

2021年度 統計データ分析コンペティション
特別賞（統計分析） [大学生・一般の部]

自治体間の人口流動性を考慮した
潜在的な人手不足の可視化

渡邊 晃大・村上 竜之介
(千葉工業大学先進工学部知能メディア工学科)

論文の概要

各都道府県の産業構造の特徴を調べるために、非負値行列因子分解によって抽出した因子データからコロプレス図を作成し、機械学習によって産業別従業者数の特徴量を抽出して結果を可視化することで、様々な業種や都道府県において潜在的な人手不足に陥る可能性を示した。

論文審査会コメント

素朴な探索的記述統計を用いた分析手順はシンプルで分かりやすく、結果も大変興味深いものが出ているが、構造化データで非負値分解する必然性の説明が必要である。

自治体間の人口流動性を考慮した潜在的な人手不足の可視化

渡邊 晃大・村上 竜之介

千葉工業大学先進工学部知能メディア工学科

1. 研究のテーマと目的

近年では都市部への人口集中が社会的な問題になっており、特に東京都では23年間連続で人口が増加し続けている。また、国土交通省-国土交通白書によると、東京圏（東京都、千葉県、埼玉県、神奈川県）には日本の人口の29%を占める約3700万人が居住していることがわかっている^[1]。2018年の東京都における転入超過数は13万6千人を記録した^[2]。

具体的な人口増加の内容を知るため、2018年度の東京圏の年齢階層別転入超過数を調査したところ、20代（20～24歳・25～29歳の合計）の占める割合が最も大きく、その割合は73%となっている^[1]。このことから若者の移動が東京圏の転入超過に多大な影響を与えていることがわかる。対照的に、環境省によると東京在住者の4割が地方への移住を検討している又は今後検討したいと考えているが、「地方へ移住する上での不安・懸念点」として雇用先の有無や、日常生活・公共交通の利便性を上げる人が多くなっていることがわかっている^[3]。

2018年の人口移動では東京圏に加えて愛知県、大阪府、福岡県で転入超過となっており、その他の道府県では転出超過である^[1]。このことから、東京圏以外の都市部においても人口集中が起きていることがわかる。地方から都市部への若者の人口移動が盛んな状況の中、地方自治体を対象にしたアンケート調査^[3]によると「現在直面している政策課題で、特に優先度の高いと考えられるもの」として主に挙げられているのは「人口減少や若者流出」、「少子化・高齢化の進行」、「中心市街地の衰退」であった。特に「人口減少や若者流出」に関しては現在地方にある産業の新しい労働者人口が不足することにつながるため、地方産業の衰退に起因することになると考えられる。

上記にも述べたように2018年度の東京圏への転出者に占める割合は20代が最も大きいため、地方産業を継いでいく若者を地方で就職したいと考えられるような対策が必要である。加えて長期的な目線で見ると、人手不足によって引き起こされる地方産業の衰退は地方産業の競争力を低下させるとともに、地方経済に悪影響を与えることが推測できることから、上記のアンケート調査^[4]にもあるように少子化・高齢化や中心市街地の衰退をさらに加速させることが考えられる。こうして地方産業が衰退し、それに伴う様々な影響により地方からの人口流出が進むことで東京一極集中がさらに深刻化し、それを繰り返すという状況を防ぐためにも早急な対策をする必要があると考える。

そこで本研究では時間変化の影響を考慮し、2014年度と2016年度の全国47都道府県の統計データを用いて分析を行なった。この分析では各都道府県における業種別従業者数の増減と全従業者数に占める業種ごとの従業者数から導いた業種ごとの重要度を調査し、それらの指標から散布図を作成した。加えて非負値行列因子分解によって抽出した因子データからコロプレス図を作成することで潜在的に人手不足に陥る可能性のある業種と地域の関係性を可視化し、人口集中の原因特定や対策の糸口を発見することを目的としている。

