

修士論文

証拠に基づく政策立案のための 潜在プロファイル分析と数法則発見法を用いた 社会実情のモデル化と可視化

Modeling and Visualization of Social Reality
Using Latent Profile Analysis and Number Law Discovery Methods
for Evidence-Based Policy Making

富山県立大学大学院 工学研究科 電子・情報工学専攻

2255013 長瀬 永遠

指導教員 奥原 浩之 教授

提出年月: 令和6年(2024年)2月

目次

図一覧	ii
表一覧	iii
記号一覧	iv
第1章 はじめに	1
§ 1.1 本研究の背景	1
§ 1.2 本研究の目的	2
§ 1.3 本論文の概要	3
第2章 金融市場予測と機械学習の動向	4
§ 2.1 金融時系列データの特性と効率的市場仮説	4
§ 2.2 経済における波及メカニズム	8
§ 2.3 経済分析と説明可能な AI	12
第3章 経済分析の数値モデル	16
§ 3.1 ridge 回帰による時系列クラスタリング	16
§ 3.2 多重共線性対策としての階層的変数クラスタリング	20
§ 3.3 Transformer ベースのシンボリック回帰について	24
第4章 提案手法	28
§ 4.1 時系列クラスタリングによる非定常性の考慮	28
§ 4.2 階層的クラスタリングおよび実験計画法を用いた変数の選択	31
§ 4.3 可読性を考慮したシンボリック回帰	34
第5章 数値実験並びに考察	37
§ 5.1 数値実験の概要	37
§ 5.2 実験結果と考察	38
第6章 おわりに	43
謝辞	44
参考文献	45

図一覽

2.1	RESAS の例（射水市） [14]	9
2.2	金融政策の波及経路 [?]	9
2.3	経済全体の波及メカニズム	11
2.4	SHAP による結果例 [14]	13
2.5	因果探索の結果例 [?]	13
2.6	SHAP による結果例 [14]	14
2.7	因果探索の結果例 [?]	14
3.1	デンドログラムの例	21
3.2	Transformer モデルの概要 [?]	25
4.1	folium ので追加できる機能の例	30
4.2	GIS を用いた潜在的クラスターの描画	30
4.3	学習データの一例	34
4.4	存在確率のカテゴリー化	35
4.5	本システムの画面遷移	36
4.6	実装した GIS	36
5.1	1～7 個のクラスターにおける LPA の BIC	39

表一覧

4.1	本システムのデータベース	34
5.1	実験環境	38
5.2	BIC が最小の回数 (RF6.4 法単体)	39
5.3	BIC が最小の回数 (提案手法)	39
5.4	BIC が最小の回数 (k-means)	40
5.5	数理モデルにおける各変数の意味	40
5.6	各手法における最小の BIC と二乗誤差の平均値	41
5.8	クラスター 3 のサンプル	42
5.9	クラスター 6 のサンプル	42
5.7	クラスター数を変化させたときの最小の BIC と二乗誤差の平均値	42

記号一覧

以下に本論文において用いられる用語と記号の対応表を示す.

用語	記号	用語	記号
LiNGAM の観測変数	x_i, x_j	質的説明変数	\mathbf{q}
x_j から x_i へのパス係数	c_{ij}	量的説明変数	\mathbf{x}
x_i に対する誤差	e_i	目的変数	y
対象 DMU の評価値	z_o	i 番目の回帰ルールのパラメータ	θ^i
各入力に対する重み	\mathbf{v}_{in}^T	RF6.4 法の誤差項	ϵ
各出力に対する重み	$\mathbf{u}_{\text{out}}^T$	回帰ルールの定数項	v_0
対象 DMU の入力	\mathbf{x}_o	回帰ルールの各項に対する係数	v_g
対象 DMU の出力	\mathbf{y}_o	回帰ルールの x_m に対する次数	v_{gm}
DMU 集合の入力	\mathbf{X}	質的説明変数にかかる係数	v_{kl}
DMU 集合の出力	\mathbf{Y}	学習パラメータ	ϕ
x_{n_i} に対する改善案	\hat{x}_{n_i}	学習パラメータの要素	$c_0, c_{gd}, c_{dkl}, v_{gm}$
DMU_{r_d} の n_i 番目の入力	$x_{n_i r_d}$	RF6.4 法の最適探索幅	α
DMU_{r_d} に対する重み	λ_{r_d}	3.3 節の BIC のデータ数	n
y_{n_j} に対する改善案	\hat{y}_{n_j}	3.3 節の BIC の残差平方和	RSS
DMU_{r_d} の n_o 番目の出力	$y_{n_o r_d}$	3.3 節の BIC のパラメータ数	k_p
DMU_{r_d} に対する重み	μ_{r_d}	重み減衰法の正則化係数	λ
混合分布モデルの観測変数	y_i	k-means 法の係数値ベクトル	\mathbf{w}^n
混合分布モデルのパラメータ	Θ	スケーリング後の目的変数	\tilde{y}
各分布における平均	μ_k	目的変数の平均値	$mean(y)$
各分布における分散	σ_k	目的変数の標準偏差	$std(y)$
各分布の構成割合	π_k	スケーリング後の量的説明変数	\tilde{x}
観測変数ベクトル	\mathbf{y}_i	各量的説明変数の最大値	$max(x)$
各分布における共分散行列	$\Sigma_{\mathbf{k}}$		
分布の平均のベクトル	μ		
分布の共分散行列	Σ		
各分布の構成割合のベクトル	π		
LPA における潜在変数	z_{ik}		
LPA における潜在変数ベクトル	\mathbf{z}		
3.1 節の BIC の最大対数尤度	L		

はじめに

§ 1.1 本研究の背景

近年、世界各国の政府を中心に証拠に基づく政策立案（Evidence-Based Policy Making: EBPM）に対する取り組みの重要性が説かれている。EBPMとは、政策の立案をその場限りのエピソードに基づいて行うのではなく、政策によって改善したい対象を明確化したうえで、対象に関するデータを可能な限り収集し、合理的根拠に基づいて意志の決定を行うという考え方である [1]。EBPMを推進することは、政策の有効性を高め、国民の行政への信頼確保につながるとされる。

現在、日本政府におけるEBPMの取り組みとして、2017年の官民データ活用推進戦略会議の決定のもと内閣府によってEBPM推進委員会が発足され、内閣府の各部局によってEBPMの推進が図られている。また、EBPMを「科学的根拠に基づいた政策立案を推進する、アカデミズムと政治領域にまたがった運動」[2]と定義する論文もあることから、EBPMは単に行政のみが取り組むべき事柄ではなく、大学や民間の研究機関などと連携し、専門知識を活用しながら解決すべき課題であると考えられる。

特に効果的なデータ分析や適正な政策評価という観点では大学等の研究機関の寄与するところが大きく、現在の日本におけるEBPMに対する取り組みについての考察 [3] やエビデンスの質について言及し、システマティック・レビューを最も重要と位置づける書籍 [4] などEBPMに関する文献はさまざまな研究分野に属する研究者から出版されている。

また、特定の地域に対して議論を行う際にその地域の特色を正確に把握するためには、その地域が日本全体の中でどう位置づけられるかという視点を用いることが重要であるとされ、統計指標を用いて各地域の特徴を分析している研究もある [5]。

以上のように、近年、日本において政府が積極的に推進し、研究機関においても多くの分野で多面的に考察がなされているEBPMであるが、現在でも全ての自治体、全てのケースにおいてEBPMに基づく意思決定を行うということは極めて困難である。そのため、現場における政策決定のいくらかは住民から行政機関に寄せられる問題に対して対面処理的な対応を行うエピソードベースの意思決定が用いられる。

このような背景には様々な課題が考えられるが、その中でも特に以下の2つが課題であると考えられる [6]。一つは政策における目的と手段の間に成り立つ関係を明確化することである。解決すべき目的とそれに対する手段である政策との論理的な関係性を示すことができない場合、政策の実施が課題解決にどうつながるのかを議論することが難しく、住民からの理解も得られにくい。

もう一つは、収集したデータを統計的手法に基づいて分析し、政策の実施と無関係の要

因を取り除いた政策本来の効果を求めることである。政策立案の対象となるフィールドは様々な社会情勢の影響を受けているため、それらの影響を可能な限り排除した政策本来の効果を求めることは政策の有効性を議論するうえで非常に重要である。

§ 1.2 本研究の目的

1.1 節では、EBPM における研究動向と課題について述べた。これらに対し、日本では政府によってその推進に向けたアウトラインが作成されるとともに、様々な取り組みがなされている。しかし、政策立案の分野が扱う対象は非常に広範であり、様々な視点からの分析や知見が必要とされる。

そこで、本研究では前述の要素のうち、地域特性の分類と目的・手段間の関係性の2点に着目し、統計分析の手法を用いた新たなデータ分析システムを提案することでその解決を目指した。各要素に対する本研究のアプローチの概要を以下に示す。

1. 地域特性を捉える手段として、行政が持つ統計データを用いて、その背景に存在する潜在的要素を分析し、その結果を考慮して自治体をいくつかのクラスターに分ける手法を適用する
2. 目的・手段間の関係性を明確にするための指針の一つとして、データ間に成り立つ関係性を数理モデルとして表す手法を提案する

具体的な手法として、潜在的な要素を考慮した自治体のクラスタリングには潜在プロフィール分析 (Latent Profile Analysis: LPA) [7] という手法を用いた。LPA では、実世界において観測可能なデータ (観測変数) がその背後に存在する観測不可能なカテゴリー変数 (潜在変数) の影響を受けて顕在化していると仮定する。そのうえで、尤もらしい潜在変数の特徴を観測変数から推測することでデータのクラスタリングを行う。

次に、データ間に成り立つ関係性を数理モデルとして表す手法として、パーセプトロンの学習を用いた多変量多項式回帰の一つである RF 法 (Rule extraction method from Fact) [8] を用いた。RF 法では、複数の説明変数と単一の目的変数の間に成り立つ関係を多項式の数理モデルとして表すことができる。また、行政が持つ統計データの特徴をより正確に捉えるために、前述の LPA の結果を考慮して RF 法の精度を向上させる手法について提案する。

また、これら二つの手法における分析結果を表現する手法として、地理情報システム (Geographic Information System: GIS) [9] を用いた Web アプリケーションシステムの開発を行った。ただし、システムの基本的な構造は本研究と同様に EBPM の推進に向けたデータ分析手法の提案を行った先行研究 [10] のものを踏襲し、そのシステムに機能を追加するという形で開発を行った。

最後に、本研究で提案する手法の精度を検証するために、オープンデータサイトを用いて収集した実際の統計データを用いて数値実験を行った。また、その結果について考察を行い、提案手法をより効果的に用いるために考慮すべき事柄について言及する。

§ 1.3 本論文の概要

本論文は次のように構成される。

- 第1章** 本研究の背景と目的について説明する。背景では、EBPMの重要性と日本国内での広がり、適切に導入する際に障壁となる課題について述べる。目的では、背景で述べた課題の中で本研究の対象としたいものを取り上げ、その解決に向けた新しいデータ分析手法を提案することについて述べる。
- 第2章** EBPMの概要とよく用いられる手法について解説し、それらを効率的に行う上でのICTの重要性に言及する。また、EBPMの推進に向けたデータ分析・可視化システムに関する先行研究を挙げ、その課題について述べる。加えて、本研究の目的である自治体のクラスタリングとデータの関係性のモデル化について、代表的な手法についてその概要を示す。
- 第3章** 本研究の提案手法を構築するにあたって参考としたデータ分析手法の先行研究を挙げ、一般的な理論について解説する。
- 第4章** 第3章で挙げた分析手法を用いて自治体が属する潜在的なクラスターを分析する手法を提案する。また、その結果を考慮した数法則発見法を提案し、統計データの間に成り立つ関係を数理モデルとして表す手法を示す。加えて、これらの結果をGISで提示するシステムの開発について、システムの概要と画面遷移を説明する。
- 第5章** 提案手法における精度の検証を目的として行った数値実験について、その概要と結果を示す。また、結果に対して考察を行い、提案手法において考慮すべき事柄について言及する。
- 第6章** 本研究に関する内容を簡潔にまとめ、本研究において実現できたことと今後の展望を示す。

金融市場予測と機械学習の動向

§ 2.1 金融時系列データの特性と効率的市場仮説

経済学およびファイナンスの分野において、実証分析の基礎となるのは、現実の市場活動から生成されるデータの性質を正確に理解することである。特に為替レートや株式収益率といった金融時系列といった金融時系列データは、物理法則に従う自然科学のデータとは異なり、人間の意思決定の集積によって生成される社会科学的データである。これらは非定常性やボラティリティの変動、ファットテールといった、一般的な統計モデルとは異なる特有の性質を有している。本節では、経済変数の基礎的な物理的定義から始め、時系列データの分解手法、金融市場データ特有の統計的性質および市場価格の形成メカニズムに関する理論的枠組みである効率的市場仮説について、先行研究に基づき包括的に述べる。

経済変数の数理的定義と分類

経済時系列データは数理的に、ある経済変数 X について時間 t を添字として観測された一連のデータ列 $\{X_t\}_{t \in T}$ として定義される。数理統計学の観点から、これらは確率空間 (Ω, \mathcal{F}, P) 上で定義された確率過程の一つの実現値として捉えることができる。ここで、 Ω は標本空間、 \mathcal{F} は事象の集合体、 P は確率測度を表す。確率過程 X は、時間 t と事象 ω の関数 $X(t, \omega)$ として記述され、これは無数にあり得た可能性の中から確率的に選択された一つの結果であると解釈される [?]. 時系列分析における本質的な課題として、推測したい母集団の性質が確率過程全体の性質であるのに対し、利用できる情報が単一の実現値から得られる時間平均に限定される点にある。本来、ある時点 t における確率過程の期待値 μ_t は、すべての事象 $\omega \in \Omega$ にわたる積分として以下のように表現される。

$$\mu_t = E[X_t] = \int_{\Omega} X(t, \omega) dP(\omega) \quad (2.1)$$

一方、我々が実際のデータ解析において算出可能な統計量は、観測期間 T に基づく時間平均 \bar{X}_T である。

$$\bar{X}_T = \frac{1}{T} \sum_{t=1}^T X(t, \omega_0) \quad (2.2)$$

確率過程がエルゴード性を満たす場合、観測期間 T が十分に大きければ、この時間平均はアンサンブル平均に確率収束することが保証され、以下の関係が成り立つ。

$$\lim_{T \rightarrow \infty} \bar{X}_T = E[X_t] = \mu \quad (\text{a.s.}) \quad (2.3)$$

このエルゴード性が成立して初めて、過去の時系列データから母集団の統計的性質を推定し、将来予測を行うことが数理的に正当化される。しかし、経済システムにおいては構造変化などが頻繁に生じるため、この性質が常に保障されずこれが経済予測を困難なものにしている。

経済変数の分類と定義

経済変数を時系列モデルに導入する際、その変数が持つ時間的な属性と評価基準を厳密に区別することは、モデルの整合性を保つ上で不可欠な要件である。

ストックとフローの数理的關係

最も基礎的な分類はストックとフローの二分法である。これらは連続時間において、互いに微分・積分の關係にある変数として定義される。

マネーストックや対外純資産などに代表される、ストック変数 $S(t)$ は、ある特定の一時点 t において計測される貯蔵量であり、状態変数としての性質を持つ。一方、フロー変数 $F(t)$ は GDP や投資額など、ある期間 Δt 内に生じた流量であり、ストックの時間変化率として定義される。数理的には、時刻 t_0 から t までのストックの変化は、その期間のフローの累積として記述される。

$$S(t) = S(t_0) + \int_{t_0}^t (F(\tau) - \delta(\tau)) d\tau \quad (2.4)$$

ここで $\delta(\tau)$ は減価償却などの減耗分を表す。離散時間の実証分析においては、これは以下のような差分方程式として表現されることが一般的である []。

$$K_t = (1 - \delta)K_{t-1} + I_t \quad (2.5)$$

時系列分析において、これら次元の異なる変数を同一の連立方程式系に組み込む際は、階差をとってすべてをフローに統一するか、あるいはフローを蓄積してストック化するなどの前処理を行い、変数の定常性と次元を整合させる必要がある。

名目値と実測値

多くの経済変数は貨幣単位で測定されるが、貨幣自体の購買力はインフレーションやデフレーションなどの物価変動によって時間とともに変動する。したがって、観測されたままの名目値 Y_t^{nom} は数量的な実態と価格変動が混在したデータである。経済構成や生産活動の実態的規模を評価するためには、物価指数をデフレーター P_t として用い、価格変動の影響を除外した実測値 Y_t^{real} を以下の式で算出する必要がある。

$$Y_t^{real} = \frac{Y_t^{nom}}{P_t} \quad (2.6)$$

新古典経済学の標準的な家庭によれば、長期的には貨幣は中立であり、経済主体の意思決定は名目額ではなく、実質額に依存しているとされる。経済主体が名目値の変化を実測値の変化と混同して行動する現象は貨幣錯覚と呼ばれ、短期的な不均衡の原因となりうるが、合理的な予測モデルの構築においては、実質変数または対数差分を採用し、貨幣的ノイズを排除することが一般的であるとされている [?]

