

卒業論文

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

Data Fusion through Web-GIS Visualization
Using Open Data for Evidence-Based Policy Making

富山県立大学 工学部 電子・情報工学科

1815048 長瀬 永遠

指導教員 奥原 浩之 教授

提出年月: 令和4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 EBPM と ICT の有用性	4
§ 2.1 ICT とデータの利活用による行政施策	4
§ 2.2 GIS による地域的な政策決定支援	7
§ 2.3 WebGIS によるデータフュージョン	10
第 3 章 データ間の因果性と改善値の導出	13
§ 3.1 因果探索によるデータ間の関係性	13
§ 3.2 DEA による効率値と入力・出力改善値の導出	16
§ 3.3 Folium を用いた Web-GIS の開発	18
第 4 章 提案手法	22
§ 4.1 データのスクレイピングと因果探索によるデータの選定	22
§ 4.2 選定されたデータに基づく DEA 分析	24
§ 4.3 Web-GIS を用いたデータフュージョンのシステム開発	27
第 5 章 数値実験並びに考察	29
§ 5.1 数値実験の概要	29
§ 5.2 実験結果と考察	30
第 6 章 おわりに	31
謝辞	32
参考文献	33

図一覧

2.1	EBPM のロジックモデル	5
2.2	GIS における 4 つの特徴	8
2.3	GIS を用いることの 3 つの利点	8
2.4	ポテンシャルマップの重ね合わせ	11
2.5	身体地図情報システムの構成	11
3.1	因果グラフのパス図	15
3.2	誤差同士の因果関係	15
3.3	CCR の効率フロンティア	18
3.4	Folium による Web-GIS 実装例	21
4.1	csv のフォーマット	23
4.2	システムのフロントページ	23
4.3	因果探索の結果とアウトプットの例	24
4.4	各都道府県の DMU 内訳	25
4.5	DEA 部分のフロー	25
4.6	DEA のアウトプットの例	27
4.7	改善の必要あり	28

表一覧

3.1	代表的な GIS ソフトウェア	19
4.1	データベース内のデータ	23

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
LiNGAM における観測変数	x_i, x_j
	b_{ij}
LiNGAM における誤差 (非観測変数)	e_i
主問題における各入力・出力に対する重み	v^T, u^T
主問題における対象 DMU の評価値	z
CCR モデルにおける対象 DMU の入力・出力	x_o, y_o
CCR モデルにおける全入力・出力	X, Y
双対問題における対象 DMU の評価値	w
入力指向モデルにおける対象 DMU の評価値	θ
入力指向モデルにおける各 DMU に対する重み	λ
出力指向モデルにおける対象 DMU の評価値	η
出力指向モデルにおける各 DMU に対する重み	μ
入力指向モデルにおける対象 DMU の改善値	$\hat{x}_i$
入力指向モデルにおける各 DMU の入力	x_i
出力指向モデルにおける対象 DMU の改善値	$\hat{y}_j$
出力指向モデルにおける各 DMU の出力	y_j

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。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、平成29年の統計改革推進会議の決定のもと内閣府によってEBPM推進チームが発足され、内閣府の各部局によってEBPMに用いるためのデータを公開するという形でEBPMの推進を図っている。また、EBPMを端的に「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」[4]と定義する著書もあることから、EBPMは単に行政のみが取り組むべき問題ではなく、研究機関にとっても関係の深いものである。

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[4]やシステムティック・レビューを最も重要度の高いエビデンスであると位置づける書籍[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。

以上のように、近年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ではあるが、現在でも地方自治体における政策決定の多くには、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。

このような従来用いられているエピソードベースの政策決定では、その政策の効果によって問題が解決されるケースも存在するが、場合によっては施行した政策の影響で今まで発生していなかった別の問題が発生することもある。また、政策によって別の問題が発生した場合にその意思決定に至った根拠に信頼性が乏しい場合、行政が住民の信頼を得ることは困難となる。

これらの原因の一つとして、政策の対象となる問題に関する要因の複雑性が挙げられる。つまり、ある事柄に対して問題が顕在化する前に政策を講じようと考えたときに、周囲のどの要因が問題に対して影響を及ぼすことによって問題が発生するのかを特定することが困難なのである。

§ 1.2 本研究の目的

1.1 節で取り上げた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对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、上記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、EBPM を取り入れる上での負担を軽減しうるものと考えた。

本研究では、政策決定における対象の複雑性という課題に対して、政府や地方自治体が収集し、一般に提供するという形でサイバー空間上に存在する多種多様なオープンデータを用いていくつかのデータ分析を行うことでその解決を支援する手法を提案する。本研究の提案手法では、はじめにサイバー空間上に存在するオープンデータから項目の分類を問わず複数のデータを集める。

そのうえで、政策の対象としたい事柄に関するデータを基準として線形非ガウス非巡回モデル (Linear non-Gaussian acyclic model: LiNGAM) を用いた因果探索 [4] を行い、政策対象のデータと因果関係があるデータを抽出する。

また、抽出されたデータと政策対象のデータとの間の成り立つ因果関係の方向に着目して抽出されたデータを行政機関における活動の入力・出力に振り分けることでそれらを用いて包絡分析法 (Data Envelopment Analysis: DEA) [4] による分析を行い、行政機関における活動の評価値および入力・出力の改善値を算出する。

加えて、これらの結果を地理情報システム (Geographic Information System: GIS) によって可視化し、DEA による分析に用いたデータの中で地理的な特徴を持つデータと同一プラットフォーム上で視覚的な重ね合わせを可能にすることで行政が新たな政策に対する知見を得ることを支援するアプリケーションの開発を行う。

本論文では、まず、政策決定において ICT 技術やデータマイニングが効果的に用いられた例、地方自治体を中心に取り入れられている GIS を用いた政策決定の例、GIS を用いてデータフュージョンが行われた例などを挙げることで本研究の意義を示す。

次に、前節で挙げた課題を解決するためのオープンデータに対して因果探索や DEA を用いたデータ分析を適用する手法および GIS を用いた効果的な結果の提示手法、データの重ね合わせに関する手法を提案する。

最後に、地方自治体における政策決定について仮想的な問題を設定することで提案手法のデータ分析における各段階の結果 (因果探索で導かれたデータ間の因果関係、DEA における評価値、入力・出力の改善案の試算) を個別で示し、それらをもとに GIS による結果の可視化とデータの重ね合わせを行う。また、それらをもとに本研究の有効性および今後の展望について言及する。

§ 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の重要性と地方自治体におけるEBPM導入の困難さについて述べる。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** EBPMに対する内閣府の取り組みについてまとめ、EBPMとICTの関係性について述べる。また、地域的な政策決定とデータの重ね合わせの二つを取り上げて、本研究におけるGISの重要性について述べる。
- 第3章** 本研究の提案手法に用いるデータ分析手法の理論について解説する。また、提案手法のシステム部分に用いるWeb-GISの作成方法を述べる。
- 第4章** 提案手法中の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。その後、それらのプログラム間における関係性についてデータの送受信などを示しながら説明することで本研究の提案手法全体の流れについて述べる。
- 第5章** 地方自治体における政策決定について仮想的なケースを1つ取り上げてその問題に対して本研究の提案手法を適用することで提案手法における各分析の結果を示し、そこから得られる知見について考察することで本研究の有効性を示す。
- 第6章** 本論文における前章までの内容をまとめつつ、本研究で実現できたことと今後の展望について述べる。

EBPMとICTの有用性

§ 2.1 ICTとデータの利活用による行政施策

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・分析する必要がある、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。

本節では、はじめに平成30年に内閣官房行政改革推進本部事務局より公開されたEBPMのロジックモデルをもとに日本においてEBPMに関する取り組みのベースとなっている思考のながれを解説する。その後、EBPMの推進に対する各機関の取り組みをいくつか提示し、その解説を行うことで日本におけるEBPMの現状、EBPMとICTの関連性、ICTを用いたEBPMの有効性を示す。また、本研究においてEBPMを取り扱う理由について言及する。

