卒業論文

政策課題解決に向けた効果的な施策立案のための意思決定支援システム

Decision Support System for Effective Planning of Measures to Solve Policy Issues

放送大学　教養学部情報コース
2010022122　奥原由利恵
指導教員

提出年月: 令和7年 (2025年) 月

目 次

図一覧

表一覧

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
j 人目の使用者の名前	ϵ_j
j 人目の身長	α_j
j 人目の体重	β_j
j 人目の基礎代謝量（下限）	B_j^L
j 人目基礎代謝量（上限）	B_j^H
j 人目のアレルギー情報	x_j
j 人の有する生活習慣病	z_j
対象の日数	D
レシピの数	R
食材の数	Q
栄養素の数	N
データベース上の食材数	S
データベース上の食材番号	$d : 1, 2, 3, \cdots, S$
日の番号	$k : 1, 2, 3, \cdots, 3D$
栄養素の番号	$l : 1, 2, 3, \cdots, N$
材料の番号	$m : 1, 2, 3, \cdots, Q$
レシピの番号	$i : 1, 2, 3, \cdots, R$
i 番目のレシピの名前	y_i
i 番目のレシピの献立フラグ	r_{ki}
i 番目のレシピの主菜フラグ	σ_i
i 番目のレシピの調理時間	T_i
i 番目のレシピの摂取カロリー	C_i
i 番目のレシピの調理コスト	G_i
i 番目のレシピの m 番目の材料の名前	q_{im}
i 番目のレシピの m 番目の材料量	e_{im}
i 番目のレシピの l 番目の栄養素の名前	n_{il}
i 番目のレシピの l 番目の栄養素の量	f_{il}
d 番目の食材名	Z_d
d 番目の食材の販売単位	W_d
d 番目の食材の値段	M_d

第1章 はじめに

§ 1.1 本研究の背景

政策課題とは，行政が対応すべき社会的な問題のことである．戦後半世紀の急速な経済成長を経て，新たな時代の半世紀へと移行し，20 世紀から 21 世紀へと時代が大きく転換する中で，社会の潮流にも大きな変化が生じており，それに応じた対応が強く求められている．総務省では，重要分野として積極的に取り組むべき施策について,「新しい資本主義」や「デジタル田園都市国家構想」に基づき，我が国を取り巻く環境の変化や国内の構造的課題に対応し，活力ある地域づくりや地方行財政基盤，持続可能な社会基盤を確保しながら，地域課題の解決を通じて持続可能な地域社会の実現を目指すと取りまとめている [1].

また,，世界経済の中で大きな比重を占める我が国は，世界 の平和と繁栄の目立たない受益者に留まることはできない．自らが経済 システムを改革し，その国際的調和を図るとともに，世界の望ましいあり方や，その実現 のための道筋，すなわち，グランド・デザインについて，自らの考え方を世界に向けて提 示し，地球社会の発展に積極的に参画していく必要がある．

地域が直面する課題は極めて複雑で多岐にわたり，その多くは景気の循環的要因と構造的要因が複雑に絡み合っている．情報通信技術 (Information and Communication Technology: ICT) の進化，少子高齢化，地域からの人口流出に加え，新型コロナウイルス感染症の影響を経て，地域経済の構造も急速に変化している．しかし，現在の経済社会構造はこうした変化に十分に対応しきれておらず，様々な構造的課題が表面化し，人々の将来に対する漠然とした不安感を助長している．

構造改革の過程においては，こうした改革なしに現在感じている将来への不透明感を払拭し，中長期的な発展を実現することは困難とされている．このため，これまで以上に地域経済の動きを迅速かつ的確に捉え，景気動向と構造的問題の双

方を統合的に分析することが重要である．内閣府においても，経済財政や地域創生，外交・安全保障などの各分野において政策を展開し，それぞれに対応する施策が準備されている．

地域課題は政策課題の一つとして位置づけられ，地域ごとに異なる特性や背景を持ち，それが地域活性化や持続可能な発展を妨げる要因となっている [2]．したがって，地域の特性や課題を正確に把握し，効果的な対策を講じることが不可欠である．特に，人口減少や少子高齢化は重要な地域課題の一つであり，生産年齢人口の減少や若年層の都市部流出は，地方自治体の税収低下を招き，行政サービスの地域格差を拡大させる要因となっている．

人口減少社会の本格的な到来，地方創生を契機とした地域の特性に応じたまちづくり，激甚化する自然災害，デジタル・トランスフォーメーション (Digital Transformation: DX) の推進，感染症対策など，自治体を取り巻く環境は大きく変化している．

§ 1.2　本研究の目的

政策課題は行政・自治体が政策によって問題解決を図る課題を意味される.「政策」と聞くと国が対応するイメージがありますが，現在は地方分権社会であるため身近な行政サービスは自治体が運営しており，自治体には政策の遂行能力が求められている．政策課題の設定の流れは，課題と目標を設定する段階の政策目的の設定，目標を実現するための政策手段の立案である．

政策目的の設定は問題の発見，問題の検討，課題の選定，目標の設定の 4 つの軸がある．政策手段の立案は手段の探索，手段の発想，手段の選定の 3 つの軸がありこれらの 7 つの工程を経て完成させることができる．

問題の発見では，受動的に問題を発券する場合と能動的に発見する場合があり，環境の変化から考えてみる方法がある．問題の検討では，問題の原因を分析・整理しておくと後で役に立つという点がある．課題の選定で公共性の基準と必要性の基準が考えられ，課題を 1 つに絞る必要がない．

政策手段は「ある目的や問題を『政策課題』として決定した上で，課題を解決するための政策目標の達成に向かって，その政策目標の実現のために行なう活動の総称 [3]」を政策手段と呼ぶ．ある政策課題を解決するための政策手段は 1 つに限定されず複数考えられる．多くの手段により複合的に政策課題を解決しようとすることが重要である．政策手段の立案にあたっては，コスト意識をもつことが重要である．

政策決定における対象となる問題には原因となる事柄が一対一ではなく複数個

存在し，なおかつそれらが複雑に影響しあっているため，それらを正しく見通すことが困難であるという課題を解決するためには関連性が示唆される単一項目のみのデータを別々に見るのではなく，一見関係のない項目も含めたより広い範囲のデータを統合的に考慮することが重要であると考える．

しかし，このような作業を人力のみで行うのは非常に困難であり，その他にも多くの業務を抱える地方自治体のような組織では現実的ではない．そこで，これらの作業を ICT 技術やデータ分析手法を用いて適切に処理することは政策における物事の難解さの解決と効率的な意思決定の支援に効果的であり，負担を軽減しうるものと考えた．

本研究の目的は，地域が抱える多様かつ複雑な政策課題に対して，客観的かつ科学的に分析を行い，課題の本質的要因を把握し，効果的な対策の立案を支援する意思決定支援システムを構築することである．現代社会における自治体は，人口減少，少子高齢化，地域経済の停滞，産業空洞化，自然災害や感染症のリスク拡大など，従来以上に構造的かつ多元的な課題に直面している．こうした背景により，地域の持続可能性を確保するためには，これまで以上に地域経済や社会構造の動きを正確に把握し，実情に即した合理的な政策立案が求められている．

現行の政策立案手法では，行政の経験則や過去の事例に依存する傾向が強く，地域課題の構造的な分析や施策の効果検証が十分に行われているとは言いがたい．そこで本研究では，データ駆動型の分析アプローチとして，因果探索アルゴリズムを用いた要因構造の可視化と，データ包絡分析による施策の効率性評価を組み合わせることで，政策課題に対する分析と対策立案を体系的に支援するフレームワークを提案する．

具体的には，統計データやオープンデータ等の社会経済情報をもとに，地域ごとに異なる変数間の因果関係を導出し，政策的に介入可能な要素と成果指標を明確化する．そのうえで，データ包絡分析により複数地域間のパフォーマンス比較を行い，現状の効率性を定量的に評価する．このような分析結果をもとに，地域特性に応じた最適な施策群の提示や，改善点の抽出を支援することで，現場での政策判断に資する知見を提供する．

さらに，得られた知見を自治体職員や政策立案者が実務で活用できるよう，操作性に優れた支援ツールとしてシステム化を図る．本研究を通じて，科学的根拠に基づく政策立案の促進と，地域の自立的・持続的な発展に資する意思決定支援の新たな枠組みを構築することを目的とする．

§ 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 背景では，政策課題や地方自治体における課題への重要性について述べる. 目的は背景で挙げた課題に対して，ICT とデータ分析手法を用いて解決するアプローチについて提案することを述べる.

第 2 章 使用する対象データの収集とデータ結合を行い，単一代入法や多重代入法などの欠損値の補完や機械学習モデルの解釈手法を使用した前処理について述べる. また，ロジックモデルの解説や政策課題への内閣府の取り組み, 地方自治体の取り組みについて述べる,

第 3 章 本研究の提案手法に用いる因果探索を活用し収集したデータの入出力の分類の方法について述べる．データ包絡分析の効率の評価方法や改善策の導出について解説する.

第 4 章 提案手法ではデータの収集方法やデータの結合方法、それらのデータの補完方法を述べる。因果探索で使用した手法やデータ包絡分析での方法を解説する。システムのプログラムとそれらに用いるデータベースの作成方法を理論の区切りごとに説明する.

第 5 章 提案手法に基づいて意思決定システムを構築して，実際に施策立案の改善策の導出を行った結果を示す．そして，本研究の提案手法によって得られた結果が有意であることを示す.

第 6 章 本研究で述べている提案手法をまとめて説明する．また，今後の課題について述べる.

第2章

サイバー空間からのデータ収集と処理

§ 2.1　多様な要因を考慮したデータセットの作成

政策課題の解決のためには，数多くの要因が考えられる．そのため要因に着目し、改善案を作成する。そのモデルを作成するためには，それらを表現する説明変数を多く考慮する必要がある．しかし，我々が一般に取得できるデータ，すなわちオープンデータには，そのアクセスに限界がある．国勢調査の結果など，統計的なデータは比較的公開されているものの，土地価格の要因として重要視される地理的なデータ，たとえば，特定の施設の位置などといったものは，依然として取得が容易ではない.

そこで本研究では, 対象データの収集を異なるウェブサイトから行った. これらのサイトは, 情報の網羅性，更新頻度，データの信頼性を考慮した選定した. 国土交通データプラットフォーム，e-Stat，不動産情報ライブラ，犯罪データの概要について説明する.

国土交通データプラットフォーム

国土交通省は、保有する多くのデータと民間等のデータを連携し、Society 5.0 が目指すフィジカル空間をサイバー空間に再現するデジタルツインにより、業務の効率化やスマートシティ等の国土交通省の施策の高度化、産学官連携によるイノベーションの創出を目指し、国土交通データプラットフォームの構築を進めています。

デジタルツインの実現を目指し、３次元データ視覚化機能、データハブ機能、情報発信機能を有するプラットフォームの構築を進めており、「国土交通データプラットフォーム」を公開している。これらの機能は地方自治体をはじめ，その他地域の活性化に関心を持つ人々に対して一般に公開されており，地方自治体における

図 2.1: 国土交通データプラットフォーム [?]

