

2024年度 統計データ分析コンペティション

## 審査員奨励賞 [大学生・一般の部]

XAI を用いた介護業界における  
地域別の従業者数の就業要因に関する一考察

宮内 弘太

(一般財団法人計量計画研究所研究本部)

# XAI を用いた介護業界における地域別の従業者数の就業要因に関する一考察

宮内弘太\*1

\*1: 一般財団法人 計量計画研究所 研究本部

## 1. 背景と目的

近年、わが国では医療技術の目覚ましい発展による死亡率の低下や、少子化の進行による若年人口の減少などにより超高齢化社会が深刻化している。超高齢化社会とは、65歳以上の割合が全体の21%を超える場合を指し、2025年には全体の30%に達すると言われている<sup>(1)</sup>。高齢化に伴う働き手となる労働人口の減少は介護業界をはじめ、社会全体に大きな影響を及ぼしている。

介護業界では昨今、介護サービスを担う人材が不足し、介護を必要とする高齢者が適切なサービスが受けられない介護難民や、高齢者が高齢者を介護する老々介護が顕在化している。厚生労働省の報告によると<sup>(2)</sup>、介護保険を利用する被保険者数は年々増加しており、近年そのペースは急激に拡大している。介護業界に従事する従業者数も年々増加しているが、被保険者数の増加度合いに追いついていないのが現状である。我が国では、この解決策として外国人介護人材の受け入れ施策を実施<sup>(3)</sup>しており、2025年からは、これまで外国人労働者は施設系介護業務しか従事できなかったのが、訪問介護業務への従事も可能になるように検討されている<sup>(4)</sup>。

積極的な人材確保や育成の検討はされている一方、介護業界における地域の従業者数がどのような要因で決定されているのかを定量的に検討した事例は未だないことを踏まえ、本研究では介護業界の地域別の従業者数に着目し、従業者数を予測するモデルの構築を試みる。モデルを構築する際は、今日様々な分野で活用されているXAI (Explainable AI)<sup>(5)</sup>を適用し、構築したモデルの定量的な解釈を試みる。

モデルの解釈に着目する理由は、介護業界の人材不足等の現状を踏まえると、今後の政策を検討するにあたり、地域の従業者数の大小に影響する要因を把握することが重要だからと考えたからである。そこで本研究では、多様なデータが揃えられているSSDSE-A (各種のデータが市区町村別に分かれた教育用標準データセット)を用いることで、その技術的問題の解決を試みる。

なお構築したモデルは、今後想定される高齢者人口の増加度合いや政策状況等をインプットすることで、将来必要となる地域別の介護業界の従業者数を推計することが可能で、今後起こりえる人材不足への対策検討に寄与できる。

## 2. 分析の流れと適用手法の概要

### 2.1 分析の流れ

分析の流れを以下の図1に示す。SSDSE-Aには市区町村別の介護業界の従業者数のデータがないため、本研究ではまずSSDSE-Aと、令和3年度の経済センサス活動調査の地域別の社会保険・社会福祉・介護事業の従業人口データ<sup>(6)</sup> (以後、本研究ではこのデータを介護業界の従業者数と定義)を市区町村コードと結合し、分析用のデータセットを作成する。

次に、結合したデータを基に基礎分析を行う。本研究では二つの観点から基礎分析を行う。一つは地域別の従業者数について整理を行い、現在の状況を考察する。もう一方は、従業者数を予測するモデルを構築する際に組み込むべき特徴量の検討を行う。

次に、基礎分析の結果を踏まえ、従業者数を予測するモデルの構築とモデルの解釈を行う。モデルに適用す

る手法の詳細は次項で述べるが、本研究では XGBoost (Extreme Gradient Boosting) (7)を用いた回帰モデルに SHAP(8)を適用することで、構築したモデルの精度検証の結果と解釈を行う。

モデルの構築後は、今後のわが国で起こり得る高齢者人口の増減や、政策状況を踏まえた特徴量をモデルにインプットすることで、将来必要となる地域別の従業者数の推計を行う。そして、基礎分析で整理した現行の地域別の従業者数と推計結果を踏まえ、今後の介護業界における人材不足問題の課題解決に向けた検討を行う。