EBPMのロジックモデル

日本におけるEBPMのロジックモデルおよびその評価方法を図2.1に示す。図2.1からも分かるようにEBPMは行政施策を5つの段階に切り分け、それぞれの段階において次の段階に移行する際に発生する問題に対してそれらに即した計画および効果の検証を行う。

行政施策における5つの段階とは、インプット（資源）を用いることで何らかのアクティビティ（活動）を行うことができ、その活動によってアウトプット（活動目標）に即した結果を得ることができ、その結果がアウトカム（成果目標）に従った利益につながり、その利益がインパクト（社会への影響）を与えるというものである。

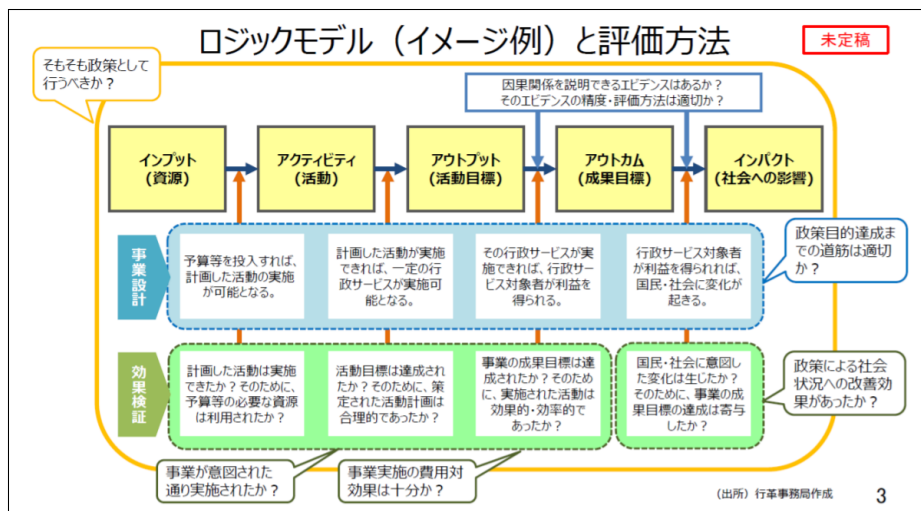


図 2.1: EBPM のロジックモデル

これら 5つの段階における取り組みのそれぞれで適切な計画を設計することで行政での活動を企画することで目的を達成するためにその政策が妥当であるかを議論する。また、それぞれの段階に切り分けて社会状況に対する政策の効果を検証することで政策が意図した改善効果を生むかどうかを評価する。以上の 2つの結果をもって、その政策を施行することが適切であるかを客観的に検討する。

エビデンスを用いた評価に関しては全ての段階において、それぞれの段階で適切なデータおよび手法を用いて分析を行い、行政における活動が社会に対して有効な影響を与えるか、また、それらは投入した資源に見合ったものであるかについて吟味することが EBPM の基本的な考え方であるが、その中でもアウトプットとアウトカム、アウトカムとインパクトの間の分析におけるエビデンスの精度や分析の手法が確からしいか否かが重要視される。

ICT を用いた EBPM

各自治体における EBPM の推進に向けて、内閣府では各部局によって様々な取り組みがなされている。ここでは、EBPM の推進を目的にしたデータの共有や分析に ICT を活用した事例をいくつか取り上げ、管轄する部局やシステムの目的、システムの概要等を解説する。

内閣府エビデンスシステム (e-CSTI)

e-CSTI とは、科学技術・イノベーション推進事務局が提供するシステムで、特に大学等の研究機関における「研究力」、「教育力」、「資金獲得力」を分析することに特化したものである。

e-CSTI では、研究機関におけるデータを収集し、インプットとアウトプットに分類することでそれらの関係性を分析・「見える化」する各種機能を提供する。エビデンスに基づいた研究機関の運営を支援するという目的のもと、2020 年 3 月から運用が開始している。

地域経済分析システム（RESAS）

RESASとは、地域創生の実現に向けて、内閣府地方創生推進室ビッグデータチーム、内閣官房まち・ひと・しごと創生本部事務局と経済産業省地域経済産業調査室が提供しているWEBアプリケーションである。日本全国の各自治体区分における人口、産業、観光、まちづくり、医療・福祉、財政など幅広い分野のビッグデータを官民より集約し、表、グラフ、マップなどのフォーマットを用いてデータを可視化するシステムとして提供されている。

また、データを可視化するだけでなく、集積されたデータをもとに各テーマに沿った分析を行うことや、APIを利用することによってシステムにデータを直接取り込むことも可能である。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。

V-RESAS

V-RESASとは、前述のRESASを提供する内閣府地方創生推進室と内閣官房まち・ひと・しごと創生本部事務局が提供しているWEBアプリケーションであるが、2020年頃から現在まで世界各地で爆発的な感染拡大を引き起こしているCOVID-19と関連する情報のみに焦点を当てているという点で前述のRESASとは異なる。

V-RESASでは、各自治体区分における人流、消費、飲食、宿泊、イベント、雇用などの情報をグラフ等を用いてリアルタイムに発信し、前年比、週平均など過去のデータとの比較をフィードバックすることによって、COVID-19が地域経済に与える影響を可視化することで地方創生の様々な取組を情報面から支援することを目的としている。

内閣府経済社会総合研究所（Economic and Social Research Institute: ESRI）

経済社会総合研究所は2001年の省庁再編に伴って設立され、それ以来、内閣府の活動における理論と政策の橋渡しを行うためのシンクタンクとして機能している。具体的には、経済政策・社会政策に関する理論の研究を行うことに加え、政策研究を担う人材の育成・研修に取り組むことで現状のみならず将来的な観点でも政府に貢献している。また、国民経済計算、景気動向指数や機械受注などの統計使用を作成・公表するなどの役割も担っている。

ESRIのホームページでは、主に経済に関するデータに焦点を当てて統計データ、研究結果やそれらをまとめた論文などが公開されており、経済・財政分野におけるより詳細なデータ収集・分析に活用されることが期待される。

以上の事例に代表されるように、昨今、内閣府によってEBPMが積極的に推進されている。そのため、EBPMは研究機関においても積極的に扱われ、議論されるべき対象であると考えられる。また、その具体的な取り組みの多くにICTが活用されていることから、EBPMの効率的な実施には、ICTの活用が必要不可欠であり、情報工学との親和性は高いものと思われる。これらのことから、本研究は情報技術を用いたEBPMの支援を目的とする。

§ 2.2 GIS による地域的な政策決定支援

GIS とは、位置に関する情報を持ったデータ（地理空間データ）を総合的に管理・加工し、地理的位置とデータを結び付けることによって視覚的に表示することで、空間データの高度な分析や迅速な判断を可能にする技術である。地理空間データの例として、主題図（土地利用図、地質図、ハザードマップ等）、都市計画図、地形図、地名情報、台帳情報、統計情報、空中写真、衛星画像が挙げられるが、これらの情報はその汎用性の高さから、多様な分野で利用される。

また、GIS はいくつかの点において優れた特徴を持つため、GIS を用いてデータの分析を行う利点は複数あるが、その中でも代表的なものとして図 2.3 が示すような 4 つの特徴と図 2.4 が示すような 3 つの利点 [4] が挙げられる。

GIS における 4 つの特徴

- データの可視化

2D や 3D、アニメーションなど多様な表現方法で地図上にデータを可視化することによって、数値のみを見るだけでは気づくことができないようなデータの傾向やデータ間の関連性など様々な情報を一目で把握できるようになる。

- データ間の関係性の把握

複数のデータによる重ね合わせを行うことでデータ間の対比が容易になり、データ同士の関係性を直感的に把握することができる。また、地図上の位置関係からデータを特定することによって、定量的な情報を把握することができる。

