

地域間産業連関表の空間詳細化 にむけた統計的機械学習の応用

村上 大輔¹・高山 雄貴²

¹統計数理研究所 データ科学研究系 (〒190-8562 東京都立川市緑町 10-3)

E-mail : dmuraka@ism.co.jp

²金沢大学 理工学研究域 (〒305-8506 石川県金沢市 920-1192)

E-mail : ytakayama@se.kanazawa-u.ac.jp

地域ブロック間の産業動態を記述する表である地域間産業連関表は約半世紀にわたって整備され政策検討・立案などに役立てられてきた。しかしながら同表には次の課題がある：(i)公式な表は平成17年までしか存在しない；(ii)地域ブロック単位のため都道府県間や市町村間のような空間詳細な産業動態は把握できない。幸い、近年発展の著しい統計・機械学習モデルの将来推計や空間詳細化に対する有用性が示されてきた。また、過去の地域間産業連関表や、都道府県・市町村内の産業連関表、事業所の空間分布といった幅広い関連データが利用でき、それらを用いたモデルの精緻化が可能である。以上を踏まえ、統計・機械学習モデルを地域間産業連関表の空間詳細化に応用する。それにより、同表の推計における統計・機械学習モデルの有用性や限界を明らかとする。

Key Words : *Input-Output table, Downscale, Gradient boosting*

1. はじめに

地域間産業連関表は、地域間の経済的なつながりや地域毎の経済活動状況を把握する目的で昭和45年以降、経済産業省によって5年毎に作成されてきた。同表は地域間波及効果の分析や政策検討を行う上で有用である。

一方で、同地域間表には次の課題がある。まず(i)業務の合理化、公表の早期化等の観点から平成12年以降、地域間表の作成・公開・中止が検討されはじめ、平成17年を最後に地域間産業連関表は公開されていない。また(ii)同地域表は全国9ブロック毎のマクロな取引を要約したものであり、より詳細な都道府県間の取引をこの表から把握することはできない。

幸い、課題(ii)に関しては宮城ら(2003)、菅(2019)、新井(2020)らによって地域間産業連関表から都道府県毎や市区町村毎の産業連関表の推計が試みられてきた。例えば宮城ら(2003)と新井(2020)は、上記地域間表と各都道府県が公開する産業連関表を組み合わせることで都道府県間の産業連関表を推計した。一方、菅(2019)は産業連関表と経済センサスに基づいて市区町村毎の産業連関表を推計した。以上の研究では(地域間)産業連関表に対して何らかの按分を行うことで都道府県間や市区町村間の取引量を推計している。例えば宮城ら(2003)は取引の発地と着地のポテンシャルを表す量を繰り返し計算によって

求め、それらの積で按分比を与えている。新井(2020)は商品流通調査で得られた取引額に基づいて按分比を与えている。

按分比の与え方によっては推定交易量にバイアスがかかることから、この按分比は慎重に与える必要がある(中澤2002)。幸い、按分を含む予測/補間のための手法は統計学や機械学習分野で開発が進められてきた。例えば按分に関しては、面積や人口を用いる基本的な按分法(例えばFischer and Landford, 1995)、回帰を用いて按分比を決める方法(例えばCromley et al., 2012)、空間パターンを元に按分比を決める方法(Kyriakidis, 2004)などが提案されてきた。宮城ら(2003)や新井(2020)でもまた部分的にここで挙げたような按分法が用いられている。

以上に加え、機械学習分野では幅広い関連データとの相関関係を柔軟に考慮した予測手法が幅広く提案されてきた。例えばブースティングやニューラルネットはその一例である。特に前者は比較的小標本でも高い予測精度が期待でき産業連関表のような比較的小規模な標本のモデル化にも向いている。ブースティングを用いることで、例えば別の地域間取引データとの相関を考慮した予測や、幅広い説明変数(都道府県別の従業員数や事業所数など)を考慮した柔軟な予測・按分が行える可能性がある。実際、商品流通調査や物流センサスのような関連が深いデータの中には都道府県毎の情報を持つものもあり、

それらとの相関関係を活用することで、高い精度で都道府県間の産業連関表が推計できる可能性がある。

以上を踏まえ、本研究では地域間産業連関表の空間詳細化に対する有用性検証の第一歩として、ブースティングを都道府県間の取引額（産業部門別）の推計に応用し、その精度や有用性・限界を検証する。

2. データ

2.1. 地域間産業連関表

本分析では平成17年地域間産業連関表 (https://www.meti.go.jp/statistics/tyo/tiikiio/result/result_02.html) を用いる。同データは全国9地域（北海道、東北、関東、中部、近畿、中国、四国、九州、沖縄）間・産業部門間の取引関係を要約した表である（詳細は宮城ら2003参照）。今回は、同表内にある同一部門内の9地域間取引額データを用いて、都道府県間取引額（内生部門）を推計する。部門間の取引額の推計については発表会にて報告したい。