図 2.2: e-Stat [4]

地域課題の抽出， 地域版総合戦略の立案といった活用法に加えて，地方創生に関心のある民間の団体・個人による活用も期待される．

e-Stat

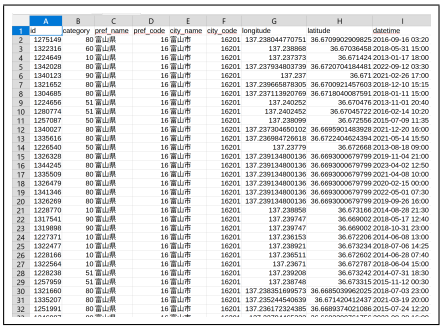
日本の政府統計に関する情報のワンストップサービスを実現することを目指した政府統計ポータルサイトです. これまで各府省等が独自に運用する WEB サイトに散在していた統計関係情報を本サイトに集約, 社会の情報基盤たる統計結果を誰でも利用しやすいかたちで提供することを目指し, 各府省等が登録した統計表ファイル, 統計データ, 公表予定, 新着情報, 調査票項目情報, 統計分類等の各種統計関係情報を提供していきます.

e-Stat で提供されている統計データは，集計されている区分ごとに，全国ごと，都道府県ごと，市区町村ごと，“ ··· 丁目” といったの小地域ごと，グリッドセルごとの 5 種類が存在する．

不動産情報ライブラリ

不動産情報ライブラリとは, 不動産の取引価格, 地価公示等の価格情報や防災情報, 都市計画情報, 周辺施設情報等, 不動産に関する情報をご覧になることができる国土交通省の WEB サイトである. 利用にあたって特別なソフトを必要としない WebGIS を採用し，スマートフォンでも閲覧可能である．表示するデータについては，民間事業者等とのシステム連携を可能としており，新たなサービスの基盤となることが期待されている．

犯罪データ



	A	B	C	D	E	F	G	H	I
1	id	category	pref_name	pref_code	city_name	city_code	longitude	latitude	datetime
2	1275149		80 富山県		16 富山市	16201	137.238044770751	36.6709902909825	2016-09-16 03:20
3	1322316		60 富山県		16 富山市	16201	137.238868	36.67036458	2018-05-31 15:00
4	1224649		10 富山県		16 富山市	16201	137.237373	36.671424	2013-01-17 18:00
5	1342028		80 富山県		16 富山市	16201	137.237934803739	36.6720704184481	2022-09-12 03:30
6	1340123		90 富山県		16 富山市	16201	137.237	36.671	2021-02-26 17:00
7	1321652		80 富山県		16 富山市	16201	137.239665878305	36.6709921457603	2018-12-10 15:15
8	1304685		80 富山県		16 富山市	16201	137.237113920769	36.6718040087591	2018-01-11 15:00
9	1224656		51 富山県		16 富山市	16201	137.240252	36.670476	2013-11-01 20:40
10	1280774		51 富山県		16 富山市	16201	137.2402452	36.67045722	2016-02-14 10:20
11	1257087		50 富山県		16 富山市	16201	137.238099	36.672556	2015-07-09 11:35
12	1340027		80 富山県		16 富山市	16201	137.237304650102	36.6695901483928	2021-12-20 16:00
13	1335616		80 富山県		16 富山市	16201	137.236984729618	36.6722404624394	2021-05-14 15:50
14	1226540		50 富山県		16 富山市	16201	137.23779	36.672668	2013-08-18 09:00
15	1326328		80 富山県		16 富山市	16201	137.239134800136	36.6693000679799	2019-11-04 21:00
16	1344245		80 富山県		16 富山市	16201	137.239134800136	36.6693000679799	2023-04-02 12:50
17	1335509		80 富山県		16 富山市	16201	137.239134800136	36.6693000679799	2021-04-08 10:00
18	1326479		80 富山県		16 富山市	16201	137.239134800136	36.6693000679799	2020-02-15 00:00
19	1341346		80 富山県		16 富山市	16201	137.239134800136	36.6693000679799	2022-05-01 07:30
20	1326269		80 富山県		16 富山市	16201	137.239134800136	36.6693000679799	2019-09-26 16:00
21	1228770		10 富山県		16 富山市	16201	137.238958	36.673166	2014-08-28 21:30
22	1317541		90 富山県		16 富山市	16201	137.239747	36.669002	2018-05-17 12:40
23	1319898		90 富山県		16 富山市	16201	137.239747	36.669002	2018-10-31 23:00
24	1227371		10 富山県		16 富山市	16201	137.236153	36.672206	2014-06-08 13:00
25	1322477		10 富山県		16 富山市	16201	137.238921	36.673234	2018-07-06 14:25
26	1228166		10 富山県		16 富山市	16201	137.236511	36.672602	2014-06-28 07:40
27	1322564		10 富山県		16 富山市	16201	137.23671	36.672787	2018-06-04 15:00
28	1228238		51 富山県		16 富山市	16201	137.239208	36.673242	2014-07-31 18:30
29	1257959		51 富山県		16 富山市	16201	137.238748	36.673315	2015-11-12 00:30
30	1321660		80 富山県		16 富山市	16201	137.238351699573	36.6685039962025	2018-07-03 23:00
31	1335207		80 富山県		16 富山市	16201	137.235244540639	36.6714204124372	2021-03-19 20:00
32	1251991		80 富山県		16 富山市	16201	137.236172324385	36.6689374021086	2015-07-24 12:20

図 2.3: 犯罪データからの変数抽出

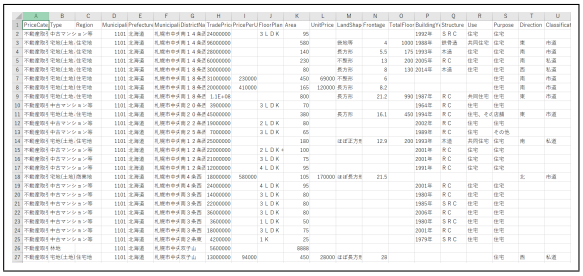


図 2.4: 不動産情報のデータセット

今回の数値実験で使用する犯罪発生データは，富山県警が公開している犯罪発生マップから取得したものを用いる．犯罪発生データには，データ項目として発生時刻，罪種，発生場所の緯度や経度が含まれており，自転車盗難，ひったくり，車上ねらい，部品ねらい，わいせつ，声かけ，つきまとい，タイヤ盗難の 7 種類の路上犯罪が収録されている．また，なお，犯罪発生マップは，このような分析を目的としておらず，掲載されている情報に誤差や欠損が存在する可能性があることに注意されたい．

§ 2.2　データクリーニングによる前処理

データセット内での欠損データは多くの分析にとって課題となっている．欠損データを適切に扱うことは分析の正確性を保証し，より信頼性の高い結果を得るためには不可欠となる．欠損データの補完にはデータの特性を考慮し，適切な手法を選択する必要がある．欠損データを補完することで正確な統計分析を行うことが可能となる．

欠損値には 3 つのパターンがあるとされており，欠損データの割合や欠損のパターンに基づいて様々なデータの補完方法を行う（図 2.5）．

- 完全無作為欠測（Missing Completely At Random: MCAR）
  欠損が完全にランダムである状況を指している．データが欠損している理由が観測されたデータや欠損しているデータ自体とは無関係である．この場合，欠損データを除外してもデータ分析においてバイアスが生じることは少ないと考えることができる．

- 無作為欠測（Missing At Random: MAR）
  欠損がランダムであるが，他の観測されたデータによって説明可能である状

況を指す．他の変数を使い欠損データのパターンをモデル化しその影響を緩和することができる．

- 無作為でない欠測（Missing Not At Random: MNAR）
  欠損が観測されていないデータ事態に依存している状況を指す．この場合は，欠損データは分析にバイアスをもたらす可能性が高く，これを考慮するために特別な統計的手法が必要となる．

欠損値が発生したとして対処法として大きく 5 つある．欠損しているデータにより補完方法を使い分ける．

- 欠損データの削除
- 単一代入法
- 多重代入法
- 完全情報最尤推定法

欠損データの削除は，欠損値を含むデータポイントをデータセットから完全に除去するアプローチである．これはリストワイズ法と呼ばれ，欠損が MCAR であれば推定結果は不偏であるが MAr の場合には偏りが出ることがある．この方法は欠損が少量であり，その除去がデータセット全体に大きな影響を与えない場合に適している．メリットとしては基本的な操作が簡単なことがあげられるが，デメリットとしてはデータ数が少ないときに制度に大きな影響を及ぼすことがある点である．

単一代入法

単一補完とは何かしらの 1 つの値で欠損値を補完するという方法の事である．欠測している状態より前のデータの中から，最後に観測された値を使って欠測値補完をする方法を LOCF 法（Last Observation Carried Forward）という．これは最後に得られた値がそれ以降ずっと続くという仮定がある．また，平均値補完や中央値補完などがある．メリットとしては列での集計量を代入できるため極端に不適な値をとることがなくなる．逆にデメリットとしてあり得る値を撮るが故欠損値という情報は損失し訳す妥当とは言えない値を代入する場合があるという点である．この方法は欠損データがランダムに分布している MCAR に最適であり，簡単に実装することができる利点があるが，元データの分布を変える可能性がある．

多重代入法

1つの欠損部分に複数の値を代入する方法の事である．欠損値を代入したデータセットを複数作成し，その結果を統合することで欠損値のデータの統計的推定を行う．データセットを複数作成することで，欠損値による推定の不安定でを結果に反映させている．

多重代入法による分析の流れは以下の通りである．N 個の代入済みのデータを生成し，それぞれのデータでモデル作成や分析を行い，その結果を統合する．統合方法は次のように行われる．

統合されたパラメータ θ_M は m 番目の代入済みデータセット推定されたパラメータ $\hat{\theta}_m$ を算術平均で求める

$$\theta_M = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_m \tag{2.1}$$

多重代入法におけるパラメータの分散は2種類存在する．代入内分散と代入間分散である．代入内分散はパラメータ $\hat{\theta}_m$ の分散 $var(\hat{\theta}_m)$ の算術平均である式（2.2）．代入間分散は $\hat{\theta}_M$ に自由度が1つ減っているため次のように求める式（2.3）．

$$\hat{W}_M = \frac{1}{M}\sum_{m=1}^{M} var(\hat{\theta}_m) \tag{2.2}$$

$$\hat{B}_M = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\theta}_m - \hat{\theta}_m)^2 \tag{2.3}$$

また $\hat{\theta}_m$ の分散 T_M は代入内分散と代入間分散と合算して求める．

$$T_M = \hat{W}_M + (1 + \frac{1}{M})\hat{B}_M \tag{2.4}$$

機械学習，たとえば，近年急速に注目されている深層ニューラルネットワークといったアルゴリズムは，複雑かつ非線形な性質であってもモデリングすることができる．すなわち，より予測精度の大きいモデルを作成することができる．しかしながら，一般にモデルの精度が大きくなるほど，その解釈性は小さくなる性質がある．

機械学習モデルの解釈手法の概要

データ分析で機械学習を使用する機会が増えており，適切な機械学習モデルの選択はより正確な分析結果を導くうえで重要である．特徴量が多いと計算負荷が高

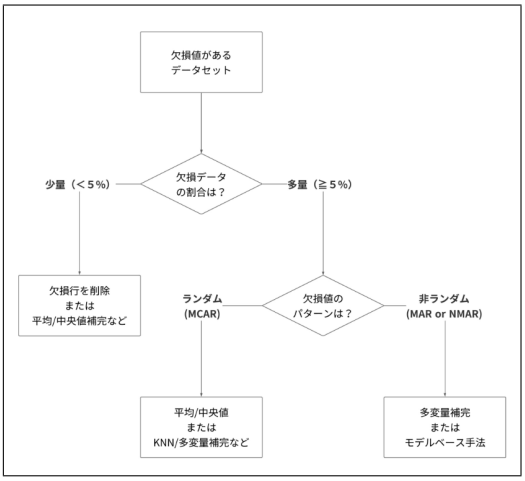


図 2.5: 欠損値の処理のフローチャート [?]