2. 研究の方法と手順

本研究では、各都道府県の産業構造の特徴を調べるために産業別従業者数の特徴量を機械学習によって特徴量を抽出し、その結果を可視化することによって知見の獲得を試みる。解析手順は以下の手続きに従って進

められる。

手順1：データセットの加工

ここでは、都道府県別の特徴を把握するために特徴量を計算する。産業別に年度間において増加傾向にあるのか、減少傾向にあるのかを調査するために年度間の差分を計算し「従業員の増加数」を求める。次に、その地域において中心的な役割を果たす産業が何かを調べるために、各産業に対する「産業の重要度」を調べる。この指標は地域内における従業員数全体に占める割合で計算している。詳細については次節で述べる。

手順2：特徴量を計算し、産業別のポジショニングマップを作成

潜在的に人手不足に陥る可能性をはらんでいる地域を調べるために、手順1で求められた特徴量の大小によってポジショニングマップを作成する。各産業において横軸に「増加数」を、縦軸にその産業の「重要度」を割り当てた散布図を作成し、各軸の平均値に基づいて4象限に分割することでその産業の位置づけを調べる。この場合、第二象限は産業従事者（横軸）が減少傾向にあり、かつその地域では重要度（縦軸）が高い中心的な産業であるため将来的には人手不足に陥る可能性が高いことを表している。このような性質に着目し解析を進めていく。

手順3：非負値行列因子分解によって因子データを抽出し、コロプレス図を作成

多数の変数から構成されるデータであり、全体的な傾向や地域性を把握するために教師なし学習手法である非負値行列因子分析^[5]によって次元を削減し、抽出された変数に基づいて考察する。この結果として得られた結果を把握しやすくするために分解された行列の重みを都道府県別に彩色することでデータを可視化する。

3. データセットの加工

本研究では各業種の増加数とその業種が各地域において重要であることを調べるために、年度をまたいだデータを利用している。2021年度版のSSDSE-Aには2016年の産業別従業員数が掲載されており、2020年度版のSSDSE-Aには2014年の産業別従業員数が掲載されている。これを用いることで年度をまたいだ各産業の増減の傾向を調べている。

$$\text{産業 A の増加率} = \frac{\text{2016 年の産業 A の従業員者数} - \text{2014 年の産業 A の従業員者数}}{\text{2014 年の産業 A の従業員者数}}$$

これを用いて業種別の増加数と、その業種が当該都道府県においてどの程度重要かを示す指標を次式に基づいて計算している。なお、2020年度版のデータには漁業従事者数が掲載されていないため、解析対象からは除外している。

$$\text{産業 A の重要度} = \frac{\text{2016 年の産業 A の従業員者数}}{\text{2016 年の全事業従業員者数}}$$

上記の計算を各業種に対して適用することで、各業種における従業員数の重要度や増加率を計算するわけだが、これらの変数が多くなり、全体てきな傾向をはあくすることができない。そこで、本研究では先述のとおり非負値行列因子分解によって特徴量を変換し、合成変数を作成することによってデータの可視化を試みる。データのスケールの違いが学習結果に及ぼす影響を回避することと非負値の制約を満たすために各変数を次

式によって0以上1以下の範囲に変換したものを学習している。

$$\text{対象都道府県の特徴量 } X = \frac{\text{対象都道府県の指標 } A - \min(\text{全都道府県の指標 } A)}{\max(\text{全都道府県の指標 } A) - \min(\text{全都道府県の指標 } A)}$$