- データの統合と分析

位置情報をキーとしてそれぞれ別の特徴を持つデータを統合したり、複数のデータを重ね合わせて分析することによって、独自のアプローチで課題の解を導き出すことができる。

- データの作成と更新

今日の日本では新しいビルの建設や合併による行政界の変更など、現実世界が日々変化しているため地理情報データもそれらに基づいて定期的に更新していく必要がある。この際、GIS を利用することでデータの作成・更新における負担が軽減され、常に鮮度を保ってデータの管理・提供を行うことができる。

GIS を用いることの 3 つの利点

- 業務効率化によるコスト削減

GIS は日常の業務を最適化するために幅広く利用されている。紙地図から GIS を利用したデジタルな地図へ移行することで、現地調査や設備管理、統計分析などを行う際の物質的な制約が減少し、より簡単かつ効率的な作業の実現が可能になる。これにより作業時間や人員など業務にかかるコストを大幅に削減することができるという利点を持つ。

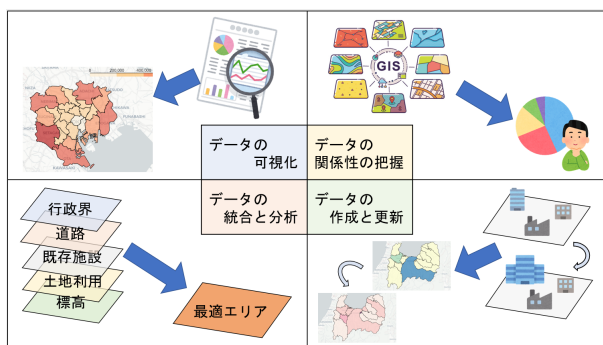


図 2.2: GIS における 4 つの特徴

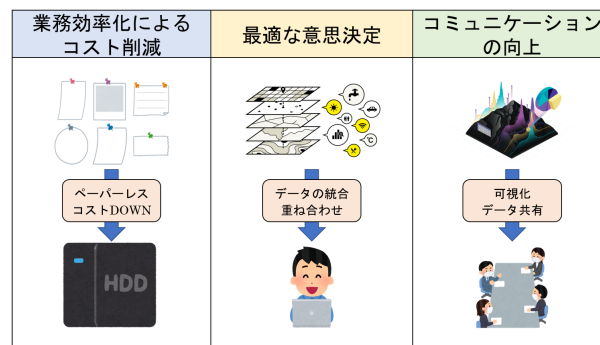


図 2.3: GIS を用いることの 3 つの利点

- 最適な意思決定

組織における活動において、場所に関する意思決定を正しく行うことは成功のための重要な要素の1つといえる。場所に関する決定に視覚を用いることができるという利点を持つ GIS は店舗の出店場所や配送ルート、避難地域・経路、天然資源の採取地点など多様な分野で最適な場所を策定するために使用されている。

- コミュニケーションの向上

GIS を使用することでさまざまな表現方法を用いて位置情報を地図上に可視化することができる。可視化された位置情報は状況を効果的に伝え、的確な理解を促すことができ、グループや組織間、社会におけるコミュニケーションの向上を図ることが可能になるという利点を持つ。

本節では、以上のような特徴・利点を持つ GIS によってデータが分析・加工されることによって、政策における意思決定の支援に用いられた事例を挙げることで、政策決定に GIS を利用することの有効性を提示するとともに、本研究における GIS の位置づけを示す。

東日本大震災における福島県相馬市

太平洋沿岸部に位置する福島県相馬市は、平成 23 年に発生した東日本大震災によって甚大な被害を受けた地区の一つである。相馬市では、震災発生前の平成 21 年より情報政策課を中心に GIS の利用環境が整備されており、震災発生初期から現在も続く復興活動にわたって GIS が活用されている。相馬市が震災発生から被災の各段階においてどのような用途について、いかにして GIS を活用してきたかを時系列順に解説する。

被災初期

GIS を利用して航空写真判読による家屋被害の状況把握や人的被害の全容把握のための全住民位置情報の作成が行われ、罹災証明発行業務の円滑化など業務の効率化を可能にした。大規模震災による被災という非常時において人手不足の解消に役立てられた。

復興活動

地盤高情報を用いて津波被害の特徴を明確化したうえで、津波のリスクを織り込んだ今後の土地利用に関する資料が作成された。また、放射線量マップを作成し、当該地図や全住民位置情報、土地利用地図、地形図等と重ね合わせることによって、放射線対策、及び、今後予見される被害状況の広がり対策に関する意思決定に用いられた。

相馬市では、以上に挙げたように震災という非常事態における意思決定に GIS が有効活用された。また、これらの活動に際して、使用された GIS のメーカーから資料や技術力の提供、新潟大学の学生ボランティアから人員の供給が行われ、実際の運用や意思決定を相馬市が行ったことから、産学官の連携による活動の例でもある。

奈良県橿原市の庁内における情報共有

橿原市で用いられている行政基本 GIS は、行政基本図データを共通基盤にして各課で個別に運用している都市計画、道路管理、下水道管理、上水道管理などの業務 GIS の中から全庁における利用頻度の高い情報を共用主題データとして、庁内すべての職員が参照できるシステムで、各課の垣根を超えた情報共有を可能にしている。

また、共用主題データの中から一般に公開する必要があると考えられるデータに関しては国土地理院が提供する電子国土 Web システムを用いて大きな手を加えることなくスムーズに一般公開がなされている。一般に公開されるデータ例として、選挙期間中には投票所 MAP などが随時公開されている。

各課における GIS の導入をバックアップするために共有データは庁内で共通した製品仕様書を定めることで作成・運用方法の統一が図られている。全庁において統一したフォーマットで GIS を運用することで庁内の情報流通の促進、行政事務の効率化・高度化、市民サービスの向上などの利点も期待することができる。

各課における GIS に対する知識の醸成と実際の利用によって生まれる好循環の確立により更新される基盤地図情報を使って、これまで以上に行政事務の効率化・高度化や住民サービスの向上を図ることが可能となる。

また、市域・県域をまたいだ広範囲の多くの機関で情報共有が必要となる防災分野においては橿原市のみに限らず全国を均質に扱う基盤地図情報が共通基盤として活用されること望まれる。

新潟県魚沼市の業務効率化とコスト削減

魚沼市には平成 16 年に 6 つの町村が合併することによって誕生したという経緯があり、合併前には GIS を導入している町村と紙図面で管理する町村がどちらも存在するという状況だった。そのため、合併を機に県の交付金を活用するなどして全市域に対応した GIS を導入することでデータの電子化がなされた。

庁内共有のデータとして、市保有の地図情報レベル 2500 の地形図をベースに道路線を地図情報レベル 1000 で作成したハイブリッド地形図があり、都市計画分野や道路管理等の庁内業務には、この大縮尺地形図データを用いた管理が用いられている。

一方、庁内業務では施設の配置情報などで広域の管理を行う業務があり、従来は縮尺 1/25,000 の紙の地形図を使って管理がなされてきた。これに対して、合併を契機に導入することになった庁内共用 GIS でも広域での施設管理が行えるようにするためには地図情報

レベル 25000 程度の地形図データも必要となるが、庁内には地図情報レベル 25000 程度の地形図データは存在しなかった。

そのため、国土地理院から提供され、無料で利用できる縮尺レベル 25000 基盤地図情報を採用した。無料のため初期投資はかからず、更新費用もなく適時更新され、新しい背景地形図データを利用することが可能となった。

以上に挙げた事例からもわかるように、近年、GIS は地方自治体における業務に積極的に活用されており、先述のような特徴と利点を持って多くの業務の効率化、意思決定の補助としての役割をはたしている。これらのことから、GIS は本研究の提案手法において分析結果を感覚的に理解することや非常に多様なデータを 1 つのプラットフォーム上にまとめ、新たな発見を生み出すことに有効であると考ええる。よって、本研究の提案手法では最終的なユーザへのフィードバックとして GIS を採用する。