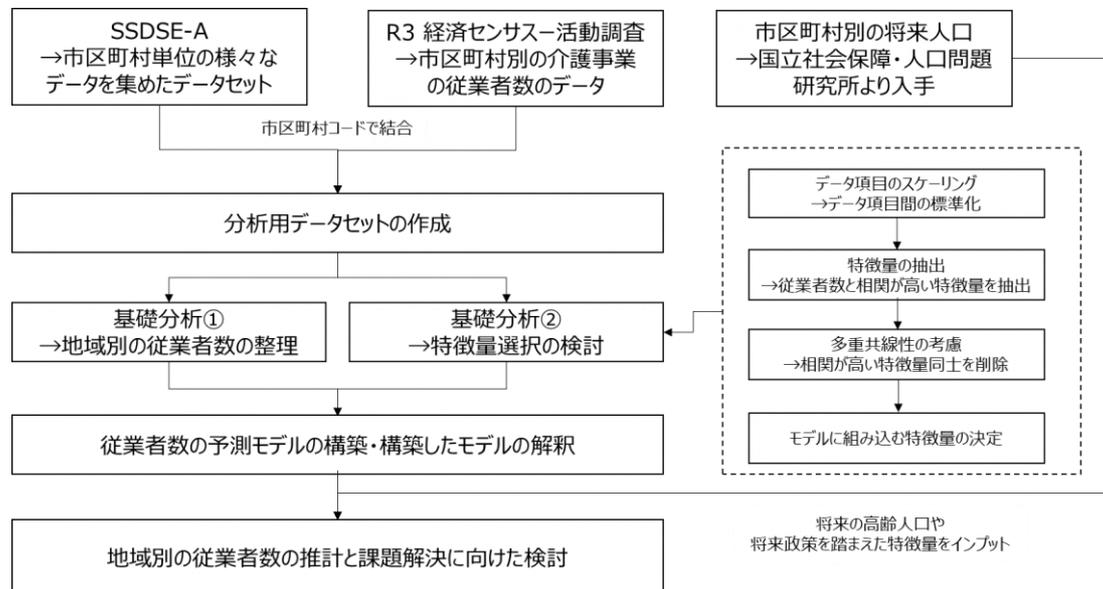


図1 本分析のフロー

## 2.2 適用手法の概要

本研究では XGBoost を用いた回帰モデルに SHAP を適用することで、従来の (AI・統計による) 予測モデルに比べて高精度かつブラックボックスの透明化を図った予測モデルの構築を試みる。

XGBoost は、決定木による勾配ブースティングアルゴリズムを採用したモデルである。決定木とはデータ項目をツリー状に整理し大小を比較していくアルゴリズムである。勾配ブースティングは、複数の弱学習器 (先述の決定木) を逐次的に学習させることで、それぞれのモデルの誤差を次のモデルに補正する学習手法である。すなわち、ステップが更新されるたびに前のモデルの残差を最小化するように新しいモデルが追加されるため、より高い精度で予測ができる手法である。

そして SHAP は機械学習によるモデルで予測された結果を解釈することに秀でた手法の一つである。結果の解釈に秀でている理由は、ゲーム理論で使われる Shapley 値を用いて、各特徴量が予測に与える影響を定量的に評価できるからである。Shapley 値は、共同作業の成果に対して、貢献した価値を公平に分配するための概念で、これを予測結果に応用することで、各特徴量が予測値にどれだけ貢献しているかを正負の観点も考慮して算出できる。本来、機械学習を用いた手法はその解法の性質上、ブラックボックスに陥りやすく、結果の解釈が難しいとされていたが、SHAP などの XAI の出現により、以前と比べて結果の解釈も容易になっている。

本研究では、上記の二つの手法を組み合わせることで、地域別の介護業界の従業者数を高精度に予測することに加え、設定した特徴量が従業者数の増減のどちらに起因するかを把握する。

## 3. 分析用データの概要

### 3.1 地域別における従業者の介護者数の整理

地域別の介護分野の従業者数を整理した結果を図 2 に示す。なお、ここでは地域間の従業者数の違いを分

かりやすくするために、対象地域の75歳以上の高齢者人口を従業者数で除する事で、従業者1人当たりが最大で介護すると思われる高齢者数を整理する。なお、75歳以上の高齢者人口に着目した理由は、内閣府の報告書<sup>9)</sup>によると75歳以上の要介護認定の割合が特に増えており、一般的に介護が必要とし始める年齢と思われるからである。

図中の左上の円グラフは、従業者1人あたりが介護する高齢者数を階級別（2人未満～8超の5区分）に整理した結果である。最も多い階級が従業者1人に対し、高齢者2～4人を介護する地域が約半数を占めている。なお、地域によっては8人以上の高齢者を1人の介護従業者が介護しなければならない可能性も存在する。

