロジェクトとして実施した。

参考文献

- 1) 国土交通省：都市と地方の新たなモビリティサービス懇談会中間とりまとめ、
<https://www.mlit.go.jp/common/001279833.pdf> (2023. 2. 14 閲覧確認)
- 2) 国土交通省：地域の自立的発展のためのモビリティ確保に向けた検討の手引き、
<https://www.mlit.go.jp/kisha/kisha08/15/150327/08.pdf> (2023. 2. 14 閲覧確認)
- 3) OpenAI, <https://chat.openai.com/chat> (2023. 2. 14 閲覧確認)
- 4) 松尾豊：深層学習と人工知能, 認知科学, Vol. 28, No. 2, pp. 299-307, 2021.
- 5) 森田哲夫, 入澤覚, 長塩彩夏, 野村和弘, 塚田伸也, 大塚裕子, 杉田浩：自由記述データを用いたテキストマイニングによる都市のイメージ分析, 土木学会論文集D3(土木計画学), Vol. 68, No. 5, pp. 315-323, 2012.
- 6) 大塚裕子, 森田哲夫, 吉田朗, 小島浩, 塚田伸也：テキストマイニングによる都市・景観イメージ分析 一水緑環境に着目して一, 土木計画学研究・講演集, Vol. 41, CD-ROM, 2010.
- 7) 中津川市：付知地区の概要,
<https://www.city.nakatsugawa.lg.jp/soshikikarasagasu/tsukechi/local/1217.html> (2023. 2. 14 閲覧確認)
- 8) 中津川市次世代交通研究室：自動運転技術の社会実装を見据えた人・物の輸送社会実験,
<https://www.city.nakatsugawa.lg.jp/soshikikarasagasu/jisedai/keikaku/jisedai/12627.html> (2023. 2. 14 閲覧確認)
- 9) 樋口耕一：テキスト型データの計量的分析-2 つのアプローチの峻別と統合-, 理論と方法, No. 19, No. 1, pp. 101-115, 2004.

- 10) KH Coder, <http://khc.sourceforge.net/> (2023. 2. 14 閲覧確認) 時布置の仕組みと読み取り方を中心にー, コンピュータ&エデュケーション, Vol. 47, pp. 18-24, 2019.
- 11) 佐々木靖弘, 佐藤理史, 宇都呂武仁: 関連用語収集問題とその解法, 自然言語処理, Vol. 13, No. 3, pp. 150-175, 2006.
- 12) 樋口耕一: 計量テキスト分析における対応分析の活用ー同 **(Received?? , 2023)**

TEXT MINING ANALYSIS OF OPEN-ENDED QUESTIONNAIRE AFTER SOCIAL EXPERIMENT
OF ASSUMING AUTONOMOUS DRIVING

Reiko MAKINO, Takahiro SHIGA and Eiji HATO