§ 2.3 WebGIS によるデータフュージョン

2.2 節では、GIS について政策支援の観点から有効性を示したが、本節では、地理空間データの重ね合わせによるデータフュージョンとそれによる新しい知見の発見という観点から GIS の有効性を示す。また、本研究におけるデータフュージョンの重要性を提示する。

先述のとおり、GIS は地理空間データに対する多様な加工によって、それらを高度に分析し、地図上に視覚化することでそのままでは概形のとらえにくい地理空間データに対する判断を迅速にすることができるという利点に加えて、地図上に視覚化したデータを重ね合わせて表示できるという特性を持つ。

このことによって、単一のデータの可視化のみでは表面化してこなかった課題をくみ取ることや、逆に、課題に対する解決策を一見関係のなさそうな分野から発見するといったことが可能となる。

このような GIS でのデータフュージョンによって他分野での知見を合わせることで新たな発見に至った事例やデータの重ね合わせによって新たに知見を得ることを目的にして行われた開発の事例を以下に挙げる。

都市計画と生態学

1992 年 6 月にブラジルで開催された国連環境開発会議で条約加盟のための署名が開始され、1993 年 12 月 29 日に発行された「生物多様性条約」に代表されるように 20 世紀後半から現在にかけて自然環境や生物多様性の保護に関する機運は世界中で高まっている。

その動きはわが国においても例外ではなく、2008 年には「生物多様性基本法」が成立するなどしている。そのような状況の中で大学等の研究機関においても生物多様性の保護に向けた様々な研究がなされている。

そのうちの一つに、生態学・ランドスケープ分野における植物の生育環境と農村計画分野における将来的に都市化の可能性がある地帯の分析とを重ね合わせることで新たな知見を得ようとする研究 [4] がある。

同研究の具体的な内容は福島県旧原町市域を対象として今後都市化する可能性のある地域を可視化したマップ（都市化ポテンシャルマップ）と二次草原（人為的な干渉によって維持されている草原）が生育する可能性のある地域を可視化したマップ（二次草原の潜在

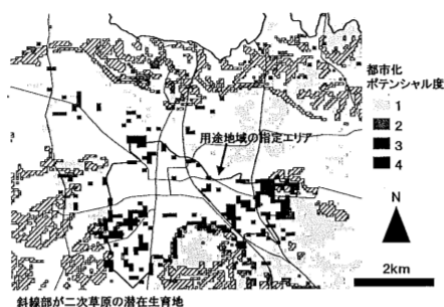


図 2.4: ポテンシャルマップの重ね合わせ

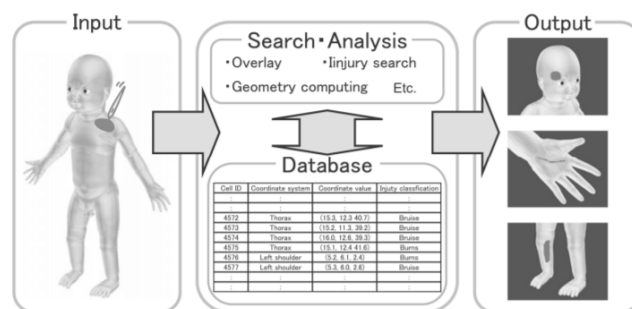


図 2.5: 身体地図情報システムの構成

生育地マップ) とを GIS 上で重ね合わせることで今後の都市化によって人知れず損なわれる可能性のある二次草原の生育地を分析するというものである。都市化ポテンシャルマップと二次草原生育地マップを重ね合わせた様子を図 2.4 に示す。

同研究内では従来別々の分野で進められてきた研究の成果を GIS を用いて巧みにオーバーレイし、普段の研究では同一プラットフォーム上に並べられることのないデータをかけ合わせることで、新たな知見として、都市化ポテンシャルの高い場所と二次草原の潜在生育地が重なる場所の環境特性を明らかにした。

また、研究を通じて多くの学術知見を統合するというケースにおける GIS の有意性が言及されており、1つのプラットフォーム上で様々な情報を共有できることの効率性が取り上げられている。

身体情報に対する GIS の転用

GIS を用いるうえでの利点であるデータの可視化および重ね合わせを地理情報以外に応用した研究の中で非常に興味深いものとして GIS における地理的な位置関係という特徴を人体における各部位の位置関係に置き換えることで身体情報をモデル化した研究である [] を挙げることができる。

この研究では、現在、一般に研究がなされている分野の中で医学、工学、産業など人間にかかわるものは数多く存在するにもかかわらず、身体情報に関してそれぞれが収集した情報や分析結果などが異分野もしくは同分野においても効率的に共有されていないことを問題として取り上げている。

また、これらの問題が発生することによって分野間にまたがった研究の連携を阻害すると考え、原因として人間に関する情報の表現方法が統一されていないことを挙げてその解決方法を模索・提案している。

人体におけるデータを蓄積共有するシステムは論文内で「身体地図情報システム」と名付けられている。有効性を示すための適用事例として対象を子どもとし、14 歳以下の子どもの死因として最も割合の高い不慮の事故を予防するために傷害情報を統合的に管理、収集、分析を行うようなシステムとして実装されている。身体地図情報システムの構成を図 2.5 に示す。

身体地図情報システムは解剖学の知見に基づいた 38 箇所の主要体節に座標系を設定し、その座標系を参考に傷害情報を登録するものとなっている。また、傷害情報は位置・形

状・属性の3つの要素を持ち3次元人体モデルにおける体表面上を隙間なくセルで区切ることで作成されたラスタモデルによって記述される。

身体地図情報システムには、入出力機能・検索機能・解析機能の3つの機能が搭載されており、3次元人体モデル上をマウスでペイントとすることによって傷害情報が正規化、データベースに登録され、データベース内のデータを検索またはそれらを用いた解析を行うことができる。データ解析には3つの機能があり、それぞれ以下のようにになっている。

1. オーバーレイ解析

複数の傷害情報をラスタレイヤとして重ね合わせることで同じ位置にあるセルの値同士に四則演算を実行し、新たなセルを作成する。この機能を用いて傷害レイヤ同士を重ね合わせることで各セルにおける傷害の頻度をもとめることができる。

2. 受傷面積算出機能

ラスタレイヤにおいて傷害として塗りつぶされた範囲の面積を自動的に算出し、その値が体表全体の何%にあたるかを表示する。この機能は実際の医療現場において熱傷の患者の重症度を概算するために用いられる患部が体表の何%にあたるかという判断基準をもとに実装された。

3. ICD-10 コード変換機能

ICD-10 コードは世界保健機関によって定められた傷害の種類および受傷した部位に基づく傷害の分類コードである。身体地図情報システムでは頭部、頸部、胸部など全部で10の領域を用いて部位を分類した身体部位定義レイヤに対して傷害情報および受傷位置情報を持つレイヤを重ね合わせることで入力された傷害情報をICD-10 コードに変換する。

これらの機能の中でも、オーバーレイ機能は異なるデータ同士を重ね合わせて位置情報に基づいた判定を行ったのち、位置が等しいデータのみに対して演算を行うことで新たなデータレイヤを作成する点、ICD-10 コード変換機能は異なるデータの重ね合わせによって新たに生まれる領域を用いて別のデータに変換するという点でGISによる重ね合わせを用いたデータフュージョンの有効性が示されている。

以上のようなGISによる複数の研究分野での研究結果の統合を扱った先行研究やGISによる重ね合わせの有用性を示した先行研究からも読み取れるとおり、GISにおける同一プラットフォーム上でのデータの重ね合わせは問題を引き起こす原因が1つのみに限らないような複雑な問題における新知見の発見に対して大いに役立つものと考えられる。

よって、本研究で目指すデータフュージョンによる政策決定支援を行う上でもGISは大いに有効なものであると考える。そのため、本研究ではデータを用いた分析の結果と地理空間データをGISによって同一プラットフォーム上にオーバーレイするという形でデータフュージョンを行うこととする。

データ間の因果性と改善値の導出

§ 3.1 因果探索によるデータ間の関係性

因果探索とは、観測データを用いて、そのデータ群の因果グラフ（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もの）を導出するための教師なし学習のことである。

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。

例えば、「ある小売店 A 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 A 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での研究に用いられている。

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、

$$x_i = \sum_{j \neq i} b_{ij} x_j + e_i \quad (i = 1, \dots, p) \quad (j = 1, \dots, p) \quad (3.1)$$

以下のような仮定を用いることで因果グラフを導出する手法である。

1. 外生変数と内生変数をつなぐ関数は線形関数とする。（内生変数とは実際に観測されている変数、外生変数とは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。）

2. 外生変数の分布は非ガウス連続分布とする.
3. 因果グラフは非巡回とする.
4. 外生変数は互いに独立とする.