また中央の図は、階級値の結果を地域別に着色した結果で、地方によってその傾向が異なる。特に、北海道地方では他の地方に比べ地域間のばらつきが大きい。例えば、芦別市では従業者1人あたりの介護人数が8人超であるのに対し、芦別市に隣接する南富良野市は2人未満であることから、隣接地域でも従業者1人当たりの負担が大きく異なる。また、東日本と西日本でも違いは生じており、東日本では黄色い着色がされた地域（4～6人）が多い一方、西日本では緑色の着色がされた地域（2～4人）が多い。

以上を踏まえると、介護従業者1人当たりが介護する高齢者は4～6人程度の地域が多い一方で、隣接する地域間でも人数に偏りが生じていることから、地方別や都道府県単位で従業者の就業負担を見るのではなく、市区町村別などの細かい解像度で見ると思われる。その上で、介護従業者の負担が大きい地域に対し、優先的な介護従業者の人材支援や人材育成などが必要と思われる。

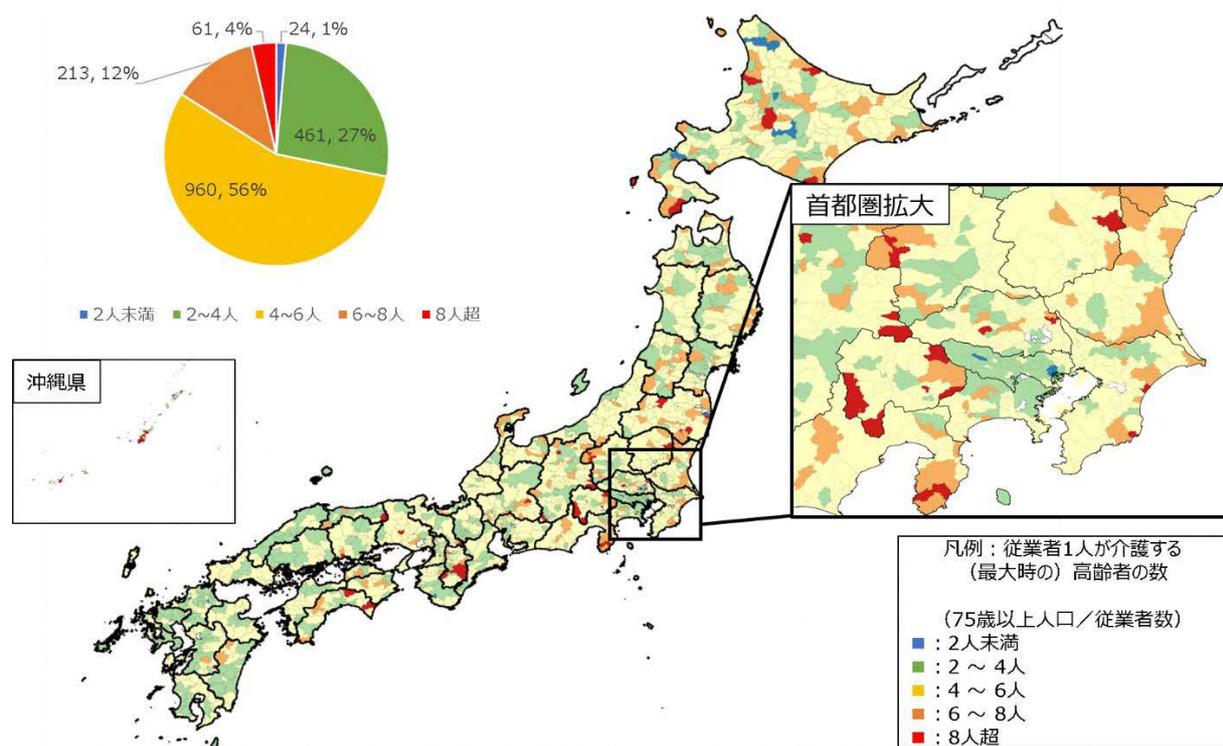


図2 地域別の従業者数に対する75歳以上人口の割合

### 3.2 特徴量選択の検討

SSDSE-Aで用意されているデータセットは、約100種類のデータ項目が含まれており、全てのデータ項目をモデル内に組み込むことは困難である。また、データ項目によっては相関が高い特徴量同士をモデルに組み込むことで、回帰の結果を歪めてしまう多重共線性に陥る可能性があるため、特徴量の選択は必要不可欠である。なお、SSDSE-Aに含まれるデータ項目はHP内<sup>(10)</sup>に記載されているため、説明を割愛する。

