

Kanazawa University,
Faculty of Economics and Management

Discussion Paper Series

No. 048

主成分分析に基づく地域クラスタリング
と産業構造の可視化

山本けい子
原田魁成
寒河江雅彦

Sagae.masahiko@gmail.com

4th April 2019



金沢大学経済学経営学系
〒920-1192 金沢市角間町

Faculty of Economics and Management,
Kanazawa University

Kakumamachi, Kanazawa-shi, Ishikawa, 920-1192, Japan

<http://econ.w3.knazawa-u.ac.jp/DP/>

主成分分析に基づく地域クラスタリング と産業構造の可視化

金沢大学大学院 人間社会環境研究科

山本けい子

金沢大学大学院 人間社会環境研究科

原田魁成

金沢大学 人間社会研究域 経済学経営学系

寒河江雅彦

要旨

地域内の経済構造の把握や経済波及効果分析に用いられる産業連関表の新たな活用法を提案する。我々は、産業連関表が、調査対象地域によらず統一かつ定期的に作成されることから、地域間の比較や産業構造の可視化、経年変化の抽出においても有用な統計表として利活用できると考えている。そこで、都道府県単位で作成された産業連関表（取引基本表にある内生部門データ）を多変量データとみなして、いくつかのデータ解析手法を適用した。本稿では、類似性比較や可視化の観点から、複雑な高次元データである産業連関表データに対して、主成分分析を用いて、低次元特徴空間へデータを縮約し、クラスタリングやマッピングを試みた結果について示す。クラスター分析では、縮約した次元を特徴付ける産業群に関する都道府県の分類が得られ、産業部門別のマッピングでは、都道府県間の相対的な位置関係を可視化することができた。また、都道府県ごとの主成分分析からは、都道府県内の産業構造の概観や年次推移を読み取ることができた。最後に、これらの分析手法と得られた結果を通して、多変量解析の枠組みを用いた産業連関分析の可能性について総括する。

キーワード

主成分分析, 産業連関表, クラスタリング, 産業構造の可視化

Regional Clustering and Visualization of Industrial Structure based on Principal Component Analysis for Input-output Table Data

YAMAMOTO Keiko

HARADA Kaisei

SAGAE Masahiko

ABSTRACT

We discuss the new possibility of Input-Output(I/O) data analysis based on multivariate analysis framework, for understanding the industrial structure and analyzing the economic ripple effect in the region. We consider that it can be used as a statistical table that is useful for regions comparisons, visualization of the industrial structure, and extraction of secular change. Because the I/O table is created uniformly and regularly regardless of the study area. We applied multivariate analysis methods to the I/O table (endogenous sector data in the basic transaction table), created in prefectural units. In this paper, the principal component analysis is applied to reduce complex high-dimensional I/O data to low-dimensional feature space and we express them by using the clustering and scatter plot. The clustering is used to classify prefectures according to industry groups characterizing the reduced dimension, and the scatter plot is also used to visualize the relatively positional relationship between prefectures. In addition, we extract the overview and the industrial structure change from I/O data of the prefecture by using the principal component analysis.

Keywords

Principal Component Analysis, Input-output table data, Clustering, Visualization of Industrial Structure

1. はじめに

産業連関表は、国内経済において一定期間（通常1年間）に行われた財・サービスの産業

間取引を1つの行列に示した統計表であり、5年ごとに作成・公開されている。一般的に、作成対象地域における経済構造の把握、経済波及

効果の計算、各種経済指標の基礎資料として用いられているが、我々は、恣意的な選択なく、統一的に作成されるという特性から、従来のような地域内の分析だけでなく、地域間の比較や年次推移の可視化などにも用いることができるのではないかと考えた。産業連関表を多変量データとして扱い、分析を試みた既存研究としては、山本・寒河江(2016)¹⁾、入江(2018)²⁾、原田(2019)³⁾がある。山本・寒河江(2016)では、2005年の産業連関表の内生部門データを用いて都道府県のクラスター分析や産業構造の可視化を行っている。また、入江(2018)では、秋田県を除く46都道府県の2011年産業連関表のクラスター分析によって都道府県の類型化とそれらの集団の特徴づけを行っている。原田(2019)は、2011年都道府県産業連関表に非負値行列因子分解を適用することで、地域類似性及び地域特性抽出を行っている。

本稿では、2005年と2011年の2時点における都道府県版産業連関表（取引基本表内の内生部門）データを用いて、産業取引の構造に関する都道府県比較や経年変化の抽出および可視化を試みる。特に、産業部門間の複雑な取引を数値化した内生部門データは、高次元（部門数次元）データとなることから、比較や可視化のために、主成分分析を行い、低次元の特徴空間へ縮約して解析を行う。これらの分析とその結果から、産業連関表の新たな活用法について議論する。