以上の整理からもわかるように、観測された経済データは、必ずしもそのままの形で経済の実態を表しているわけではない。蓄積量としてのストックと流量としてのフローの区別、あるいは貨幣価値の変動を除去した実質値への変換は、数値の背後にある経済活動の本質的な規模や変化を捉えるために不可欠な手続きである。これらの処理を経て初めて客観的かつ合理的な比較・分析が可能となる。

経済時系列の成分分解

経済時系列データの変動は、単一の要因ではなく、性質の異なる複数の要因が複合的に作用することで生成される。Box and Jenkins(1970)や肥後・田中(1998)に代表される伝統的な時系列解析の枠組みにおいて、時系列 t における観測値 Y_t は主に以下の4つの成分に分解して考えられる。

傾向変動 経済の長期的な成長や技術進歩、人口動態などを反映し、データの長期的な方向性を決定する成分。

季節変動

気象条件や決算期や盆正月などの社会的習慣に起因し、1年周期で規則的に繰り返される成分。

循環変動

景気循環に代表される、数年から十数年の周期をもつ波動成分。

不規則変動

上記のいずれにも属さない突発的な事象や測定誤差などのノイズ成分。

多くのマクロ経済変数は、経済規模の拡大に伴ってその変動幅も増大する傾向を持つ。そのため、各成分が加法的に和として構成されるのではなく、以下の式のような乗法モデルとして記述されることが一般的である。

$$Y_t = T_t \times C_t \times S_t \times I_t \quad (2.7)$$

この乗法モデルに対して自然対数をとることで、以下に示すような加法モデルへと変換することが可能となる。この対数変換は変動の振幅を安定化させると同時に、各成分の変化率を線形の関係として扱える利点がある。

$$\ln Y_t = \ln T_t + \ln C_t + \ln S_t + \ln I_t \quad (2.8)$$

このように、時系列データを成分に分解することでデータの構造を把握することが可能となる。しかし、モデルの構築の観点からは、これらの成分が時間を通じて安定しているか、あるいは確率的に変動するかという統計的性質がより本質的な問題となる。

金融時系列データの統計的特性

先に、一般的な経済時系列の成分分解について述べたが、金融資産価格の時系列データは、物理学や一般的な信号処理で扱われるデータとは異なり、複雑かつ特異な統計的性質を有している。これらは金融計量経済学において様式化された事実として総称されており [?], 予測モデルの構築やリスク評価において無視することのできない経験則となっている。以下に最も考慮すべき3つの性質を記述する。

非定常性と単位根過程

金融時系列解析において対処すべき課題の一つが非定常性である。統計的推論の多くはデータの平均や分散、自己共分散が時間推移に依存せずに一定である弱定常性を前提としている。しかし、資産価格の時系列データは、効率的市場仮設が示唆するランダムウォークに近い挙動を示し、短遺恨を持つ非定常過程に従うことが一般的である。Nelson and Plosser(1982)は、主要な米国マクロ経済変数が、確率的なショックが蓄積して形成される確率的トレンドを持つことを実証した[?]。これは、数理的には単位根を持つ家庭として表現され、一度生じたショックの影響が減衰せずに永続的に残ることを意味する。Granger and Newbold(1974)は、このような単位根を持つ非定常な変数同時を用いて回帰分析を行うと、両者に因果関係がなくとも決定係数が高く算出され、統計的優位性が誤認される見せかけの回帰が発生することを指摘した。この問題を回避し、統計的な定常性を確保するため、解析においては現系列 P_t の階差、あるいは対数収益率 $r_t = \Delta \ln P_t$ を用いることが標準的な手続きとされている。

分布の非正規性とファットテール 資産収益率の頻度分布は、多くの古典的な金融理論が仮定してきた正規分布とは明確に異なる形状を示す。Mandelbrot(1963)やFama(1965)は、実際の市場データにおける収益率分布が、正規分布に比べて平均値周辺の頻度が鋭く、かつ分布の裾野が厚いファットテール特性を持つことを示した。この特性は標準的な正規分布モデルが予測するよりも冒頭や暴落といった極端な変動がはるかに高い頻度で発生することを意味する。したがって、平均二乗誤差のような外れ値に敏感な評価指標を用いる場合、極端な値の影響を重度に受けるリスクがあり、ロバストな統計手法や損失関数の設計が最重視される。

ボラティリティ・クラスタリング 収益率系列 r_t そのものの自己相関は、効率的市場仮設が示唆するようにほぼゼロであり、過去の情報から将来の価格変動の方向を予測することは困難である。しかし、収益率の絶対値 $|r_t|$ や二乗値 r_t^2 については、長期にわたって有意な正の自己相関が観測されることが知られている。これは大きな変動の後には大きな変動が続き、小さな変動の後には小さな変動が続くという現象であり、ボラティリティ・クラスタリングと呼ばれる。Engle(1982)はこの現象を、誤差項の分散が時間とともに変動する条件付き不均一分散としてモデル化した。この事実は、市場予測において短夏価格水準の推計だけではなく時間変動するリスクの構造的返還を考慮する必要性を示唆している。

効率的市場仮説と予測可能性

金融市場における価格形成メカニズムと予測可能性を論じるうえのパラダイムとして効率的市場仮説 (Efficient Market Hypothesis: EMH) が存在する。Famaによって体系化されたEMHは、市場価格は利用可能なすべての情報を完全に反映しているということを示すものであり、EMHが成立する市場において、現在の価格は将来の収益に影響を与えうる情報をすでに織り込んでいるため、投資家が過去のデータや公開情報を分析して市場平均を上回る超過収益を継続的に獲得することは不可能とされる[?]。Famaは価格に反映される情報の範囲に基づき、市場の効率性をウィーク型、セミストロング型、ストロング型の3つに分類した。この仮説が強固に成立する場合、将来の価格変動は予測不可能なランダムウォーク

に従うこととなり、過去のデータを用いた予測モデルによる超過収益の獲得は原理的に否定される。しかし、完全な効率性は論理的な矛盾を孕んでいる。Grossman and Stiglitz(1980)は情報の収集や分析にコストがかかる以上、市場が常に完全に効率的であれば、だれもコストを分析しなくなり、結果として価格は情報を反映しなくなるという Grossman-Stiglitz のパラドックスを提唱した。この議論は市場機能が維持されるためには、情報の分析に対する報酬としての均衡のとれた非効率性が常に存在しなければならないことを示唆している。さらに近年では、Lo(2004) が市場を生態系としてとらえる適応型市場仮説を提唱しており、機械学習などの高度な計算手法を用いることで従来の線形モデルでは捉えきれない局所的な非効率性や裁定機会を発見できる可能性が理論的に支持されている。

§ 2.2 経済における波及メカニズム

現代の経済システムは、実体経済変数と金融市場変数の多くが相互に依存しあう複雑なネットワーク構造を有しており、ある市場で生じた変動は必ずしもその市場内では完結するわけではない。したがって、特定の市場や変数で発生した外生的なショックは、その発生源にとどまることなく、直接的または間接的な経路を通じてシステム全体へと伝播する。この現象は経済における波及と呼ばれ、政策立案やリスク管理における中心的な課題となっている。

Sims(1980) は、経済変数の動学的な相互依存関係を記述するためにベクトル自己回帰モデル (VAR) モデルを提唱し、ある変数へのショックが時間の経過とともに他の変数へどのように波及するかを示すインパルス応答関数の概念を確立した [?]. この枠組みによれば経済システムは動的な均衡状態にあるのではなく、常に新たなショックを受け、それに対する調整過程を繰り返す動的なシステムとして捉えることができる。

本節では、代表的な経済波及メカニズムとして、金融政策の実体経済への波及プロセス、企業間の生産ネットワークを通じた構造的波及について述べる。また、これらの波及メカニズムが為替レートなどの金融市場にどのような影響を与えるのかについても議論する。

金融政策の波及経路と構造変化

こうした動的な依存関係が最も顕著に現れるのが、金融政策の実体経済への波及プロセスである。日本銀行 (2021) は、金融緩和の効果が企業行動や実体経済へ波及する主要なメカニズムとして、アベイラビリティ、金利、為替レートという 3 つのチャンネルを提示している。

波及経路の全体像と主要チャンネル

図 2 に示すように、金融政策は複数の経路を通じて、設備投資や個人消費に代表される国民民間需要を刺激する構造を持つ。まずはアベイラビリティ・チャンネルについて述べる。これは、情報の非対称性や契約の不完備性が存在する現実の金融市場において、企業が調達可能な外部資金の量が投資行動を制約するメカニズムを指す。Bernanke and Blinder (1988) の信用チャンネル理論によれば、金融引き締めは銀行の貸出供給能力を低下させ、銀行借入に依存する中小企業の資金調達を制約することで実体経済を押し下げる [?]. 日本企業の財務データを用いた分析によれば 2000 年代以降、企業の内部留保の蓄積により平時におけるこのチャンネルの重要性は相対的に低下しているも

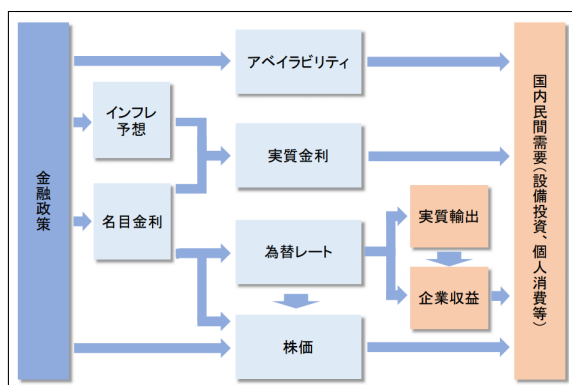


図 2.1: RESAS の例（射水市） [14]

の、リーマン・ショックのような金融危機時には不確実性の増大に伴う予備的流動性需要に対し、中央銀行が流動性を供給することで企業の資金繰りを支える決定的な役割を果たすとされる。これは、市場環境によって波及経路の有効性が変化することを示唆している。 [?].

金利チャネルと資産価格形成の多層的メカニズム

金利チャネルは経済学における伝統的な金融政策の波及メカニズムであり、政策金利の変更が市場金利や予想物価上昇率を実質金利を変化させ、企業の資本コストに影響を与える経路である。日本銀行の分析によれば、金利低下が経済活動を刺激する効果のうち、資金調達コストの低下を通じた直接的な効果は約3割であり、5割強は株価や為替レートといった金融資本市場を経由した間接効果であると推計されている。すなわち、金利というマクロ変数は資産価格という中間変数を媒介として多層的に実体経済へ波及する構造を有しているということがわかる。また、イールドカーブ・コントロールにより長期金利を低位安定させる政策は、回収期間の長い研究開発投資やインフラ投資を下支えし、長期的な洗剤成長率に働きかける経路としても機能する。

為替レート・チャネル

開放経済モデルにおいて看過できない経路として、為替レートチャネルが挙げられる。金融緩和に伴う金利低下は、裁定行動を通じた資本流出を招き、自国通貨安をもたらす。この為替レートの変動は、相対価格の変化を通じて実体経済に複合的な影響を及ぼす。自国通貨安は輸出材の現地通貨建て価格を引き下げ、価格競争力を高めることで輸出数量を増加させる。同時に、外貨建て資産の円換算額が増加することで、企業の場合タンスシートや収益を改善させる効果も期待される。一方で、輸入材価格の上昇は、家計の実質購買力の低下や内需型企業のコスト増大を招くため、為替レートチャネルの総体的な効果は輸出の数量効果と交易条件の悪化による所得効果のバランスによって決定される。

しかし、日本銀行の実証分析によれば、近年の日本経済において、この波及メカニズムの構造変化が観測されている。図3に示す事変パラメータVARモデルの推計結果は、為替レートの変動が実質輸出に与える感応度が2008年のリーマンショック以降、構造的に低下していることを示している。かつては円安が輸出数量の増加を強く

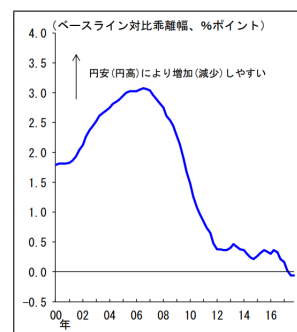


図 2.2: 金融政策の波及経路 [?]

牽引したが、2010年代以降、その波及効果は減衰傾向にある。この背景には日本企業のグローバル・サプライチェーン戦略の転換がある。円高局面での海外生産シフトや、輸出品の高付加価値化が進んだ結果、企業は為替変動に対して輸出数量を調整するのではなく、現地通貨建て価格を据え置く行動をとるようになった。その代わりとして海外現地法人からの配当金受取や直接投資収益の拡大を通じた波及経路が台頭している。

これによって、円安は輸出数量の増加よりもむしろ海外収益の円換算額の増加を通じて企業収益を押し上げるメカニズムへと変化している。

すなわち、為替レートという入力変数が経済システムへ及ぼす出力の形式は、単純な数量効果から金融的な所得効果へと構造的に変化しているのである。これは経済予測において為替レートの影響をモデル化する際、過去の相関関係かを単純に適用する線形モデルでは捉えきれない非定常性が存在することを示唆している。

生産ネットワークを通じた構造的波及

金融政策のようなマクロ的なショックに加え、実体経済内部の取引関係を通じたミクロ的な波及メカニズムもまた、経済理解するうえで不可欠な要素である。企業の生産活動は、原材料の調達から加工、組み立てに至るまで、複雑なサプライチェーン・ネットワークの上に成り立っている。そして、特定の企業や産業で発生した局所的なショックは、このネットワークを通じて増幅および拡散される可能性がある。

マクロ経済学において、個別の企業や産業に生じた微小なショックは、経済全体でみれば互いに相殺されるか、軽微なものとして無視できるという前提が一般的である。これは、Hulten(1978)の定理として知られ、完全競争市場と一定の規模の収穫を仮定する場合、ある産業への生産性ショックがGDPに与える影響は、その産業の売り上げシェアに等しいとされる。つまり、シェアの小さな産業の変動がマクロ経済全体の変動要因にはなりえないという立場である。しかし、近年のネットワーク理論は、この従来の見解に対して重要な修正を加えている。Acemoglu et al.(2012)は、産業連関ネットワークにおいて特定のハブ産業が存在する場合、微小なショックが相殺されずに伝播し、マクロ経済全体に無視できない変動を引き起こすことを理論的に示した。これは粒状性とネットワーク効果によって説明され、経済システムの構造そのものが引き起こすことを理論的に示すものである。ここで粒状性とは、経済システムにおいて大規模な企業や産業が存在し、それらがシステム全体の動態に対して非対称な影響力を持つ性質を指す。さらに、Baqae and Farhi(2019)はこの議論を拡張し、生産関数に代替の弾力性が低い投入物が含まれる場合、ボトルネックとなる産業への負のショックがネットワークを通じて非線形に増幅され、経済全体に甚大な損失をもたらすことを数学的に証明した。

こうした理論的予測は実際の災害データを用いた実証研究によっても強く支持されている。Carvalho et al.(2021)は、2011年の東日本大震災における詳細な企業間取引データを用いて被災地の企業から直接的な取引関係のない地域の産業へとショックが連鎖的に波及する様子を分析した。これによると、震災による直接的な被害が日本経済全体の成長率を押し下げた効果に加え、サプライチェーンを通じた間接的な波及効果がその被害をさらに拡大させたことが確認されている[?]. 具体的には、被災企業方の部品供給が滞ることで、直接の取引先だけでなく、その先の2次、3次の取引先の生産活動まで連鎖的に停滞する現

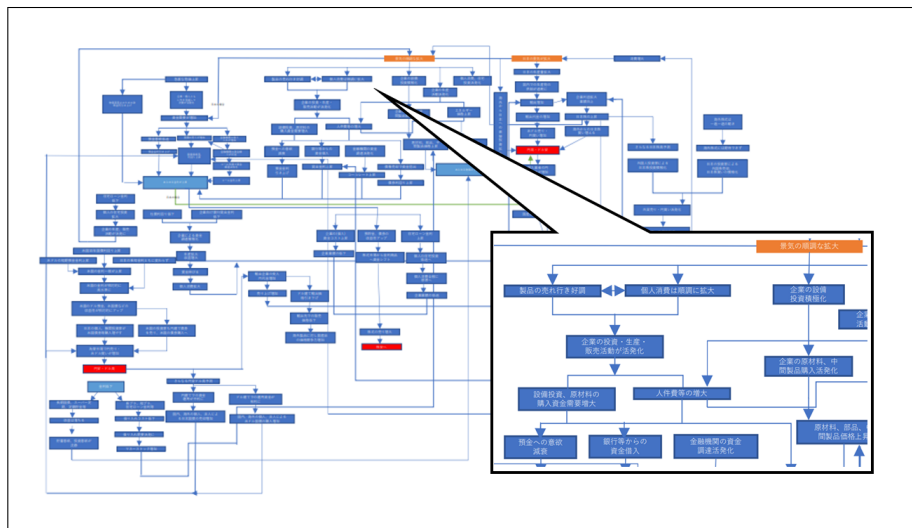


図 2.3: 経済全体の波及メカニズム

象や、逆に納入先の被災によって生産調整を余儀なくされる現象が観測された [?]. これらの事実は、経済予測モデルの構築において、単にマクロ指標の時系列変化を追うだけでは不十分にあることを示唆している。

生産ネットワークのグローバル化とマクロ経済への帰結

生産ネットワークの重要性は、国内市場にとどまらず、国境を越えたグローバル・バリュー・チェーン（GVC）の拡大という形でも現れている。このミクロレベルでの生産構造の変化は、結果としてマクロ経済変数である為替レートの波及メカニズムそのものを変質させている。

伝統的なマクロ経済理論では、円安は輸出価格の低下を通じて競争力を高めるとされる。しかし、Amiti et al. (2014) が指摘するように、現代の企業行動においては、大規模な輸出企業ほど原材料や中間財の輸入比率が高いという「自然ヘッジ」の構造が存在する [?]. すなわち、円安で輸出競争力が高まっても、同時に輸入コストが増加するため、企業レベルでの利益マージンや価格設定は為替変動に対して不感応となる。