特徴量選択の検討は図1の点線内に示すフローで実施する。まず、各項目のデータは地域の規模や特徴量間の大小を考慮するために地域の可住地面積で除し、1m<sup>2</sup>当たりのデータとして算出し、その値を基に標準化（平均0、標準偏差1に従う分布）を行う。なお可住地面積を採用した理由は、可住地面積はその地域において人が住み得る地域面積のことを指し、人口や地域面積で除するよりも地域の現状を反映できると考えたからである。次に、フィルターメソッドによる相関係数を用いた方法で、従業者数と相関が高い特徴量を抽出する。ここでは抽出の閾値を0.5に設定し、0.5以上となった特徴量のみを抽出する。最後に、フィルターメソッドで抽出された特徴量同士の相関係数を算出し、高い変数同士の組み合わせは一方を除外する。ここでは相関係数が±0.8以上の特徴量の組み合わせを削除する。

最終的に選択された特徴量を表1に示す。本研究では11の特徴量が選択され、その傾向に着目すると、地域の人口、財政、施設を示すデータと従業者数には相関があり、かつ特徴量間での相関はない。なお、参考図に特徴量間の相関関係を示した図を示す。図中の数字は表1の特徴量と対応している。

表1 選択された特徴量

人口関連	財政関連	施設関連
①75歳以上人口	⑥土木費（市町村財政）	⑨公民館数
②従業者数（生活関連サービス業、娯楽業）	⑦民生費（市町村財政）	⑩図書館数
③従業者数（金融業、保険業）	⑧教育費（市町村財政）	⑪義務教育学校数
④従業者数（不動産業、物品賃貸業）		
⑤一般行政部門職員数		

## 4. 分析結果

### 4.1 従業者数の予測モデルの精度検証

まず、従業者数の予測モデルの構築方法について述べる。ターゲット変数は従業者数の実数値、特徴量は3.2で選択されたデータ項目（可住地面積1m<sup>2</sup>当たりで算出し、さらにデータ項目ごとに標準化）を設定する。また、モデルの性能評価は過学習による汎化性能の低下を防ぐために、交差検証法（k=5）による方法で実施する。なお、XGBoostはハイパーパラメータが必要で表2に示す条件を設定する。

交差検証法による精度検証の評価指標は、kを1から5まで変化させた時のR<sup>2</sup>（coefficient of determination）の平均値とする。R<sup>2</sup>は回帰モデルの性能を評価するための指標である。0から1の間で値が決まり、1に近いほどモデルの説明力が高い。なお、一般的に0.5以上であると予測モデルとして良好とされている。精度検証の結果を表3に示す。表中の平均値を見るとR<sup>2</sup>は0.56であることから、本研究で構築した従業者数の予測モデルとして概ね妥当である。

表2 XGBoostの主なハイパーパラメータの設定条件

項目	説明	設定値
n_estimators	ブースティングの数	100
learning_rate	各ブースティングステップでの学習率	0.3
max_depth	各決定木の最大深度	6
subsample	各ブースティングで使用するデータの割合	1
objective	使用する損失関数	“squareerror”

※その他の設定条件については本稿では割愛

表3 従業者の予測モデルの精度検証の結果

交差検証 (k)	K=1	K=2	K=3	K=4	K=5	(5回の) 平均値
R <sup>2</sup>	0.420	0.655	0.611	0.408	0.683	0.555

#### 4.2 構築したモデルの特徴量の解釈

前項で精度検証した結果を基に特徴量の解釈を行った。なお、SHAP による特徴量の解釈は局所的な解釈と、大局的な解釈の二通りがある。局所的な解釈では、特定のレコード（本研究では特定の地域）に対する特徴量の影響を評価することが可能である。一方、大局的な解釈では、データ全体に対する解釈を指し、従業者数を予測するモデルを構築した際の特徴量の影響を評価することが可能である。

本研究では、大局的な観点からの特徴量の解釈に着目する。図3に構築したモデルの特徴量を解釈した結果を示す。図の見方について述べる。まず図中の点は個々のサンプル（本研究では地域単位）を示している。

縦軸は予測全体への重要度を示しており、重要度が大きい特徴量から順に並んでいる。したがって本モデルの場合、「②従業者数（金融業、保険業）」、「⑤一般行政部門職員数」、「⑥土木費（市町村財政）」、「⑩図書館数」が設定した特徴量の中で、特に影響力を与えている。