2. 産業連関表

産業連関表は、総務省を中心に各府省庁において、西暦年の末尾が0又は5の年次（昭和30年(1955年)表以来）を対象として5年ごとに作成される。調査対象地域によって、全国版・地域版・都道府県版などがあり、基本的な構成要素は、取引基本表、投入係数表、逆行列係数表である。対象となる産業部門は、分類の粒度によって小分類(190)・中分類(108)・大分類(37,平成17年は34分類)に分けられる（なお、平成23年は、13分類も存在する）。取引基本表は、各産業間で取引された財・サービスを金額で表示したもの、投入係数表は、取引基本表の内生部門における原材料等の投入額と粗付加価値を含めた当該産業の生産額で除した係数、逆行列係数表は、ある産業に対して、1単位の最終需要があった場合、各産業の生産が究極的にどれだけ必要となるかを示す係数となっている。

本稿で用いる産業連関表は、都道府県単位で作成された平成17年版(2005年)と23年版(「経済センサス-活動調査」の調査対象年次に合わせて例外的に2011年で作成)の取引基本表内の内生部門データ(大分類)を分析対象としている。

3. 分析方法

産業連関表（正確には、取引基本表の内生部門データ）を、分類部門数×分類部門数の多変量データとみなした場合、もっとも分類数の少

ない大分類であっても、34あるいは37次元の高次元小標本データとなる。我々の分析目的が、個々の産業間取引ではなく、全体の取引構造に着目した比較や可視化であることから、主成分分析を用いて、高次元小標本データを直感的理解が容易な低次元の特徴空間へ縮約する。特徴空間上のデータの作成や分類、可視化のために、本稿で採用した主成分分析とクラスター分析の2つの分析手法⁵⁾について以下に示す。

3.1. 主成分分析

主成分分析は、多変量データの次元縮約手法として知られており、データの次元数が多い場合に、少ない情報の損失で少数の合成変数にデータを変換して分析する手法である。一般には多変量データ X に対する合成変数 Z は(1)式のように \bar{X} と A の2つの行列の積の形で与えられる。

$$Z = \bar{X}A \quad (1)$$

ただし、 \bar{X} はデータの偏差行列、 A は主成分と呼ばれる係数行列である。通常、データの分散(相関)が最大になるような主成分 A を求めるが、これは、分散共分散行列の固有ベクトルに相当する。元のデータの情報を多く含む(固有値の大きさの)順に、第1主成分 Z_1 、第2主成分 Z_2 、第3主成分 Z_3 、・・・となる。本稿では、第1主成分と第2主成分を用いて2次元データに変換し、比較や可視化の分析を行う。

3.2. クラスター分析

クラスター分析は、個々のデータを1つのクラスター(まとめり)としてスタートし、クラスター間の距離(類似度)に基づいて、2つ

のクラスターを逐次的に結合しながらデータを分類する手法である。クラスター分析には、階層的クラスタリングと非階層的クラスタリングの2種類があるが、非階層的クラスタリングでは、あらかじめクラスター数が必要となるため、本稿では、階層的クラスタリングを使用し、樹形図を用いて階層構造を表現する。類似度の指標としては、ユークリッド距離、マンハッタン距離、マハラノビス距離、コサイン類似度など数多くあり、クラスター間の結合方法としては、ウォード法、群平均法、最短距離法、最長距離法などがある。

主成分分析を用いて高次元小標本データの次元を縮約することで、産業間取引の全体構造に着目した地域間での比較や年次推移の可視化が容易となる。また、クラスター分析では、地域間の類似性を樹形図によって視覚的に確認することができる。なお、解析は、オープンソースの統計解析システム R⁴⁾を用いて実装した。

4. 分析対象データの収集とその概要

分析対象として、以下のデータを収集した。

- 対象地域：都道府県
各都道府県のホームページからダウンロードした
- 対象年：平成17年版と平成23年版
分析時の最新データは23年版であった
- 対象表：取引基本表内の内生部門内データ

(中間供給×中間需要)

図 2

● 対象分類：大分類

平成 17 年版は 34 分類，平成 23 年版は 37 分類であったため，23 年版を 34 分類になるよう人手により再構成した

統一的に作成されている産業連関表であるが，都道府県によって，産業部門の設定に相違がみられたため，できる限り全国表の産業部門に合わせる形で人手によりデータを加工して用いた。