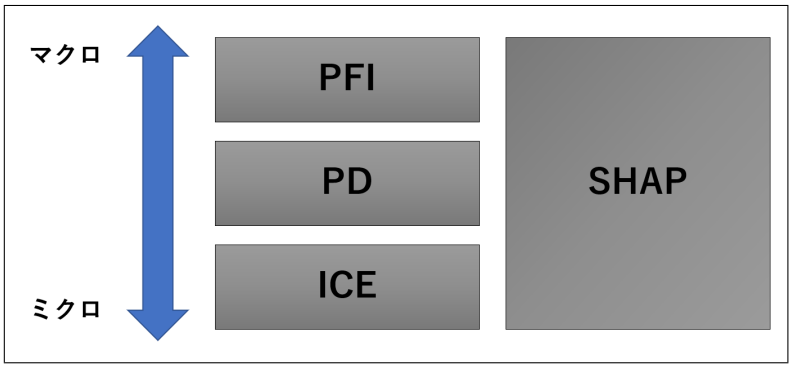


図 2.6: 機械学習モデルの解釈手法

まる．そこで本質的な特徴だけにあらかじめ絞ることが出来ると効果的であると考えられる．

機械学習によって作成されたモデルに対して，何らかの解釈を与える手法はいくつか存在するが，特に有用なものとして，以下の 4 つが挙げられる．

- Permutation Feature Importance（PFI）
- Partical Dependence（PD）
- Individual Conditional Expectaion（ICE）
- Shapley Additive Explanations（SHAP）

それぞれは何を解釈できるのかが異なり，用途によって使い分ける必要がある（図 2.6 参照）．例えば，モデル全体の傾向など，マクロな視点から解釈する場合は PFI を，出力されたひとつの予測値に対する根拠など，ミクロな視点を知りたい場合は ICE を用いるべきだろう．本研究では，ミクロな視点から解釈できるものの，マクロな視点からの解釈も可能な SHAP [8] を用いることとする．

SHAP

これまでに存在した解釈手法 Additive Feature Attribution Methods に協力ゲーム理論の Shapley Values を導入して改良したものであり，機械学習モデルの解釈手法の 1 つである．

Additive Feature Attribution Methods は特徴量の線形関数を用いた機械学習の解釈手法の事である．元の特徴量 x_i から，その特徴量の存在有無を示すバイナリ値 z'_i に代えて扱う．定義式は以下の通りである．

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i \tag{2.5}$$

ここで $g(z')$ は元の予測モデル $f(x)$ に対する局所的な代替モデルを表す．ϕ_i は特徴量 i の有無によってどれだけ最終的な予測結果に影響があるかを示している．各特徴量の予測への寄与度が ϕ_i で，特徴を使ったかどうかのバイナリベクトル空間上の関数で，説明したい関数 f に近似する．

これを以下の 3 つの条件 [9].

1. Local accuracy
   貢献度モデルの出力と貢献度の合計が一致する．

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i x'_i \quad x = h_x(x'_i) \tag{2.6}$$

2. Missingness
   真のモデルに存在しない特徴は出力に影響を与えない．

$$x'_i = 0 \Rightarrow \phi_i = 0 \tag{2.7}$$

3. Consistency
   ある特徴量 i の影響が新しいモデル f' で，増加または同等である場合その特徴量の貢献度も減少してはならない．$f_x(z') = f(h_x(z'))$ と，i 番目の要素を 0 にする z'_1 は $z'_i = 0$ を表す．任意の 2 つのモデル f と f' に対して，全ての入力 $z' \in \{0,1\}^M$ 対して

$$f'_x(z') - f'_x(z'\ i) \geq f_x(z') - f_x(z'\ i) \tag{2.8}$$

   ならば $\phi_i(f', x) \geq \phi_i(f, x)$ である．

のもとで，式（2.5）に従い，それぞれの条件のすべて満たす説明モデル g はただ 1 つだけ存在する．

$$\phi_i(f, x) = \sum z' \subseteq x' \frac{|z'|!(M - |z'| - 1)!}{M} [f_x(z') - f_x(z'\ i)] \tag{2.9}$$

$|z'|$ は z' におけるゼロでない要素の数を表し $z' \subseteq x'$ は z' のゼロでない要素が x' のゼロでない要素の部分集合であるようなすべての z' ベクトルを意味する．式（2.9）により，各特徴量の寄与度 ϕ_i は一意に求まることがゲーム理論により示されている．

予測精度と解釈性のトレードオフ

以下のような線形回帰モデルを考える．

$$f(X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \tag{2.10}$$

このとき，X_1 が 1 だけ大きくなると，$f(X_1, X_2)$ は β_1 倍だけ大きくなることが明示的に分かる．このように，線形回帰モデルは，目的変数と説明変数とのあいだに単純な関係を仮定しており，モデルに対する透明性が高いと言える．これを，一般に解釈性が高いと言う．

一方で，比較的近年に発表されたアルゴリズム，例えば深層ニューラルネットワークやランダムフォレストなどは，目的変数と説明変数とのあいだに線形性などの仮定を置いていない．よって，より複雑な関係をモデリングできるようになり，一般に線形回帰モデルよりも予測精度は大きくなりやすい．しかしながら，線形回帰モデルと違い，その複雑さから，なぜその予測値を出力するのかを理解することができず，その中身はブラックボックスとなりやすい．これを，一般に「解釈性が低い」と言う．

SHAP

$\mathbf{X} = (X_1, \ldots, X_J)$ を説明変数とする学習済みのモデルを $\hat{f}(\mathbf{X})$ とする．インスタンス i の説明変数が $\mathbf{x_i} = (x_{i1}, \ldots, x_{iJ})$ とすると，インスタンス i の予測値は $\hat{f}(\mathbf{x}_i)$ である．ここで，予測の期待値を $\mathbb{E}\left[\hat{f}(\mathbf{X})\right]$，インスタンス i の説明変数 x_{ij} の貢献度 ϕ_{ij} としたとき，

$$\hat{f}(\mathbf{X}_i) - \mathbb{E}\left[\hat{f}(\mathbf{X})\right] = \sum_{j=1}^{J} \phi_{ij} \tag{2.11}$$

のように，期待値からの差分を貢献度の総和で表現できるように，貢献度を分解することが，SHAP の基本的な考え方である．線形モデルであれば，比較的容易に分解することができるが，非線形モデルではこのままでは難しい．そのため，SHAP では，協力ゲーム理論の Shapley 値の考え方を用いて，貢献度を分解する．

§ 2.3　政策課題解決と多様な要因の関係

経済社会構造が急速に変化するわが国において，限られた資源を有効活用しながら国民に信頼される行政を展開するために，政策の対象に関するデータを収集し，それに基づいて政策における意思決定を行うという考え方である証拠に基づく政策立案（Evidence Based Policy Making: EBPM）を推進することが重要視されている．

しかし，全ての政策において効果的な EBPM を適用するためには，膨大かつ多種多様なデータを収集・保存・管理し，それらのデータを適切かつ高速に高い信頼度を保って選択・統合・分析する必要があり，担当者に対する大きな負担となるため人手のみでそれらを行うことは困難となる．

そのため，特に地方自治体において EBPM を政策の広範囲に適用することは人員の観点から見ても難しい課題であると考えられる．これらのことから，EBPM において適切なエビデンスの収集・分析をおこなうには，ICT を用いることが欠かせない．

また，そういった場合では，一般に専門的な ICT との接点が少ないと考えられる地方自治体の職員に対して感覚的に理解しやすいシステムを提供するか，庁内全体で講習会を開催するなどして ICT に関する知識を醸成する必要があるとも考えられる．

ロジックモデル

ロジックモデルとは，プログラム理論に基づいている．プログラム理論とは，プログラムの実施と成果の出現の間に介在するメカニズムであり，プログラムを実施することで，それがどのようなプロセスを経て成果が表れるのかを表す仮説のことである．そのプログラム理論を踏まえ，その仮説を明確に示すため政策課題とその現状に対し，政策手段から政策目的までの経路を端的に図示化したものをロジックモデルという．

行政機関が行う政策の評価に関する法律の第 3 条に政策に基づき実施し，又は実施しようとしている行政上の一連の行為が国民生活および社会経済に及ぼし，又は及ぼすことが見込まれる影響のことを表す政策効果という言葉がある．ロジックモデルは政策の実施により，その目的が達成差荒れるまでの理論的な因果関係を明示したものであり，政策効果と近いものであると考えることが出来る [13].

ロジックモデルの作成には 4 つのステップを踏むことで作成される．

1. 現状の把握や課題の吟味
   手段の検討やデータ分析をいきなり行うのではなく，前提となる背景や解決したい社会問題等の政策課題の精緻化や目的の明確化を行う必要がある．

2. 論理の流れの検討
   政策目的から手段に至るまでの，論理的なつながりは，事業実施に必要な予算等の投入する資源をあらわすインプット，事業の実施内容等の政策手段による活動のアクティビティ，政策手段による活動目標や実績であるアウトプット，事業を実施し期待される変化である成果目標を表すアウトカムをの 4 つである．これらを検討し明確にする．
3. アウトプットやアウトカムをはかる指標の設定
   指標を設定する際，ムリな定量化や取得しやすい数字の設定等によって戦略をゆがめないように注意することが大事である．
4. ロジックと指標の再吟味
   改めて，2 の手順を両方向でロジックを確認する．また，ロジックモデルの用途や効果検証をすべきタイミング，周辺の事情を踏まえたうえで各項目や指標の内容を調整する．

これらのモデルを作成するにあたり，問題深掘り型と仮説思考型の 2 つのアプローチがある．

1 つ目の「問題深掘り型」は生じている問題とその根本の原因を構造的な整理を行うことで解決すべき課題を特定したうえで最適な解決手法を検討する思考の手順の事である．これは一般的に，新規で政策課題を探索し吟味する場合に適している．要因や対策を網羅的に検討するメリットがあるが，目的から過度に逸脱してしまうケースが見られる点がある．

2 つ目の「仮説思考型」はあるべき姿と現状との差を特定したうえで，あるべき姿から逆算を行い実現すべき状況を検討する思考手順の事である．こちらは一般的に政策課題や政策方針について既に議論がなされている場合に適している [14].

EBPM の活用

EBPM とは，政策目的を明確化させその目的のため本当に効果が上がる手段を考え，政策の基本的な枠組みを証拠に基づいて明確にするための取り組みである．そこでは，限られた資源を有効に活用し信頼される行政を展開するため EBPM を推進する必要があると考えられる．

平成 30 年に政府が発表した「内閣府本府 EBPM 取組方針」で政策の企画立案をその場限りのエピソードに頼るのではなく，政策目的を明確化したうえで政策効果の測定に重要な関連を持つ情報やデータ（エビデンス）に基づくものとすることが求められていると定義している．

政策の立案や推進の過程においてエビデンスは 3 つのタイミングで登場すると考えられる．1 つ目は，政策の必要性が問われた際その解答となる課題設定のエビ

図 2.7: ロジックモデルの構造 [15]

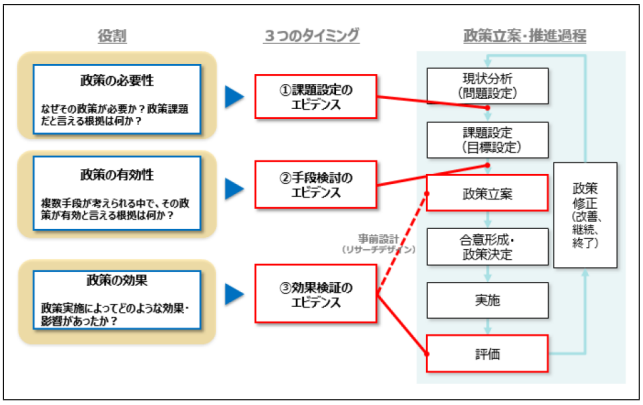


図 2.8: EBPM [16]

デンスである．2つ目は，解決すべき課題が明確になると次に検討しなければならないのは手段である．複数の手段が考えられる中で複数の政策手段を比較検討し判断する手段検討のエビデンスである．3つ目は，政策実施によりどのような効果や影響，また副作用的な影響も踏まえそれらががあったかを判断する効果検証のエビデンスである．

これらのように3つのエビデンスそれぞれで役割は異なる．3段階の度のエビデンスにも困難が伴う上に，情報や時間も限られた中で意思決定を行う必要がある．こうした中で，正しい意思決定を行い続けることは難しいと考えられる．エビデンスには課題・手段・効果の3つのタイミングがあることを認識し，それらを切り分けたうえでエビデンスをとらえることが重要である．