横軸は各特徴量の大小が、従業者数の予測の増減のどちらに寄与しているかを示す。赤色のプロットが多い場合は、そのデータの値が大きいと増減に影響があることを示し、青色のプロットが多い場合は、そのデータの値が小さいと増減の影響があることを示す。横軸の SHAP 値が右側に伸びている場合は、その特徴量は従業者数の増加要因として寄与し、反対に SHAP 値が左側に伸びている場合は、従業者数の減少に寄与する。

したがって、特徴量の重要度が大きい上位4つに着目すると、「②従業者数（金融業、保険業）」と「⑥土木費（市町村財政）」は、各データの値が大きくなると、従業者数の増加に寄与し、「⑤一般行政部門職員数」と「⑩図書館数」は、従業者数の予測の減少に寄与する。

なお、介護分野の従業者数の増減に直結すると思われる「①75歳以上人口」は、特徴量の重要度が特段高い度合を示していないことから、必ずしも従業者数の増減に起因しない。一方で、他の業種の従業員数や、地域の財政状況などの地域の社会基盤や活性状況を示す特徴量の方が、従業者数を予測する上で重要である。

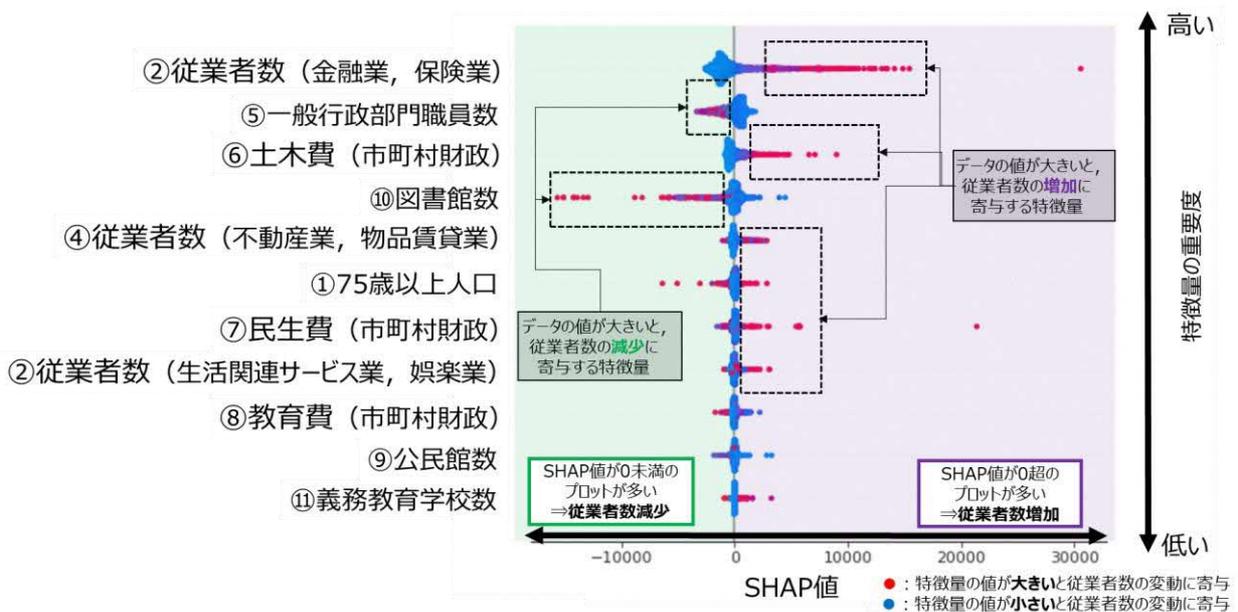


図3 SHAP による特徴量の解釈

## 5. 推計結果を踏まえた課題解決に向けた検討

### 5.1 将来の地域別の従業者数の推計

モデルの精度検証を踏まえ、全地域のデータセットを用いて予測モデルを構築し、将来（本研究では 2050 年を想定）のわが国における高齢者人口の変化等を踏まえた従業者数を推計する。なお、本研究では厳密な根拠に基づいた推計ではなく、一参考例としての位置づけである。