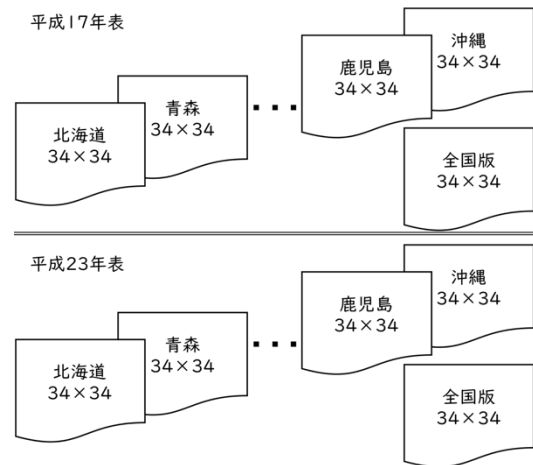


図 2. 収集データのイメージ図

図 1

需要部門 (買い手)		中間需要				
		1	2	3	...	計
供給部門 (売り手)	1 農林水産業	A
	2 鉱業
	3 製造業
	計 D

図 1. 取引基本表内の内生部門（赤枠）

内生部門データは，34 に分類された産業（行方向：供給部門）がどの産業（列方向：需要部門）にいくら販売したか，逆に言えば，34 の産業（列方向：需要部門）は，どの産業（行方向：供給部門）からいくら購入したかを表している。

比較のために，全国表についても同様に収集した。これにより，全国版および 47 都道府県版のそれぞれについて，34×34 の行列データが得られた。図 2 は収集データのイメージである。

本稿では，これらの収集データを多変量データとみなし，分析を実施する際には，供給部門の各産業データが需要部門の産業の金額で与えられると解釈することとする。内生部門データ行列を転置した場合も同様に分析が可能である。

取引基本表は各都道府県内で取引される金額データであることから，都道府県によって金額の単位が大きく異なる。分析の目的において，金額に着目する場合は，収集データそのものを利用し，産業間の取引（需要/供給）構造に関心がある場合は，収集データを標準化（行列総和で除する/列和・行和で除する/平均 0 分散 1 の正規化など）したものを利用する。なお，本稿で用いる標準化データは，列ごとに平均 0，分散 1 となるよう正規化したデータを表す。

5. 分析結果

5.1. 主成分分析による次元縮約

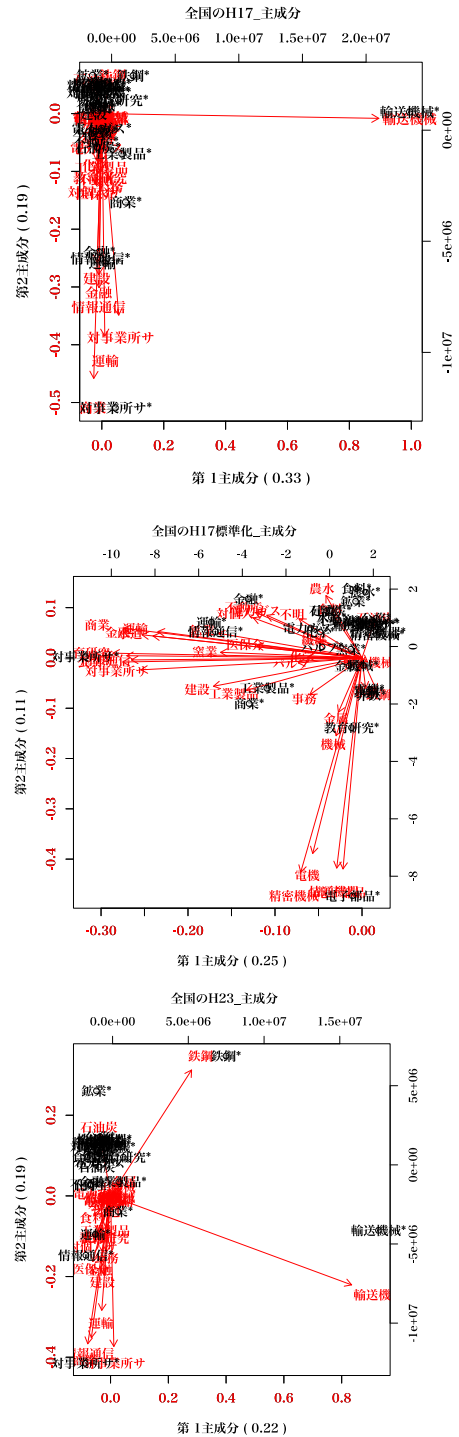
3. で述べたように、47 都道府県ごとに存在する 34×34 の行列データに対して、主成分分析を用いて 34×2 の 2 次元データへ縮約する。その際、変換基準が異なると、同一平面上での議論ができなくなるため、全国表の主成分分析で得られた第 1 主成分と第 2 主成分の各軸上に统一的に各都道府県データを変換した。