この式を各都道府県の各指標に適用することによって、非負値行列因子分解の特徴量を算出している。

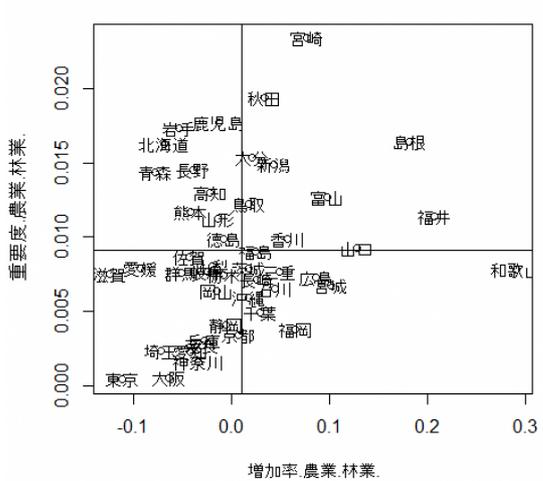
表1 使用した変数とその出典

使用変数	出典
A.従業者数（民営）	市区町村別、多分野データ（2014・2016）
B.従業員数（民営）（農業、林業）	市区町村別、多分野データ（2014・2016）
C.従業者数（民営）建設業）	市区町村別、多分野データ（2014・2016）
D.従業者数（民営）（製造業）	市区町村別、多分野データ（2014・2016）
E.従業者数（民営）（電気・ガス・熱供給・水道業）	市区町村別、多分野データ（2014・2016）
F.従業者数（民営）（情報通信業）	市区町村別、多分野データ（2014・2016）
G.従業者数（民営）（運輸業、郵便業）	市区町村別、多分野データ（2014・2016）
H.従業者数（民営）（卸売業、小売業）	市区町村別、多分野データ（2014・2016）
I.従業者数（民営）（金融業、保険業）	市区町村別、多分野データ（2014・2016）
J.従業者数（民営）（不動産業、物品賃貸業）	市区町村別、多分野データ（2014・2016）
K.従業者数（民営）（学術研究、専門・技術サービス業）	市区町村別、多分野データ（2014・2016）
L.従業者数（民営）（宿泊業、飲食サービス業）	市区町村別、多分野データ（2014・2016）
M.従業員数（民営）（生活関連サービス業、娯楽業）	市区町村別、多分野データ（2014・2016）
N.従業者数（民営）（教育、学習支援業）	市区町村別、多分野データ（2014・2016）
O.従業者数（民営）（医療、福祉）	市区町村別、多分野データ（2014・2016）
P.従業者数（民営）（複合サービス事業）	市区町村別、多分野データ（2014・2016）
Q.従業者数（民営）（サービス業（他に分類されないもの））	市区町村別、多分野データ（2014・2016）

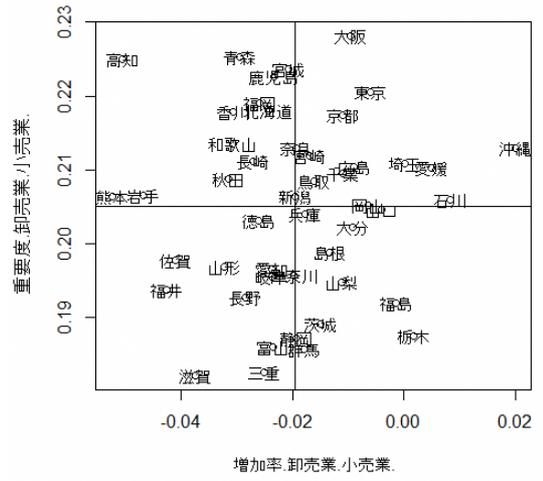
4. データ分析の結果

ここでは、先に述べた解析手順に基づいて得られた結果ならびにそこから分かった結果について述べる。まずは産業ごとの増加数と重要度の2つの指標を用いて作成した15個の散布図から特に特徴的であった5つを以下の図1(a)~(e)に示す。