設定条件を述べる。まず、「①75 歳以上人口」は今後の更なる高齢化を反映するため、社会人口問題研究所で公表されている 2050 年時の 75 歳以上高齢人口<sup>(1)</sup>を設定する。「②従業者数（生活関連サービス業、娯楽業）」、「③従業者数（金融業、保険業）」、「④従業者数（不動産業、物品賃貸業）」、「⑤一般行政部門職員数」の従業者数は、雇用年齢の拡大等を想定し、現行の 1.25 倍とする。「⑥土木費（市町村財政）」、「⑦民生費（市町村財政）」、「⑧教育費（市町村財政）」は、国からの補助金の増加等を想定し、現行の 1.25 倍とする。「⑨公民館数」、「⑩図書館数」は現行のままとする。最後に、「⑪義務教育学校数」は少子化による影響を踏まえ、現行の 0.75 倍とする。したがって今回の推計では、全ての地域において同一の変化量である。

推計結果を図 5 に示す。現況の結果と同様に、従業者 1 人あたりが介護する 75 歳以上の高齢者数を地域別に整理している。本研究で設定した条件で予測を行うと、現況時に比べ多くの地域で、赤色の着色（従業者 1 人あたりが 8 人超の高齢者を介護）となる。

また図 6 に現況と将来の従業者数の散布図を示す。図中の緑色の範囲に含まれる地域は、将来の従業者数が現況に比べて増加する地域、オレンジ色の範囲に含まれる地域は、現況に比べて減少する地域である。結果として、緑色の範囲に含まれる地域が多いことから、本研究で想定した条件で推計すると、将来の従業者数は多くの地域で増加する。なお、現況と将来の差分が大きい地域を旗揚げすると、特に政令指定都市において増加が顕著である。

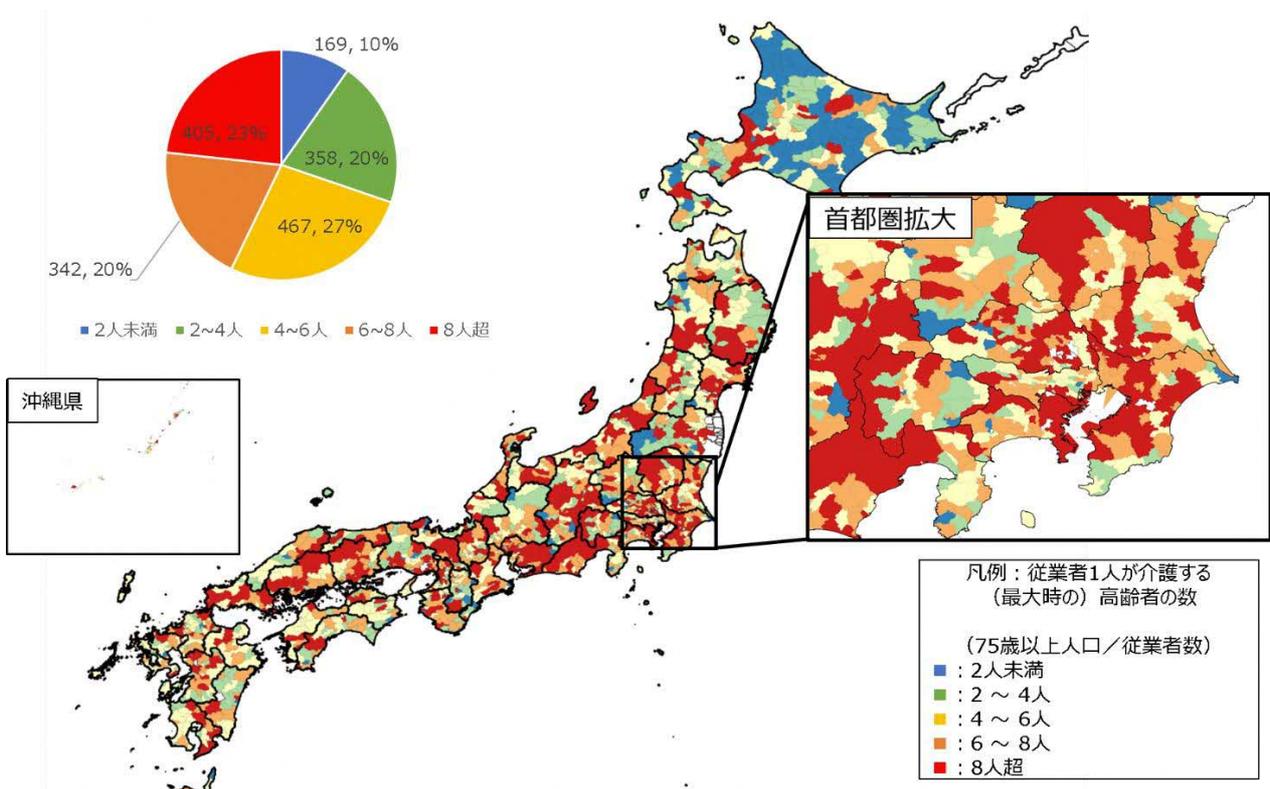
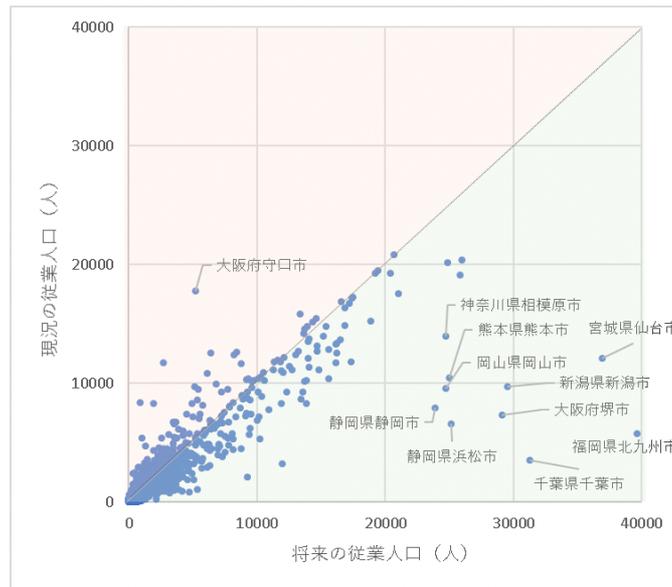