図 3 は各都道府県データを変換するための全国表の主成分分析に関するプロットであり、第 1 主成分と第 2 主成分に寄与する固有ベクトル（赤色）と主成分得点（黒色）の散布図を示したものである。各軸の()内の数値は寄与率である。図 3 の上から順に、平成 17 年データ、平成 17 年標準化データ、平成 23 年データ、平成 23 年標準化データを用いた場合の主成分に関するプロットである。

例えば、図 3 に示した平成 17 年のプロットから、寄与率に関しては、第 1 主成分と第 2 主成分の合計が 0.52 であることから、縮約した第 1, 第 2 主成分得点によって、元の 34 次元データの約 52% の情報を保持しているとみなすことができる。また、輸送機械が固有ベクトルおよび主成分得点ともに第 1 主成分に大きな値として特出しており、第 2 主成分では、固有ベクトルとして商業、主成分得点として対事業所サービスが特出している。標準化によって、取引金額（生産額）の情報ではなく、取引構造に着目すると、第 1, 第 2 主成分については、いくつかの産業部門がまとまった方向にプロッ

トされ、主成分得点としては対事業所サービスや電子部品が特徴的に現れている。

図 3



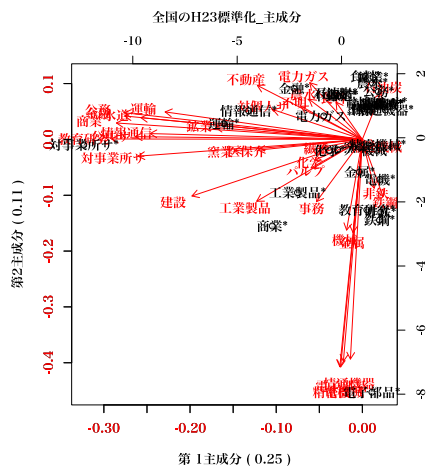


図3. 全国版の主成分分析（第1・第2主成分の固有ベクトルと主成分得点）

5.2. クラスター分析による都道府県クラスタリング

47都道府県ごとの34×34の産業連関表データを、5.1.で示した全国表主成分分析で得られた第1,第2主成分軸へ統一的に変換し、各都道府県のデータを34×2の2次元データへ縮約する。これらを元に、クラスター分析を用いて、都道府県間の類似度に基づくクラスターを形成し、樹形図によって可視化する。

前述のように、データの性質上、分析の視点が取引金額（生産額）の場合は収集データを、産業間の取引構造の場合は、標準化したデータ（本稿では、列ごとに平均0,分散1に正規化）を用いる。例えば、 $n \times m$ 行列で表されるA県のデータ X^A とB県のデータ X^B の類似度 $d(X^A, X^B)$ の指標は、(2)式で与えられるユークリッド距離で統一した。

$$d(X^A, X^B) = \sqrt{\sum_{i=1}^n \sum_{j=1}^m (X^A_{ij} - X^B_{ij})^2} \quad (2)$$

先に示した第1,第2主成分へ変換した場合は、 $n = 2, m = 34$ となる。また、主成分ごとに類似度を測る場合は、 $n = 1$ とすればよい。本稿では、図3に示したように、全国表の特徴が主成分ごとに大きく異なることから、第1主成分と第2主成分に分けて類似度を算出し、クラスタリングを行った。

図4は平成17年版データ、図5は平成23年版データで、さらに標準化の有無と第1主成分得点および第2主成分得点に分けて図示したものである。なお、クラスターの結合方法は、最短距離法を用いた。