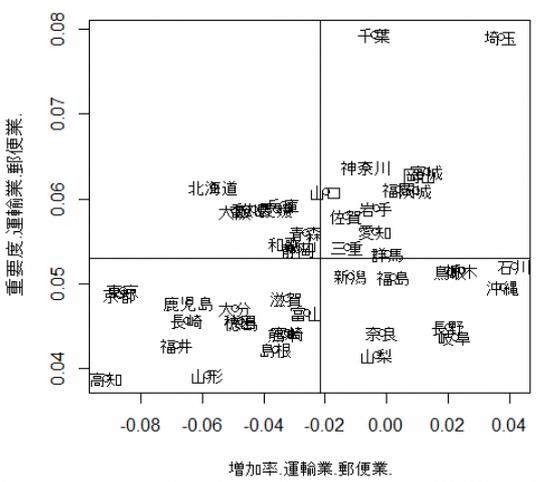
図1-(a)の散布図では宮崎県、秋田県、島根県などが重要度の高さに応じて増加率も高い位置にあり、今後の農業や林業において期待が持てることが読み取れる。反対に北海道、岩手県、青森県、長野県、鹿児島県などは重要度が高い割に、増加率が低いことから、農業や林業においてこのままの状態では人手不足に陥るであろうことが読み取れる。図1-(b)の散布図では都市部に加えて沖縄県が重要度の高さと人手の増加率が合致しており、今後卸売業小売業において伸びが予想される。反対に同業種では北海道、鹿児島、熊本、和歌山、長崎、秋田などは重要度の高さに割に、増加しておらず今後の人手不足が危惧されることがわかる。図1-(c)の散布図では埼玉県と千葉県が主に今後も運輸業、郵便業において上昇していくことが予想され、重要度も高い。反対に北海道、大阪府などは重要度が高いにもかかわらず、従業者数が減少していることがわかる。よって今後の人手不足が危ぶまれることがわかる。図1-(d)の散布図では東京が情報通信業において極端に上昇傾向にあることがわかる。そしてその他の県はほぼ同位置に存在しているため情報通信業は東京だけが抜き出ていることが読み取れる。このことから東京一極集中の原因の1つとなっているのではないかと考えられる。図1-(e)の散布図は不動産業、物品賃貸業を表しており、ここでも東京が頭抜けており、他都道府県に比べ圧倒的に差がある。関西圏や北海道沖縄県は不動産において重要度が高いにもかかわらず増加率が低く、観光都市は今後人手不足に陥る可能性をはらんでいる状況であると考えられる。この不動産業、物品賃貸業も東京が極端な値をとっていることから一極集中の要因の可能性があると考えられる。



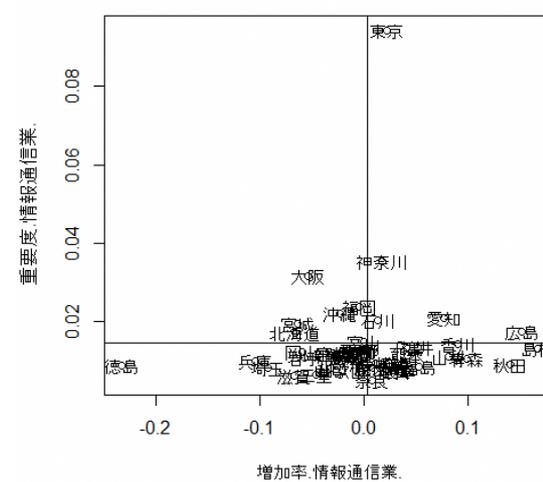
(a) 農業・林業



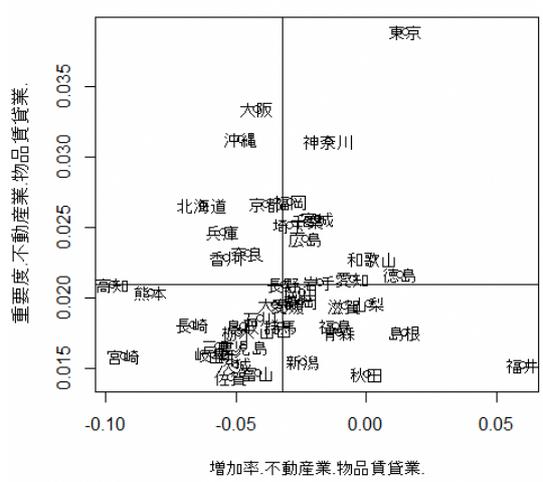
(b) 卸売行・小売業



(c) 運輸業・郵便業



(d) 情報通信業



(e) 不動産業・物品賃貸業

図1 業種別増加率と重要度の関係性を示す散布図
(第二象限が高重要度にもかかわらず低増加率または減少傾向な地域を表している)