地域課題

地域課題とは，地域が抱える経済や社会，環境などの問題の事である．地域経済は地域ごとに異なる特徴や要因を持ち地域の活性化や持続的な発展を妨げる要因となる．これらの解決には地域の特性や課題を的確に把握し，効果的な対策を講じることが重要である．現在，多くの自治体や政府が主導する地方創生を目指しているが，根本には以下の3つの課題の是正が求められている．

- 人口減少
- 少子高齢化
- 地域経済の衰退

2020 年の日本人口は約 1 億 2369 万人で前年の 2019 年比で約 19 万人減少しており，29 の都道府県で人口減少が観測された．人口減少が進むと地域の将来性や活力の低下や労働力不足，税収減少などの課題が引き起こされることが懸念される．65 歳以上の人口の割合が集落の半数を超えたり人口が減少し地域としての昨日がうまく回らず，社会的共同生活が困難な限界集落ができてしまう．また，住民の人口が 0 となってしまう消滅集落も増加する．山間地や離島に位置し現在も消滅集落になりかけている集落が約 63,237 もあるとされている．これらの問題は人口減少が影響していると考えられる．

高齢者（65 歳以上）の割合は約 29.3 %，若年層（0〜14 歳）の割合は約 12.3 %となっている．少子高齢化は，介護や医療サービスの不足や労働不足などの課題を引き起こす．若年層の減少により，地域経済活動が低下する可能性がある，自治体や企業は地域課題を解決するために若年層の定着を促す取り組みや，高齢者の社会参加支援等を行っている．

地方都市や田舎地域を中心に企業倒産や雇用機会の減少，人口減少や高齢化が起こっている地域経済の衰退が起こっている．地域の産業構造が単一化し競争力が低下すると，地域内での就職や雇用が困難になる可能性がある．地域経済の活性化を目指し，自治体や企業は新しい産業の導入や地意識資源の活用など様々な施策を展開している．

地域課題への取り組み

地域課題解決のための取り組みは，観光振興や産業支援，公共サービスなど多岐にわたる．地域課題の解決は地域のインフラ・経済の発展促進や住民の暮らしやすさの改善，地域コミュニティの強化などの影響をもたらす．しかし，解決のためには先ほども述べたように少子高齢化や東京一極集中などの人口減少を解決することが大切である．

和歌山県では，和歌山県及び県内 30 市町村では，県内の中小企業等における人手不足の解消，県内での起業及び移住・定住の促進を目的に，移住支援事業，マッチング支援事業及び起業支援事業を実施している．また，移住者向けの暮らしと仕事だけでなく，空き家の家財撤去費用の補助や空き家の回収などを行うための住まいのサポート，保育料や在宅育児支援，転入学や編入学等の教育についての情報などの子育てのサポートも行っている．

鳥取県では，日本で人口が 1 番少ない県であるため，第 1 次産業をはじめ後継者難という問題に直面している．そこで，鳥取県ではデジタル推進が取り組まれている．中山間地問題，農林水産や産業，観光の振興や医療福祉，防災などにデジタル活用をする地域 DX を行っている．また，県職員の業務を自働化や効率化を推進し，県民への行政サービスの質やスピードを向上させるだけでなく，生み

出された時間を創造性が必要な業務や人でないとできない企画を考える時間に活用することを促す県庁 DX がある [17].

新しい施設を建設する場合，人手や工事費・維持費の問題が生じる．和歌山県の取り組みのように，地域資源である未利用の土地や建物を新たな使い方で活用することや自然や歴史・文化等の観光資源を活かす取り組みも有効である．また，協働できる事業やプロジェクトの情報を発信することにより熱意やスキルを持った人材と企業がマッチする機会を作り出し，地域の関係人口を増やすことが可能である．デジタル推進を行っている鳥取県のように，自治体の業務改善を行い人口減少による労働力定価の対策のため，ICT の活用もポイントとなってくる．

第3章

要因間の因果関係と主体の効率

§ 3.1　因果探索に基づく入力と出力の分類

因果探索とは観測データを用いて，そのデータ群の複数の観測データにおいて，それぞれの値がお互いに及ぼしあっている影響の度合いを構造的に示した因果グラフを導出するための教師なし学習のことである．因果探索には 3 つのアプローチが存在する．1 つ目は線形性や誤差項に分布を仮定しないアプローチであるノンパラメトリックである．

また，類似する手法として因果推論が挙げられるが，因果推論では因果関係の向きが既知である場合にその因果関係が本当に有意であるのかをデータから分析する手法であるのに対し，因果探索は因果関係が不明かつ因果関係の向きも不明であるデータ群に対して，それらの間に因果関係が成立するかを導く手法である．その 1 つに Linear Non-Gaussian Acyclic Model（LiNGAM）がある．

例えば,「ある小売店でアイスクリームの安売りを行った際にアイスクリームの売り上げが向上した．また，同日の小売店の来客数は前日より 100 人多かった」というケースがあったとする．このとき，アイスクリームの安売りを行ったことが売り上げの向上につながったかどうかを調べるのが因果推論である．これに対して，アイスクリームが安かったから来客数が増加したのか，来客数が多かったためにアイスクリームの売り上げが向上したのかという因果の方向性も含めて分析を行うのが因果探索である．

このような特徴を持つため，因果探索は適用されるデータの分野に対しての制約が少なく，様々な分野のデータに適用することができる．それゆえ，因果探索を用いた応用研究も盛んにおこなわれており，疫学，経済学，神経科学，化学，医学をはじめとした幅広い分野のほか土木計画学の研究でも用いられている．

LiNGAM

近年，因果探索の手法における研究が活発化したことで，因果探索における様々なモデルが提唱されている．代表的なものとしては独立主成分分析の手法を用いたセミパラメトリックなもので，非時系列データに対しても適用可能な LiNGAM が挙げられる．LiNGAM とは，一般的に以下の式（3.1）のように定式化される．

$$x_i = \sum_{i \neq j} b_{ij} x_j + e_i \qquad i, j = 1, \cdots, p \tag{3.1}$$

LiNGAM は 4 つ仮定を置くことで因果ダイアグラムを特定することが可能である．[?].

- 実際に観測されている変数である内生変数と内生変数以外の変数で内生変数のそれぞれに関する未知の値である外生変数をつなぐ関数は線形関数とする．
- 外生変数の分布は非ガウス連続分布とする．
- 因果グラフは非巡回とする．
- 外生変数は互いに独立とする．

ここで，図??のような内生変数が 4 つの因果グラフがあると考える．このとき，仮定 3 があることによって 4 つの内生変数のうち，どの変数からも因果的影響を受けない変数が少なくとも 1 つ必ず同定される．

$$\begin{bmatrix} x_1 \\ x_2 \\ x_4 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_4 \\ x_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_4 \\ e_3 \end{bmatrix} \tag{3.2}$$

つまり，式（3.2）のように因果的影響を受ける場合にその値，受けない場合には 0 を入れたパス係数行列を考えると必ず右肩が逆三角形に全て 0 になる行列となる．そのため，式（3.2）における x_1 のように全てのパスに対する係数が 0 となる内生変数を因果グラフから除外し，再度パス係数行列を求めるという操作を繰り返すことによって未知である因果グラフを同定することが可能になる．

また，上記の同定法を成立させるにあたって，仮定 4 がなくてはならない．前述のとおり，LiNGAM における因果関係の同定では内生変数同士の因果関係のみに着目して因果グラフの最も外側に位置する内生変数を順に除外する方法をとるため，内生変数同士の間に成立する因果関係以外の因果関係が内生変数間に発生してはならない．

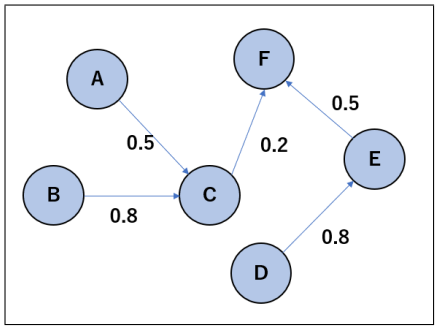


図 3.1: 因果グラフの例

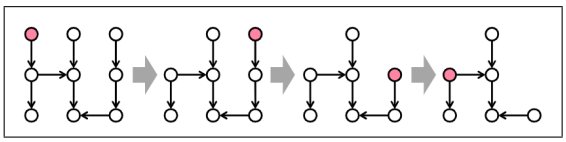


図 3.2: DirectLiNGAM のアルゴリズム

ここで，もし外生変数同士が独立ではなければ外生変数同士の間に因果関係が生じてしまい，内生変数同士がそれぞれの内生変数に関わる外生変数同士の因果関係を介在として内生変数間に存在しない新たな因果関係を持ってしまう．このような場合には前述のような因果関係の同定法が成り立たなくなるため，LiNGAM における仮説 4 は必ず必要となる．

Direct-LiNGAM

前述のようなアルゴリズムによって内生変数間の因果関係を推定する LiNGAM であるが，推定時の計算方法の違いによって現在までにいくつかのアプローチが提唱されている．代表的な例として，独立成分分析によるアプローチである ICA-LiNGAM や回帰分析と独立性評価によるアプローチである Direct-LiNGAM などが挙げられる．その中でも，本研究で取り扱う Direct-LiNGAM に関する解説を行う．DirectLiNGAM は 2 変数間の単回帰を繰り返して因果グラフにおいて親が存在しない外生変数を 1 つずつ順番に検出して取り除く作業を繰り返すことで，因果順序を推定する手法である．

Direct-LiNGAM によるアプローチの基本的な考え方は

- 観測変数群から 2 変数を取り出しそれらの変数間に成り立つ因果関係を同定することを繰り返して観測変数群全体における因果の始まりとなる変数を探す．
- その変数を観測変数群から除外し，残った変数のみで再度，観測変数群を形成する．

という 2 つの操作を観測変数群に属する変数が存在しなくなるまで繰り返すことによって元の観測変数群の因果グラフを同定するというものである．例として，観測変数群から 2 変数 x_1, x_2 を取り出し，以下の構造方程式モデルが背後にあるものと想定する．

$$\begin{cases} x_1 = e_1 \\ x_2 = b_{21}x_1 + e_2 \end{cases} \tag{3.3}$$

ここで e_1, e_2 は互いに独立かつ非ガウス分布に従い，$b_{21} \neq 0$ とする．これについて単回帰分析を行うことによって，因果順序の同定を行う．まず x_2 を目的変数，x_1 を説明変数として回帰する場合を考える．この場合元の構造方程式モデルの第 2 式がそのまま成り立つことになる．そのため，回帰残差は e_2 となりこれは $x_1 = e_1$ と独立となる．一方，x_1 を目的変数，x_2 を説明変数とした場合，回帰残差 r_1 は

$$r_1 = \left\{1 - \frac{b_{21}cov(x_1, x_2)}{var(x_2)}\right\} e_1 - \frac{b_{21}var(x_1)}{var(x_2)} e_2 \tag{3.4}$$

となり e_2 の項が出てくる．一方冒頭の構造方程式に戻ると，x_2 は式として e_2 を含むので，この回帰残差と説明変数 x_2 とは従属する．この従属性の成立に関しては前述の仮定 2 にて示した「外生変数が非ガウス分布とする」というきまりに基づいており，以下に示すダルモア・スキットビッチの定理を用いている．

ダルモア・スキットビッチの定理

2 つの確率変数 y_1, y_2 が，互いに独立な確率変数 $s_i (i = 1, ..., q)$ の線形和で下記のように表されているとする．この時，もし y_1, y_2 が独立なら，$\alpha_j \beta_j \neq 0$ となるような変数 s_j はガウス分布に従う

$$y_1 = \sum_{i=1}^{q} \alpha_i s_i \tag{3.5}$$

$$y_2 = \sum_{i=1}^{q} \beta_i s_i \tag{3.6}$$

上記の考察から，両方のパターンで回帰分析を行い残差と説明変数の独立性を判定することで因果の向きを推定することが可能となる．なお独立性の評価には相互情報量という量を用いる．この量が 0 となるときに独立であると判定するが，実際には推定誤差があり正確には 0 にはならないため，相互情報量が 0 に近い方を独立とみなして因果の順序を決定する [20].