図4

図5

H17_PC1

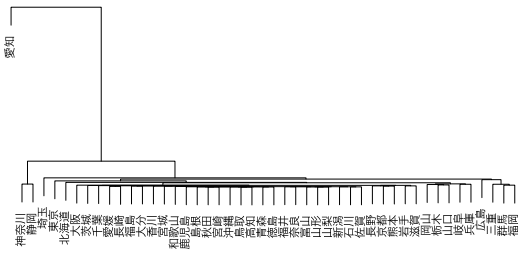


図 4-1. H17 年（標準化なし/第 1 主成分得点）

H23_PC1

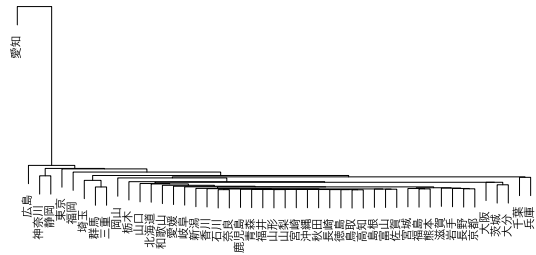


図 5-1. H23 年（標準化なし/第 1 主成分得点）

H17_PC2

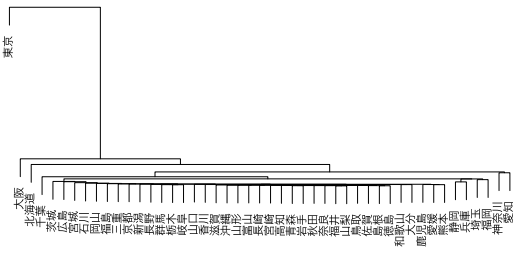


図 4-2. H17 年（標準化なし/第 2 主成分得点）

H23_PC2

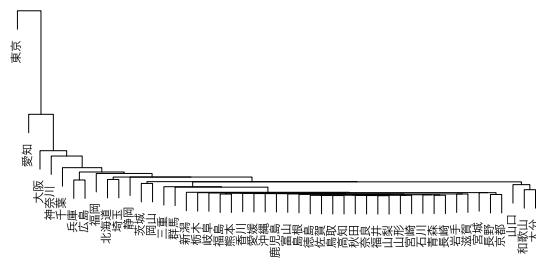


図 5-2. H23 年（標準化なし/第 2 主成分得点）

H17標準化_PC1

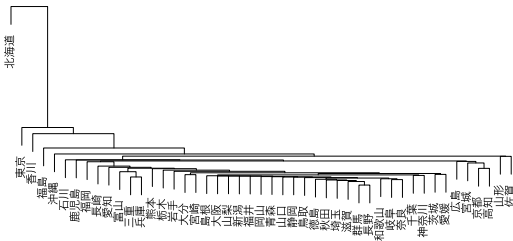


図 4-3. H17 年（標準化あり/第 1 主成分得点）

H23標準化_PC1

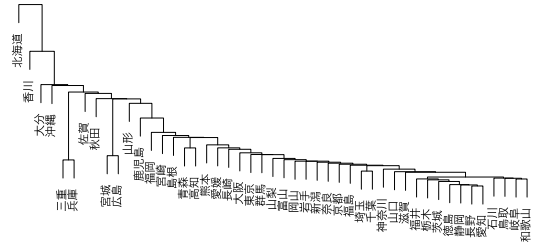


図 5-3. H23 年（標準化あり/第 1 主成分得点）

H17標準化_PC2

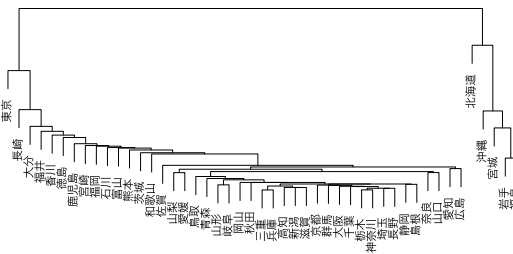


図 4-4. H17 年（標準化あり/第 2 主成分得点）

H23標準化_PC2

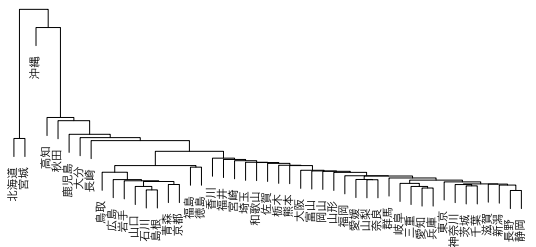


図 5-4. H23 年（標準化あり/第 2 主成分得点）

図 4-1 では、愛知県が 1 つのクラスターを形成しているが、全国表の図 3 上図から、第 1 主成分に影響ある産業部門として「輸送機械」がみられることから、大手自動車産業を有する愛知県が特出していると考えられる。また、図 4-2 の第 2 主成分では、東京都がクラスターとして出現しているが、図 3 の第 2 主成分は、「商業」、「運輸」、「対事業所サービス」などであり、それらの特徴は東京都の産業として関連づけることができる。標準化データを用いた場合は、生産額の影響が現れないことから、各主成分を特徴づける産業群に対して、他の都道府県とは特異な構造を有する都道府県がクラスターとして出現しているように見える。階層の低いクラスターに関しては、クラスターの結合方法などにも依存して変化するため、ここでは言及しない。