ここで、図 3.1 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.

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

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.

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、図 3.2 の例に見られるように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つてしまう. 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. 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.

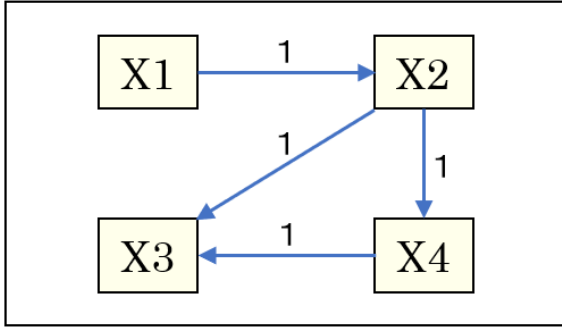


図 3.1: 因果グラフのパス図

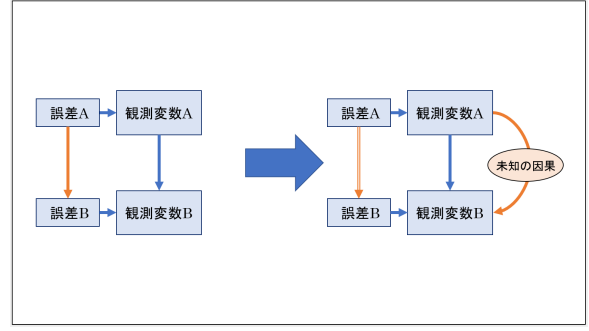


図 3.2: 誤差同士の因果関係

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

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二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、回帰残差は

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

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 $s_i (i = 1, \dots, q)$ 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

$$y_1 = \sum_{i=1}^q \alpha_i s_i \quad (3.5)$$

$$y_2 = \sum_{i=1}^q \beta_i s_i \quad (3.6)$$

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 となる時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．

§ 3.2 DEA による効率値と入力・出力改善値の導出

DEA とは、ある分野における組織の集合において、対象の組織の業績を評価するために生み出されたデータ指向のアプローチで 1978 年に charnes, cooper と Rhodes によって提唱された。ここでいう組織とは、その活動においていくつかの種類の入力（投入）をいくつかの出力（産出）に変換することに携わる生産体（Decision Making Unit: DMU）のことである。DEA での分析の利点の一つとして、複数の入力・出力があるデータを扱うことができることが挙げられる。

DMU における活動の例として、半導体生産工場における毎期の生産活動や、様々な市場に対する宣伝や販売活動等が挙げられる。DEA はこれらの活動に対して、自らを基準とした集合内の他の DMU との比較によって、その業績を評価することが可能である。

1978 年に提唱されて以来、DEA に関する研究、応用は世界中の研究機関で積極的に行われており、CCR, BCC などの基本的なモデルをはじめとして、現在までに様々なモデルが発表されている。本節では、多数存在する DEA のモデルの中で最も基本的であり、本研究の分析にも用いる CCR モデルについて、そのアルゴリズムを解説する。

また、本研究で DEA におけるデータの分析を行う際に前述のとおり多数存在するモデルの中から CCR モデルによる分析を選択した理由については次のようなことが挙げられる。CCR モデルは最もはじめに考えられた DEA のモデルであり、原点からの距離の比を用いて値の最小を定める「比率尺度データ」を対象とした分析法である。

そして、それに次いで考えられたモデルは CCR モデルに各種制約を追加することで「感覚尺度データ」など、対象のデータに適用できるように拡張されたものである。ここで、本研究で分析に用いるオープンデータを考えると、すべてが「比率尺度データ」であったため、本研究では CCR モデルを適用することとした。

DEA (CCR モデル)

DEA における DMU の評価法の基本的な考え方は「どれだけ少ない入力を用いてどれだけ多くの出力を生み出しているか」である。そのため、DEA の評価値は次のように定義される。

$$\left[\frac{\sum (\text{各 DMU の出力} \times \text{出力のウェイト})}{\sum (\text{各 DMU の入力} \times \text{入力のウェイト})} \right]$$

上記の式にいくつかの制約式を追加し、それらに基づいて入力・出力のウェイトを最適化する線形計画問題を解くことによって DEA における評価値を算出することができる。CCR モデルにおける制約条件は以下の二つである。

- すべての DMU に対する評価値はいずれも 1 を超えない。
- 入力・出力に対するウェイトはいずれも 0 以上である。

これらをもとに CCR モデルを線形計画問題として定式化すると以下ようになる。

<CCR モデルの主問題>

$$\text{Objective} \quad \frac{u^T y_o}{v^T x_o} = z \rightarrow \max \quad (3.7)$$

$$\text{Constrained} \quad -v^T X + u^T Y \leq 0 \quad (3.8)$$

$$u \geq 0 \quad (3.9)$$

$$v \geq 0 \quad (3.10)$$

ここで、DEA における分析が持つもう一つの利点を考える。それは、算出された各入力・出力に対するウェイトを用いることで、対象の DMU の入力・出力をどのように増減させれば評価値がより優れたものになるかを数学的に示すことができることである。この DMU における評価値をより良くするための入力・出力の値（本論文では以後、改善案と呼ぶ）を算出するためには、評価値における入力（分母）もしくは出力（分子）を 1 と仮定し、前述の線形計画問題を主問題とする双対問題を考える必要がある。

これらをもとに、入力に対する改善値を算出することを目的とした入力指向モデルと出力に対する改善値を算出することを目的とした出力指向モデルを定式化すると以下のようになる。

< 入力指向モデル >

$$\text{Objective} \quad w = \theta \rightarrow \min \quad (3.11)$$

$$\text{Constrained} \quad Y\lambda \geq y_o \quad (3.12)$$

$$-X\lambda + x_o\theta \geq 0 \quad (3.13)$$

$$\lambda \geq 0 \quad (3.14)$$

< 出力指向モデル >

$$\text{Objective} \quad w = \eta \rightarrow \max \quad (3.15)$$

$$\text{Constrained} \quad X\mu \leq x_o \quad (3.16)$$

$$-Y\mu + y_o\eta \leq 0 \quad (3.17)$$

$$\mu \geq 0 \quad (3.18)$$

ここで、入力指向モデルを取り上げて解説を行うと、その線形計画問題は対象の DMU における出力の大きさを維持しつつ、最小の入力を達成することができる DMU の集合（参照集合）をすべての DMU の中から探すという操作ととることができる。つまり、線形計画問題の解においてその重みが正の数となる DMU が参照集合に属し、その場合における効率的な DMU ととらえることができる。

これらのことから、参照集合とは対象の DMU が見本とすべき DMU の集合であり、参照数値に属する DMU におけるそれぞれの要素の値とその重みの大きさをを用いることによって、対象の DMU に対する入力および出力の改善案を算出することができる。

参照集合における DMU の数を K 、入力・出力の項目数をそれぞれ m 、 n とすると対象の DMU のある項目 i, j における入力・出力における改善案はそれぞれ次のようにしてとめることができる。