こうしたミクロレベルでの企業行動の変化は、マクロレベルでの統計的性質にも明確に表れている。Ahmed et al. (2015) は 46 カ国のパネルデータ分析により、GVC への参加度が高い国ほど、為替レート減価による輸出数量の押し上げ効果が有意に低下することを実証した [?]. これは、前述した日本銀行（2021）が示した「輸出の為替感応度の低下」というマクロ的な構造変化が、企業のグローバルな生産ネットワーク構造に起因する必然的な帰結であることを裏付けている。

もっとも、ここまで述べたサプライチェーンの深化による感応度の低下は、現代の為替波及メカニズムが抱える複雑性の一側面に過ぎない。学术界や実務においては、これ以外にも企業の価格設定行動の変化や、市場の不確実性が投資行動に与える非線形な影響など、為替レートの波及効果を左右する多岐にわたるメカニズムが議論されている。こうした多様な論点と、為替レートが実体経済および金融市場へ及ぼす複合的な波及経路を体系的に整理したものを図に示す。

§ 2.3 経済分析と説明可能な AI

先に述べたように、現代の為替市場を含む経済システムは高度な非線形性と構造変化を有しており、従来の線形モデルではそのダイナミクスを十分にとらえきれない。これに対応するため、ディープラーニングなどの高度な機械学習モデルの活用が不可欠となる。しかし、金融分野への適用に当たっては、単なる予測精度の向上だけでなく、その判断根拠に対する透明性が求められる場面も多い。

本節では、金融実務及び規制の観点から、AI モデルにおける説明責任の重要性と、従来の機械学習モデルが抱えるブラックボックス問題について述べる。次に、モデルの予測精度と解釈性の間に存在するトレードオフの関係について整理する。そのうえで、近年注目されている説明可能な AI の手法についても概説する。

金融における説明責任

近年、ディープラーニングに代表される近年の機械学習モデルは、膨大なデータから複雑な非線形パターンを学習し、従来の手法を凌駕する高い予測精度を実現している。しかし、これらの高度なモデルは、数百万から数十億ものパラメータを持つ多層的な演算構造を有しており、入力データがどのようなプロセスを経て特定の予測値に至ったのかを人間が認知・追跡することは極めて困難である。このモデルの不透明性はブラックボックス問題と呼ばれ、高い信頼性が求められる金融分野への適用において最大の障壁となっている。

画像認識や広告配信といった分野とは異なり、金融分野においてはこのブラックボックス性は許容され難い。その主たる理由は、金融機関が顧客や市場に対して負う説明責任の重さにある。金融庁（2021）は「金融業における AI 活用の促進とガバナンスのあり方」と題する報告書の中で、AI の判断プロセスがブラックボックス化することのリスクを指摘し、金融機関に対して説明可能性および透明性を確保するためのガバナンス体制の構築を求めている [?]. 例えば、AI を活用した与信審査や投資判断において、顧客からその理由を求められた際、「AI がそう判断したから」という回答は、金融機関としての受託者責任を果たしたことにはならない。顧客の利益を守り、不当な差別や誤った判断を防ぐためには、予測の根拠となる変数の寄与度や論理的背景を人間が理解可能な形で提示する必要がある。

また、リスク管理の観点からも説明可能性は不可欠である。日本銀行（2018）が指摘するように、リーマン・ショックのような金融危機の局面では、市場環境が過去の学習データとは全く異なる挙動を示すことが多い [?]. ブラックボックス化したモデルは、平時には高いパフォーマンスを発揮したとしても、未知の局面でどのような挙動をとるか予測できず、意図しない巨額損失を引き起こす危険性を孕んでいる。モデルが「なぜその予測を出したのか」というメカニズムが可視化されていれば、専門家である人間がその妥当性を検証し、異常な挙動を事前に察知して運用を停止するなどの介入が可能となる。すなわち、金融分野における説明可能な AI (XAI) の導入は、単なる技術的な改良ではなく、金融システムの安定性と信頼性を担保するための必須要件であると言える。

可読性に優れた分析手法

先に述べたブラックボックス問題と説明責任のジレンマを解消するために、近年、機械学習および統計学の分野では説明可能な AI に関する研究が急速に発展している。

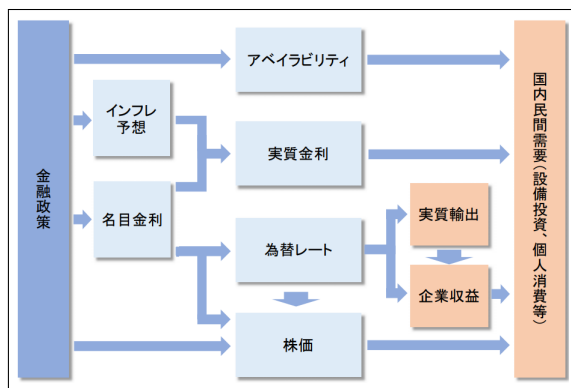


図 2.4: SHAP による結果例 [14]

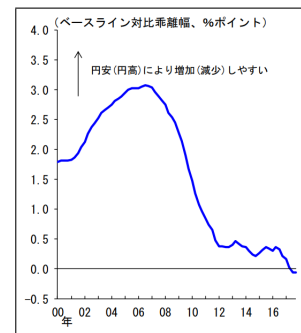


図 2.5: 因果探索の結果例 [?]

Arrieta et al.(2020) によると、ブラックボックスモデルに対する解釈手法はモデルの内部構造を直接ホワイトボックス化する透明性のあるモデル構築と、構築済みのモデルに対して外部から解釈を与える事後的な説明の2つに大別される [?]. 金融時系列解析においては、予測精度を維持しつつ透明性を確保するための主要なアプローチとして採用されてきた。その代表的なものが複雑なモデルの入出力関係を観察し、各変数が予測にどの程度寄与したかを定量化する特徴量重要度に基づくアプローチである。また、因果推論による変数間の因果関係の特定も、単なる相関関係を超えた解釈性を提供する手法として注目されている。以下で、それぞれの分析手法について解説する。

SHAP による要因分解

経済学の実証分析において、ある経済変数の変動を複数の要因に分解して解釈することは、経済における透明性を確保するうえで、重要な役割を果たす。歴史的には、線形近似に基づく分解手法や、中島 (2007) に代表される VAR モデルを用いた時変インパルス応答関数による動的な波及効果の検証が標準的な手法として定着してきた [?]. これらの手法は、パラメータが全期間を通じて一定であるか、あるいはモデル構造に対して強い線形性の仮定を置いている。

しかし、2.2 節でも述べたように、現代の為替市場やサプライチェーンは高度な非線形性と構造変化を有しており、線形モデルに基づく分解では、急激な変動局面における変数の寄与を正確に捉えられきれない可能性が指摘されている [?]. こうした線形モデルの限界を克服し、非線形なブラックボックスも出に対して解釈を与える手法として、Lundberg らが提唱した Shapley Additive exPlanatinos が経済分野で急速に普及している [?].

SHAP は 1953 年に Lloyd Shapley によって提唱された協力ゲーム理論におけるシャープレイ値を機械学習に応用したものである。この理論では、予測モデルへの入力変数をゲームのプレイヤー、予測値をゲームの利益とみなす。そしてプレイヤーがゲームに参加した場合としなかった場合で利得がどれだけ変化するかをあらゆるプレイヤーの組み合わせについて計算し、その加重平均をとることで、各変数の貢献度を算出する [?]. この手法の利点は、加法性という数学的性質を有している点にある。すなわち、全変数の SHAP 値を合計すると、モデルの予測値と完全に一致する。これにより、ブラックボックスモデルの複雑な非線形予測であっても、各変数のプラス・マイナスの寄与として厳密に分解すること

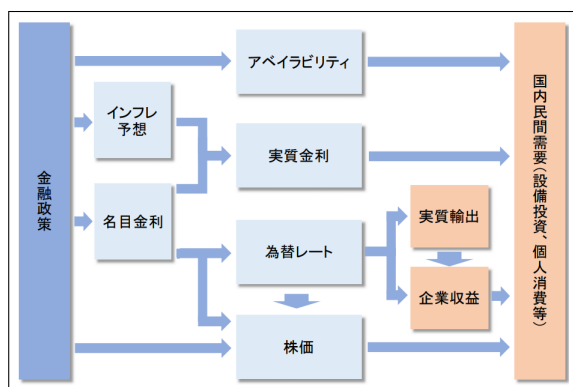


図 2.6: SHAP による結果例 [14]

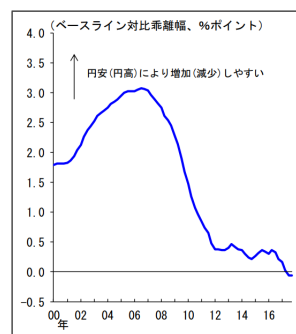


図 2.7: 因果探索の結果例 [?]

が可能となる。

SHAP は経済分野において様々な活用事例が存在するが、その代表的な例として、マクロ経済予測における非線形性の可視化についての研究がある。Joseph らは、英国中央銀行の研究において、インフレ率や金融危機予測に機械学習モデルを適用し、SHAP を用いて変数の重要度を分析した [?]. この研究では、ある変数の値が極端に大きい場合と通常時とで、予測に対する感応度が非線形に変化する様子が SHAP 値の分布によって明らかになっている。

図 12 は、SHAP による要因分解の結果の一例であり、横軸は予測への寄与度、縦軸は各変数を示す。各点はデータサンプルに対応しており、色の濃淡が変数の値を示している。この図を用いることで、単なる変数の重要度ランキングだけでなく、金利が高いとくに予測値を押し下げる傾向があるなどの変数の値の水準に応じた非線形な影響を視覚的に把握することが可能となる。

因果探索による変数間の関係性の特定

また、可読性に優れた経済における要因分析の手法として、統計的因果探索も注目されている。変数間の因果関係をデータのみから推定する統計的因果探索は、長らく困難な課題とされてきた。特に、Shimizu らが提案した Linear Non-Gaussian Acyclic Model (LiNGAM) は変数の誤差項が非正規分布に従うという仮定の下で、因果の方向を一位に識別できる手法として知られている [?].

図 13 は、LiNGAM を用いて因果関係を推定した例である。LiNGAM では出力結果として、観測された経済変数間の因果構造を推定し、有向非循環グラフ (DAG) として可視化されたものを示す。このグラフにおいて、各ノードは分析に用いた経済変数を表し、ノード間を結ぶ有向エッジは変数間の因果的順序を示している。矢印の始点は原因となる変数、終点は結果となる変数を意味しており、従来の相関分析では識別不可能であった変数間の方向性が明示されている。そして、エッジに付された数値は線形構造方程式モデルにおけるパス係数を表している。この値は原因変数が一単位変動した際に、結果変数が直接的に何単位変動するかを示す因果効果の大きさと符号に対応する。様々な分野で活用がなされている LiNGAM だが、経済学の分野においても Moneta らなどがこの手法を用い、SVAR モデルにおける変数の因果的順序をデータ駆動的に決定する研究などが行われている [?].

シンボリック回帰による数式モデルの発見

しかしながら、上記で述べた2つの手法には共通する問題がある。SHAP はあくまでモデルがどの変数に反応したかを示すだけであり、経済構造そのものを特定するものではない。また、変数間に強い相関がある場合、多重共線性により寄与度の配分が不安定になり、誤った解釈を導くリスクがある。また、因果探索は因果の方向を特定できるものの、変数間が具体的にどのような非線形関数で結ばれているかまでは明示しない。

すなわち、SHAP も因果探索も、複雑な経済システムのデータ生成プロセスそのものを記述する数式モデルを提示するには至らない。したがって、経済理論との整合性を数理的に検証し、堅牢な予測を行うためには、これらの手法の利点を踏まえつつ、データから直接的に最適な関数形を特定するシンボリック回帰のアプローチが不可欠である。

シンボリック回帰は、与えられたデータセット (X, y) に対して、その関係性を最もよく記述する数学的な関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ を、関数形式そのものを探索対象として特定する機械学習手法である。一般的な線形回帰や非線形回帰では、モデルの構造を分析者が事前に仮定し、その内部パラメータのみを最適化する。これに対し、シンボリック回帰はモデル構造に対する事前過程を置かず、演算子、変数、定数を構成要素とする数式の空間そのものを探索する。すなわち、最適化の対象は有限個のパラメータではなく、無限の自由度を持つ関数空間となる。遺伝的アルゴリズム (GP) に基づくアプローチでは、数式を演算子をノード、変数や定数を葉とする木構造として表現する。この手法は交叉や突然変異などを用いることで、生物の進化を模倣したプロセスを通じてデータへの適合度が高い数式の集団を世代ごとに進化させることで、最適な関数形を探索する。GP は、関数の形状を事前に仮定しないという点で協力的な手法であるが、実用上はいくつかの重大な課題を抱えている。まずは探索空間が広大であるため、収束までに膨大な計算時間を要する。また、GP は数式の形を見つけることに優れているが、定数や係数の値を正確に決定することが苦手である。そのため、従来の手法では GP で数式の構造を探索したあと、BFGS 法などの非線形最適化手法を用いて係数を別途フィッティングする必要があった。近年、これらの課題を克服するため、深層学習を用いて数式探索を行うアプローチが急速に進展している。初期の研究では、リカレントニューラルネットワークを用いて数式の表現を生成する手法などが提案されたが、これらも依然として係数の最適化には別途プロセスを必要とするものが多かった。

こうした中で、シンボリック回帰の新たなスタンダードとして注目されているのが、Transformers を用いたアプローチである。自然言語処理の分野で革命的な進展をもたらした Transformer は、数式生成タスクにおいても、入力データから数式への翻訳問題として捉えることで、高い性能を発揮することが明らかになった。

経済分析の数理モデル

§ 3.1 ridge 回帰による時系列クラスタリング

時系列データの構造変化を検知し、背後にある潜在的な状態であるレジームを識別する研究は、信号処理、制御工学、計量経済学などにおいて長い歴史を持つ。従来、これらの問題に対してはハミルトンに代表されるマルコフ・スイッチ・モデルや、状態空間モデルにおけるカルマンフィルタリングなど確率的な遷移プロセスを事前仮定するアプローチが主流であった。これらはモデル駆動型のアプローチと呼ばれ、データの生成過程に対して強い仮定を置くことで、パラメータ推定の効率性を担保することができる。しかし、2000年代以降、機械学習技術の発展に伴い、事前の確率分布過程を排除し、データの幾何学的構造から直接的に状態を学習するデータ駆動型のアプローチが台頭した。特にローリング回帰による時変パラメータの推定と、高次元軌道に対するクラスタリングによる離散化を組み合わせる2段階の手法は、その柔軟性と解釈性の高さから、複雑系システムの解析において地位を確立しつつある[?, ?, ?]。しかし、このデータ駆動型アプローチを実際の金融時系列データに適用するうえでは、データ固有の数理的な課題に対処する必要がある。金融・経済データは説明変数間に強い相関を持つことが多く、通常の実最小二乗法をそのまま適用するとパラメータ推定が極めて不安定になるためである。

そのため、ここではこのデータ駆動型アプローチの一例として、Ridge 回帰を用いた時系列クラスタリング手法の数理的性質および定式化について論じる。

Ridge 回帰による時変パラメータの推定

時系列解析において、説明変数間の相関が高い場合や、観測データ数に対して、変数の次元が大きい場合、通常の実最小二乗法による推定は破綻する。本節では、この逆問題としての不安定性を数理的に解説し、正則化手法がもたらす安定化について述べる。

逆問題としての線形回帰と不良設定性

時刻 t における目的変数を $y_t \in \mathbb{R}$ 、説明変数ベクトルを $\mathbf{x}_t \in \mathbb{R}^p$ とし、以下の線形モデルを考える。

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \quad (3.1)$$

観測データ行列 $X \in \mathbb{R}^{T \times p}$ および目的変数ベクトル $\mathbf{y} \in \mathbb{R}^T$ が与えられたとき、最小二乗法による推定量 $\hat{\boldsymbol{\beta}}_{OLS}$ は、残差平方和 $RSS(\boldsymbol{\beta}) = \|\mathbf{y} - X\boldsymbol{\beta}\|_2^2$ の最小化解として、正規方程式 $(X^\top X)\hat{\boldsymbol{\beta}} = X^\top \mathbf{y}$ を解くことで得られる。

$$\hat{\boldsymbol{\beta}}_{OLS} = (X^\top X)^{-1} X^\top \mathbf{y} \quad (3.2)$$

この解が一意かつ安定に存在するためには、グラム行列 $G = X^\top X$ が正定値であり、かつその条件数が十分に小さいことが要求される。条件数 $\kappa(G)$ は、最大固有値 λ_{\max} と最小固有値 λ_{\min} の比で定義される。

$$\kappa(G) = \frac{\lambda_{\max}}{\lambda_{\min}} \quad (3.3)$$

時系列データにおいて説明変数間に多重共線性が存在する場合、説明変数空間の退化により最小固有値は $\lambda_{\min} \rightarrow 0$ に近づき、その結果、条件数は $\kappa(G) \rightarrow \infty$ と発散する。これはジャック・アダマールが提唱した良設定問題の要件のうち、第3条件である、解のデータに対する連続的依存性が満たされないことを意味し、線形回帰問題は数学的に不良設定問題となる。具体的には観測データ y に微小な観測ノイズ ϵ が加わった際の推定量の変動 $\delta\hat{\beta}$ を考えると、その相対誤差の上界は条件数を用いて以下のように評価される。

$$\frac{\|\delta\hat{\beta}\|}{\|\hat{\beta}\|} \leq \kappa(G) \frac{\|\epsilon\|}{\|y\|} \quad (3.4)$$