ここでは、(2)式のように全 34 産業部門を合算した類似度を用いてクラスター分析を行ったが、産業部門ごとの類似度を用いてクラスター分析することも可能である。

5.3. 主成分分析を用いた都道府県マッピング

低次元特徴空間（2次元の場合は平面）へ縮約した産業連関表データを用いて、平面上に各都道府県のデータをマッピングし、相対的な関係の可視化を試みた。

5.1.に示したように、図 3 の全国表における主成分軸上へ各都道府県のデータを変換する。例として、石油・石炭製品(石油炭)と電気機械

(電機)、電力・ガス・熱供給(電力ガス)、情報通信部門に対し、対象年と標準化の有無によってプロットしたものが、図 6 から図 9 である。横軸は、第 1 主成分得点、縦軸が第 2 主成分得点を表す。

産業全体における都道府県の間を示すには、変換後の各都道府県 34×2 の主成分得点データの総和を取ることで 47×2 データを生成し、図示する方法も考えられるが、都道府県間の関係は、産業部門ごとに特徴が異なることから、あえてまとめない方がよいだろう。

- 図 6
- 図 7
- 図 8
- 図 9

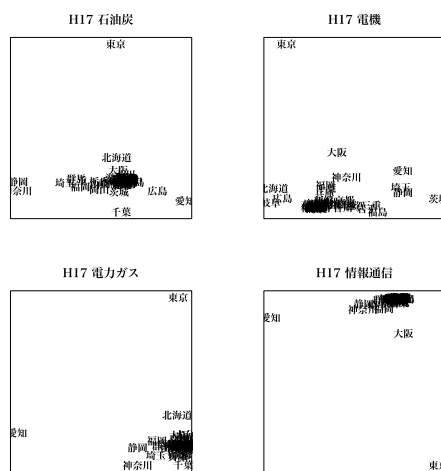


図 6. 都道府県マッピング/H17 年

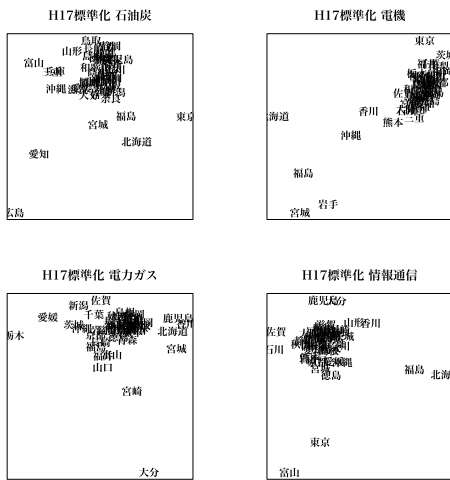


図 7. 都道府県マッピング/H17年標準化あり

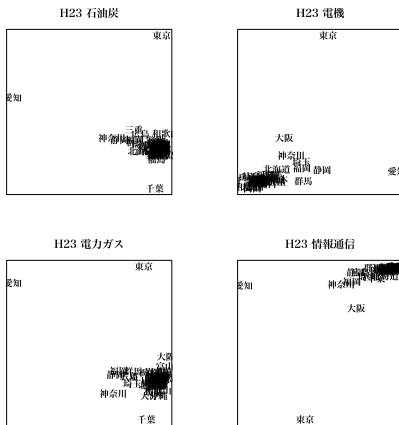


図 8. 都道府県マッピング/H23年

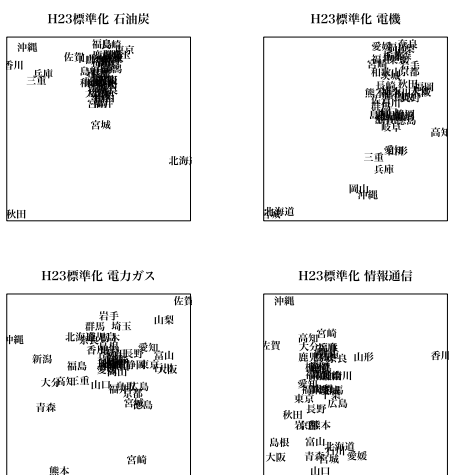


図 9. 都道府県マッピング/H23年標準化あり

これらのマッピングでは、都道府県ごとに、本来は 34 次元で表される供給部門のデータを同じ基準を用いて 2 次元に縮約することで、供給部門における都道府県の関係性を同一マップ上に表現することができる。