< 入力改善案 >

$$\hat{x}_i = \sum_{k=1}^K x_{ik}\lambda_k \quad (i = 1, 2, \dots, m) \quad (3.19)$$

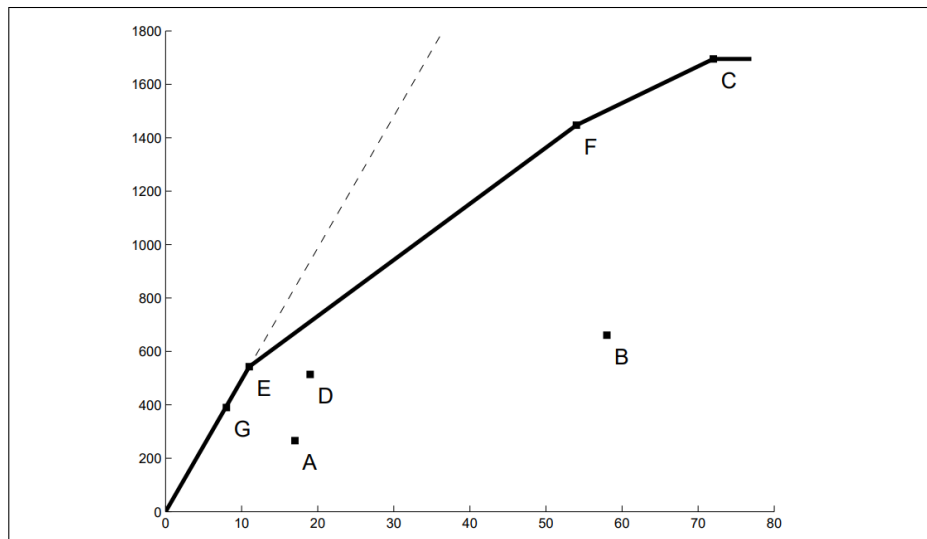


図 3.3: CCR の効率フロンティア

< 出力改善案 >

$$\hat{y}_j = \sum_{k=1}^K y_{jk} \mu_k \quad (j = 1, 2, \dots, n) \quad (3.20)$$

本研究における提案手法において、入力指向モデルおよび出力指向モデルによって算出された対象の市区町村における評価値と式 (3.19)、式 (3.20) によって算出された入力および出力の改善案を提示することを政策に対する意思決定を支援する手法の1つとして用いることとする。

§ 3.3 Folium を用いた Web-GIS の開発

2.2 節, 2.3 節では、すべての GIS に対して一般的な名称として GIS と表記したが、GIS には動作するプラットフォームや形態、提供している団体によって複数種類のソフトウェアが存在する。代表的な GIS アプリケーションを表 3.1 に示す。また、本節では、Web アプリケーションとして World Wide Web 上で機能する GIS を Web-GIS と表記するものとする。

html 形式で実装され、Web-GIS は多様なプラットフォーム上で動作する GIS の中でも代表的なフォーマットであるといえる。計算機を用いて広く利用することができ、処理の大部分は html が置かれているサーバ上で行われることで参照する端末自体のスペックに依存しづらいことから、利用者も多い。また、html 形式で実装されるがゆえに作成者が直接作成する方法のほか、いくつかのプログラミング言語を用いて自動的に生成することも可能である。

本節では、以上のような特徴を有する Web-GIS の開発において、現在、一般に広く用いられているプログラミング言語の一つである Python のライブラリを用いた方法を解説する。また、Web-GIS において実装することができる代表的な機能とその役割を解説する。

Web-GIS の実装方法

表 3.1: 代表的な GIS ソフトウェア

GIS ソフトウェア	無料	オープン ソース	Windows	MacOS	Linux	BSD	Unix	Web
Microsoft MapPoint	×	×	○	×	×	×	×	○
ArcGIS	×	×	○	○	×	×	○	○
GRASS GIS	○	○	○	○	○	○	○	○
QGIS	○	○	○	○	○	○	○	○
MapInfo	×	×	○	×	×	×	×	○
TNTmips	×	×	○	○	○	×	○	×

Python を用いた Web-GIS の開発には「Folium」という Python 用のライブラリを用いる。Folium を用いてメソッドに対して初期位置，ベーススタイル，初期縮尺などを引数として与えてプログラムを実行することで Web-GIS のベースとなるマップを表示する html が自動的に生成される。ベーススタイルとして指定することの出来るマップスタイルには代表的な例として以下のようなものがある。

- CartoDB (positron and dark_matter)
- OpenStreetMap
- Mapbox Bright
- Cloudmade
- Mapbox

Folium による Web-GIS の開発はこのベースマップに対して Folium のライブラリ内に含まれる様々なメソッドを用いることで Web-GIS における各種機能や実際に表示した情報を追加するという形で行われる。ここからは，Folium によって実装することの出来る各種機能とその内容について代表的なものを取り上げる。

ベースマップの切り替えとレイヤの重ね合わせ

2.3 で言及したような GIS によるデータの重ね合わせは FeatureGroup 関数によるレイヤの作成とそれらを制御する LayerControl メソッドによって実現される。また，レイヤとは Web-GIS 上でマップの重ね合わせを行う際のそれぞれの層のことを表す。

まず，FeatureGroup 関数によってベースマップとは別のマップ（レイヤ）を任意の個数作成する。次に，後述する様々なメソッドを用いてベースマップや各レイヤに対して各種機能や情報を追加する。

この時，常に表示され，レイヤの切り替えに左右されないようにする必要のある情報に関してはベースマップに，それ以外の情報に関しては切り替えによって表示したい各レイヤに追加するようにする。最後に，LayerControl メソッドを用いてレイヤを管理する機能を Web-GIS に付与することによって，html を生成した際に自由にレイヤを切り替える機能を持ったものが生成される。

マーカーを置く

`folium.Marker` メソッドに対して引数としてマーカーを置く位置の座標を与えることで地図上の任意の位置にマーカーを立てることができる。なお、マーカーは `Folium` 内に組み込まれているものの中から色やアイコンのマークを自由に切り替えて使用することができるほか、`CustomIcon` 関数によってアイコンのマークを自作し、独自のマーカーとして使用することもできる。また、任意のマーカーに対して `popup` 機能を追加し、テキストを付与しておくことでマーカーをクリックした際にポップアップとしてテキストが表示されるようになる。

ヒートマップを描く

ヒートマップとは、二次元データの数値を色やその濃淡で表したものである。広義におけるヒートマップは「マップ」と付いてはいるが必ずしも地図で表現する訳ではなく、テーブルを値で色分けしたものなど数値データを色分けによって可視化したものすべてがこれにあたる。ただ、Web-GIS におけるヒートマップは地図上にプロットされた色の濃淡で数値の大小を示すものである。

また、地図の色分けによって数値の大小を表現する方法として、ヒートマップとは様式が異なるものとして、コロプレス図 (階級区分図) がある。コロプレス図とは、例えば地図を都道府県ごとに境界線で分けて、各都道府県における統計データの大きさによって色分けするなどのものがある。具体的な事例としては、アメリカ大統領選の際の州ごとに赤と青で色分けされた地図などが挙げられる。

`folium.plugin.HeatMap` メソッドを用いることで地図上の任意の位置を中心としたヒートマップを作成することができる。引数として値を与えることで半径や色の透明度、グラデーション、ぼかしの程度が設定できるほか、前述の `LayerControl` と組み合わせることで表示の切り替えも行うことができる。

大量のマーカーをまとめて表示

前述の `folium.Marker` メソッドでは、1つのマーカーに対して多くの情報を付与する方法について解説したが、`Plugins.MarkerCluster` メソッドでは、大量のマーカーを点として地図上にプロットし、一定の閾値を定めることで地図の縮尺によってその付近にあるマーカーを1つのマーカーとして表示することができる。これによって、大量のマーカーを一度にプロットした際でも見やすく情報を提供することができる。また、このような機能はマーカーの密度という観点でヒートマップのような表し方と考えることもできる。