対象とする部門は飲食料品、繊維製品、輸送機械とした。各部門の地域間取引額は図1にプロットした通りである。各行列の行は供給側であり、列は需要側である。従って*i*行*j*列は地域*i*で生産されて地域*j*で消費された取引額である。この図から、飲食料品は関東における取引が支配的、繊維製品は関東、中部、近畿の各地方とそれらの間の取引が活発であること、輸送産業は関東、中部、ならびにそれら2地域間の取引が比較的活発であることが確認できる。

2.2. その他の関連データ

地域間産業連関表が都道府県間の取引に関する情報を持たないのに対し、商品流通調査 (<https://www.meti.go.jp/statistics/tyo/ryuutuu/index.html>) では(a)9地域で生産された商品毎に消費地割合（都道府県別）データが公開されている。従って、このデータを用い、9地域から47都道府県への取引

に関する事前学習を行う。この事前学習は産業連関表の飲食料品、繊維製品、輸送機械の各部門に対するものである。飲食料品、繊維製品、輸送機械に関しては、それぞれ商品流通調査の「食料品・たばこ」、「繊維工業製品、衣服・その他の繊維既製品」、「乗用車、その他の自動車、その他の輸送機械・同修理」の各消費割合データを用いることとする。なお、今回はデータが取得できた平成23年の商品流通統計データを用いることとした。

上記に加え、平成17年工業統計調査で公開されている次の変数を都道府県毎の説明変数として用いることとした：(b)事業所数、(c)従業員数、(d)現金給与総額、(e)原材料使用額、(f)製造品出荷額。各変数は供給地域と需要地域の両方について整備する。

さらに次の変数を地理的要因を捉えるための説明変数として用いることとした：(g)供給地域と需要地域が同一であることを示すダミー変数(内々)、(h)地域毎のダミー変数(供給側)、(i)地域毎のダミー変数(需要側)、(j)供給地域と需要地域の地理的重心間の直線距離。

(a)は9地域と47都道府県の間をモデル化するための事前学習の教師データとして用い、(b)~(j)は同事前学習ならびに産業連関表データを用いた学習における説明変数（特徴量）として使用する。具体的な推定手順は次章で説明する。

3. 推計手法

ブースティングとは予測や分類のための学習アルゴリズムの一つであり、誤差が小さくなるようにモデルを逐次的に更新していくことで精度の良いモデルを推定する。数あるブースティングアルゴリズムの中でもExtreme Gradient Boosting (XGB; Chen and Guestrin, 2016) は特に性能の良いアルゴリズムとして知られており、本研究でも同手法を推定に用いる。予測モデルには回帰または決定木を用いることができるが、一般に後者の方が予測性能が良いと言われ

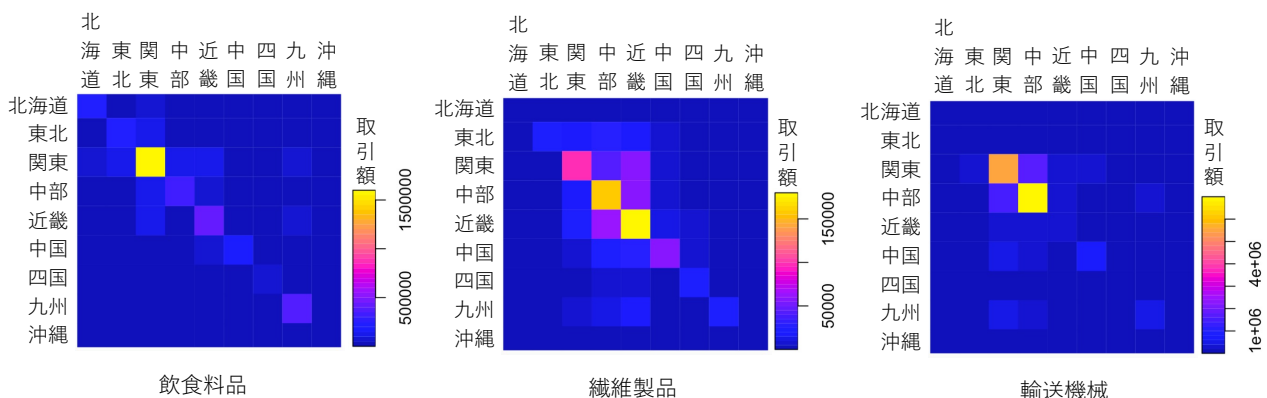


図1：9地域間の飲食料品、繊維製品、輸送機械の各取引額（行：供給側、列：需要側）

ていることから、ここでも決定木を用いることとした (XGBの詳細についてはChen and Guestrin 2016参照)。

本研究では地域間産業連関表の取引額を教師データとした決定木の学習 (逐次更新) をXGBを用いた二乗誤差最小化によって行うことで、都道府県間の取引額の予測モデルを構築する。そのために、まずは9地域と47都道府県の間をモデル化するための事前学習(1)を行う：