標準化しないもの(図 6,図 8)は、各供給産業部門における取引金額の大きい都道府県が特出する傾向にある。また、標準化したもの(図 7, 図 9)では、供給産業部門における取引構造の性質が類似する都道府県は近くに、異なる場合は他から離れて配置されている。

本稿では、変換基準として全国表の主成分(各産業の固有ベクトル)を用いたが、全国表に合わせることで、全国表と似た産業構造の都道府県は過大に、そうでない都道府県は過少に特徴を抽出している可能性もあることから、分析の目的に応じて基準を設定するとよい。

5.4. 都道府県ごとの産業構造の可視化

都道府県ごとの主成分分析を用いて、当該都道府県における産業間取引の全体構造の把握と可視化を行う。

図 10 から図 12 は、例として福島県、東京都、石川県のデータに関する主成分分析の結果として、第 1 主成分・第 2 主成分として寄与する固有ベクトル(赤色)と主成分得点(黒色)の散布図を示したものである。左上図が平成 17 年データ、右上図が平成 17 年標準化データ、左下図が平成 23 年データ、右下図が平成 23 年標準化データを用いた場合である。これまでと同

様に、横軸が第1主成分、縦軸は第2主成分に
対応している。

図 10

図 11

図 12

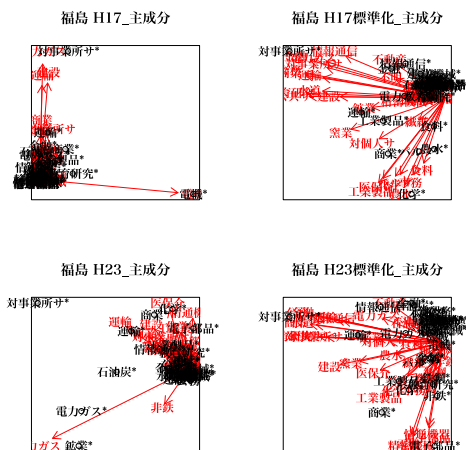


図 10. 福島県データの主成分分析

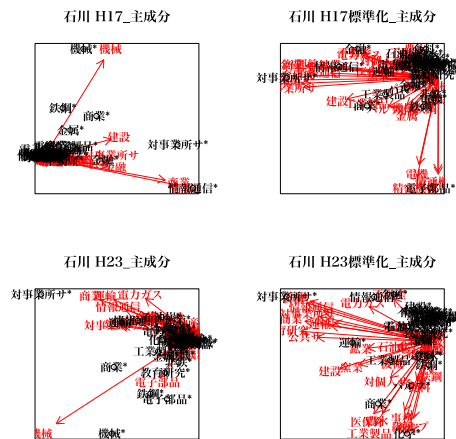


図 12. 石川県データの主成分分析

図中の主成分得点で表される産業部門（黒色）は相対的な位置関係を示しているため、上下左右の反転（正負）の違いには意味がないことに注意が必要である。

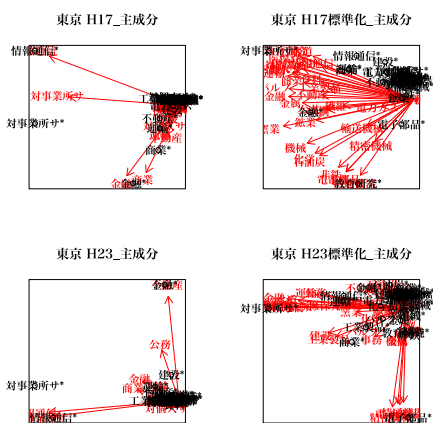


図 11. 東京都データの主成分分析

5.5. 産業構造の推移の可視化

平成 17 年と平成 23 年の 2 時点のデータをそれぞれで分析してきたが、例えば、図 10～図 12 ごとに上図と下図を比較することで、対象都道府県における産業構造の推移を確認することもできる。見やすさのために、取引金額における、第 1 主成分、第 2 主成分について絶対値の大きい順に 6 産業部門についてのみプロットしたものを図 13 に示す。各軸の () 内の数値は寄与率である。