次にデータセットの加工にて記載した手順に則り実行した非負値行列因子分解によって得られた10通りの結果の中から、特に特徴的であった3つを表2及び図2-(a)～(c)に示す。

まず第2因子の特徴としては情報通信業、金融業、不動産業・物品賃貸業などで重要度が高く、増加数も多いことが表2からわかる。そしてその傾向が表れている都道府県が東京都であることが図2-(a)から読み取れる。図2-(a)を見てもわかる通り、この傾向はほぼ東京都のみに集中していることから、情報通信業、金融業、不動産業・物品賃貸業などの産業においては東京都への一極集中が観測できる。

次に第3因子の特徴として、農業・林業の重要度が高いのにもかかわらず増加数が低いことが表2から読み取れる。また図2-(b)から、その特徴が顕著にあらわれている都道府県は北海道や青森県、岩手県などの東北地方であることから、これらの都道府県では農業・林業が潜在的に人手不足に陥る可能性があることがわかる。対照的に近畿地方より西側に位置する府県ではあまり農業・林業に関して人手不足に陥る可能性はないという結果が得られた。

最後に第10因子では表2から卸売業・小売業の重要度が本研究で最も高い値を示したにもかかわらず、その増加数は低かった。図2-(c)から、その特徴は北海道や青森県、岩手県、福島県などの東北地方の他に大阪府、京都府、高知県などを含めた西日本において顕著に現れていることがわかる。これらのことから卸売業・小売業においては上記の道府県で潜在的に人手不足に陥る可能性があることがわかる。また、東京、神奈川に隣接する千葉や埼玉の値が著しく大きいことから、東京近郊では卸売業・小売業に関して人手不足に陥る可能性が高いことがわかる。

表2 因子の内容（各因子において重みが大きい変数）

第2因子		第3因子		第10因子	
変数名	加重値	変数名	加重値	変数名	加重値
重要度.学術研究.専門.技術サービス業.	0.36739343	重要度.建設業.	0.5356286	重要度.卸売業.小売業.	0.56618684
重要度.サービス業.他に分類されないもの.	0.32444269	重要度.農業.林業.	0.33801349	重要度.運輸業.郵便業.	0.4292525
重要度.情報通信業.	0.32381878	増加率.製造業.	0.30763833	増加率.教育.学習支援業.	0.32767879
増加率.学術研究.専門.技術サービス業.	0.31054455	増加率.情報通信業.	0.25947901	重要度.不動産業.物品賃貸業.	0.27947752
重要度.不動産業.物品賃貸業.	0.2983465	増加率.運輸業.郵便業.	0.23711129	増加率.金融業.保険業.	0.23403147
重要度.金融業.保険業.	0.29748402	重要度.複合サービス事業.	0.23109245	増加率.宿泊業.飲食サービス業.	0.22820069
増加率.教育.学習支援業.	0.22893321	増加率.電気.ガス.熱供給.水道業.	0.22784976	増加率.サービス業.他に分類されないもの.	0.21910902
増加率.不動産業.物品賃貸業.	0.2175095	増加率.建設業.	0.22743937	重要度.教育.学習支援業.	0.20361729
増加率.卸売業.小売業.	0.19239316	増加率.宿泊業.飲食サービス業.	0.18750768	重要度.サービス業.他に分類されないもの.	0.15616972
増加率.電気.ガス.熱供給.水道業.	0.18974958	増加率.金融業.保険業.	0.18173957	増加率.農業.林業.	0.14095098
増加率.情報通信業.	0.18496254	増加率.卸売業.小売業.	0.1694533	増加率.医療.福祉.	0.13821762
重要度.卸売業.小売業.	0.17916788	増加率.不動産業.物品賃貸業.	0.1504013	重要度.医療.福祉.	0.13006988