以上で解説した各種機能の実装例については図 3.4 にて示す。また、以上のような機能のほかにも、`Folium` によって実現可能な Web-GIS の機能は多くあるが、本研究では特にマーカーによる情報のフィードバックおよび重ね合わせを中心に行っていく。システムの詳細については 4 章にて示す。

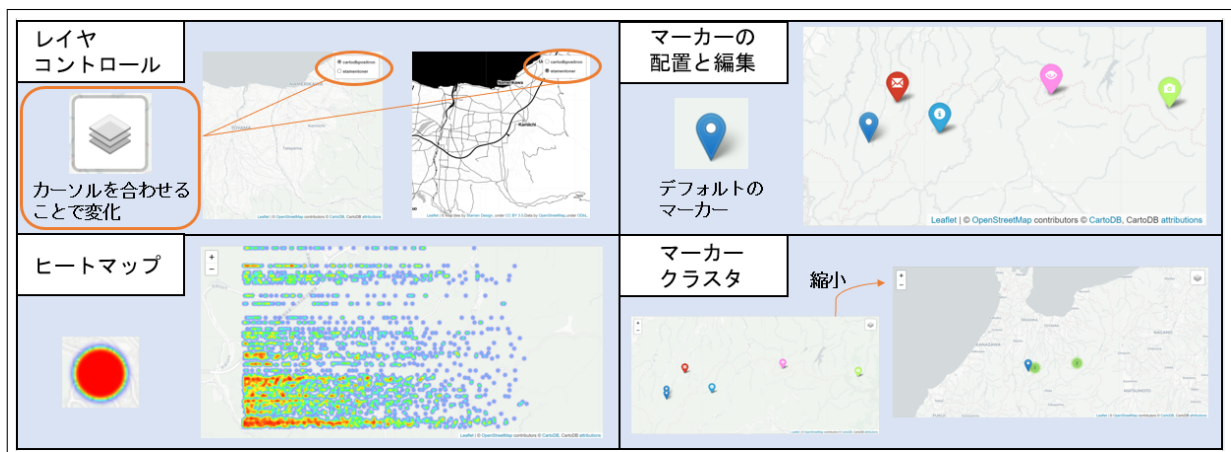


図 3.4: Folium による Web-GIS 実装例

提案手法

§ 4.1 データのスクレイピングと因果探索によるデータの選定

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 上からあらかじめ収集し、サーバ上のデータベースに保存されたデータの中から利用者が任意のデータを選択する。
2. 選択されたデータとの間に因果関係が成立するデータを 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、評価値および、改善値を算出する。
4. DEA の結果を Web-GIS 上に可視化し、対象のデータと因果関係があり、地理的要素を持つデータについてはレイヤーを作成して重ね合わせが行えるようにする。

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. 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. 結果の表示および重ね合わせ

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表 4.1 に示す。データの属性としては地理的情報を持つか否かで大きく二つに分けることができる。地理的情報を持たないデータの例には人口、製造品出荷額などが挙げられ、数年に一回、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

表 4.1: データベース内のデータ

GIS ソフトウェア	無料	オープン ソース	Windows	MacOS	Linux	BSD	Unix	Web
Microsoft MapPoint	×	×	○	×	×	×	×	○
ArcGIS	×	×	○	○	×	×	○	○
GRASS GIS	○	○	○	○	○	○	○	○
QGIS	○	○	○	○	○	○	○	○
MapInfo	×	×	○	×	×	×	×	○
TNTmips	×	×	○	○	○	×	○	×

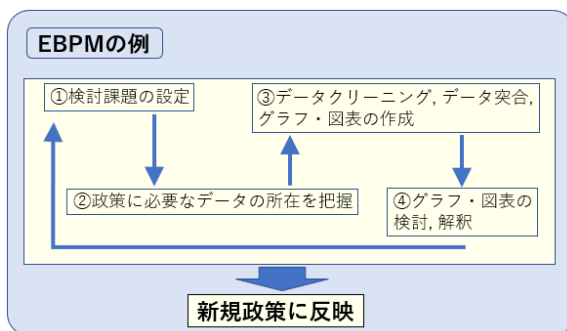


図 4.1: csv のフォーマット

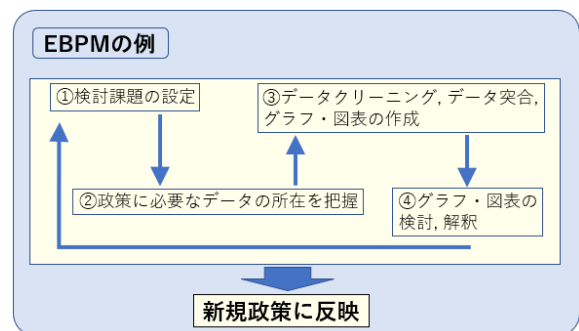


図 4.2: システムのフロントページ

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 RESAS 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 4.1 に示す。RESAS 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。施設の位置データについては

システムのフロントページおよび対象の選択

提案手法においてユーザサイドに提示されるフロントページを図 4.2 に示す。図 4.2 に示した通り、システムのはじめにユーザには前述のデータベース内に格納されている全てのデータの項目名が提示される。また、データの項目名をクリックすることによってデータの中身を参照することができる。このページにおいてユーザは自身が対象としたい市区町村およびデータ項目を選択する。

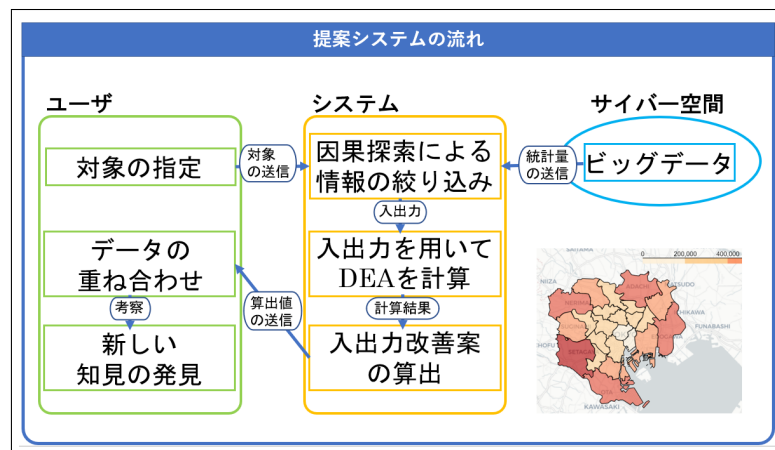


図 4.3: 因果探索の結果とアウトプットの例

因果探索のよるデータの選定

本研究では、1.1 節で挙げた政策の対象における要因の複雑性という課題の解決を支援する手法として 3.1 節で挙げた LiNGAM による因果探索を提案する。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。システムのアプローチとしては、前述のフロントページにてユーザが選択した政策の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表 4.1 に示したデータのうち、グループ 1 およびグループ 2 に属するすべてのデータであり、これらすべて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を DEA の入力・対象のデータから出る矢印が向くデータを DEA の出力として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 4.3 に示す。

§ 4.2 選定されたデータに基づく DEA 分析

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

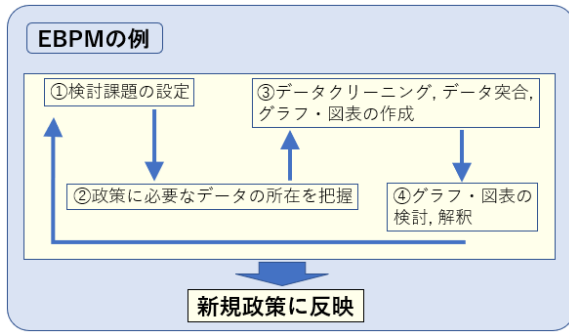


図 4.4: 各都道府県の DMU 内訳

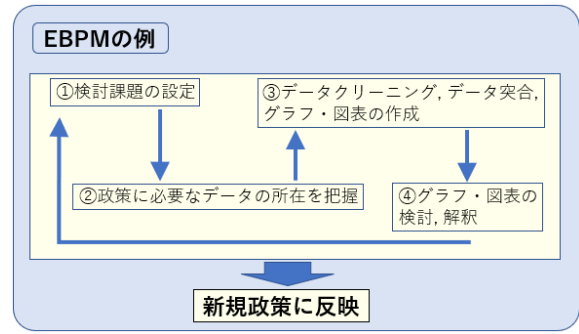


図 4.5: DEA 部分のフロー

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 4.3 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 (3.19) および式 (3.20) 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〇〇個の市、〇〇個の区、〇〇個の町、〇〇個の村の合計〇〇個を 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もとめられる。