パラメータ推定を行う DirectLiNGAM.fit メソッドは，標準化を行う処理が含まれているため，前処理で標準化を行う必要はない．しかしその場合でも前処理での標準化の有無によって推定された隣接行列の意味合いが変わるため，分析の目的に応じて標準化の有無を使い分ける必要がある．

まず前処理で標準化を行わない場合，推定された隣接行列の第 (i,j) 成分 a_{ij} は因果グラフにおける辺 $x_j \to x_i$ の重み，すなわち「$x_j \to x_i$ の直接効果」を表す．例えば $x_j \to x_i$ の直接効果や総合効果を知りたいときは，標準化を行わずに推定することになる．

一方で前処理で標準化を行う場合，推定された隣接行列の第 (i,j) 成分 a_{ij} は元の $x_j \to x_i$ ではなく標準化後の変数 $z_i = \frac{(x_i - \mu_i)}{\sigma_i}$，$z_j = \frac{(x_j - \mu_j)}{\sigma_j}$ 間の直接効果を表すことになる．標準化ありで推定を行うケースとしては例えば，x_i の原因となる変数の重要度を知りたいときが挙げられる．変数の重要度を比較する際は隣接行列の成分の絶対値の大小を比較するので，スケールの違いで結果が変化しないように標準化を行う [21].

§ 3.2　データ包絡分析の効率の評価

データ包絡分析（Data Envelopment Analysis: DEA）の分析対象となる事業体は銀行，デパート，病院，都道府県，学校，等のように多種多様である．事業体が n 個存在するとし，それらを DMU_1,$\mathrm{DMU}_2 \ldots \mathrm{DMU}_n$ と番号づける．次に，それぞれの生産活動に共通した投入（入力）項目と産出（出力）項目を選ぶ．ごく一般的な選び方として次のような方針をとる

- まず，n 個の事業体はある程度独立して経営を行っていると考える．数学的にいうと，全ての事業体が一時独立性を満たしていると考える．
- 投入項目，産出項目とも数値データを準備する．一般に，入出力値は原則としてすべて正とする．また，すべての事業体は同数の入出力を持っていると考える．入出力が各事業体によって違う場合は双対比較ができないので，DEA では通常取り扱わない．
- 投入項目，産出項目の選定にあたっては，企業のトップや政策決定者に直接聞いてみるのが一番よく，経営者が現実の意思決定に際して，どのような項目に注目してそれぞれの事業体の評価を判断しているのかを確かめる必要性がある．もしそのような機会が得られない場合は，自分が試してみたいと思う入出力項目を選べばよい
- 原則として入力に関していえば，値の小さいものほど好ましく，出力に関しては大きいものほど好ましいものを選ぶ．
- 投入項目，産出項目の数値の単位は任意に取ってよい．

いま，m 個の投入項目と s 個の産出項目が選定され，DMU_j の投入データを $x_1j, x_2j, \ldots x_nj$ 産出（出力）データを $y_{1j}, y_{2j}, \ldots y_{nj}$ とする．各事業体の生産活動に関するデータを縦に並べて行列を作り，入力データ行列 X と出力データ行列 Y とする．それらは次のような $(m \times n)$ 型，$(s \times n)$ 型の行列で表される．

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}, \quad Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{s1} & y_{s2} & \cdots & y_{sn} \end{bmatrix} \tag{3.7}$$

複数の事業体それぞれについて比率尺度で効率性を測定していくが，対象になっている事業体を k 番目とし，DMU_k と書くことにする．以下，添字 k は $1, 2, \ldots n$ のどれかを指すものとする．入力につける重みを $v_i (i = 1, 2, \ldots, m)$ 出力につける重みを $u_r (r = 1, 2, \ldots, m)$ として，その値を次の分数計画問題を解くことによって定める．

$$\begin{aligned} &\text{目的関数} \qquad Max \quad \theta = \frac{u_1 y_{1k} + u_2 y_{2k} \ldots + u_s y_{sk}}{v_1 x_{1k} + v_2 x_{2k} + \ldots + v_m x_{mk}} \\ &\text{制約式} \qquad \frac{u_1 y_{1j} + \ldots + u_s y_{sj}}{v_1 x_{1j} + \ldots + v_m x_{mj}} \leq 1 (j = 1, 2, \ldots, n) \\ &\qquad\qquad v_1, v_2, \ldots, v_m \geq 0, u_1, u_2, \ldots, u_s \geq 0 \end{aligned} \tag{3.8}$$

この式（3.8）は，その制約式で，仮想的に考えられた総入力と総出力の比をすべての事業体の生産活動において，1 以下に抑えるようにモデル化されている．その上で，k 番目の事業体の効率値 θ を最大化するように重み v_i と u_r を決めている．したがって，最適な θ の値 (θ^*) は 100% かそれ以下である．下限は 0% と考えてよい．

上の分数計画問題に対して次の線形計画問題を考える．

＜入力指向＞

$$\begin{aligned} &\text{目的関数} \quad Max \quad \theta = u_1 y_{1k} + \ldots + u_s y_{sk} \\ &\text{制約式} \qquad v_1 x_{1k} + \ldots + v_m x_{mk} = 1 \\ &\qquad\qquad u_1 y_{1j} + \ldots + u_s y_{sj} \quad - (v_1 x_{1j} + + \ldots + v_m x_{mj}) \leq 0 \quad (j = 1, 2, \ldots, n) \\ &\qquad\qquad v_1, v_2, \ldots, v_m \geq 0, u_1, u_2, \ldots, u_s \geq 0 \end{aligned} \tag{3.9}$$

＜出力指向＞

目的関数　$Min \quad \theta = v_1 x_{1k} + v_2 x_{2k} + \ldots + v_m x_{mk}$

$$
\begin{aligned}
\text{制約式} \quad & u_1 y_{1k} + u_2 y_2 k \ldots + u_s y_{sk} = 1 \\
& u_1 y_{1j} + \ldots + u_s y_{sj} - (v_1 x_{1j} + \ldots + v_m x_{mj}) \leq 0 \quad (j = 1, 2, \ldots, n) \\
& v_1, v_2, \ldots, v_m \geq 0, u_1, u_2, \ldots, u_s \geq 0
\end{aligned} \tag{3.10}
$$

線形計画問題は普通の線形計画法のアルゴリズムによって解くことができる．したがって，実用的な解法としては分数計画問題より線形計画問題を解く方が良い．以下は式（3.8）の性質である．

- 式（3.10）の最適解を (v^*, u^*) とし，目的関数値を θ^* とする．そのとき $\theta^* = 1$ ならば DMU_k は効率的であるという．逆に $\theta^* < 1$ の場合でも，スラックと呼ばれる入力の余剰や出力の不足が発生していることがあるので注意を要する．
- DMU_k が $\theta^* = 1$（非効率）のとき，式（3.10）の制約の中には重み付け (v^*, u^*) に対して，次のような等式が成立する j が存在する．

$$
R_k = \{j : \sum_{r=1}^{s} u_r^* y_{rj} = \sum_{i=1}^{m} v_i^* x_{ij}, \qquad j = 1, \ldots, n\} \tag{3.11}
$$

$$
\tag{3.12}
$$

この集合 (R_k) は DMU_k を非効率と判定させるもとになっている事業体群である．その意味で，DMU_k に対する参照集合という．この参照集合 (R_k) は効率的フロンティアの一部を形成している．容易にわかるように，この集合に属する事業体はそれ自体が効率的である．

よって，式（3.8）は以下のようになる．

$$
\begin{aligned}
\text{目的関数} \quad & Max \quad \sum_{r=1}^{s} u_r y_{rk} \\
\text{制約式} \quad & -\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} \leq 0 (j = 1, 2, \ldots, n)
\end{aligned} \tag{3.13}
$$

$$
\begin{aligned}
& \sum_{i=1}^{m} v_i x_{ik} = 1 \\
& v_i \geq 0 (i = 1, 2, \ldots, m), u_1 \geq 0 (r = 1, 2, \ldots, s)
\end{aligned}
$$

となる．双対形は

$$
\begin{aligned}
&\text{目的関数} \qquad Min \quad \theta \\
&\text{制約式} \qquad -\sum_{j=1}^{n} x_{ij}\lambda_j + \theta x_{rj} \leq 0 (i = 1, 2, \ldots, m) \\
&\qquad\qquad \sum_{j=1}^{n} y_{rj}\lambda_j \geq u_{rk} \\
&\qquad\qquad \lambda_j \geq 0 (j = 1, 2, \ldots, n), \qquad\qquad \theta : \text{制約なし}
\end{aligned}
\tag{3.14}
$$

となる．双対問題とは，線型計画法においての主問題に対して行ベクトルを変数とする問題のことを言う．双対問題で注意することは以下の 3 点である．

- 最小化問題は最大化問題に置き換わっている
- 制約式の負統合の向きが逆になっている
- 統合で表された制約式に関する双対変数は制約がなくなっている．

上記の式（3.14），（3.15）が CCR モデルと呼ばれる DEA モデルである．目的関数値は双対定理により一致し，DEA 効率値を示している．なお，式（3.15）において θ を制約なしとしたのは，式（3.14）において $\sum_{n=1}^{m} v_i x_{ik} = 1$ が等式となっているため，$\theta \geq 0$ としても問題ではない．

次に $\theta^* = 1$ の場合でも，入力の余剰 $d_i^k (i = 1, 2, \ldots, m)$ と出力の不足 $d_r^y (r = 1, 2, \ldots, s)$ はそれぞれ

$$
\begin{aligned}
d_i^x &= \theta x_{ik} - \sum_{j=1}^{n} x_{ij}\lambda_j (i = 1, 2, \ldots, m) \\
d_r^y &= \sum_{j=1}^{n} y_{rj}\lambda_j - y_{rk} (r = 1, 2, \ldots, s)
\end{aligned}
\tag{3.15}
$$

で表される．d_i^k と d_r^y は次の式を解くことにより得られる．

$$
\text{目的関数} \quad Max \quad \sum_{i=1}^{m} d_i^x + \sum_{r=1}^{s} d_r^y
$$

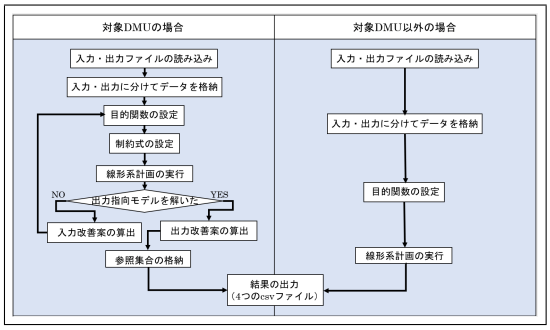


図 3.3: DEA のフローチャート

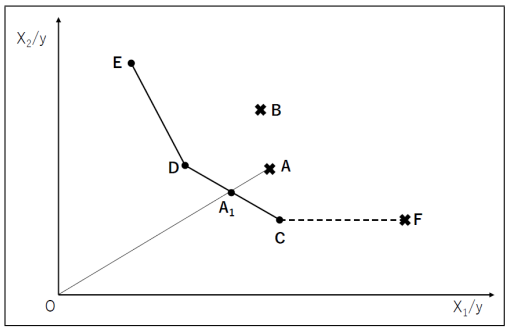


図 3.4: 効率的フロンティア

制約式

$$\sum_{j=1}^{n} x_{ij}\lambda_j + d_i^x = \theta^* x_{ik} (i = 1, 2, \ldots, m)$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j - d_i^y = y_{rk} (r = 1, 2, \ldots, s) \tag{3.16}$$

$$\lambda_j \geq 0 (j = 1, 2, \ldots, n), d_i^x \geq 0 (i = 1, 2, \ldots, n) \qquad d_r^y \geq 0 (r = 1, 2, \ldots, s)$$