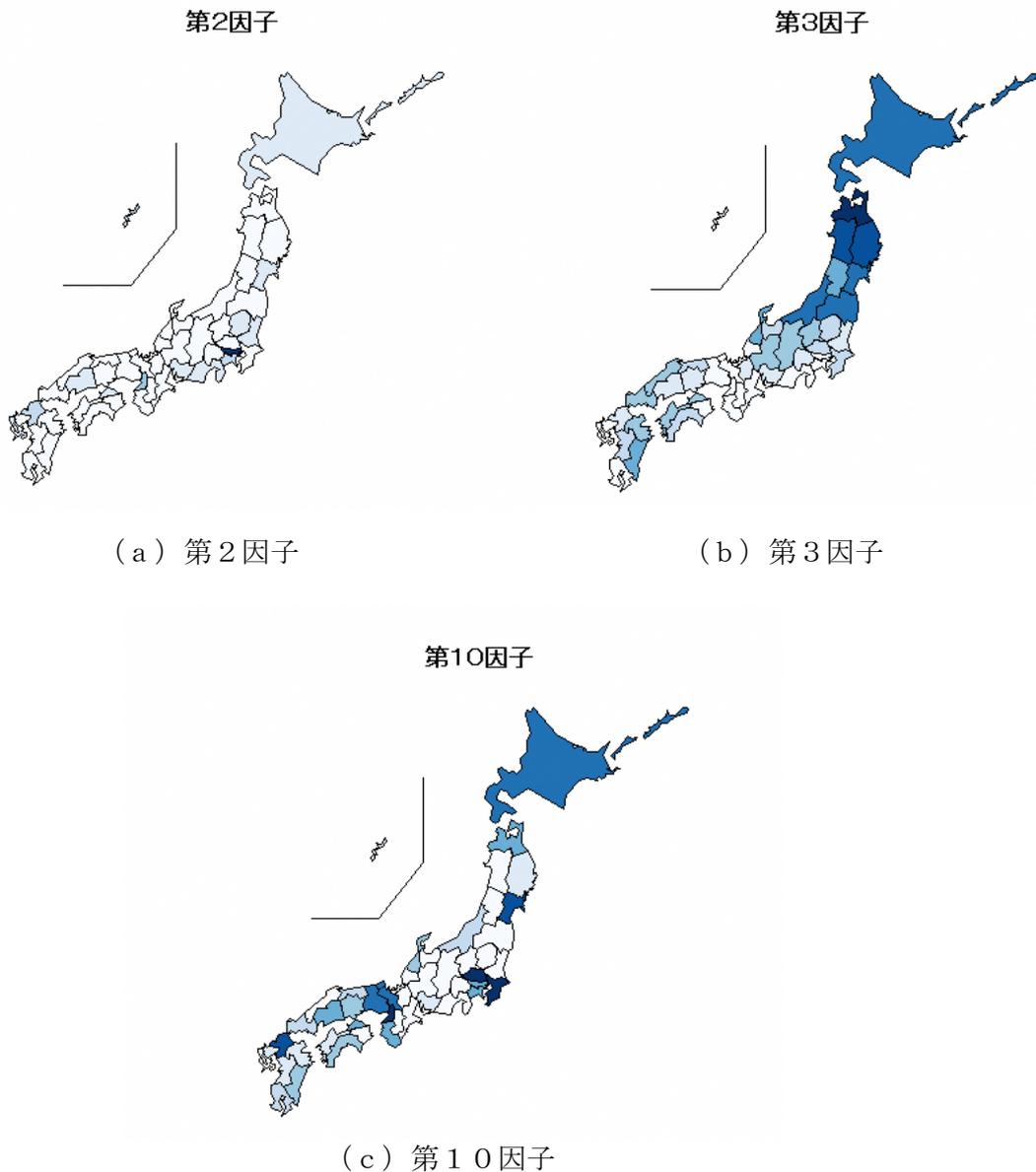


図2 因子の反応が強い都道府県を彩色したコロプレス図

5. 結果の解釈

図 1-(d) 及び表 2、図 2-(a) から読み取れる結果として情報通信業、金融業、不動産業・物品賃貸業が東京都に集中していることがわかった。この結果に対する考察として不動産業・物品賃貸業に関しては東京都の人口が多いことから住宅を必要とする人も必然的に多くなるため、東京都に不動産業・物品賃貸業が集中することが考えられる。金融業に対しても同様のことが考えられる。情報通信業に関しては日本の主要企業が東京に集まり、その企業を相手にする産業であることや、その影響で最新技術が東京都に集中することが原因で東京都に情報通信業が集中することが考えられる。また、大阪、愛知、宮城、広島、福岡、関東圏、が薄く彩色されていることから、この因子が都市部を特徴づける要因であることがわかる。

次に図 1-(a) 及び表 2、図 2-(b) から読み取れる結果として、北海道や青森県、岩手県などで農業・林業が潜在的に人手不足に陥る可能性があることがわかった。この結果からは、北海道や青森県、岩手県など与其他

の地域を比較しても、北海道、青森県、岩手県が潜在的に人手不足に陥る可能性があると言えるだけの決定的な要素が見つかることができなかった。そのため、今後の対策として市区町村レベルの粒度の高いデータを扱うなどしてこの3県が農業・林業において潜在的に人手不足に陥る可能性があることを説明できるようにしておく必要がある。

最後に図 1-(b)及び表 2, 図 2-(a)から読み取れる結果として、北海道や青森県、岩手県、福島県などの東北地方の他、大阪府、京都府、高知県などを含めた西日本などの多くの道府県で卸売業・小売業が潜在的に人手不足に陥る可能性があることがわかった。卸売業・小売業は第 10 因子においては従業員の絶対数が大きいことから重要度が高い値となっているが、昨今の日本においてはネットショッピングが普及に伴う都市部での需要増がその周辺における卸売業・小売業に影響を与えた結果になったと考えられる。