図 13

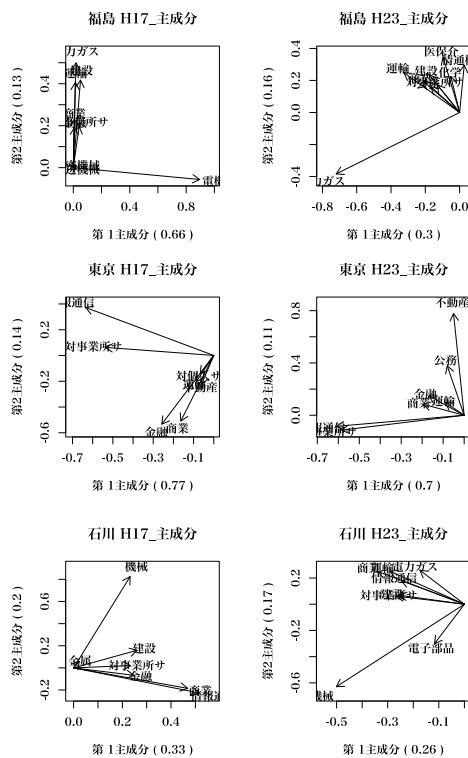


図 13. 産業構造の推移の可視化
(上から福島県，東京都，石川県)

ベクトルの長さに着目して、福島県の場合は、平成 17 年に比べて平成 23 年では、「電力ガス」の影響が大きいことがわかる。東京都の場合は、平成 17 年では「情報通信」と「金融」、「サービス」系に特徴的な傾向がみられたが、平成 23 年では、「サービス」系の増加に加えて、さらに「不動産」も一つの大きな特徴として現れていることがわかる。また、石川県においては、基盤産業の「機械」に加えて、「電子部品」や「電力ガス」も長さが増している。

6. 考察

産業連関表における取引基本表内の内生部

門データに対して、データ解析手法を適用した都道府県間の関係や産業構造の変化について可視化を軸に考察する。

5.2.のクラスター分析では、都道府県をデータ間の距離に基づいて分類し、樹形図で表現した。産業を軸としたこの分類では、34 の産業部門間の取引金額(収集データ)と取引構造(標準化データ)をもとに都道府県の関係性を可視化した。収集データを用いて分析する場合、生産額の大きな東京や愛知、大都市を有する都道府県がまとまったクラスターとして存在する形となった。標準化したデータで分析する取引構造の場合は、産業構造の特異な都道府県など、何らかの共通解釈ができるクラスターもあればそうでないクラスターもあった。今回、クラスター間の結合方式として、最短距離法を用いたが、連結方法によってクラスターの違いが若干見られることから、分析の際にいくつか試行して、総合的に判断するとよいだろう。

5.3.の主成分分析を用いた都道府県マッピングでは、取引金額でみた場合、産業部門ごとに額の大きな都道府県が特出するため、全体がわかりやすい。また、取引構造でみた場合は、各産業部門で同じ傾向のある都道府県は近くに、異なる場合は単独で配置される結果となった。

5.4.で示したように各都道府県のデータを主成分分析することによって、産業構造を概観することができ、様々な可視化や都道府県間での比較も可能となる。また、5.5.のように年次で比較すれば、経年変化などに関する直感的な把握

も容易となり、そのポイントから、産業構造の変化に関する詳細な要因について探っていくことも可能なのではないだろうか。

7. おわりに

データ解析者の立場から、産業連関表の新たな活用法として、主成分分析やクラスター分析を適用し、それらの可視化に関して分析の有効性や可能性を検証した。分析結果における詳細な解釈は、経済学の立場から行うべきであり、現段階では、まだ十分できていない。しかし、本研究では、産業連関表に関する個々のデータに興味があるのではなく、全体の経済構造や都道府県間の関係性に関心があるため、厳密な数値よりは、可視化というツールで全体の把握に徹してきた。今後は、産業連関表からさらに有益な情報を抽出するために、経済や産業に関する専門家の意見や知識を取り入れつつ、適切な分析手法を選択あるいは確立できるよう研究を進めていきたい。

【参考文献】

- 1) 山本けい子・寒河江雅彦「産業連関表に基づく地域クラスタリング分析の試み」『2016年度統計関連学会連合大会講演報告集』,2016,pp.176.
- 2) 入江啓彰「2011年産業連関表から見た都道府県の産業構造」『産研論集』第45号,2018,pp.35-43.

- 3) 原田魁成「産業連関構造における地域特性抽出のための非負値行列因子分解法に関する研究」『金沢大学大学院人間社会環境研究科修士学位論文』2019

- 4) R Core Team "R: A language and environment for statistical computing", *R Foundation for Statistical Computing*, 2018,Vienna,Austria.

URL <https://www.R-project.org/>.

- 5) 永田靖, 棟近雅彦『多変量解析法入門』,サイエンス社, 2001.

【参考資料】

各都道府県産業連関表ホームページ

総務省産業連関表ホームページ