$\kappa(G)$ が過大である場合、右辺の係数が非常に大きくなるため、わずかな観測ノイズ ϵ が逆行行列演算によって $\frac{1}{\lambda_{\min}}$ 倍に増幅され、推定量に甚大な誤差をもたらす。統計的観点からは、これは推定量の分散 $V[\hat{\beta}_{OLS}] = \sigma^2(X^\top X)^{-1}$ のトレースが発散することを意味し、推定量がデータに対して過学習（Overfitting）を起こし、汎化性能を著しく損なう主因となる。

Ridge 回帰の導出

上述した不良設定性を緩和するため、Hoerl and Kennard は、損失関数に L_2 ノルム制約を付与する Ridge 回帰を導入した [?]. これは、ラグランジュの未定乗数法を用いた以下の制約付き最適化問題として定式化される。

$$\hat{\beta}_{Ridge} = \underset{\beta}{\operatorname{argmin}} (\|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2) \quad (3.5)$$

ここで、 $\lambda > 0$ は正則化パラメータであり、バイアスと分散のトレードオフを制御するハイパーパラメータである。この目的関数 $J(\beta)$ は β について展開し、偏微分を行うと以下の勾配が得られる。

$$\frac{\partial J(\beta)}{\partial \beta} = -2X^\top(y - X\beta) + 2\lambda\beta \quad (3.6)$$

目的関数は凸関数であるため、大域的最適解において勾配はゼロベクトルとなる。 $\frac{\partial J}{\partial \beta} = 0$ と置くことで、以下の正規方程式が得られる。

$$(X^\top X + \lambda I_p)\beta = X^\top y \quad (3.7)$$

これにより、Ridge 回帰推定量は以下の閉形式解として一意に決定される。

$$\hat{\beta}_{Ridge} = (X^\top X + \lambda I_p)^{-1} X^\top y \quad (3.8)$$

この解の幾何学的意味を理解するために、データ行列 X の特異値分解を導入し、 $X = UDV^\top$ と分解する。ここで $D = \operatorname{diag}(d_1, \dots, d_p)$ は特異値 $d_i = \sqrt{\lambda_i}$ を対角成分に持つ行列である。このとき、グラム行列は $X^\top X = VD^2V^\top$ となり、その固有値は特異値の二乗 d_j^2 となる。

これを正規方程式に代入して整理すると、OLS 推定量は正規直交基底 v_j を用いてそれぞれ以下のようにスペクトル分解できる。

$$\hat{\beta}_{OLS} = \sum_{j=1}^p \frac{1}{d_j} \mathbf{v}_j (\mathbf{u}_j^\top \mathbf{y}) \quad (3.9)$$

$$\hat{\beta}_{Ridge} = \sum_{j=1}^p \frac{d_j}{d_j^2 + \lambda} \mathbf{v}_j (\mathbf{u}_j^\top \mathbf{y}) \quad (3.10)$$

上式における係数に注目すると、Ridge 回帰は OLS の各成分に対して、以下の縮小係数を乗じていることと等価である。

$$\frac{d_j}{d_j^2 + \lambda} = \frac{1}{d_j} \times \left(\frac{d_j^2}{d_j^2 + \lambda} \right) \quad (3.11)$$

ここで、括弧内の項 $\frac{d_j^2}{d_j^2 + \lambda}$ は常に 1 以下の値をとり、特異値 d_j が小さい次元ほどその値は 0 に近づく。すなわち、Ridge 回帰はデータ構造における主要な成分はそのまま残しつつ、微小な特異値を持つ不安定な成分に対して選択的に減衰させるローパスフィルタの役割を果たしている。極限として $d_j \rightarrow 0$ の場合、OLS 係数は $1/d_j \rightarrow \infty$ となり発散するが、Ridge 係数は 0 に収束する。この性質により、多重共線性を含む悪条件なデータに対しても数値的に安定した解が得られることが保証される。

ローリングウィンドウ法による動的推定

金融時系列データのように、そのデータ生成過程が経時的に変化する非定常過程においては、全期間を通じて単一のパラメータ β を仮定することは不適切である。そこで、系が短い時間枠内では定常であるとみなす局所定常性を仮定し、固定長の観測窓を用いたローリング過程を適用する。観測期間 T に対し、ウィンドウ幅 $W \in \mathbb{N}$ ($W < T$) を設定する。時刻 $t \in \{W, W+1, \dots, T\}$ における局所的なデータセット \mathcal{D}_t は、直近 W 個の観測値によって構成される。

$$\mathcal{D}_t = \{(\mathbf{x}_{t-i}, y_{t-i}) \mid i = 0, \dots, W-1\} \quad (3.12)$$

この \mathcal{D}_t に対応する計画行列を $X_{[t]} \in \mathbb{R}^{W \times p}$ 、目的変数ベクトルを $\mathbf{y}_{[t]} \in \mathbb{R}^W$ と定義する。時刻 t における時変パラメータ推定量 $\hat{\beta}_t$ は、この局所データに対する Ridge 回帰問題の解として定式化される。

$$\hat{\beta}_t = \underset{\beta}{\operatorname{argmin}} (\|\mathbf{y}_{[t]} - X_{[t]}\beta\|_2^2 + \lambda \|\beta\|_2^2) \quad (3.13)$$

先に導出した正規方程式により、この局所最適解は以下の閉形式で与えられる。

$$\hat{\beta}_t = (X_{[t]}^\top X_{[t]} + \lambda I)^{-1} X_{[t]}^\top \mathbf{y}_{[t]} \quad (3.14)$$

この推定プロセスを時間軸に沿ってスライディングさせることで、パラメータの動的な軌跡を表す時系列集合 \mathcal{B} が生成される。

$$\mathcal{B} = \{\hat{\beta}_W, \hat{\beta}_{W+1}, \dots, \hat{\beta}_T\} \subset \mathbb{R}^p \quad (3.15)$$

行列 B の各要素は、対象システムの構造変化を p 次元の特徴量空間内の遷移として表現したものであり、クラスタリングアルゴリズムへの入力ベクトルとなる。なお、本手法における推定精度はハイパーパラメータ W および λ の設定に強く依存する。

ウィンドウ幅 W は、推定の時間分解能と統計的安定性のトレードオフを決定する。 W が小さいほど急激な構造変化への追従性は向上するが、サンプルサイズ不足により推定分散が増大する。逆に W が大きいほど推定は平滑化され安定するが、変化に対する遅延が生じ、バイアスが増大する傾向にある。したがって、分析の目的に応じて適切なパラメータ選定を行うことが、動的解析の信頼性を担保する上で不可欠となる。

時系列係数ベクトルのクラスタリング手法

前節までに得られた時変パラメータ系列 $B = \{\hat{\beta}_W, \dots, \hat{\beta}_T\}$ は、各時点における市場のリスク変数への感応度を表す p 次元の連続ベクトルである。ここでは、この連続的な軌道を解釈可能な離散状態へと変換するための数理的手法として、K-means 法を用いたベクトル量子化について述べる。

K-means 法は、データ空間を K 個の排反なクラスタ C_1, \dots, C_K に分割し、各領域の代表点（セントロイド）によりデータを近似する手法である [1]。これは、以下の目的関数 J を最小化する最適化問題として定式化される。

$$J = \sum_{t=W}^T \sum_{k=1}^K r_{tk} \|\hat{\beta}_t - \mu_k\|_2^2 \quad (3.16)$$

ここで、 $r_{tk} \in \{0, 1\}$ は帰属変数であり、時刻 t のパラメータベクトル $\hat{\beta}_t$ がクラスタ k に属する場合に 1、そうでない場合には 0 をとる。また、 μ_k はクラスタ k の重心ベクトルである。この大域的最適解を求める問題は NP 困難であることが知られているが、K-means アルゴリズムでは Lloyd のアルゴリズムを用いることで、以下の反復の手順により局所最適解を効率的に求める [2]。

1. 初期化: K 個の初期重心 $\mu_1^{(0)}, \dots, \mu_K^{(0)}$ をランダムに選択する。
2. 割り当てステップ: 各データ点 $\hat{\beta}_t$ を、ユークリッド距離が最も近い重心のクラスタに割り当てる。

$$r_{tk}^{(i)} = \begin{cases} 1 & \text{if } k = \underset{j \in \{1, \dots, K\}}{\operatorname{argmin}} \|\hat{\beta}_t - \mu_j^{(i-1)}\|_2^2 \\ 0 & \text{otherwise} \end{cases} \quad (3.17)$$

ここで、上付き文字 (i) はアルゴリズムの反復回数を表す。

3. 更新ステップ: 各クラスタに割り当てられたデータ点の算術平均を計算し、新たな重心とする。

$$\mu_k^{(i)} = \frac{\sum_{t=W}^T r_{tk}^{(i)} \hat{\beta}_t}{\sum_{t=W}^T r_{tk}^{(i)}} \quad (3.18)$$

4. 収束判定: 重心の変化量が閾値以下になるか、割り当てが変化しなくなるまで割り当てステップと更新ステップを繰り返す。

このプロセスにより、連続値であるパラメータベクトルは、最も近い代表状態 k へと量子化される。得られた重心ベクトル μ_k は、そのレジームを象徴する市場構造を表している。

例えば、ある重心 μ_k が市場ファクターに対して大きな正の値を持つ場合、その期間は市場変動への感応度が高いリスク選好レジームとして解釈することができる。逆に、重心の値が 0 に近い、あるいは低い値を示す場合は、市場変動の影響を受けにくいリスク回避的なレジームとして解釈することが可能になる [?].

以上のように、Ridge 回帰と K-means クラスタリングを組み合わせることで、非定常時系列データに潜む構造変化を効果的に検出し、解釈可能な離散状態へと変換する枠組みが提案されている。

§ 3.2 多重共線性対策としての階層的変数クラスタリング

前節では、観測時点 t 方向の特異性に着目し、パラメータの時変構造を捉えるための時系列クラスタリングについて論じた。一方で、高次元の多変量データを扱う際には、説明変数 j 方向の構造的冗長性を把握することもまた、モデルの堅牢性を確保するうえで不可欠である。

経済・金融時系列データにおいては、多数のマクロ経済指標や市場変数が互いに類似した挙動を示すことが一般的であり、これは設計行列 X の列ベクトル間に強い線形従属関係、すなわち多重共線性をもたらす。数理統計学の観点からは、多重共線性はグラム行列 $X^T X$ の最小固有値を 0 に近づけ、その逆行列 $(X^T X)^{-1}$ の対角成分を発散させる効果を持つ。これにより、パラメータ推定量の分散 $V[\hat{\beta}]$ が増大し、モデルの予測精度および係数の解釈性が損なわれることは 3.1 節で述べたとおりである。

本節では、Ridge 回帰のような正則化項による数値的な安定化とは異なるアプローチとして、変数の従属関係そのものを構造化し、情報の冗長性を排除する階層的クラスタリングの理論的枠組みについて述べる。この手法は、変数間の類似度に基づいて説明変数を階層的にグルーピングし、各グループから潜在的な代表成分を抽出することで、多重共線性の影響を回避しつつ、元の変数空間の情報を効率的に縮約することを目的とするものである。

変数空間における凝集型階層クラスタリング

先に述べた目的を達成するため、確率空間 (Ω, \mathcal{F}, P) 上で定義された p 個の確率変数からなる確率ベクトル $\mathbf{X} = (X_1, \dots, X_p)^T$ を考える。ここで、各変数 X_j は有限の分散を持つ。要するに $X_j \in L^2(\Omega)$ であると仮定する。通常のクラスタリングが観測個体の分割を目的とするのに対し、変数クラスタリングは、確率変数空間 $L^2(\Omega)$ における変数の従属構造に基づいた分割を目的とする。

変数のグルーピングを行うための標準的な手法として、各変数 X_j を初期状態で独立したクラスター $C_j = \{X_j\}$ とみなし、類似したクラスターを順次統合していく凝集型階層クラスタリングが広く用いられる。このアルゴリズムは、以下の反復的プロセスにより定式化される。

1. 初期化:

各変数 X_j ($j = 1, \dots, p$) を単一の要素を持つ初期クラスター $C_j = \{X_j\}$ とし、初期クラスター集合 $\mathcal{S}^{(0)}$ を定義する。

$$\mathcal{S}^{(0)} = \{C_1, C_2, \dots, C_p\} \quad (3.19)$$

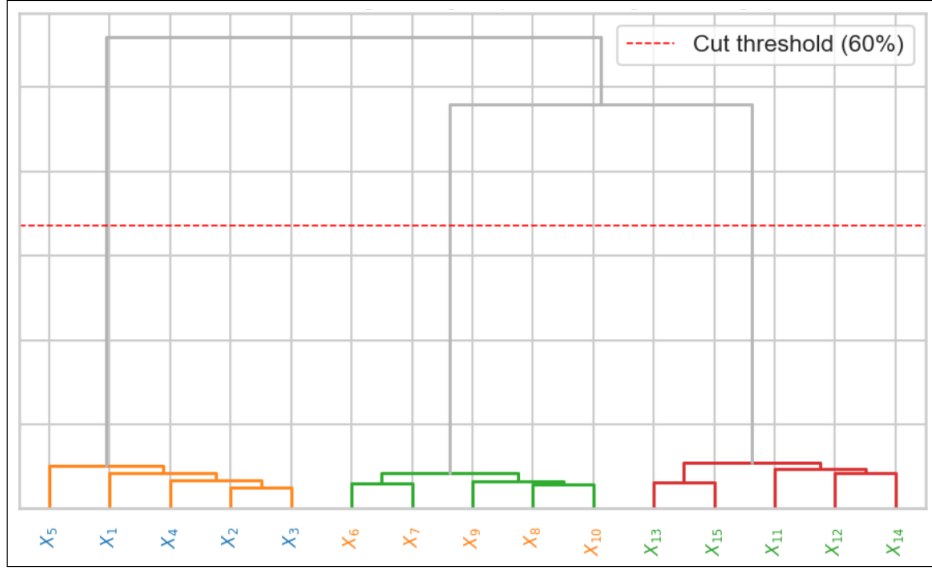


図 3.1: デンドログラムの例

2. 最小距離ペアの探索:

第 k ステップ ($k = 1, 2, \dots$) において, 現在のクラスタ集合 $\mathcal{S}^{(k-1)}$ に含まれる全てのクラスタ対 (C_a, C_b) 間の距離 $d(C_a, C_b)$ を計算し, その距離を最小化する対 (C_{a^*}, C_{b^*}) を特定する.

$$(C_{a^*}, C_{b^*}) = \underset{C_a, C_b \in \mathcal{S}^{(k-1)}, a \neq b}{\operatorname{argmin}} d(C_a, C_b) \quad (3.20)$$

3. クラスタの統合と集合の更新:

特定された対を統合し, 新たなクラスタ $C_{new}^{(k)} = C_{a^*} \cup C_{b^*}$ を生成する. これに伴い, クラスタ集合を以下のように更新する.

$$\mathcal{S}^{(k)} = (\mathcal{S}^{(k-1)} \setminus \{C_{a^*}, C_{b^*}\}) \cup \{C_{new}^{(k)}\} \quad (3.21)$$

4. 終了判定:

クラスタ集合の要素数が 1, すなわち $|\mathcal{S}^{(k)}| = 1$ となるまで, 上記 2 および 3 のプロセスを繰り返す.

この一連の操作により, 個々の変数から全変数を包含する単一のクラスタに至るまでの階層構造が形成され, その履歴はデンドログラム \mathcal{T} として記録される. 変数の階層的クラスタリングによって得られるデンドログラムの例を図に示す. 図中の縦軸はクラスタが結合される際の距離を表しており, 結合位置が低いほど, その変数群が互いに強い相関を持ち, 統計的な性質が類似していることを示す. 逆に上層で結合するクラスタ間は独立性が高い.