図 5 予測モデルによる地域別の従業者数に対する 75 歳以上人口の割合



※現況と将来の従業人口の乖離が大きい地域を旗揚げ

図6 地域別の現況と将来の従業人口の散布図

## 5.2 今後の課題解決に向けた検討

今回設定した条件では、多くの地域において従業者1人あたりが介護する高齢者が8人超であることから、従業者の負担は今よりも大きくなることが示唆された。ただ一方で、予測モデルで推計した従業者数を見ると、現況の従業者数よりも増加している地域が多い。これらを踏まえると、深刻な高齢化に従業者数が追いついていないことを示しており、従業者の確保は引き続き喫緊の課題であることが示唆された。

本モデルの構築を通して明らかになった、従業者の確保を促進するための要因の一つとして、介護が必要される75歳以上の高齢者数よりも、その地域の社会基盤や活性状況の発展に関する要因の方が大きいことから、我が国が今後行っていく対策方針の一つとして、介護分野の直接的な補助や整備だけでなく、地域の社会基盤や活性化を踏まえた間接的な補助や整備も併せて行っていくことが重要であることが確認された。

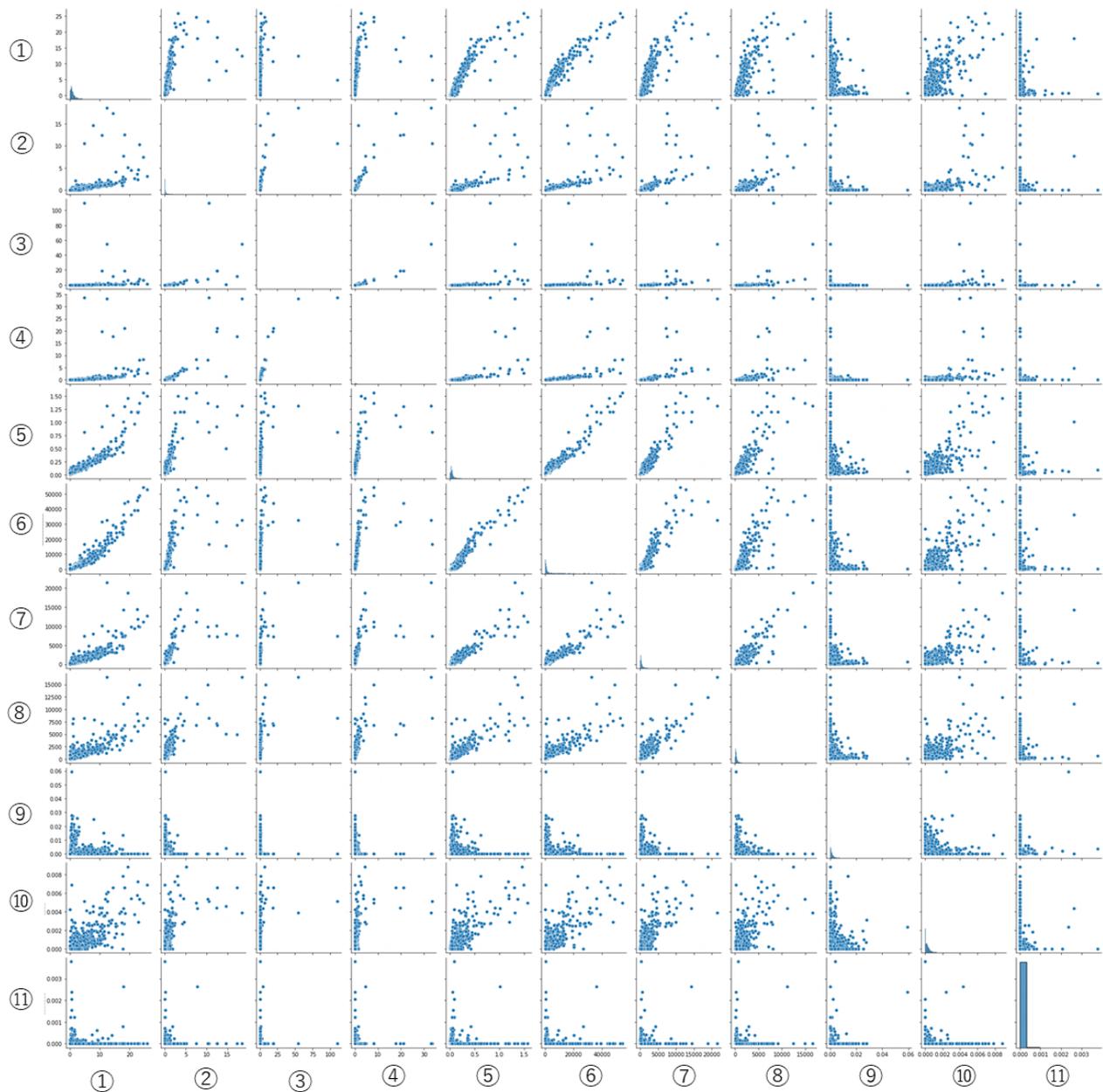
なお、本研究で設定した仮想条件により地域ごとの現実的な数値をインプットすることで、それぞれの地域において必要となる従業者数の推計が可能である。

## 参考文献

- (1) 総務省：総務省統計局統計トピックス No. 132, 1. 高齢者の人口,  
<https://www.stat.go.jp/data/topics/topi1321.html>
- (2) 厚生労働省：介護分野をめぐる状況について,  
<https://www.mhlw.go.jp/content/12300000/000608284.pdf>
- (3) 厚生労働省：外国人介護人材の受入れの仕組み,  
<https://www.mhlw.go.jp/content/12000000/000994004.pdf>