ここで θ^* は式（3.15）の最適目的関数値を示している．よって第 1 段階として式（3.15）を解き θ^* を求めた後，第 2 段階として式（3.16）を解くことになる．式（3.15）より得られる最適解 $(\theta^*, \lambda_1^*, \ldots, \lambda_n^*)$ と式（3.16）から得られるスラックの最適解 $(d_1^{x*}, \ldots, d_m^{x*}, d_1^{y*}, \ldots, d_s^{y*})$ を合わせて DEA 最適解と呼ぶ．

DEA の CCR 効率性の定義として，DEA 最適解において $\theta^* = 1$ かつ全てのスラックが 0 の場合，効率的である．それ以外のときは，非効率的である．

§ 3.3　データ包絡分析による改善策の導出

DEA は事業体の効率性を示すだけでなく，同時に非効率的な事業体の改善案についても示すことができる．DMU_k が非効率であるとき式（3.15）に対応する参照集合は

$$R_k = \{j \mid \lambda_j^* > 0, j = 1, \ldots, n\} \tag{3.17}$$

として表される．この式（3.17）は式（3.12）と同等である．このことは線形計画法の相補性によって証明される．相補性とは，実行可能解 x, y が最適ならば，以下の 2 つの条件が成り立つ．

事業体	A	B	C	D	E	F
入力 x_1	4	4	4	3	2	6
入力 x_2	2	3	1	2	4	1
出力 y	1	1	1	1	1	1

表 3.1: 6 事業体の入力・出力

事業体	DEA 効率値	参照集合	v_1^*	v_2^*	u^*	d_1^{x*}	d_2^{x*}	d_1^{y*}
A	0.833	$C(\lambda_C^*=0.333), D(\lambda_D^*=0.677)$	0.167	0.167	0.833	0	0	0
B	0.727	$D(\lambda_D^*=0.909), E(\lambda_E^*=0.091)$	0.182	0.091	0.727	0	0	0
C	1	$C(\lambda_C^*=1)$	0.200	0.200	1	0	0	0
D	1	$D(\lambda_D^*=1)$	0.250	0.125	1	0	0	0
E	1	$E(\lambda_E^*=1)$	0.500	0	1	0	0	0
F	1	$C(\lambda_C^*=1)$	0	1	1	2000	0	0

表 3.2: DEA の分析結果

- すべての $i=1,\ldots,n$ に対し，$x_i=0$ または $(y^{\mathrm{T}}A)_i=c_i$
- すべての $j=1,\ldots,m$ に対し，$y_j=0$ または $(Ax)_j=b_i$

逆に実行可能解 x,y が相補性条件を満たすならば $c^{\mathrm{T}}x=y^{\mathrm{T}}Ax=y^{\mathrm{T}}b$ を満たすことがわかる．

参照集合 R_k に属する事業体は効率的である．R_k に属する事業体の存在が DMU_k を非効率とさせる原因であり，この集合を非負結合させたものが DMU_k の効率値を決定する効率的フロンティアを形成する．

式 (3.15) において DMU_k の効率値は DEA 最適解を $(\theta^*,\lambda_1^*,\ldots,\lambda_n^*,d_1^{x*},\ldots,d_m^{x*},d_1^{y*},\ldots,d_s^{y*})$ とすると，この効率的フロンティアの上のダミー事業体 $(\theta^*x_{ik},\ldots,\theta^*x_{mk},y_{1k},\ldots,y_{sk})$ と双対比較することで決定される．ここで θ^*x_{ik} と u_{rk} は次のように表される．

$$\theta^*x_{ik}=\sum_{j\in R_k}\lambda_j^*x_{ij}+d_i^{x*}(i=1,2,\ldots,m)$$
$$y_{rk}=\sum_{j\in R_k}\lambda_j^*x_{rj}-d_r^{y*}(r=1,2,\ldots,s)$$

すなわち，DMU_k は入力を θ^* 倍に縮小し，さらに余剰を除去する．出力は不足を補うことで，効率的な事業体となる．DMU_k の入力の過剰量 (Δx_{ik})，出力の不足量 (Δy_{rk}) は次のように表される．

$$\Delta x_{ik}=x_{ik}-(\theta^*x_{ik}-d_i^{x*})=(1-\theta^*)x_{ik}+d_i^{x*}(i=1,2,\ldots,m)$$
$$\Delta y_{rk}=d_r^{y*}$$

よって，

$$x_{ik}\Rightarrow x_{ik}-\Delta x_{ik}=\theta^*x_{ik}-d_i^{x*}(i=1,2,\ldots,m)$$

$$y_{rk} \Rightarrow y_{rk} + \Delta y_{rk} = y_{rk} + d_r^{y*}$$

とすれば，DMU_k は効率的な事業体となる．ただ，非効率な事業体を効率的に改善する方法は 1 つではないことに留意する．

表 3.1 に示すように A F の 6 事業体があり，それぞれが 2 入力・1 出力をもつとする．ただし，例を簡略化するため，出力値はいずれも 1 である．事業体 A の効率性を求める CCR モデルは，v_1, v_2, u を入力と出力の重み付けとして，次のように定式化される．

目的関数	$Max \quad u$	
制約式	$-4v_1 - 2v_2 + u \leq 0$	(A)
	$-4v_1 - 3v_2 + u \leq 0$	(B)
	$-4v_1 - v_2 + u \leq 0$	(C)
	$-3v_1 - 2v_2 + u \leq 0$	(D)
	$-2v_1 - 4v_2 + u \leq 0$	(E)
	$-6v_1 - v_2 + u \leq 0$	(F)
	$4v_1 + 2v_2 = 1$	
	$v_1 \geq 0, v_2 \geq 0, u \geq 0$	(3.18)

この最適解は，$v_1^* = 0.167, v_2^* = 0.167, u^* = 0.833$ であり，目的関数値，すなわち A の DE 効率値は 0.833 となる．A の参照集合は $R_A = \{C, D\}$ であり，C,D の存在が A を非効率にしいることがわかる．次に，θ, λ を双対変数として双対形をとると，

目的関数	$Min \quad \theta$
制約式	$-4\lambda_A - 4\lambda_B - 4\lambda_C - 3\lambda_D - 2\lambda_E - 6\lambda_F + 4\theta \geq 0$
	$-2\lambda_A - 3\lambda_B - \lambda_C - 2\lambda_D - 4\lambda_E - \lambda_F + 2\theta \geq 0$
	$\lambda_A + \lambda_B + \lambda_C + \lambda_D + \lambda_E + \lambda_F \geq 1$
	$\lambda_j \geq 0 (j = A, B, \ldots, F), \theta$: 制約なし

のようになる．この最適解は，$\lambda_A^* = \lambda_B^* = 0, \lambda_C^* = 0.333\lambda_D^* = 0.677, \lambda_E^* = \lambda_F^* = 0, \theta^* = 0.833$ となる．これらの最適から次の関係が分かる．

$$0.833 * (\text{A の入力}) = 0.333 * \text{C の入力} + 0.677 * (\text{D の入力})$$
$$(\text{A の入力}) = 0.333 * (\text{C の入力}) + 0.677 * (\text{D の入力})$$

図 3.4 からわかるように，事業体 A は OA に直線を引いたとき，効率フロンィアである CD の直線と交わる点 A_1（A の入力を一様に 0.833 倍に縮小した点）において効的になる．この点は CD 間の比率を表している．すなわち，効率的な事業体 C と D をそれぞれどの度参照しているかがわかる．次に，スラックの最適解を求めるために，以下の式を解く．

目的関数	$Max \quad d_1^x + d_2^x + d_1^y$
制約式	$4\lambda_A + 4\lambda_B + 4\lambda_C + 3\lambda_D + 2\lambda_E + 6\lambda_F + d_1^x = 4 * 0.833$
	$2\lambda_A + 3\lambda_B + \lambda_C + 2\lambda_D + 4\lambda_E + \lambda_F + d_2^x = 2 * 0.833$
	$\lambda_A + \lambda_B + \lambda_C + \lambda_D + \lambda_E + \lambda_F - d_1^y = 1$
	$\lambda_j \geq 0 (j = A, B, \ldots, F), d_1^x \geq 0, d_2^x \geq 0, d_1^y \geq 0$

この最適解は，$d_1^{x*} = d_2^{x*} = d_1^{y*} = 0, \lambda_A^* = \lambda_B^* = 0 \lambda_C^* = 0.333, \lambda_D^* = 0.677, \lambda_E^* = \lambda_F^* = 0$ となる．このよに，事業体 A の DEA 最適解は $\theta^* = 0.833, d_1^{x*} = d_2^{x*} = d_1^{y*} = 0 \lambda_A^* = \lambda_B^* = 0, \lambda_C^* = 0.333, \lambda_D^* = 0.677 \lambda_E^* = \lambda_F^* = 0$ となる．なお，A と同様に，事業体 B も非効率的であることがわる．事業体 C,D,E は，A や B の参照集合になっているので，効率的である．例えば C の場合 $\lambda_A^* = \lambda_B^* = 0, \lambda_C^* = 1, \lambda_D^* = \lambda_E^* = \lambda_F^* = 0$ となり，C 自身が参照されている．これは C 自身が効率的であり，参照集合も C となることを意味ている．D と E の場合も同様である．F に関する DEA 最適解は $\theta^* = 1$ であるが，入力 1 に対するスラック d_1^{x*} が 2 であるで，DEA 効率性の定義により F は非効率であると判定する．F を効率的にするには，C のレベルま入力 $1(x_1)$ を減らす必要がある．
最後に，DEA による分析結果を表 3.2 にまとめた．

第4章 提案手法

§ 4.1 着目要因への入力による分析データの構築

入力されるものでプラスとマイナスにわけて考える．プラスは分子に，マイナスは分母に．本研究では，

本研究の提案手法は，大別すると以下のような 4 つの工程からなる．

1. Web 上からあらかじめデータを取得し，データベースを作成する．
2. 目的とデータとの間に因果関係が成立するデータを Direct LiNGAM によってデータベース上から抜き出し，DEA の入力と出力に振り分ける．
3. 振り分けられたデータを用いて DEA を解き，改善値を算出する．

そのため，提案手法は以下のような 4 つのプログラムに分けることができる．

1. 対象のデータの受け取り
2. Direct LiNGAM による因果分析およびデータの振り分け
3. DEA によるデータ分析
4. 改善策の提案

本節では，ユーザが対象データを指定するにあたって前段階として必要なデータベースの作成方法とデータベースに保存されているデータの種類，対象の指定から LiNGAM による因果探索でのデータの選定，DEA の入力・出力に対するデータの振り分けまでの提案手法のしくみを解説する．

データベースに用いた統計データ

データベースに用いたデータの項目については、データの属性としては位置データか数値データか，地理的情報を持つか否かで大きく 3 つに分けることができる．

地理的情報を持たない数値データの例には人口，製造品出荷額などが挙げられ，数年に 1 回，全国統一でデータの収集が行われるものである．地理的情報を持つ数値データとしては，公園や医療機関の数が挙げられ，これらのデータに関してはどこにその施設が存在するかの位置データが紐づけられている．

本研究に用いるデータベースを作成するにあたって，政策の対象となる事柄について，その原因となるものを探すという目的を達成するために，ある程度広い分野の情報がそろっていること，日本全国に対応しており市区町村単位の粒度のデータであること，各地方や都道府県などでデータにおいて大きな欠損がないことなどの条件を満たす必要があった．

これらの条件から，地理的情報を持たないデータに関しては 2.1 節で挙げた e-Stat の API を用いてデータの収集を行った．

★また，地理的情報を持つデータに関しては国土交通省の国土数値情報ダウンロードにおけるデータを収集し，それらに基づいて各市区町村ごとの施設数を自動的に数え上げることで施設数のデータとした．★

全てのデータは csv 形式でサーバに保存し，各処理ごとに取り出して参照，編集，加工できるようにした．データベースとして用いた csv のフォーマットについて，各市区町村における人口および公園の数，位置のデータを例に挙げて図**??**に示す．e-Stat によって収集したデータと施設の数を表すデータに関しては，1 列目に総務省から各市区町村に対して割り当てられている全国地方公共団体コードをキーとして格納し，2 列目に対象のデータを格納するという形をとった．

データの前処理

本研究で分析に用いるデータは単位が異なり，値の大きさも広範囲にわたるため，分析を行う前にデータの正規化を行う．まず，本研究で用いるデータは正規分布に従わないため，そのようなデータに対しても適用することが可能な robust Z-score を用いる [[?]]．robust Z-score について以下に示す．

<robust Z-score>

$$\iota = \frac{x - median(x)}{NIQR} \tag{4.1}$$

robust Z-score では，式 (4.1) に示すように各データとデータ集合の中央値との差をとり，その値とデータ集合の正規四分位範囲との商を求めることによって値を正規化する．正規四分位範囲とは，四分位範囲と 1.3489 の商である．また，robust Z-score の結果では各値が 0 を中心に正規化されるため負の値となるという場合が発生する．しかし，CCR モデルによる分析では負の値を持つデータを扱うことが

出来ないため，さらに値がすべて 0〜1 の範囲に収まるように正規化を行う．用いる処理を以下の式に示す.