ワード法による情報の損失最小化

先に示したアルゴリズムにおいて, 形成されるクラスタの品質およびデンドログラムの幾何学的構造は, 変数間の非類似度 $d(X_i, X_j)$ およびクラスタ間の結合基準の定義に強く依存する. ここでは, 確率変数空間をヒルベルト空間とみなし, 内積構造に基づいた距離尺度

と、情報の損失を最小化するワード法について論じる。

まず、変数間の基本距離について述べる。各確率変数 X_j が平均 0、分散 1 に標準化されていると仮定すると、任意の 2 変数 X_i, X_j 間のピアソン積率相関係数 ρ_{ij} は、確率空間上の内積として一意に定義される。

$$\rho_{ij} = \mathbb{E}[X_i X_j] = \langle X_i, X_j \rangle \quad (3.22)$$

このとき、ベクトル空間における変数間のユークリッド距離は、余弦定理により相関係数を用いて以下のように導出される。

$$\|X_i - X_j\|_2^2 = \langle X_i - X_j, X_i - X_j \rangle = \|X_i\|^2 - 2\langle X_i, X_j \rangle + \|X_j\|^2 = 2(1 - \rho_{ij}) \quad (3.23)$$

この式は、相関が高いほど距離が 0 に収束することを意味し、幾何学的な整合性を持つ。さらに、金融時系列解析においては、負の完全相関もまた強い従属関係として解釈される。したがって、相関の符号を無視した絶対値を用いることで、情報の絶対的な重複度を評価する距離関数 d_{ij} として、一般に以下が定義される。

$$d_{ij} = d(X_i, X_j) = \sqrt{2(1 - |\rho_{ij}|)} \quad (3.24)$$

この定義により、正負の方向性にかかわらず、強く連動する変数は距離空間上で近傍に配置されることとなる。

次に、クラスタ間の統合基準について述べる。最短距離法や最長距離法など多様な定義が存在する中で、統計的な情報圧縮の観点から優れた性質を持つ手法として、ワード法が知られている。ワード法は、クラスタ統合に伴う情報の損失をクラスタ内誤差平方和の増加量として定義し、これを最小化するように階層構造を構築する手法である。

ある変数クラスタ C_k に属する変数の集合を $\{X_{k,1}, \dots, X_{k,n_k}\}$ とし、そのクラスタを代表する幾何学的重心であるセントロイドを \bar{X}_k と定義する。

$$\bar{X}_k = \frac{1}{n_k} \sum_{m=1}^{n_k} X_{k,m} \quad (3.25)$$

このとき、クラスタ C_k が保持する情報の凝集度 E_k は以下のように定式化される。

$$E_k = \sum_{m=1}^{n_k} \|X_{k,m} - \bar{X}_k\|^2 \quad (3.26)$$

ここで、アルゴリズムの第 k ステップにおいて、2つのクラスタ C_a と C_b を統合して新たなクラスタ $C_{new} = C_a \cup C_b$ を生成する過程を考える。統合前の系全体の SSE は $E_a + E_b$ であり、統合後の SSE は E_{new} となる。ワード法におけるクラスタ間距離 $D(C_a, C_b)$ は、この統合操作によって不可避免的に生じる SSE の増加分 (ΔE) として定義される。

$$D(C_a, C_b) = E_{new} - (E_a + E_b) \quad (3.27)$$

さらに、分散の加法性定理を用いることで、この距離関数は各クラスタの重心間距離とサイズを用いて、以下の閉形式で表現できることが示される。

$$D(C_a, C_b) = \frac{n_a n_b}{n_a + n_b} \|\bar{X}_a - \bar{X}_b\|^2 \quad (3.28)$$

上式は、ウォード法における距離が、単なる重心間のユークリッド距離 $\|\bar{X}_a - \bar{X}_b\|^2$ ではなく、調和平均的な重み係数 $\frac{n_a n_b}{n_a + n_b}$ によって調整されていることを示している。この数理的性質により、ウォード法は要素数の小さいクラスタ同士を優先的に結合するバイアスを持ち、結果として極端にサイズの異なる鎖状クラスタの形成を抑制し、解釈性の高い球状のクラスタ構造を抽出することが可能となる。

ランス・ウィリアムズ式による再帰的更新

階層的クラスタリングのプロセスにおいて、クラスタ C_a と C_b が統合され新たなクラスタ $C_{new} = C_a \cup C_b$ が生成されたとする。このとき、任意の第3クラスタ C_c と C_{new} との間の距離 $D(C_{new}, C_c)$ を、元の距離 $D(C_a, C_c), D(C_b, C_c), D(C_a, C_b)$ の線形結合として表現することができれば、元のデータを参照することなく、距離行列のみを用いて高速に更新が可能となる。Lance and Williams(1967) は、多くの階層的クラスタリング手法を包括する以下の一般化更新式を提案した。

$$D(C_{new}, C_c) = \alpha_a D(C_a, C_c) + \alpha_b D(C_b, C_c) + \beta D(C_a, C_b) + \gamma |D(C_a, C_c) - D(C_b, C_c)| \quad (3.29)$$

ここで、係数 $\alpha_a, \alpha_b, \beta, \gamma$ はクラスタ結合の基準に応じて定義されるパラメータである。先で述べたウォード法の場合、距離は分散の加法性に基づいているため、各クラスタの要素数 $n_k = |C_k|$ を用いてパラメータが一意に決定される。具体的には、統合後の全要素数を $N = n_a + n_b + n_c$ と置くと、各係数は以下のように導出される。

$$\alpha_a = \frac{n_a + n_c}{N}, \quad \alpha_b = \frac{n_b + n_c}{N}, \quad \beta = \frac{-n_c}{N}, \quad \gamma = 0 \quad (3.30)$$

$\gamma = 0$ であることから、ウォード法における更新式は、距離の差分項を含まない以下の形式に表される。

$$D(C_{new}, C_c) = \frac{n_a + n_c}{n_a + n_b + n_c} D(C_a, C_c) + \frac{n_b + n_c}{n_a + n_b + n_c} D(C_b, C_c) - \frac{n_c}{n_a + n_b + n_c} D(C_a, C_b) \quad (3.31)$$

この更新式は、新たな距離が既存の距離の重み付き和から、統合されたクラスタ間の内部距離の一部を減産することで産出されることを示している。この漸化式を用いることで、距離行列の更新計算は $O(1)$ で完了し、アルゴリズム全体の計算量は $O(p^2)$ まで削減される。これにより、高次元の金融変数空間においても効率的な階層構造の探索が可能となる。

最終的に得られたデンドログラムを切断することで、変数は K 個のクラスタ G_1, \dots, G_K に分割される。これにより、互いに類似した変数が同一のグループに集約される。Vigneau et al. (2003) が示すように、クラスタリングによって得られたグループ構造は変数の冗長性を反映している。一般的に、このようなグループ構造が得られた後の次元収縮プロセスとして、各クラスタから代表変数を一意に選定する手法が用いられる。変数クラスタリングによって相関の高い変数群をまとめ、各クラスタの重心に最も近い変数を代表として抽出することで、情報の損失を抑えつつ多重共線性の問題を効果的に回避できることが示されている。これらのアプローチは、線形回帰モデルの安定性を確保するうえで、標準的かつ有効な手段として知られている。

§ 3.3 Transformer ベースのシンボリック回帰について

抽出された変数群から具体的な関数形式を同定するためのアルゴリズムとして, Kmienny et al. により提案された Transformer ベースの End-to-End シンボリック回帰モデルを挙げられる. 同モデルは, 数式の構造探索と定数最適化を分離せず, 単一のモデルで同時に推論を行う点に最大の特徴がある.

シンボリック回帰の分野において, 従来支配的であった遺伝的プログラミングに基づく手法は, 高い散策能力を持つ一方で, 解の進化と評価に膨大な計算コストを要し, 推論時間が長時間に及ぶという課題があった. また, 近年提案されている深層学習を用いたアプローチの多くは, 数式の演算子と変数の組み合わせである骨格のみを予測し, 定数は別途 BFGS 法などの数値最適化でフィッティングするという 2 段階の手順を踏襲していた. しかし, この 2 段階アプローチには, 骨格予測の段階で定数の値が考慮されないため, 最適な関数形を見逃すリスクや, 非凸な損失関数に対する最適化の困難さが存在していた.

これに対し, Kamienny らが提案した End-to-End モデルは, 数式の構造と定数を同時にトークンとして生成することで, これらの問題を解決している. 本章では, この手法の数理的基盤と, Transformer アーキテクチャを用いたシンボリック回帰モデルの設計について述べる.

Transformer モデルの概要

ここで, End-to-end シンボリック回帰について述べる前に, その手法の基盤となる Transformer モデルについて簡単に説明する. Transformer は Attention 機構のみに基づいたニューラルネットワークアーキテクチャである. 従来のリカレントニューラルネットワークや Long Short-Term Memory が系列データを逐次的に処理するのに対し, Transformer は系列全体を並列に処理することが可能であり, 長距離の依存関係を効率的に学習できるという特徴を持つ.

Transformer の核となる構成要素は, Scaled Dot-Product Attention と呼ばれる機構である. 入力として, クエリ Q , キー K , バリュース V の 3 つの行列を受け取り, クエリとキーの類似度に基づいてバリュースの加重和を計算する. 具体的には, 入力の次元を d_k とすると, Attention は以下の式で定義される.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.32)$$

ここで, $\frac{1}{\sqrt{d_k}}$ はスケーリング係数であり, 内積の値が大きくなりすぎてソフトマックス関数の勾配が消失するのを防ぐ役割を果たす.

単一の Attention では, 異なる位置にある情報や異なる部分空間の表現を十分に捉えきれない場合がある. そこで Transformer では, 複数の Attention 機構を並列に動作させる Multi Head Attention が導入されている. h 個のヘッドを持つ場合, 各ヘッドで独立に Attention を計算し, それらの出力を結合して線形変換することで最終的な出力を得る.

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (3.33)$$

$$\text{where head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (3.34)$$

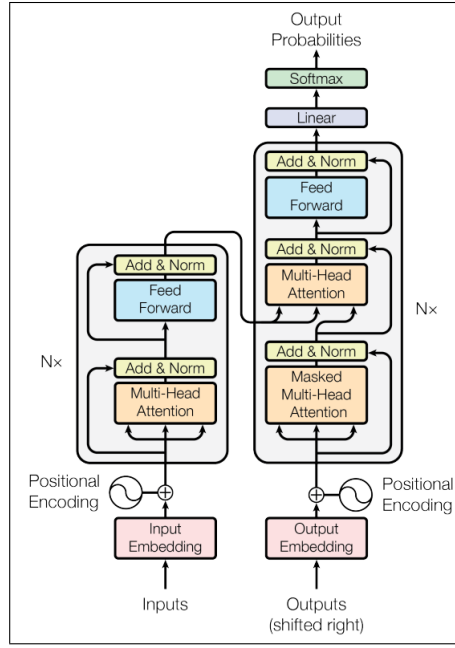


図 3.2: Transformer モデルの概要 [?]

これにより、モデルは入力系列内の多様な関係性を同時に学習することが可能となる。

Transformer は一般に、入力を連続的な表現に変換する Encoder と、その表現を用いて出力系列を生成する Decoder から構成される Sequence-to-Sequence モデルである。図 3.2 にも示した通り、Encoder および Decoder は、前述の Multi-Head Attention 層と、位置ごとの全結合フィードフォワードネットワークを積層した構造を持つ。また、各層の間には残差接続と層正規化が適用され、学習の安定化が図られている。

問題の定式化と生成モデル

シンボリック回帰は、与えられた N 個の入出力ペアからなるデータセット $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ (ここで $x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$ に対し、以下の関係を満たす解析的な関数 f を探索する問題として定式化される。

$$y_i = f(x_i) + \epsilon_i \quad (3.35)$$

ここで ϵ_i はノイズ項を表す。Kamienny らのアプローチでは、この問題を従来の遺伝的アルゴリズムのような探索問題ではなく、点群データ \mathcal{D} をソース言語、数式 f をターゲット言語とする翻訳タスクとみなし、条件付き確率分布 $P(\tau|\mathcal{D})$ を学習する。数式 f は、演算子、変数、および定数を含む語彙 \mathcal{V} 上のトークン列 $\tau = (\tau_1, \dots, \tau_L)$ として表現される。この手法は、観測データ \mathcal{D} を条件として、真の数式に対応するトークン列 τ^* が生成される事後確率 $P(\tau|\mathcal{D}; \Theta)$ を最大化させることを目的としている。Transformer モデルは、連鎖率に基づき、この同時確率を各タイムステップにおける条件付きの積として分解し、自己回帰的に数式を生成する。

$$P(\tau|\mathcal{D}; \Theta) = \prod_{j=1}^L P(\tau_j|\tau_1, \dots, \tau_{j-1}, \mathcal{D}; \Theta) \quad (3.36)$$

ここで、 Θ は Transformer の学習可能パラメータであり、Encoder および Decoder の重み行列を含む。Encoder は順序不変なアーキテクチャによって点群データ \mathcal{D} の特徴表現を抽出し、Decoder はその文脈ベクトルと過去の生成トークン系列 $\tau_{<j}$ に基づいて、次に来るべきトークン τ_j の確率分布を逐次予測する。モデルの学習は、大規模な合成データセットを用いて、政界トークン列に対する負の対数尤度を最小化することで行われる。

$$\mathcal{L}(\Theta) = - \sum_{j=1}^L \log P(\tau_j | \tau_{<j}, \mathcal{D}; \Theta) \quad (3.37)$$

この確率的な定式化により、離散的な探索空間を持つ数式の構造決定と連続的な値を持つ定数の推定を処理することが可能となる。

数値・記号トークンによる表現

End-to-End シンボリック回帰モデルにおいて、数式の生成は言語モデリングとして扱われるため、連続値をとる定数を離散的なトークン列として表現する効率的なエンコーディングが必要となる。Kamienny らは、定数 $c: \mathbb{R}$ を基数 10 の浮動小数点表記に基づき、有効数字 4 桁で丸めたうえで、以下の 3 つの独立したトークンへ分解する手法を採用している。

- 符号トークン: $s: \{+, -\}$ のいずれか。
- 有効数字トークン: $m: \{0, 1, \dots, 9999\}$ までの整数。
- 指数トークン: $e: E - 100$ から $E100$ までの範囲の離散値。

この定義に基づき、任意の実数定数 c は、写像 $\phi: \mathbb{R} \rightarrow \mathcal{V}^3$ によって以下の式で近似されるトークン列 (s, m, e) へと変換される。

$$c \approx s \times \frac{m}{1000} \times 10^e \quad (3.38)$$

数式全体は、演算子の優先順位や括弧の扱いを不要にするため、前置記法を用いて一意なトークン列に展開される。この際、モデルが扱う語彙集合 \mathcal{V} は、演算子集合 \mathcal{V}_{ops} 、変数集合 \mathcal{V}_{var} 、および上述の数値トークン集合 \mathcal{V}_{num} の和集合として定義される。

$$\mathcal{V} = \mathcal{V}_{\text{ops}} \cup \mathcal{V}_{\text{var}} \cup \mathcal{V}_{\text{num}} \quad (3.39)$$

具体的な変換例として、関数 $f(x) = \cos(1.2345x)$ を考える。この数式は、演算子 \cos 、 \times と、変数 x 、および定数 1.2345 から構成される。定数 1.2345 は $[+, 1235, E-3]$ とエンコードされるため、数式全体は以下のような単一のトークン列として表現される。

$$\tau = [\cos, \text{mul}, +, 1235, E-3, x] \quad (3.40)$$

このように、記号的特徴と数値的特徴が混在したシーケンスを生成対象とすることで、Transformer は演算子の構造決定と同時に定数のオーダーや符号を同一の埋め込み空間上で学習することが可能となる。

定数のリファインメント

先に述べたように、Transformer によって生成された数式 \hat{f} に含まれる定数は、トークンかプロセスにおいて有効数字 4 桁に離散化されている。この量子化により、生成された定数 θ_{init} は真の最適解 θ^* 近傍の値をとるものの、回帰モデルとして要求される制度には達していない場合がある。そこで、推論の最終段階として、この θ_{init} を初期値とした非線形最適化を行うことで、パラメータの微調整を実施する。具体的には、生成された数式の骨格を固定し、定数パラメータ θ のみを定数とする以下の非線形最小二乗問題を解く。

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}(\theta) = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^N \left(y_i - \hat{f}(x_i; \theta) \right)^2 \quad (3.41)$$

この最適化アルゴリズムには、準ニュートン法的一种である Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS) 法を採用している。BFGS 法は目的関数のヘッセ行列の逆行列を反復的に近似することで、勾配情報のみを用いて高速に収束することが可能な手法である。

このプロセスにおいて重要なのが初期値 θ_{init} の質である。シンボリック回帰における損失関数 $\mathcal{L}(\theta)$ の形状は一般にきわめて非凸であり、多数の局所解が存在する。従来の遺伝的プログラミングなどの手法では、数式の骨格生成と定数最適化が分離されており、定数探索をランダムな初期値から開始せざるを得なかったため、大域的最適解へ到達する確率が低く、探索に膨大な計算コストを要するという課題があった。

これに対し、本手法における Transformer は、学習データセット全体の大域的な傾向を捉え、真の値に近い定数値を初期値として提供する。これにより、最適化プロセスは真の解を含む内部から開始されることが期待できるため、局所解へのトラップを回避し、高い確率かつ少ない反復回数で大域的最適解へと収束することが可能となる。

提案手法

§ 4.1 時系列クラスタリングによる非定常性の考慮

本節では、提案手法の第一段階である、時系列データのレジーム分割手法について述べる。2.2節にも述べた通り、金融時系列データは一般に非定常であり、相関構造や因果関係といった説明変数と目的変数の関係性は時間とともに変化する。したがって、全期間を一律のモデルで学習することは異なる市場環境を混同させることになり、予測精度の低下を招く。また、可読性の観点からも、非定常性を考慮したモデルを構築することで、市場環境ごとの特徴を把握しやすくなる。そこで、本研究では、期間を重複のない区間に分割し、各区間における局所的な回帰係数をクラスタリングすることで、市場の構造変化を捉えるアプローチを採用した。

局所的線形性の仮定と Ridge 回帰による構造抽出

市場構造の変化を捉えるため、本手法では全期間における大域的なモデル構築を行わず、月次あるいは四半期といった短い時間枠ごとに局所的な線形回帰モデルを構築する。ある期間 t における説明変数を X_t 、目的変数を y_t としたとき、その期間内の市場構造は以下の線形モデルで近似できると仮定する。

$$y_t = X_t \beta_t + \epsilon_t \quad (4.1)$$

ここで、 β_t は期間 t 固有の回帰係数ベクトルであり、その時点での市場における各変数の感応度を表す。本実装では、区間内のサンプルサイズが小さい場合に生じる過学習や多重共線性の影響を抑制するため、正則化項を持つ Ridge 回帰を用いて β_t を推定している。これにより、各機関の市場構造を、価格そのものの値ではなく変数間の関係性として抽出することが可能となる。

行政が持つデータをどのように用いて具体的なモデルの推定を行うかを以降で示す。まず、行政が持つ複数種類の統計データをそれぞれ観測変数として考え、各変数が混合ガウス分布に基づいているという仮定のもと、3.1節で述べた EM アルゴリズムを用いて、データに最も適合するモデルのパラメータを求める。なお、一つの自治体を一つのサンプルとし、共分散行列には3.1節の表??に示した五つの行列をそれぞれ用いる。