- (1) 商品流通調査から得られる取引額を教師データとして9地域×47都道府県間の取引額を予測するモデルをXGBで推定

手順(1)によってモデルの構造が決定される。次に手順(2)で同モデルのパラメータを産業連関表の取引データにあうようにチューニングしたのちに都道府県間の取引額を手順(3)、(4)で推計する：

- (2) 産業連関表の取引額 (3部門の9地域×9地域) を教師データとして、(1)で推定されたモデルをXGBでチューニングする。それにより産業連関表データと整合する予測モデルを構築
- (3) (2)の予測モデルを用いて都道府県間の取引額を推定
- (4) (3)で得られた推定取引額の集計値が産業連関表の実測値 (3部門の9地域×9地域) と一致するように、推計値を地域ペア毎に定数倍

手順(4)は暗に按分比を決めるステップであり、手順(1)～(2)はこの按分比を決めるための計算とみなすことができる。以上の取引額の推定は 部門毎に個別に行った。

4. 分析結果

計算手順(2)で部門毎に評価された説明変数の重要性を表1に整理した。同表1のGainは各説明変数の二乗誤差の最小化に対する重要度を表す指数であり、ここでは上位5変数までを示した。飲食料品に関しては内々ダミーのGainが際立っており、同一都道府県内の取引が活発と推定された。また地域ダミー (需要側)のGainの大きさから取引額には地域差があることや、距離のGainから近隣都道府県との取引が活発なことなどが推定された。

繊維製品に関しても内々ダミーのGainは大きく、同一都道府県内の取引が活発と推定された。また事業所数(供給側)のGainも大きく、多数の事業所を有する都道府県からの供給が活発であることなども確認された。輸送機械に関しては地域ダミーや距離などのGainは大きくなり、地理的要因は必ずしも活発でないと推定された。一方で従業者数、製造品出荷額、原材料使用額、事業所数のGainが大きくなり生産・需要の規模が大きい地域間での取引が活発と

推定された。以上のように部門毎に要因が推定でき、その結果をもとに都道府県間の取引額が推定できる点はXGBの長所の一つである。

図2はXGBから推定された都道府県間の取引額である。参考までに新井(2019)で推定された同取引額を図3にプロットした。両図から、全体としては我々の推定結果は新井(2019)と類似していることが確認できる。

一方、飲食料品に関しては、XGBを用いた場合、東京都や大阪府、福岡県といった主要都府県における内々での推定取引額が新井(2019)よりも大きくなった。これは飲食料品の取引では内々の影響が大きいと推定されたためと考えられる (表1)。繊維製品に関しては、XGBを用いた場合は、より少数の都道府県に取引額が集中している傾向が見られた。輸送産業に関してはXGBも新井(2019)も愛知県内の取引が支配的という推定結果となった。

無論 XGB の推定精度に関しては今後詳細に検討する必要があるが、少なくとも直感に整合する形で部門内の取引額が推定できることを確認した。

5. 考察

本研究では産業連関表の空間詳細化に向けた第一歩として部門別の都道府県間取引額の推定に XGB を応用した。それにより XGB を用いることで直感に整合する形で取引額の推定が行えることを確認した。同手法を用いることで幅広い要因を考慮した柔軟な空間詳細化が可能であることから、今後は物流センサスや他年度の産業連関表など、今回考慮することのできなかつたデータも考慮したモデルの高度化が行いたい。その上で、論文提出までに実施できなかつた部門間の取引額の推定や精度検証についても今後研究を進めた上で産業連関表全体の空間詳細化について検討したい。

謝辞

本研究は国土交通省・道路政策の質向上に資する技術研究開発「公共交通ターミナル整備の空間経済分析に関する研究開発」ならびに、JSPS 科研費 18H01556, 21H03718の助成を受けたものある。

参考文献

- 1) 新井園枝 (2020) 2005年都道府県間産業連関表の作成とその概要. RIETI Discussion Paper Series, 20-J-009.
- 2) 菅幹雄 (2019) 全市区町村産業連関表 (平成23年) の推計. 研究所報, 日本統計研究所, 51, pp.1-14.