< 正規化 >

$$\iota' = \frac{\iota + \min|\iota|}{\max|\iota| - \min|\iota|} \tag{4.2}$$

因果探索のためのデータの選定

具体的には，多量かつ幅広い分野のデータに対して LiNGAM による因果探索を適用することによって，政策の対象としたいターゲットと因果関係のあるデータのみを抜き出すという手法である.

因果探索における数学的なアルゴリズムに関しては 3.1 節で示した Direct-LiNGAM に従う．システムのアルゴリズムとしては，対象に関係するデータを Direct-LiNGAM を解くコードに送信することで，そのデータをターゲットとして因果探索を行う.

この際，因果探索に用いられるデータは全てのデータであり，これらのデータを用いて一度に因果探索を行うことで政策の対象と潜在的に因果関係を持つデータのみを自動的に絞り込む.

また，これらの処理の結果を用いて 4.2 節にて後述する DEA を用いた分析を行うために，因果探索を解いた結果，対象のデータとの因果関係が示されたデータのみを結合した新たな csv ファイルを作成する．この際に DEA における入力・出力の振り分けも同時に行う.

本研究の提案手法では因果探索によって同定された因果グラフのうち，対象のデータに向く矢印を持つデータのみを分析の対象として扱うこととする．これは，本研究における因果探索の役割が DEA の入力・出力となるデータを絞り込むことにあるためである．因果探索の結果のうち，矢印の始点にあるデータが増減すると矢印の終点にあるデータにも影響がおよびその値が増減するため，政策の対象について考える場合はそのデータに影響を与えるデータのみを考慮することが妥当と考えた.

また，因果グラフにおけるパスの重みが正のときデータ同士も正の相関，負のときに負の相関をとるため，正の場合の始点側のデータを DEA の出力，負の場合の始点側のデータを DEA の入力とすることが妥当と考えた．以上のことから，それらを一つにまとめた csv ファイルを因果探索部分のプログラムのアウトプットとする．因果探索の結果と出力ファイル内のデータの例を図??に示す.

§ 4.2　データ包絡分析による主体の効率と入力要因の改善策

本節では，本研究での提案システムにおける DEA 部分の具体的なしくみ，システムの仕様，システムにおいてインプットおよびアウトプットされるデータの形式などを解説する．提案手法における DEA の役割は因果探索によって導かれた政策の対象に対する要因について，それらの数値をもとに対象としている自治体の現状を理論的に評価し，その評価を一層高めるために取り組むべき課題を明確にすることによって政策における意思決定の支援をするというものである．

DEA 部分における全体のながれ

DEA 部分では，はじめに 4.1 節での因果探索のアウトプットデータ（図??参照）を用いて全市区町村を対象に 3.2 節に示した入力指向モデルを解くことによって，各市区町村の評価値を算出する．次に，フロントページにて指定された対象の市区町村に対して，式（??）および式（??）を解くことによって，その結果から対象の市区町村における入力・出力の改善案を算出する．

改善案の算出を対象の市区町村のみに限定したのは，全国の市区町村という膨大な DMU を扱う問題において計算量を軽減するためであり，単純な評価値と比較して改善案は対象市区町村以外のものを参考にすることが少ないと考えたからである．

各市区町村に対する評価値の算出

本研究では，47 都道府県に存在する市区町村のうち 790 個の市，112 個の区，717 個の町，184 個の村の合計 1803 個を DMU の集合として，CCR モデルによる DEA 分析を行うことによってそれぞれの DMU における評価値を算出する．各都道府県における DMU の内訳を表??に示す．因果探索によって対象と因果関係が示された出力データの数を m とすると評価値は以下の線形計画問題を解くことによってもとめられる．

< 評価値の算出式 >

$$\text{minimize} \qquad \theta \tag{4.3}$$

$$\text{subject to} \qquad \sum_{d=1} y_{id}\lambda_d \geq y_{io} \qquad i = 1, 2, \cdots m \tag{4.4}$$

$$-\sum_{d=1} x_{id}\lambda_d + x_{io}\theta \geq 0 \qquad i = 1, 2, \cdots m \tag{4.5}$$

$$\lambda \geq 0 \tag{4.6}$$

対象の市区町村に対する入力・出力改善案の算出

フロントページにて対象の市区町村に指定された DMU に対してはすべての入力および出力に対して参照集合のデータをもとにした値の改善案を算出する．参照集合は前述の評価値の算出式において λ_d の値が正の数をとった DMU のみで形成され，参照集合内の DMU の数を A 個，入力の項目数を i 個，出力の項目数を j 個とすると入力および出力の改善案は式（**??**）および式（**??**）に対して $K = A$ と置くことによってもとめられる．また，4.1 節で述べた通り本研究ではデータに対して前処理を行っているため入力・出力の改善案に関しては逆変換を行い元のデータと同じ単位に戻してから結果を保存している．逆変換の式は以下のようになる．

< 正規化の逆変換 >

$$\iota = \iota' \times 2\max|\iota| - \max|\iota| \tag{4.7}$$

<robust Z-score の逆変換 >

$$x = \iota \times NIQR + median(x) \tag{4.8}$$

提案手法を実現するプログラム

本研究の提案手法では，DEA の計算に PuLP による線形計画問題のプログラムを使用している．PuLP とは，Python で数理最適化のモデルを記述するためのモジュールであり，PuLP を用いてモデルを記述することによって，混合整数最適化問題を解くことができる．混合整数最適化問題とは，以下のような特徴を持つ数理最適化問題の一種である．

- 連続（実数）変数と整数変数を使って表現される．
- 目的関数と制約条件が 1 次式である．

よって，線形計画問題を用いて定式化することが可能な DEA においても PuLP によるモデルの記述が可能で，PuLP を用いて記述したモデルは同梱される COIN プロジェクトのソルバーである CBC を用いて自動的に解くことができる．問題の定式化を行う回数や算出し，格納するデータの種類は扱う DMU がフロントページで指定された対象の市区町村であるか否かで異なるが，PuLP を用いて最適化問

題を解く部分の仕組みは共通である．提案手法の DEA 部分におけるシステムの一連のながれを図??に示す．

入力・出力のデータは図??の csv ファイルのように 1 つにまとめられた状態で読み込み，ファイル内の 2 行目に格納されている入力・出力の項目数をもとに 3 行目以降のデータを入力・出力に切り分けることによってそれぞれのデータフレームに格納する．

次に前述の PuLP を用いて入力指向モデルの目的関数，制約式を設定し実行することによって線形計画問題を解き，DMU の評価値を算出する．また，対象 DMU の場合には実行結果をもとに入力改善案を算出する．

その後，同様に PuLP を用いることで出力指向モデルの目的関数，制約式を設定し実行することによって線形計画問題を解く．こちらも対象 DMU の際には実行結果をもとに出力改善案を算出する．

加えて，入力・出力の改善案を算出する際に用いた参照集合内の市区町村名およびそれらの市区町村にかかるウェイトに関してもリストを作成して格納する．最後にこれらの結果をもとに以下に示す 4 つの csv ファイルをアウトプットする．アウトプットする 4 つの csv ファイルの例を図??に示す．

- 全市区町村を対象にしたそれぞれの市区町村に対する評価値
- 対象の市区町村における入力・出力の各項目に対する改善案
- 対象の市区町村における入力の改善案を算出した際の参照集合とそれぞれのウェイト
- 対象の市区町村における出力の改善案を算出した際の参照集合とそれぞれのウェイト

DEA による分析結果を以上のような 4 つの csv ファイルに分割したのはそれぞれのデータにおける対象の適用範囲や役割が異なり，4.3 節で後述する結果の提示の際に別の提示方法によってユーザにフィードバックされるためである．具体的には，1 つ目の csv ファイルが全ての市区町村に対してそれぞれにもとめられる評価値であり，各行に市区町村ごとの値が格納されているのに対して，その他の csv ファイルのデータはフロントページで指定された 1 つの市区町村に対する各項目のデータが格納されているという違いがある．ゆえに，システムにおけるデータの読み込みおよび取り扱いの簡便化のために個別のファイルでアウトプットすることとした．

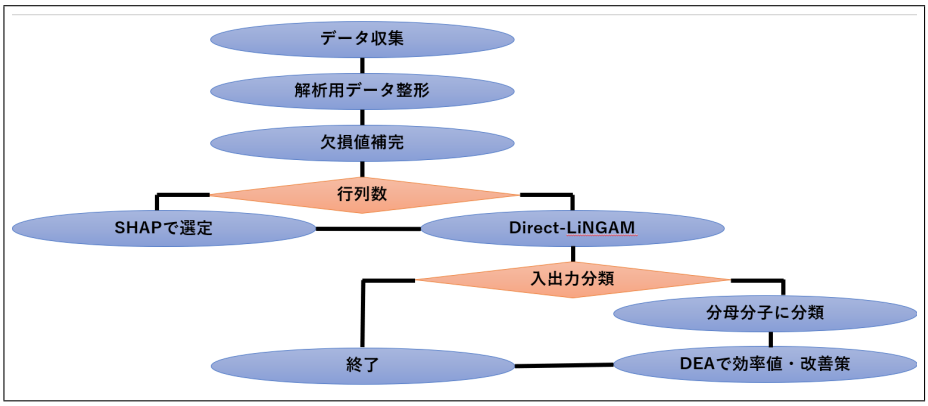


図 4.1: システムのフロー

§ 4.3　提案手法の流れ

1 章で示した政策の対象となる問題に関する要因の複雑性という課題に対して，4 章で示した各手法を統合した課題解決のための提案手法全体の流れの説明を行う．また，提案システム全体のフローを図§ 4.2 に示す．

Step 1: 要因ごとのデータベースの作成

はじめに、2.1 章で紹介した Web サイトから各データを取得しデータベースを作成する。本研究で用いるデータ取得手法として、〇〇を扱う。〇〇は

データ取得するサイトとしてまた、RESAS に記載されている不動産情報は〇〇なため、不動産情報ライブラリも活用しデータを収集している。

政策に関係した問題を抱えるユーザはその問題に関係のあるデータの項目名および自身が取り扱いたい市区町村名を各 1 つずつ選択形式でシステムに入力する．選択されたデータは次のフローへ送信される．

Step 2: 全データ結合

Step1 で取得したデータを市区町村コードを参照しデータを結合させる。その際、1 つの市区町村コードに対して複数のデータがある場合はデータによって平均値を取り 1 つの値を作り結合させる。要因に対して項目数が少なく LiNGAM や DEA にかけられる場合は、列内に欠損値が多い場合は列ごと削除し、少ない場合はその地域の行を削除する。逆に要因に対し項目数が多い場合は、列内に欠損値が多い時は列ごと削除し、SHAP にかけ要因よりも 1 個少ない項目数にする。

Step 3: Direct-LiNGAM を用いた入力と出力の分類

Step1 で選択されたデータを受信し，受信したデータに含まれる対象のデータ項目名をターゲットとしてサーバ上に存在する全ての統計データを用いた Direct-LiNGAM による因果探索を行い，対象のデータ項目と関係のあるデータのみを選別する．