これを潜在変数の数、つまり、データを分けるカテゴリーの数を変化させながら複数回行う。結果として、共分散行列とカテゴリー数ごとに最適なパラメータベクトルが得られ

る。次に、それらのパラメータを用いて式(2)の BIC を算出し、BIC の値が最も望ましいものを適切なモデルとして採用する。最後に、そのモデルを用いて各自治体の存在確率を算出する。

クラスタの可視化

前述のクラスタリング手法における結果は、各市町村がどのクラスターにどれくらいの確率で所属するかの確率によって求められる。また、日本全国の市町村を対象として分析を行った場合、サンプル数は 1700 以上となる。そのため、結果を生データから確認することも可能であるが、そのままでは情報がうまくまとまっておらず、解釈にはある程度の時間が必要になる。

クラスタリングの結果をどのような形式で表示することが適切であるかは、その後の目的によって異なるが、本研究では結果を感覚的に捉え、大まかな特徴を確認するために GIS を用いて各市町村が存在するクラスターを描画する手法を提案する。また、GIS を用いた描画では、クラスターの様子が地図上にプロットされるため、地理的な要因を含めた議論が可能になるという利点もあると考えられる。GIS を用いたクラスターの描画方法について以下で述べる。

まず、各自治体に対してそれぞれクラスターの所属確率という複数の連続値で与えられるクラスタリングの結果を地図上にどのような形でプロットするかであるが、本研究では単純に各市区町村における最も所属確率が高いクラスターを表示する方法を用いる。LPA では、各市町村が異なる複数のクラスターに属することを認めているが、多くの場合でどれか一つのクラスターに高い所属率を示すため、本研究ではそのように扱う。

次に、実装についてであるが、本研究では「folium」という Python のライブラリを使用している [34]。folium を用いると、いくつかの関数を使用することで GIS に機能を追加することができ、実行結果として作成された GIS を含む html ファイルが返される。folium によって追加できる機能には主に以下のものが挙げられる。また、機能の例を図 4.1 に示す。

- 指定した緯度・経度にマーカーを配置する
- 指定した範囲を中心にヒートマップを描画する
- 複数の緯度・経度を用いてコロプレスマップを描画する
- レイヤーを複数重ね合わせる

本研究の提案手法の場合、各市町村とそれに紐づくクラスターを GIS で可視化する必要がある。そのため、図 4.1 におけるマーカーまたはコロプレスマップを用いて各市町村の位置を示し、それを用いて各市町村が属するクラスターを表示する方法が適していると考えられる。

機能面ではどちらの方法を用いても実現できるが、コロプレスマップの作成には地図上に境界線を描画するために膨大な量の緯度・経度データが必要となる。そのため、今回のように市町村単位で細かく地図を細分する際には作成、表示ともに機器への負担が大きくなり不向きである。以上のことから、本研究ではマーカーを用いて市町村の位置を表示し、属するクラスターによってマーカーの色に変化をつけるという方法で GIS の実装を行った。

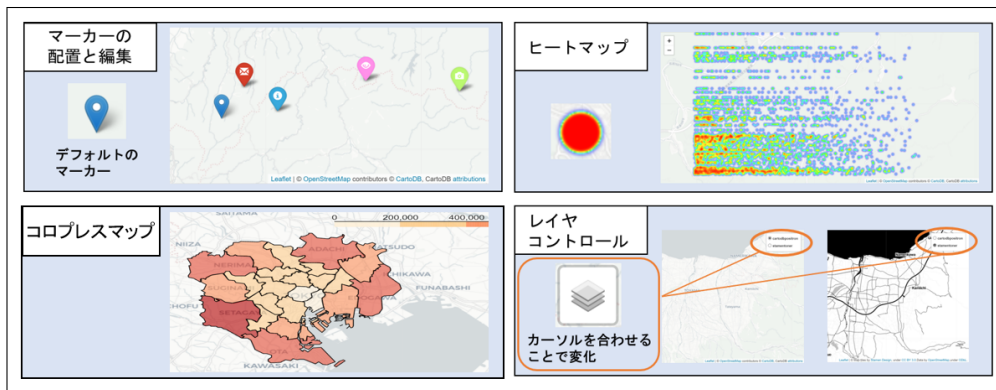


図 4.1: folium ので追加できる機能の例

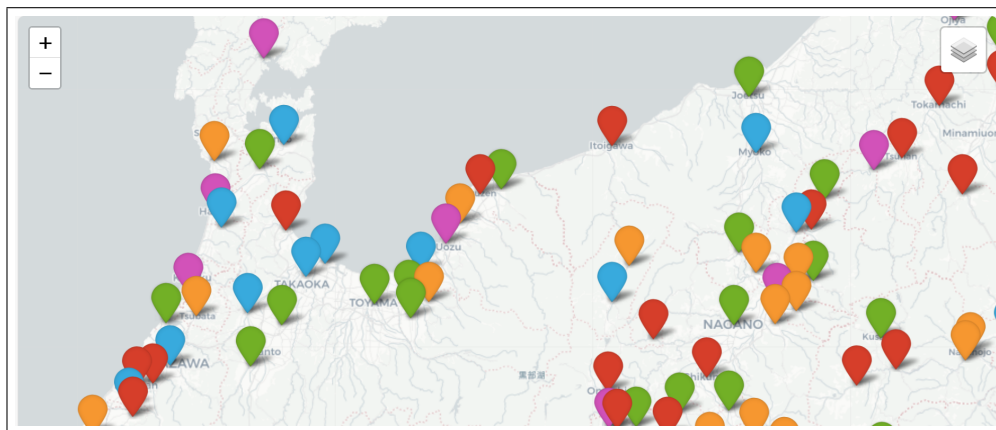


図 4.2: GIS を用いた潜在的クラスターの描画

本研究における GIS を用いた潜在的なクラスターの描画の実装例を図 4.2 に示す。この例では、市区町村が五つの潜在的なクラスターに分かれるという結果が得られた場合を描画しており、五つの色を用いてそれぞれのクラスターを表している。マーカーが配置されている位置は各市町村における代表地点である。

§ 4.2 階層的クラスタリングおよび実験計画法を用いた変数の選択

本節では、提案手法の第一段階である、時系列データのレジーム分割手法について述べる。2.2 節にも述べた通り、金融時系列データは一般に非定常であり、相関構造や因果関係といった説明変数と目的変数の関係性は時間とともに変化する。したがって、全期間を一律のモデルで学習することは異なる市場環境を混同させることになり、予測精度の低下を招く。また、可読性の観点からも、非定常性を考慮したモデルを構築することで、市場環境ごとの特徴を把握しやすくなる。そこで、本研究では、期間を重複のない区間に分割し、各区間における局所的な回帰係数をクラスタリングすることで、市場の構造変化を捉える

アプローチを採用した。

局所的線形性の仮定と Ridge 回帰による構造抽出

市場構造の変化を捉えるため、本手法では全期間における大域的なモデル構築を行わず、月次あるいは四半期といった短い時間枠ごとに局所的な線形回帰モデルを構築する。ある期間 t における説明変数を X_t 、目的変数を y_t としたとき、その期間内の市場構造は以下の線形モデルで近似できると仮定する。

$$y_t = X_t \beta_t + \epsilon_t \quad (4.2)$$

ここで、 β_t は期間 t 固有の回帰係数ベクトルであり、その時点での市場における各変数の感応度を表す。本実装では、区間内のサンプルサイズが小さい場合に生じる過学習や多重共線性の影響を抑制するため、正則化項を持つ Ridge 回帰を用いて β_t を推定している。これにより、各機関の市場構造を、価格そのものの値ではなく変数間の関係性として抽出することが可能となる。

1. 事前に式の項数などの形を指定しなくても結果を得ることができる
2. 各説明変数が目的変数に与える影響の大小の表現力が高い
3. 結果の解釈が可能であり、データ間の関係性が認識できる

各条件に対して必要と考えた理由を以下に示す。1 について、社会における事象は複雑であり、前提知識を用いてモデルの概形を定義してからモデルのパラメータを求める手法の適用は現実的ではないと考えたからである。2 について、各説明変数がどのように目的変数に寄与しているかを高い自由度で表現できることが好ましいと考えたからである。3 について、結果の解釈が難解な手法は、関係性を把握するという当初の目的にそぐわないからである。

以上のことから、本研究では求めるモデルの項数を事前に指定することなく結果を求めることができ、各説明変数における次数に制限の少ない RF 法を用いて観測データ間の関係性を求める手法を提案する。また、RF 法における結果は説明変数が目的変数に与える影響を判読可能な数式として得られるため、上記の 3 つのすべてに適している。

次に、観測された社会実情データを用いてどのように RF 法を行うかであるが、その前に観測される社会実情データにおける特徴について考える。本研究で想定する社会実情データは、各自治体における人口や税収などのため、量的変数である。そのため、一見すると量的データのみを考慮して数法則の発見を行う RF5 法を用いて分析を行うことが最適なように思える。

しかし、社会実情データは、数年に一度しか観測されないデータである。そのため、これらのデータを用いて学習を行い、各変数間の関係を数式化するためには、時間軸ではなく空間軸で大量に収集したデータを用いることになる。空間軸でのデータ収集の例を以下に示す。

単一の年の統計データにを用いて分析を行うことを考える。あるデータ A を他のデータ B, C, D で説明する数式を求めることによってそれらの間に成り立つ関係を把握したい場

合、一つの自治体における A から D の項目におけるデータは、それぞれ一つずつしか存在しない。そのため、回帰分析のように大量のサンプルが必要な場合、自治体の数を増やすという方法でしか十分な数の学習データを確保できないということになる。本研究における学習データの構造を図 4.3 に示す。

このようなデータを用いて学習を行う場合、新たに懸念される課題が発生する。それは、すべての自治体が同一な特徴を持つサンプルとして扱うことは正しいのかということである。例えば、2024 年現在、日本には約 1700 の市区町村が存在するがそれらの中には首都圏とその周辺にあるような世界有数の大規模都市から、人口の少ない地方の村までさまざまな人口規模の自治体が含まれる。

このような特徴は人口のみに限ったものではなく、そこに住む人々の構成や主要産業など扱うデータセットの項目数に応じて無数に存在し、そのすべてを人間の手によって事前に把握することは不可能といえる。

そのため、それらの自治体を対象として集められたすべてのサンプルを特に処理することなく学習サンプルとして用いて RF5 法を行った場合、得られる結果は日本全国におけるすべての自治体に対するデータを一つの数式で無理やり記述したものになる。このような手法では、社会の実情を正しくとらえることが難しいと考える。

よって、本研究では空間軸で収集した大規模な学習データに対して、各サンプルをクラスタリングし、その結果を説明変数に含めて RF 法を行うことで、より現実に即した数法則を発見する手法を提案する。

学習データに対して、それらをクラスタリングする手法だが、これには 4.1 節と同様に LPA を用いる。つまり、4.1 節の手法によって求めた日本全国における自治体ベースの潜在クラスとそれぞれの自治体における存在確率を用いて RF 法に用いる学習データに背景知識を与え、それぞれのクラスターにおける数法則を求める。

どのようにして存在確率を組み込んだ RF 法を行うかであるが、これには RF5 法ではなく RF6.4 法を用いる。3.1 節および 3.2 節でも述べたとおり、RF6.4 法では説明変数に質的変数を用いてサンプルを部分空間に切り分けることによって各サンプルの状況に応じた数法則を発見することができる。

よって、LPA によって求めた各クラスターに対するサンプルの存在確率を RF6.4 法における質的説明変数として扱う。これにより、サンプルを部分空間に切り分け、それぞれにあった数法則を発見することが可能になる。

ここで、存在確率を質的説明変数として扱うと述べたが、存在確率は 0 以上 1 未満の連続した値で与えられる。そのため、そのままでは RF6.4 法における質的変数として扱うことができない。よって、何らかの方法で実数である存在確率をカテゴリーに置き換え、質的変数として扱うことができる形にすることが必要である。

本研究の提案手法では、存在確率が保持する情報をなるべく減らさない形で量子化することを考える。まず、単一のサンプルにおける存在確率について考える。存在確率は LPA によって得られたいくつかのクラスターに対して各サンプルがどれくらいの確率で属するかを表した値であるため、単一のサンプルにおけるすべての存在確率の合計は 1 となる。

このことから、LPA によって得られたクラスター数が仮に A から E の 5 つであった場合、所属するクラスターが最もあいまいなサンプル（五つのクラスターに均等に存在するとされたサンプル）が持つ存在確率は A から E のすべてに対して 0.2 となる。よって、このよ

うな場合、RF6.4法における質的変数の個数とそれらがとりうるカテゴリー数を 5×5 とすると、質的変数によってすべての存在確率のパターンが表せることになる。存在確率のカテゴリー化の例を図4.4に示す。

本研究の提案手法では、このような方法を用いて存在確率を質的変数に変換し、それによってサンプルを部分空間に切り分けることで、学習を行う。これにより、空間軸で収集したために特徴にばらつきがあると考えられるサンプルに対しても、それらの特徴を考慮しながら数法則を発見することができると考えた。

また、従来のRF6.4法では、パーセプトロンの学習結果を用いてルールを復元する際に最適なルール数が分からない。そのため、あらかじめルール数をいくつか定め、それぞれのルール数で復元を行った後、その結果に対して交差検証法を用いることで最適なルール数を発見している。

一方、本研究の提案手法では、事前にLPAを用いてデータサンプルがいくつに分けられるかを求めているため、その結果によって得られたクラス数をルール数として採用する。これにより、大規模なデータサンプルに対しても交差検証法を用いずにルール数が決定できるため、計算量削減につながると考えられる。

最後に、提案手法におけるパーセプトロンの学習について、パラメータの初期値と各変数のスケールリングの方法を述べる。まず、初期値についてはすべてのパラメータに対して $-1 \sim 1$ の一様乱数を与えることとする。次に、各変数のスケールリングについてだが、本研究では実世界において観測された統計データを用いて分析を行うため、単位やスケールが大幅に異なる。そのため、データを何らかの形でスケールリングする必要があると考えられる。

本研究では、目的変数、質的説明変数のそれぞれに対して異なるスケールリングの方法を用いる。目的変数に対するスケールリングを式4.2、量的説明変数に対するスケールリングを式4.3に示す。目的変数と説明変数でスケールリングの方法が異なるのは、RF6.4法における説明変数が負の値を考慮していないからである。なお、スケールリング後の目的変数、量的説明変数を \tilde{y}, \tilde{x} 、元の目的変数、量的説明変数を y, x 、サンプルにおける目的変数全体の中央値、標準偏差、量的説明変数の最大値を $mean(y), std(y), max(x)$ とする。

$$\tilde{y} = \frac{y - mean(y)}{std(y)} \quad (4.3)$$

$$\tilde{x} = \frac{x}{max(x)} \quad (4.4)$$

§ 4.3 可読性を考慮したシンボリック回帰

4.1節では、行政が持つ統計データを用いて、自治体を潜在的な特徴に基づいてクラスターリングする手法とその可視化について述べた。また、4.2節では、クラスターリングの結果を考慮した多変量多項式回帰を行うことでデータ間に成り立つ関係を数理モデルによって表す手法について提案した。

本研究では、以上の結果を統合してGISを用いた可視化を行うWebアプリケーションを開発した。本節では、分析に用いたデータと提案手法について再度整理し、開発したアプ

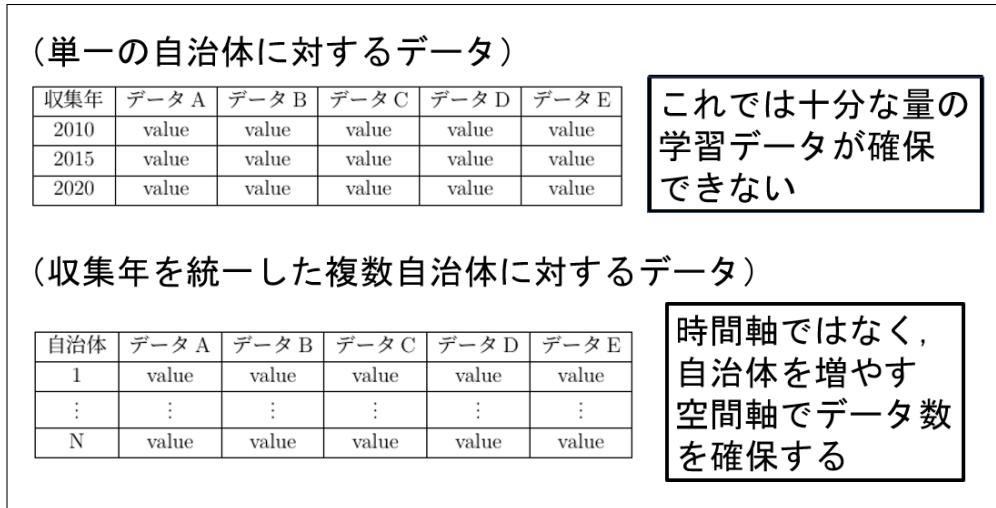


図 4.3: 学習データの一例

表 4.1: 本システムのデータベース

データ項目	単位	データ項目	単位
1人あたりの固定資産税	千円/人	総人口	人
1人あたりの地方税	千円/人	住宅用地平均取引価格	円 / m^2
1人当たりの法人住民税	千円/人	商業用地平均取引価格	円 / m^2
経営耕地面積	畝 / 経営体	農地平均取引価格	円 / m^2
製造品出荷額	万円	林地平均取引価格	円 / m^2
年間商品販売額	百万円		