- (4) 厚生労働省：開国人介護人材の業務の在り方に関する検討会, 第6回資料, 2024/3/22.  
<https://www.mhlw.go.jp/content/12000000/001231480.pdf>
- (5) 塚本 満朗, 高木 朗義：説明可能な機械学習モデルを用いた豪雨時における住民避難選択行動の要因分析, 土木計画学研究・論文集, Vol. 77, no. 5, pp. I\_181-I\_191, 2022.
- (6) e-Stat 政府統計の総合窓口, 令和3年経済センサス - 活動調査 事業所に関する集計 産業横断的集計.
- (7) 馬場 弘樹, 秋山 祐樹, 谷内田 修：自治体保有データを活用した空き家の空間分布の将来予測モデル

構築 一群馬県前橋市を対象として一, Vol. 77, no. 2, pp. 62-71, 2021.

- (8) 石嶋 悠嗣, 柳沼 秀樹, 寺部 慎太郎, 海野 遥香 : Explainable AI を活用した機械学習型交通行動モデルの解釈可能性, Vol. 78, no. 5, pp. I\_427-I\_436, 2023.
- (9) 内閣府 : 令和 3 年版高齢社会白書 (全体版) 高齢化の状況及び高齢社会対策の実施状況, <https://www8.cao.go.jp/kourei/whitepaper/w-2021/html/zenbun/index.html>
- (10) 独立行政法人統計センターHP : SSDSE-市区町村の解説, <https://www.nstac.go.jp/sys/files/kaisetsu-A-2023.pdf>
- (11) 国立社会保障・人口問題研究所 : 日本の地域別将来推計人口 (令和 5 (2023) 年推計), <https://www.ipss.go.jp/pp-shicyoson/j/shicyoson23/t-page.asp>



参考図 選択された特徴量間の相関関係の可視化