また，選別されたデータのうち，対象のデータ項目に対して因果関係が向いているデータを入力，対象のデータ項目から因果関係が向いているデータを出力とすることで選別されたデータを DEA における入力・出力データに振り分ける．選別されたデータは入力・出力の項目数とともに 1 つの csv ファイルにまとめて排出する．

Step 4: データ包絡分析による効率値の導出

Step2 にて作成された csv ファイル内の入力・出力の項目数をもとに入力・出力データを参照し，それらを対象とした DEA の CCR モデルを解くことによって全市区町村を対象にそれぞれの評価値を算出する．

また，Step1 のフロントページにて選択された市区町村に対して DEA の入力指向モデルおよび出力指向モデルを用いた入力・出力改善案の式を適用することによって対象の市区町村における入力および出力それぞれに対する改善案を算出する．

以上の操作の結果もとめられた全市区町村に対する評価値，対象の市区町村に対する入力・出力の改善案に加えて改善案を算出する際に用いた参照集合と入力指向モデル，出力指向モデルでもとめられるそれぞれの値である重みをデータの種類ごとに 4 つの csv ファイルにして排出する．

Step 6: GIS を活用したデータフュージョン

Step3 にて作成された 4 つ csv ファイルをもとに結果の表示および地理情報データとの重ね合わせによるデータフュージョンによって政策決定を支援することを目的とした EBPM-GIS を作成・表示する．4 つの csv ファイルのデータのうち，全市区町村に対する評価値のデータは GIS 上にマーカーとして表示し，その値の大小によって 3 つのレイヤに分ける．その他のデータはマーカーのポップアップ内にテキストとして表示する．

また，Step2 にて対象のデータ項目と因果関係が示されたデータのうち，施設の場所など地理的な特徴を持つデータに関してはそれらのデータ単体でマーカークラスタとして表示し，Step3 での分析結果と重ね合わせることによってデータフュージョンを行う．

システム上でこれら 4 つの Step の処理が行われることによって出力装置には最終的に EBPM-GIS が出力されることになる．つまり，ユーザサイドから見た場合，フロントページにて対象のデータ項目および市区町村を入力し実行すると GIS が表示されるという画面遷移だけが提示される．

ユーザは，これらのデータに対して切り替えや重ね合わせを用いることで考察を行い，政策における意思決定を行う．どのように考察するかはユーザの自由であるが，現段階では評価値によって自身の市区町村の現状を知り，周辺の市区町村との差などを」感覚的に理解する．改善案や優れた市区町村をもとに目標を明確化する．そのうえで，重ね合わせ等によって政策決定における新たな知見を得るという方法を想定している．

第5章

数値実験並びに考察

§ 5.1　数値実験の概要

§ 5.2　実験結果と考察

第6章 おわりに

急激な生活様式の欧米化に伴い，ジャンクフードといった，余分にエネルギーを摂取してしまうような食生活が大きく広まったことから，現在，生活習慣病を患う人々が増加している．生活習慣病を予防する一つの方法として，栄養バランスのとれた食事をとることが推奨されている．しかし，栄養バランスの取れた献立作成には，その人の身体情報，疾患情報などによってメニューや料理の分量を調整しなけらばならなく，献立作成業務の負荷は高いことがわかる．

これらの問題を解決するために，本研究では，Web サイトから得られるレシピ情報や食材価格を活用し，制約条件を考慮できる多目的遺伝的アルゴリズムによって自動的に献立を作成をするシステムを考案した．

本研究で用いるレシピデータとして，3 つのレシピサイトからスクレイピングを行うことによってレシピデータベースに多様性を持たせることができた．また，この献立作成システムは健常者だけではなく，生活習慣病を患っている人やアレルギーを患っている人でも利用できるようにした．さらに，プログラム実行に必要なすべてのプログラムをサーバーに置き，実行に必要な URL を用意することによって，ユーザはその URL をクリックするだけでプログラムを実行できるようにした．

また，プログラムの実行にはレシピデータなどの大量のデータが必要なため，プログラムの環境を整えるための手間が大変になってしまう問題があった．そのためプログラムをサーバー上に置くことでプログラム実行の環境を整える手間を省くことができた．

本研究で提案した制限食と大人数料理に対応した自動献立作成システムを実際に動作させた実験結果として，多目的最適化によって作成された献立は調理時間，料理コストを最小化しながら，設定した制約条件を満たしながら出力することができた．

本研究の課題として，摂取栄養素や摂取カロリーの上限，下限の設定などの制約条件を，ユーザ自身で決められるようにすることや，並列分散処理などを施すことにより，最適化プログラムの実行処理時間を向上し，よりユーザに快適に利用できるようにプログラムを改良する必要がある．また，ユーザが好みの料理を入力することによって，出力する料理がユーザの好みに近しいもの出るようにすることや，ユーザが現在持っている食材を入力することによって，その食材を含む料理が出力されるようにする必要があると考えられる．

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 António Oliveira Nzinga René 講師，奥原浩之教授に深甚な謝意を表します．最後になりましたが，多大な協力をしていただいた研究室の同輩諸氏に感謝致します．

2023 年 2 月

水上和秀

参考文献

[1] 内閣府, “構造改革のための経済社会計画”, https://www5.cao.go.jp/j-j/keikaku/keishin1-j-j.html, 閲覧日 2025.4.7.

[2] 総務省, “地域が抱える課題・検討の論点について”, https://www.soumu.go.jp/main_content/000919078.pdf, 閲覧日 2025.4.7.

[3] 厚生労働省, “自治体における政策づくりの意義と方法”, https://www.mhlw.go.jp/file/06-Seisakujouhou-12600000-Seisakutoukatsukan/0000114064_11.pdf, 閲覧日 2025.4.15.

[4] 政府統計の総合窓口, “e-Stat ”, https://www.e-stat.go.jp/, 閲覧日 2025.3.20.

[5] 総務省統計局, “RESAS 地域経済分析システム”, https://resas.go.jp/, 閲覧日 2025.4.3.

[6] 富山県庁, “富山県オープンデータポータルサイト”, https://opendata.pref.toyama.jp/dataset/https-opendata-pref-toyama-jp-dataset-settou2023, 閲覧日 2025.4.3.

[7] 国土交通省, “不動産情報ライブラリ”, https://www.reinfolib.mlit.go.jp/, 閲覧日 2025.4.5.

[8] 国土交通省, “不動産情報ライブラリ”, https://www.reinfolib.mlit.go.jp/, 閲覧日 2025.4.5.

[9] 橋浦亮太、 中野直人, “機械学習モデルに対する SHAP に基づく貢献度評価の検証, ” 情報処理学会第 *86* 回全国大会講演論文集, Vol. 1, pp. 287-288, 2024.

[10] , “MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition”, *IEEE Trans. Evolutionary Computation*, *Vol. 11*, No. 6, pp. 712–731, 2007.

[11] , “MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition”, *IEEE Trans. Evolutionary Computation*, *Vol. 11*, No. 6, pp. 712–731, 2007.

[12] , “MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition”, *IEEE Trans. Evolutionary Computation*, *Vol. 11*, No. 6, pp. 712–731, 2007.

[13] 総務省, “行政機関が行う政策の評価に関する法律”, https://www.soumu.go.jp/main_sosiki/hyouka/houritu.htm, 閲覧日 2025.4.16.

[14] 文部科学省, “ロジック・モデルの作成マニュアル”, https://www.mext.go.jp/content/20230410-mxt_kanseisk01-100000155-3.pdf, 閲覧日 2025.4.16.

[15] , “行政機関が行う政策の評価に関する法律”, https://www.soumu.go.jp/main_sosiki/hyouka/houritu.htm, 閲覧日 2025.4.16.

[16] みずほリサーチ＆テクノロジーズ, “EBPM 促進のための「3 つのエビデンス」の理解”, https://www.mizuho-rt.co.jp/publication/2021/articles_0033.html, 閲覧日 2025.4.16.

[17] 日経XTECH, “「デジタル先進県」がDXで成果を生み出し続けている理由”, https://special.nikkeibp.co.jp/atclh/NXT/23/hcljapan1108/, 閲覧日 2025.4.16.

[18] C. A. Coello Coello and M. S. Lechuga, “統計的因果探索による社会基盤整備のストック効果の検証, ”*Proceedings of the 2002 Congress on Evolutionary Computation (CEC'02)*, Vol. 75, No. 6, pp. 583-589, 2020.

[19] S Shimizu, PO Hoyer, A Hyvärinen, A Kerminen and M Jordan, “A linear non-Gaussian acyclic model for causal discovery, Journal of Machine Learning Research, pp. 2003-2030, 2006.

[20] Dentsu Digital Tech Blog, “Google Colabで統計的因果探索手法LiNGAMを動かしてみた”, https://note.com/dd_techblog/n/nc8302f55c775, 閲覧日 2025.5.7.

[21] ごちきか, “LinGAMによる因果探索（応用編）”, https://gochikika.ntt.com/Modeling/causal_LiNGAM_advanced.html, 閲覧日 2025.5.7.

[22] J．H. Holland，“DEA を用いた商圏属性に適合したホームセンターの部門別陳列棚数構成方法”， *Communications of the Operations Research Society of Japan*， Vol. 62， No. 10， pp. 677-684， 2017.

[23] J．H. Holland，“統計的因果探索アルゴリズム“ LiNGAM ”を用いた若手研究者支援政策に関する研究”， *Japan Advanced Institute of Science and Technology*， Vol. 36， pp. 758-763， 2021.

[24] ときわ会栄養指導課，“減塩について”，栄養指導，http://www.tokiwa.or.jp/nutrition/diet/low-salt.html， 閲覧日 2023.01.15

[25] 全国健康保険協会，“ちょっとした工夫で脂質をコントロール”，https://www.kyoukaikenpo.or.jp/g4/cat450/sb4501/p004/， 閲覧日 2023.01.15

[26] 厚生労働省，“日本人の食事摂取基準 (2020 年度版)”，https://www.mhlw.go.jp/content/10904750/000586559.pdf， 閲覧日 2023.01.15

[27] 東京医科大学病院，“カリウムは調理のくふうで減らせます”，内臓内科，https://articles.oishi-kenko.com/syokujinokihon/dialysis/05/， 閲覧日 2023.01.15

[28] 厚生労働省，“糖尿病”，https://www.mhlw.go.jp/content/10904750/000586592.pdf，閲覧日 2023.01.17

[29] 厚生労働省，“慢性腎臓病”，https://www.mhlw.go.jp/content/10904750/000586595.pdf，閲覧日 2023.01.17

[30] 腎臓内科，“慢性腎臓病の食事療法”，東京女子医科大学，https://www.twmu.ac.jp/NEP/shokujiryouhou.html， 閲覧日 2023.01.17

[31] 厚生労働省，“脂質異常症”，https://www.mhlw.go.jp/content/10904750/000586590.pdf，閲覧日 2023.01.17

[32] 厚生労働省，“高血圧”，https://www.mhlw.go.jp/content/10904750/000586583.pdf，閲覧日 2023.01.17

[33] 厚生労働省，“食べ物アレルギー”，アレルギーポータル，https://allergyportal.jp/knowledge/food/， 閲覧日 2023.01.17

[34] J. Blank, “pymoo: Multi-objective Optimization in Python ”, https://www.egr.msu.edu/ kdeb/papers/c2020001.pdf, 閲覧日 2023.1.22.

[35] 和正敏, “多目的線形計画問題に対する対話型ファジィ意思決定手法とその応用 ”, 電子情報通信学会論文誌 Vol. J 65-A, No. ll, pp. 1182-1189, 1982.

[36] 厚生労働省, “日本人の食事摂取基準 (2020 年版) ”, https://www.mhlw.go.jp/content/10904750/000586553.pdf, 閲覧日 2022.12.26.

[37] 農林水産省, “一日に必要なエネルギー量と摂取の目安”, https:/www.maff.go.jp/j/syokuiku/zissen_navi/balance/required.html, 閲覧日 2023.1.22.