リケーション全体の構成を示す。また、以下では本研究におけるシステムを本システムと呼ぶ。

まず、使用したデータの参照元について、本システムでは2.1節でも述べたRESASからAPIを使用してデータを取得している。RESASは各自治体単位の経済に関するデータを公開しているオープンデータサイトであり、人口などの基本的な項目のほかに各産業における収入などを公開している。RESASから自動的に取得し、本システムに使用しているデータの一覧を表4.1に示す。なお、表4.1の各データ項目について本システムでは2010、2015、2020年のデータを扱っている。次に、本システムにおける基本的な処理の遷移を説明する。

Step 1: 分析方法と対象年の指定

本システムは2.2節で述べたシステムと同一の枠組みによって作成されているため、システムにアクセスした際に表示されるフロントページでは、どの分析を行うか選択することになる。そのページでLPA+RFを選択すると、分析を行うデータと試行回数に関する条件を選択するページに遷移する。

分析を行いたい年と最終的に求める数理モデルにおける目的変数、RFの試行回数を選択し、その下に表示されている実行ボタンを選択すると分析待機画面に遷移する。ユーザに表示されている画面の遷移は分析が終了して結果が得られるまでその画面で

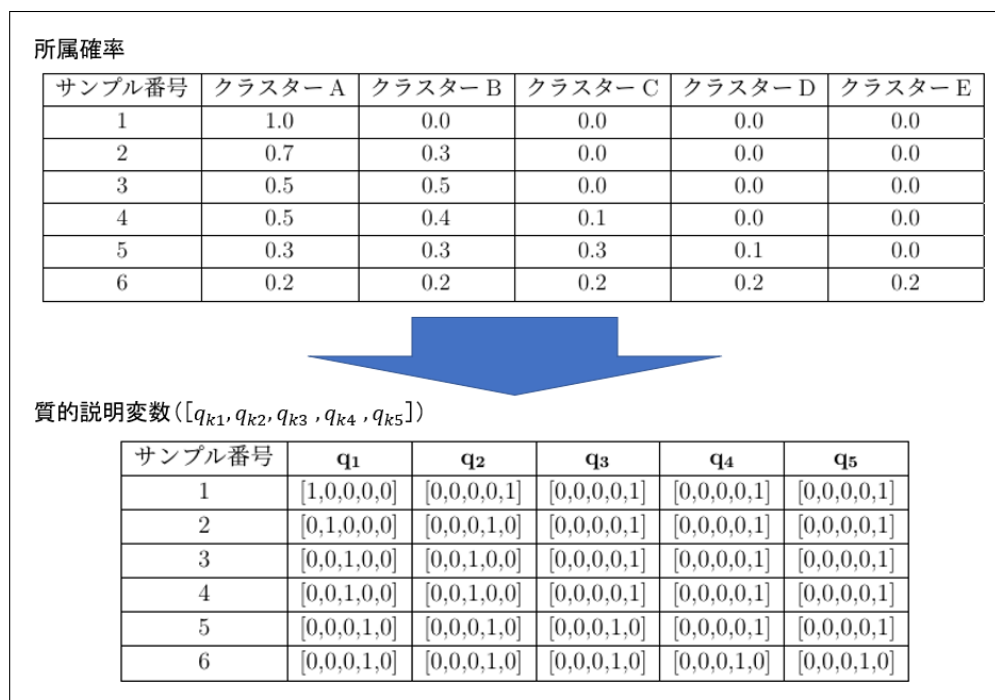


図 4.4: 存在確率のカテゴリー化

待機することになる。システム側では、指定された年におけるデータを用いて Step2 以降の分析が開始される。ここまでの画面的な遷移を図 4.5 に示す。

Step 2: LPA を用いた自治体のクラスタリング

Step2 では、Step1 で選択された年のデータに基づいて LPA を用いた自治体のクラスタリングを行う。まず、3.1 節の表??における 5 つの共分散行列を仮定してそれぞれに適したモデルを求める。なお、クラスター数の数は 1~7 個を考慮している。結果として、共分散行列の数 × クラスター数の結果である計 35 個のモデルにおける最適なパラメータが算出される。

次に、これらに対して BIC を算出し、最も良好な値を示したモデルをクラスタリングに用いるモデルとして採用する。最後に、採用したモデルを用いて各サンプルにおけるクラスターへの存在確率を求め、その結果を用いて Step3 の分析に進む。

Step 3: PF6.4 を用いたデータ間の関係性のモデル化

Step3 では、Step2 で得られた各自治体に対するクラスタリングの結果を考慮して、RF6.4 法を用いたデータ間の関係性のモデル化を行う。求める数理モデルの目的変数には、Step1 で指定されたデータ項目が用いられる。

量的説明変数には本システムのデータベースにある目的変数以外のすべての項目が使用される。質的説明変数には Step2 で得られたクラスタリングの結果がカテゴリーな形で使用される。なお、目的変数および量的説明変数に関しては、4.2 節の式 4.2 および式 4.3 に基づく正規化を施した値が使用される。

RF6.4 法におけるパーセプトロンの学習に用いる各パラメータの初期値は、-1 以上 1

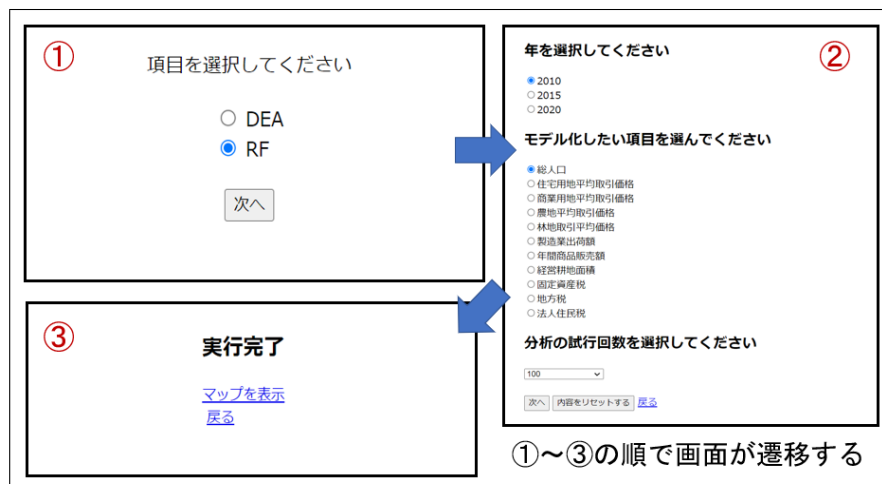


図 4.5: 本システムの画面遷移

未満の一様乱数を使用し、ユニット数 G および D の数をそれぞれ 1～5 まで変化させながらそれぞれに対するパラメータの最適化を行う。この処理をパラメータの初期値を変化させながら Step1 で選択された回数試行する。最後に、すべての結果の中で BIC が最も優れているパラメータベクトルの値を用いて数法則を復元し、データ間の関係性を表す最適な数理モデルとして採用する。

Step 4: EBPM-GIS の作成・データの重ね合わせ

Step4 では、Step2 および 3 で得られた結果を用いて GIS を作成し、結果を可視化して画面に表示する。まず、LPA を用いた自治体のクラスタリング結果について、4.1 節で述べたとおり、各自治体の所在地にマーカーをプロットし、その色を各クラスターに対応したものに塗分けるという方法で表現する。

次に、RF6.4 法の結果である数理モデルについて、それぞれの自治体とそれに対する潜在的クラスター、数理モデルはそれぞれ一対一で対応しているため、LPA の結果を表示する際にプロットしたマーカーに数理モデルを紐づけるという形で表示を行う。

具体的には、各マーカーに対してポップアップ機能を付与し、そのポップアップの中に自治体の名称と数理モデルを表示するという方法を用いる。GIS 上に表示されたマーカーのいずれかを選択すると、ポップアップが表示されるというような仕組みとなる。実装した GIS を図 4.6 に示す。

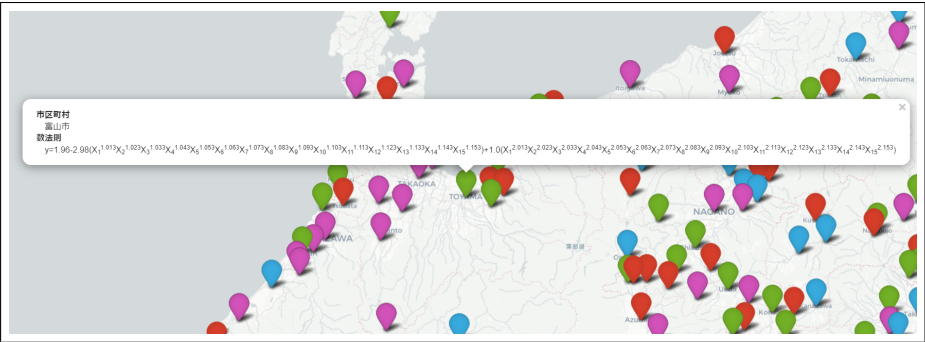


図 4.6: 実装した GIS

数値実験並びに考察

§ 5.1 数値実験の概要

4章で提案した手法のうち、4.2節の数法則発見法によって導出した数理モデルの妥当性を検証することを目的とした二つの数値実験を行った。本節では、それらの数値実験において用いたデータの詳細と初期値、実験の目的および条件をそれぞれ示す。なお、本研究の数値実験における実装に用いたプログラミング言語とパッケージは表5.1のとおりである。

実験に用いたデータと条件設定

4.3節の表4.1に示した11項目のデータのうち、総人口を目的変数、その他10項目を説明変数として数理モデルを導出した。なお、データは現在公開されている中で最新である2020年のものでデータ欠損のない650個のサンプルを使用した。

また、データのスケーリングについて、LPAによる自治体のクラスタリングの際はすべてのデータを生データで投入し、RF6.4法の際は4.2節の式4.2および式4.3を用いて変数をそれぞれスケーリングした値を用いた。

各分析に用いるパラメータの初期値はすべて-1以上1未満の一様乱数とした。RF6.4法のパーセプトロンにおける層の数である G, D は、どちらも1から5までの値で変化させた。パーセプトロンの学習はパラメータの初期値を変化させながら、それぞれ20回ずつ行い、最もBICの値が小さいモデルを学習結果とした。

実験1: 提案手法の精度検証

実験1では、LPAとRF6.4法の組合せである提案手法について、これらの組合せが統計データをモデル化するうえで有効であるかを検証した。検証は以下のような手順で行った。まず、650個のサンプルをランダムに520個の学習データと130個のテストデータに分割した。次に学習データを用いてクラスタリング、モデル化を行った。その後、テストデータを用いてモデルの二乗誤差の平均を算出した。

次に、提案手法と精度を比較する手法について示す。この実験では、以下の二つの点に着目し、それぞれに対して比較対象を設定した。

検証の目的

- クラスタリングの結果をRF6.4法の質的変数として用いることが有効であるか
- クラスタリングの手法としてLPAが適切であるか

表 5.1: 実験環境

プログラミング言語	Python 3.7.6, R 4.3.1
LPA に用いたパッケージ	depmixS4

比較対象

- クラスタリングを用いず、サンプルをランダムなクラスターに割り当てて RF6.4 法を行う (RF6.4 法単体)
- クラスタリングの手法に潜在変数を考慮しない手法である k-means を用いる (k-means 法+RF6.4 法)

これら二つの比較対象に提案手法を加えた三つの手法について、それぞれ前述の条件設定に基づくモデル化を行い、得られたモデルの検証誤差を比較した。また、それぞれの手法について、最小の BIC とそのときに得られたモデル、各 20 回の試行で BIC が最小になったときの G, D の回数を比較した。

実験 2: LPA における適切なクラスター数の検証

実験 2 では、LPA において BIC を用いたクラスター数の決定という方法が妥当であるかを検証した。この実験を行った理由として、本研究の開発したシステムでは、モデルに対する BIC のみで LPA の潜在クラスター数を決定する仕様となっていることが挙げられる。

実装の都合上、数値的な基準を用いてオートマチックにモデル選択をする手法を採用したが、分析に用いるデータの構造は非常に複雑であるため、本来は BIC 以外にも多角的な基準でモデル選択を行うべきであると考えられる。そのため、この実験ではあえてクラスター数を変化させながら分析を行うことで、今後に対する知見を広げることとした。実験の設計を以下に示す。

基本的な検証方法は実験 1 と同様である。前述のサンプルに対して LPA および RF6.4 法を行い、その結果に対して二乗誤差の平均を用いることで検証誤差を求める。ただし、実験 1 では LPA における潜在クラスター数を BIC で決定しているが、この実験では人為的にクラスター数を決定する。次に、指定するクラスター数であるが、実験 1 の LPA において BIC が最良となった 6 個とその半分である 3 個の二種類を用いた。

§ 5.2 実験結果と考察

本節では、5.1 節で述べた二つの数値実験について、それぞれの結果とそれに対する考察を示す。

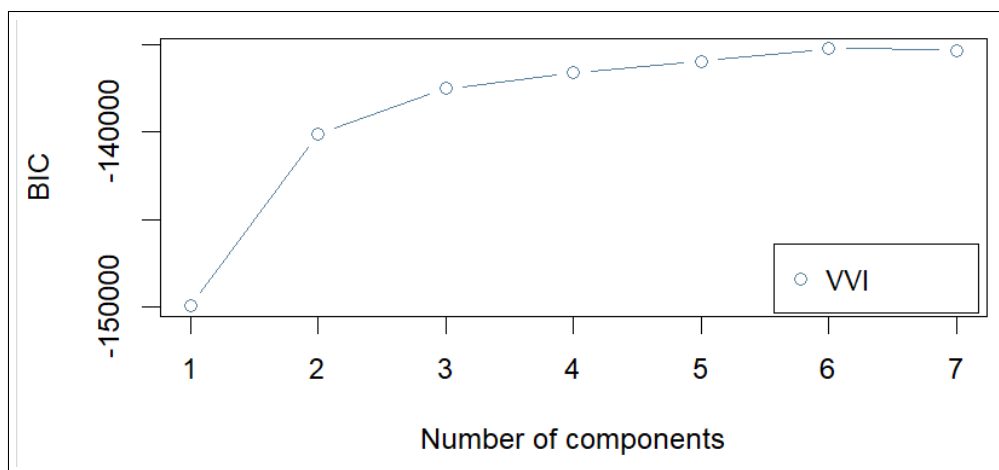


図 5.1: 1~7 個のクラスターにおける LPA の BIC

表 5.2: BIC が最小の回数 (RF6.4 法単体)

models	G=1	G=2	G=3	G=4	G=5
D=1	0	0	0	0	0
D=2	0	7	4	3	2
D=3	0	3	1	0	0
D=4	0	0	0	0	0
D=5	0	0	0	0	0

表 5.3: BIC が最小の回数 (提案手法)

models	G=1	G=2	G=3	G=4	G=5
D=1	0	0	0	0	0
D=2	0	13	2	0	0
D=3	0	5	0	0	0
D=4	0	0	0	0	0
D=5	0	0	0	0	0

実験 1 の結果

はじめに、LPA においてクラスター数を 1 から 7 まで変化させた場合の BIC の値を図 5.1 に示す。本来 BIC は小さいほどよいとされるが、本研究の LPA に用いた R 言語のパッケージである「depmixS4」では実装の関係上、大きい値をとるモデルが最良とされる。この結果から、実験 1 では三つの手法におけるクラスター数をすべて 6 と設定して比較を行った。次に、それぞれの手法における各 20 回の学習で BIC が最小となった G, D の回数を表 5.3, 5.4, 5.2 に示す。また、各手法によって得られた数理モデルを以下に示す。ただし、各説明変数とデータ項目名の対応は表 5.5 のとおりである。

最後に、三つの手法それぞれにおける最小の BIC と二乗誤差の平均値を表 5.6 に示す。提案手法と k-means 法+RF6.4 法の間には BIC、二乗誤差ともに大きな差は見られなかった。RF6.4 法のみの場合には他の二つの手法と比較して BIC、二乗誤差ともに劣る結果となった。

表 5.4: BIC が最小の回数 (k-means)

models	G=1	G=2	G=3	G=4	G=5
D=1	0	0	0	0	0
D=2	0	10	1	0	0
D=3	0	9	0	0	0
D=4	0	0	0	0	0
D=5	0	0	0	0	0

表 5.5: 数理モデルにおける各変数の意味

変数	データ項目	変数	データ項目
x_1	1人あたりの固定資産税	x_6	住宅地取引平均価格
x_2	1人あたりの地方税	x_7	商業用地取引平均価格
x_3	1人あたりの法人住民税	x_8	年間商品販売価格
x_4	経営耕地面積	x_9	農地平均取引価格
x_5	製造品出荷額	x_{10}	林地平均取引価格
y	総人口		

RF6.4 法単体で得られた数理モデル

$$\begin{aligned}
y &= -0.2592 + 0.2177h_3 + 0.8606h_4 \\
y &= -0.3305 + 0.1593h_3 + 0.8986h_4 \\
y &= -0.287 + 0.164h_3 + 0.8235h_4 \\
y &= -0.2807 + 0.1865h_3 + 0.8492h_4 \\
y &= -0.2962 + 0.176h_3 + 0.8611h_4 \\
y &= -0.29 + 0.1802h_3 + 0.8563h_4 \\
h_3 &= x_1^{-0.2172} x_2^{-0.3559} x_3^{-0.2543} x_4^{-0.1069} x_5^{0.4313} x_6^{0.1519} x_7^{-0.009} x_8^{-0.9073} x_9^{0.1422} x_{10}^{-0.1133} \\
h_4 &= x_1^{-1.4913} x_2^{-1.096} x_3^{-0.8261} x_4^{-0.0001} x_5^{0.0467} x_6^{0.2647} x_7^{0.0107} x_8^{1.0844} x_9^{0.0868} x_{10}^{0.0448}
\end{aligned} \tag{5.1}$$

k-means 法+RF6.4 法で得られた数理モデル

$$\begin{aligned}
y &= 0.3439 + 0.0301h_5 - 0.0279h_6 \\
y &= 1.7926 + 0.1595h_5 - 0.123h_6 \\
y &= 0.0021 + 0.0001h_5 - 0.0003h_6 \\
y &= 3.5654 + 0.3174h_5 - 0.2426h_6 \\
y &= 2.8009 + 0.2493h_5 - 0.1912h_6 \\
y &= 0.0373 + 0.0027h_5 - 0.0076h_6 \\
h_5 &= x_1^{0.852} x_2^{0.2726} x_3^{-0.3544} x_4^{-0.3331} x_5^{0.6123} x_6^{0.1} x_7^{0.2347} x_8^{-0.3919} x_9^{0.6471} x_{10}^{-0.497} \\
h_6 &= x_1^{-0.682} x_2^{-0.5484} x_3^{-1.7717} x_4^{0.0845} x_5^{-0.0626} x_6^{2.3711} x_7^{1.6133} x_8^{1.1976} x_9^{-0.5819} x_{10}^{-0.0135}
\end{aligned} \tag{5.2}$$