表 1：推定された説明変数の重要性

飲食料品		繊維製品		輸送機械	
説明変数	Gain	説明変数	Gain	説明変数	Gain
内々ダミー	0.28	事業所数(供給側)	0.12	従業者数(需要側)	0.26
距離	0.17	地域ダミー(需要側)	0.12	製造品出荷額(需要側)	0.24
事業所数(供給側)	0.11	内々ダミー	0.11	原材料使用額(需要側)	0.20
地域ダミー(需要側)	0.07	現金給与総額(供給側)	0.09	事業所数(需要側)	0.12
原材料使用額(需要側)	0.05	製造品出荷額(供給側)	0.06	現金給与総額(需要側)	0.04

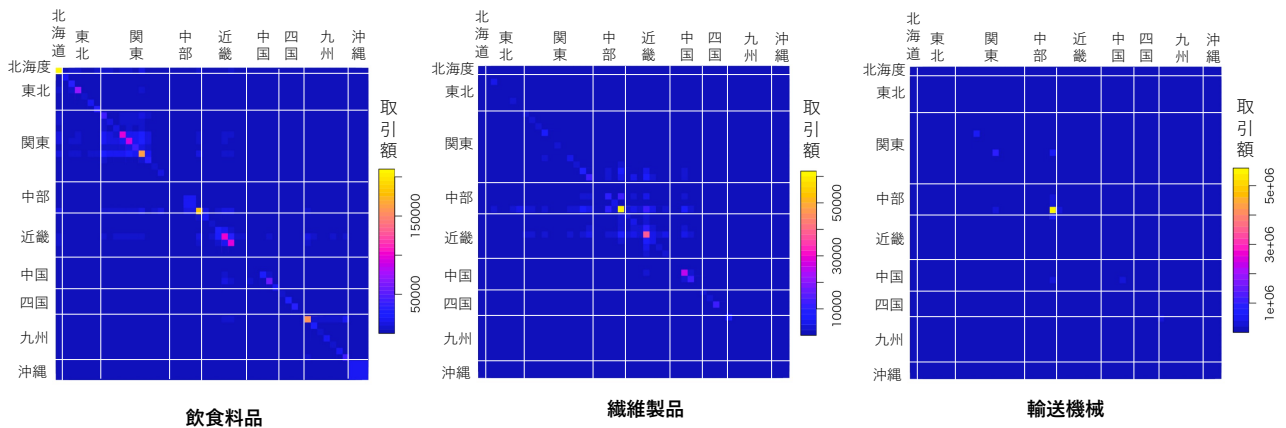


図 2: XGB による都道府県間取引額の結果 (行：供給側, 列：需要側)

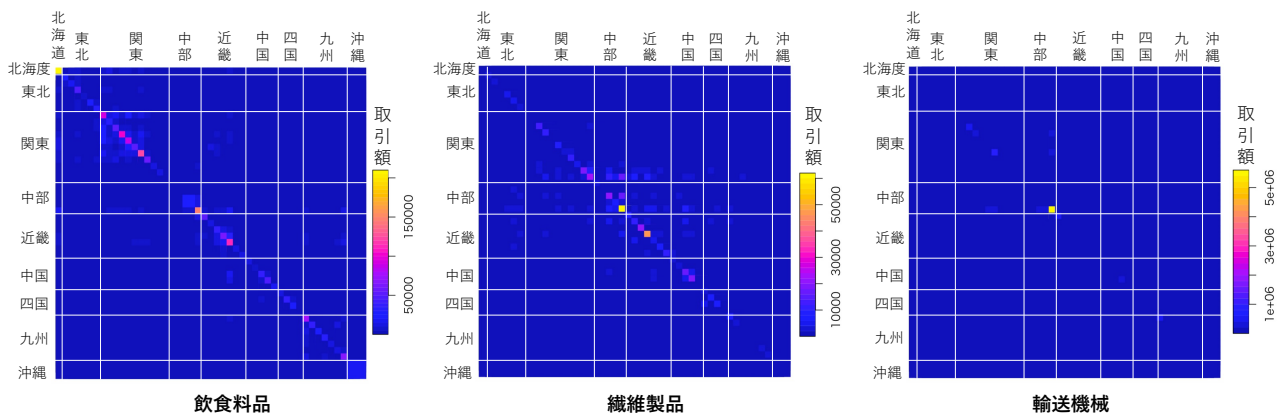


図 3: 新井(2019)による都道府県間取引額の推計結果 (行：供給側, 列：需要側)

- 3) 中澤純治[2002]「市区町村地域産業連関表の作成とその問題点」『政策科学』立命館大学政策科学会, 9(2)(通号 20), p.113~125
- 4) 宮城俊彦, 石川良文, 由利昌平, & 土谷和之. (2003). 地域内産業連関表を用いた都道府県間産業連関表の作成. 土木計画学研究・論文集, 20, 87-95.
- 5) Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- 6) Cromley RG, Hanink DM, Bentley GC (2012) A quantile regression approach to areal interpolation. *Annals of the Association of American Geographers*, 102 (4), 763-777.
- 7) Fisher PF, Langford M (1995) Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and planning A*, 27 (2), 211-224.
- 8) Kyriakidis PC (2004) A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis*, 36 (3), 259-289.