また、本研究の結果として、様々な業種、様々な都道府県で潜在的に人手不足になる可能性を可視化した。人手不足に陥る可能性のある都道府県は数多く見られたことから、地方の人手不足を解消し、東京への一極集中を是正するために、現在東京都に集中している産業を地方に分散させることや、地方で確立している産業の機能強化などの地方人口が地元での就職を希望するような政策を推進していくことが必要であると考えられる。

6. まとめ

本研究では産業別人口の増減に基づいた特徴量に対してポジショニングマップと非負値行列因子分解を用いて将来的に人手不足に陥る可能性をはらんでいる業種と地域の関係性を可視化した。この結果に基づいて地域の特性や産業が抱える課題について解析ができた。この結果ならびに本研究の解析により、地域が抱える課題の共有や産業支援施策の策定などへの貢献が期待できると考えている。

今後の課題として、非負値行列因子分解の基底数などのパラメータ設定などの学習器の設定に関する議論を深める必要があると考えられる。また、人口の比率も考慮した特徴量を扱うことや今回扱えなかった漁業などのデータの追加、市区町村レベルのより粒度の高いデータを対象とすることなど、より精緻な解析が必要であると考えている。

参考文献

- [1] 国土交通白書 <https://www.mlit.go.jp/hakusyo/mlit/r01/hakusho/r02/html/n1112000.html>
- [2] 総務省「住民基本台帳人口移動報告平成 30 年（2018 年）結果」（2019 年 1 月 31 日公表）.
- [3] 環境省 平成 27 年版 図で見る環境・循環型社会・生物多様性白書
- [4] 千葉大学 広井良典教授による「地域再生・活性化に関する全国自治体アンケート調査」（平成 22 年 7 月）
- [5] Daniel D. Lee & H. Sebastian Seung: Learning the parts of objects by non-negative matrix factorization Nature, vol. 401, pp. 788-791 (1999)