提案手法で得られた数理モデル

$$\begin{aligned}
y &= -0.2675 + 3.6088h_1 + 0.8292h_2 \\
y &= -0.3445 + 2.6431h_1 + 0.6347h_2 \\
y &= -0.2606 + 3.2077h_1 + 0.7413h_2 \\
y &= -0.2701 + 3.1969h_1 + 0.7407h_2 \\
y &= -0.1894 + 3.3758h_1 + 0.7644h_2 \\
y &= 0.0437 + 2.9772h_1 + 0.6353h_2 \\
h_1 &= x_1^{-0.3891} x_2^{-0.1932} x_3^{-0.6396} x_4^{0.8603} x_5^{0.1466} x_6^{-0.087} x_7^{0.5264} x_8^{0.0192} x_9^{0.7362} x_{10}^{0.1035} \\
h_2 &= x_1^{-0.7882} x_2^{-0.2078} x_3^{-0.651} x_4^{0.0549} x_5^{0.3084} x_6^{0.8366} x_7^{-0.069} x_8^{0.6426} x_9^{-0.0144} x_{10}^{0.0513}
\end{aligned} \tag{5.3}$$

表 5.6: 各手法における最小の BIC と二乗誤差の平均値

手法	BIC	二乗誤差
RF6.4 法単体	-46.2474	0.4378
k-means 法+RF6.4 法	-149.9973	0.1744
提案手法	-166.359	0.1776

実験 1 についての考察

まず、提案手法と RF6.4 法単体との比較について考察する。表 5.3 から、提案手法は $G=2$ のときに BIC が最小になる傾向があると考えられる。一方、表 5.2 では G が 2 から 5 まで BIC が最小になった回数が散らばっている。この原因についてであるが、量的変数と質的変数との関連性が考えられる。

提案手法における質的変数は量的変数を分析した結果から得られたものであり、その間には関連がある。そのため、質的変数によって量的変数を切り分け、学習が進んでいると考えられる。一方、RF6.4 法単体は質的変数をランダムに与えているため、その間に関連性はない。そのため、質的変数に対する重みパラメータの学習が十分に進まず、量的変数が寄与している項数が多いときに BIC が良好な値を示したと考察する。

次に、式 5.3 と式 5.1 を比較すると、提案手法の数理モデルにおける 6 本の数式は定数項、各係数ごとの大きさの差にばらつきが見られるが、RF6.4 法単体の数理モデルではその差が小さい。このことは、前述の質的変数に対する重みパラメータの学習が十分に進んでいないことと合致すると考えられる。

また、表 5.6 の BIC および二乗和誤差の平均について、RF6.4 法と比較して提案手法が良好な値を示した。以上のことから、RF6.4 法の質的変数にクラスタリングの結果を用いることは学習の精度を向上させる上で有効であることが示唆された。

次に、提案手法と k-means 法 + RF6.4 法との比較について考察する。表 5.3 と表 5.4 および表 5.6 から、モデルの精度の観点では LPA と k-means 法の間に差は見られなかった。一方、得られた数理モデルには違いが見られた。このことから、今後はクラスター数を変化させる、他のクラスタリング手法とも比較する、クラスターがある程度既知のデータに対して同様の比較を行い、サンプルがどのような特徴によってクラスタリングされているかを調査するなどより詳細な検証が必要と考えられる。

実験 2 の結果

クラスター数 3 のときに得られた数理モデルを以下に示す。各説明変数とデータ項目の対応は表 5.5 のとおりである。なお、クラスター数 6 のときの数理モデルは実験 1 の式 5.3 と同様であるため、ここには表記しない。また、各クラスター数の場合における最小の BIC と二乗誤差の平均値を表 5.7 に示す。

表 5.8: クラスタ 3 のサンプル

クラスター	1	2	3
サンプル数	220	59	371

表 5.9: クラスタ 6 のサンプル

クラスター	1	2	3	4	5	6
サンプル数	213	251	30	105	46	5

クラスタ数 3 のときに得られた数理モデル

$$\begin{aligned}
y &= -0.4449 + 3.6612h_7 + 3.8795h_8 + 1.738h_9 \\
y &= -0.0191 + 2.7273h_7 + 2.8581h_8 + 1.3415h_9 \\
y &= -0.3359 + 3.9828h_7 + 4.2052h_8 + 1.9128h_9 \\
h_7 &= x_1^{0.9928} x_2^{-0.7812} x_3^{-0.9775} x_4^{-0.2915} x_5^{-0.1333} x_6^{0.9852} x_7^{0.7399} x_8^{0.327} x_9^{0.9249} x_{10}^{0.817} \\
h_8 &= x_1^{0.2394} x_2^{1.102} x_3^{0.9017} x_4^{-0.6071} x_5^{0.3012} x_6^{1.0284} x_7^{0.377} x_8^{2.1434} x_9^{0.8348} x_{10}^{-0.0604} \\
h_9 &= x_1^{-0.6051} x_2^{-0.4339} x_3^{-0.9032} x_4^{0.0692} x_5^{0.1714} x_6^{0.3847} x_7^{0.1702} x_8^{0.8249} x_9^{0.0245} x_{10}^{0.0539}
\end{aligned} \tag{5.4}$$

表 5.7: クラスタ数を変化させたときの最小の BIC と二乗誤差の平均値

クラスタ数	BIC	二乗誤差
K=3	-235.2285	0.1147
K=6	-166.359	0.1776

実験 2 ついての考察

まず、得られた数理モデルについて比較を行う。大きな違いとして項数が挙げられる。説明変数が寄与する項について、クラスタ数 6 のときは二つであったのに対し、クラスタ数 3 のときは三つに増加した。原因として、クラスタ数の増加に対してモデルの説明力が増加していないことが考えられる。参考として、クラスタ数 3 および 6 における各クラスタに属するサンプル数を表 5.8、表 5.9 に示す。

サンプル数と式の形状から、クラスタ数 6 のときの 1, 2 と 3 と 4, 5 と 6 がクラスタ数 3 のときの 1, 3, 2 とそれぞれ対応すると考えられる。しかし、クラスタ数 6 のモデルにおける 2, 3, 4 本目の式には大きな差異が見られなかった。このことから、クラスタ数 6 では比較的特徴の近いサンプルを無理やり複数のクラスタに分割してしまっているのではないかと考えられる。この場合、項数が増えると不要なパラメータが増加するため、BIC の観点では少ない項数が好ましくなる。

次に、BIC と二乗誤差について、大きな差異は見られなかったが、クラスタ数 3 がよい値を示した。以上のことから、この二つの比較では、クラスタ数 3 の方が適当である可能性が示唆された。つまり、本研究の提案システムにおいても、今後 BIC 以外の要素を考慮する必要があると考えられる。

おわりに

本研究では、EBPMにおけるデータ分析について新たな知見をもたらすために、行政が持つ統計データを用いた新たな分析手法およびそれを組み込んだ可視化システムを提案した。

まず、統計データが混合ガウス分布に従うという仮定の下、クラスタリング手法の一つである潜在プロファイル分析を行った。それにより、統計データの背後に存在する潜在変数を分析し、それに基づいて自治体をいくつかのクラスターに振り分けた。

次に、パーセプトロンの学習によってデータ間に成り立つ関係性を数理モデル化する手法の一つであるRF6.4法を用いて統計データを数理モデル化することを考えた。また、より統計データにフィットしたモデルを求めるために、前述のクラスタリング結果をRF6.4法の質的説明変数として扱う手法について提案した。

最後に、クラスタリングの結果と求めた数理モデルをGISを用いて描画し、地図上に可視化することで、データ分析の結果と地理的特徴を重ね合わせるシステムを実装した。

数値実験では、提案手法で得られる数理モデルの妥当性を二つの方法で検証した。一つ目の実験では、提案手法、クラスタリングを用いないRF6.4法、k-means法+RF6.4法の三つの手法を用いてそれぞれ数理モデルを求め、その結果を比較した。また、学習に使用していないテストデータを用いて各モデルにおける二乗誤差の平均を算出し比較した。

二つ目の実験では、提案手法におけるLPAで自治体を振り分けるクラスターの数を変化させて一つ目の実験同様に結果の違いを比較した。クラスター数には、3個と6個の二種類を使用した。

以上の実験から、RF6.4法の潜在変数にクラスタリングの結果を使用することで得られるモデルの精度が向上することが分かった。一方、使用するクラスタリング手法において、LPAとk-means法の間には有意な差は見られなかった。ただし、統計データの構造は非常に複雑であると考えられるため、LPAで考慮するクラスター数を適切に切り替えることで得られるモデルの精度が向上する可能性がある。

今後の展望として、分析に用いる統計データを変数、サンプル共に拡張することでより深い検証と議論が可能になると考えられる。特に、本研究では変数間の関係性を考慮せずに分析を行っているが、専門家の知見やさらなる分析を用いて関連性のあるデータのみを収集し、それらを数理モデル化することで現実に即した数理モデルが得られているかや適切なクラスター数の議論ができると考えた。

また、本研究では政策分野を対象として統計データを扱ったが、金融やマーケティングなど大規模なデータの間成り立つ関係性を分析することが有効と考えられる別の分野を用いて提案手法を適用することも考えられる。

謝辞

本研究を遂行するにあたり、多大なご指導と終始懇切丁寧なご鞭撻を賜った富山県立大学工学部電子・情報工学科情報基盤工学講座の奥原浩之教授、António Oliveira Nzinga René 講師に深甚な謝意を表します。また、システム開発および数値実験にあたり、ご助力いただいた富山県立大学大学院工学研究科1年生の水上和秀氏に感謝の意を表します。最後になりましたが、多大な協力をしていただいた富山県立大学大学院工学研究科の同輩諸氏に感謝致します。

2024 年 2 月

長瀬 永遠

参考文献

- [1] 内閣府, ”内閣府における EBPM への取組”, 閲覧日 2024-02-06,
<https://www.cao.go.jp/others/kichou/ebpm/ebpm.html>.
- [2] 杉谷和哉, ”行政事業レビューにおける EBPM の実践についての考察”, 日本評価学会,
Japanese journal of evaluation studies, Vol. 21, No. 1, pp. 99-111, 2021.
- [3] 中泉拓也, ”英国の EBPM (Evidence Based Policy Making) の動向と我が国への EBPM
導入の課題”, 関東学院大学経済経営研究所年報, Vol. 41, pp. 3-9, 2019.
- [4] 井伊雅子, 五十嵐中, ”新医療の経済学: 医療の費用と効果を考える”, 日本評論社, 2019.
- [5] 山本雄三, 高見具広, 高橋陽子, ”統計指標に基づく市町村分類の試み”, 独立行政法人
労働政策研究・研修機構, 2018.
- [6] 中村圭, ”地方自治体における EBPM の進め方とは?【基本編】 地域課題の解決に繋がる
EBPM に向けて”, 閲覧日 2024-02-06,
<https://www.fujitsu.com/jp/group/fjm/business/mikata/column/local-government/fri-nakamura/>.
- [7] D. A. Pastor, K. E. Barron, B. J. Miller, S. L. Davis, ”A latent profile analysis of college
students' achievement goal orientation”, Contemporary Educational Psychology, Vol.
32, pp. 8-47, 2007.
- [8] 斉藤和巳, 中野良平, ”コネクショニストアプローチによる数法則の発見”, 情報処理学
会論文誌, Vol. 37, No. 9, pp. 1708-1716, 1996.
- [9] 国土交通省国土地理院, ”GIS とは”, 閲覧日 2024-02-06,
<https://www.gsi.go.jp/GIS/whatisgis.html>.
- [10] 長瀬永遠, ”証拠に基づく政策立案のためのオープンデータを活用した Web-GIS 可視
化によるデータフュージョン”, 富山県立大学, 2022, 卒業論文.
- [11] 國廣勇人, ”EBPM の現状と課題-国会は EBPM にどう向き合っていくべきか-”, RE-
SEARCH BUREAU 論究, Vol. 19, pp. 112-131, 2022.
- [12] 法務省矯正局, ”受刑者就労支援体制等の充実”, 閲覧日 2024-02-06,
<https://www.moj.go.jp/content/001350632.pdf>.
- [13] 田中裕太郎, 野村裕, 間真実, ”EBPM のエビデンスレベルの規定要因に関する分析”,
New ESRI Working Paper, No. 48, 2019.
- [14] 内閣府地方創生推進室ビッグデータチーム, 経済産業省地域経済産業調査室, ”RESAS
- 地域経済分析システム”, 閲覧日 2024-02-06,
<https://resas.go.jp/#/13/13101>.

- [15] 内閣官房まち・ひと・しごと創生本部事務局, 内閣府地方創生推進室, 経済産業省九州経済産業局, "観光パイオニア都市の再成長に向けた施策の検討", 別府市役所 5 階大会議室, 2016-10-18, 内閣官房まち・ひと・しごと創生本部事務局, 内閣府地方創生推進室, 2016.
- [16] S. Shimizu, T. Inazumi, Y. Sogawa, "DirectLiNGAM: A Direct Method for Learning a Linear Non-Gaussian Structural Equation Model", Journal of Machine Learning Research, Vol. 12, pp. 1225-1248, 2011.
- [17] 藤井秀幸, 傳靖, 小林里佳子, "データ包絡分析を用いたふるさと納税の戦略提案-K 市のふるさと納税への適用事例-", 日本経営工学会論文誌, Vol. 71, No. 4, pp. 149-172, 2021.
- [18] 刀根薫, "包絡分析法 DEA", 日本ファジィ学会誌, Vol. 8, No. 1, pp. 11-14, 1996.
- [19] 金成賢作, 篠原正明, "DEA における入力指向と出力指向の比較 (その 1)", 日本大学生産工学部第 42 回学術講演会, 2009.
- [20] 鈴木友彦, 後藤正幸, 俵信彦, "線形回帰モデルのベイズ最適な予測法に関する研究", 日本経営工学会論文誌, Vol. 51, No. 1, pp. 331-332, 2000.
- [21] 石渕久生, 田中英夫, "ニューラルネットによるファジィ回帰分析とその応用", 計測自動制御学会論文集, Vol. 28, No. 1, pp. 116-124, 1992.
- [22] J. Ye, T. Xiong, "SVM versus Least Squares SVM", Journal of Machine Learning Research, Vol. 2, pp. 644-651, 2007.
- [23] 田中雅史, 福間順, 鳥井清司, "GMDH (発見的自己組織化法) の概要と適用例", 農業土木学会誌, Vol. 45, No. 8, pp. 539-545, 1977.
- [24] 棚橋裕輔, "多層パーセプトロンによる可読性と汎化性に優れた回帰分析に関する研究", 名古屋工業大学, 2009, 博士論文.
- [25] 古屋秀樹, "潜在クラス分析を用いた COVID-19 流行下での旅行意向分析", 土木学会論文集, Vol. 77, No. 2, pp. 141-150, 2021.
- [26] 赤穂昭太郎, "EM アルゴリズムの幾何学", 会誌「情報処理」, Vol. 37, No. 1, pp. 43-51, 1996.
- [27] 赤池弘次, "AIC と MDL と BIC", オペレーションズ・リサーチ = Communications of the Operations Research Society of Japan: 経営の科学, Vol. 41, No. 7, pp. 375-378, 1996.
- [28] K. Saito, R. Nakano, "Partial BFGS Update and Efficient Step-Length Calculation for Three-Layer Neural Networks", Neural Computation, Vol. 9, No. 1, pp. 123-141, 1997.

- [29] lib-arts, "BFGS 法などの準ニュートン法の概要・数式理解と Python プログラムの確認", あつまれ統計の森, 閲覧日 2024-02-06,
<https://www.hello-statisticians.com/optimization/optimization3.html>.
- [30] J. A. Hartigan, M. A. Wong, "A K-Means Clustering Algorithm", Journal of the Royal Statistical Society, Vol. 28, No. 1, pp. 100-108, 1979.
- [31] 植木優夫, 笛田薫, "交差検証法のバイアス補正とモデル選択への応用", 日本計算機統計学会大会論文集, Vol. 19, pp. 143-146, 2005.
- [32] 井上寛康, "大規模データに対するべき分布性の確認方法", 大阪産業大学経営論集, Vol. 11, No. 2, pp. 165-176, 2010.
- [33] 釜国男, "人口分布の数量分析 (計算経済学の研究その 13)", 季刊創価経済論集, Vol. 46, No. 1・2・3・4, pp. 33-44, 2017.
- [34] pork_steak, "folium 事始め", 閲覧日 2024-02-06,
https://qiita.com/pork_steak/items/f551fa09794831100faa.