< 評価値の算出式 >

$$\text{Objective} \quad \theta \rightarrow \min \quad (4.1)$$

$$\text{Constrained} \quad \sum_{d=1} y_{id} \lambda_d \geq y_{io} \quad (i = 1, 2, \dots, m) \quad (4.2)$$

$$- \sum_{d=1} x_{id} \lambda_d + x_{io} \theta \geq 0 \quad (i = 1, 2, \dots, m) \quad (4.3)$$

$$\lambda \geq 0 \quad (4.4)$$

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と指定された 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以下の式を用いることでもとめられる．

< 入力改善案の算出式 >

$$\hat{x}_{io} = \sum_{a=1}^A x_{ia} \lambda_a \quad (i = 1, 2, \dots, m) \quad (4.5)$$

< 出力改善案の算出式 >

$$\hat{y}_{jo} = \sum_{a=1}^A y_{ja} \mu_a \quad (j = 1, 2, \dots, n) \quad (4.6)$$

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．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 4.5 に示す．

入力・出力のデータは図 4.3 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，入力・出力の項目数をもとにデータを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．

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の重み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 4.6 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

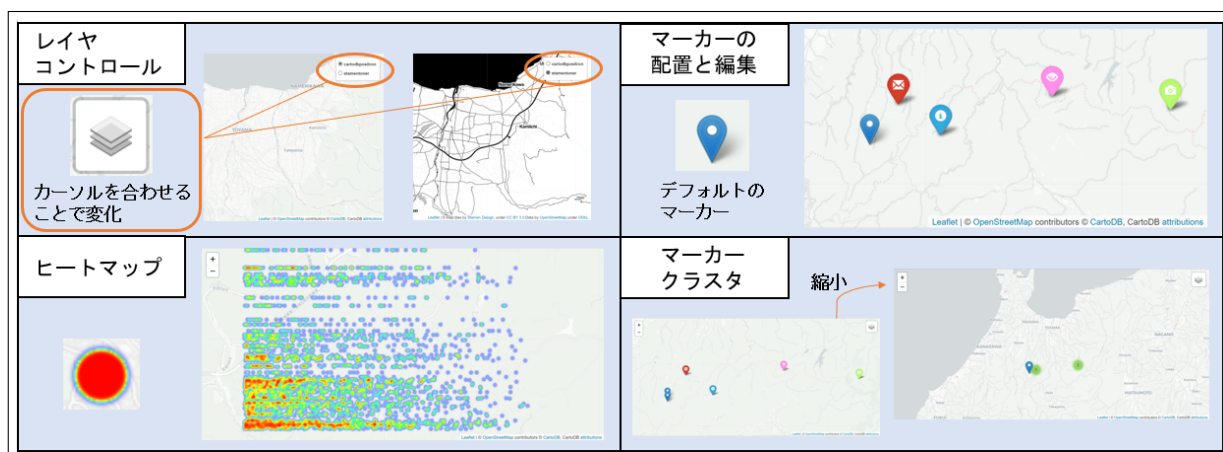


図 4.6: DEA のアウトプットの例

- 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重み
- 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重み

§ 4.3 Web-GIS を用いたデータフュージョンのシステム開発

本節では、4.1 節、4.2 節で述べたデータの選別・分析手法による結果に基づいて 3.3 節で解説した Folium による Web-GIS 作成を行うことで分析結果をユーザに提示する提案手法における結果のフィードバック部分について解説する。

また、これまでに 4 章の中で提示した技術のそれぞれがどの部分に組み込まれているかについて再度整理しながら、提案手法を組み込んだシステムの全体の流れを解説する。加えて、本研究の提案手法における WebGIS を用いたデータフュージョンの方法をシステム内の実際の仕様に言及しながら示す。

Web-GIS アプリケーションの作成

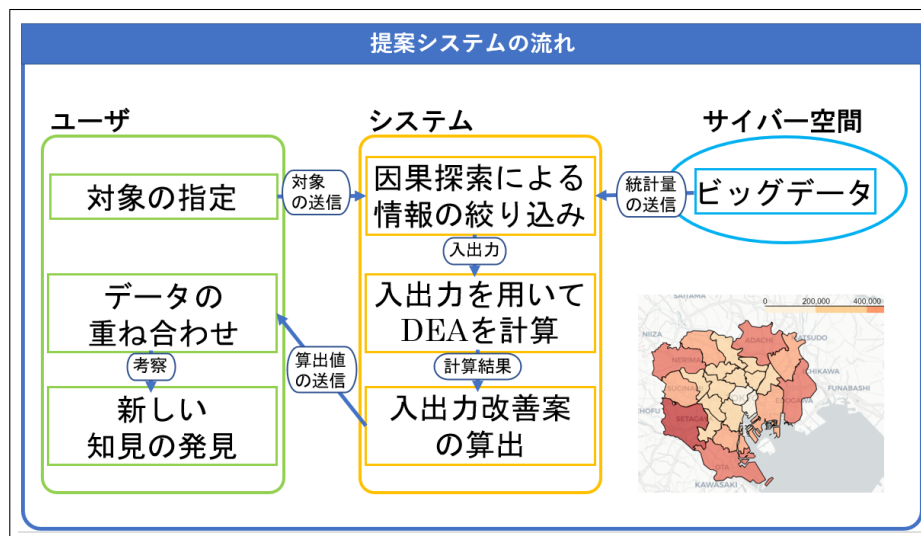


図 4.7: 改善の必要あり

数値実験並びに考察

§ 5.1 数値実験の概要

aa

§ 5.2 実験結果と考察

aa

おわりに

本研究では、市区町村単位の自治体における EBPM の支援を目的として、政策に関する意思決定のためのデータ収集および分析を行う手法を提案した。まず、不特定多数のオープンデータを用いることで、政策の対象によって収集するデータが偏ることを防ぎ、その中から LiNGAM によって対象と因果関係があるデータを選別した。

次に、それらのデータを因果関係の方向性に着目しながら二つのグループに分けることによって、DEA の入力・出力とした。そして、DEA の CCR モデルを用いて各自治体における評価値および入力・出力の改善値を算出することによって、対象の自治体が効率的に運営されているか、より効率的に運営されている自治体はないか、政策によって改善すべき分野はどれかなどを分析した。

最後に、それらの結果に基づいて Web-GIS を作成し、各自治体における評価値、入力・出力の改善値と対象の自治体に関しては、効率フロンティアに属する市区町村名とそれぞれに対する重みを地図上にプロットした。また、LiNGAM によって因果関係が示されたデータのうち、位置情報に紐づくものに関してはそれをプロットしたレイヤーを作成することで DEA による分析結果との重ね合わせを可能にし、データフュージョンによるさらなる知見の発見を可能にした。

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，ご助力いただいた島部達哉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研究室の同輩諸氏に感謝致します．

2022 年 2 月

長瀬 永遠

参考文献

- [1] “LTCM から得る教訓の重要性 - 銀行員のための教科書”, <https://www.financepensionrealestate.work/entry/2017/09/03/111854>, 閲覧日 2020.12.25.
- [2] 馬田哲次, “円安が経済に与える影響”, 山口経済学雑誌, Vol. 62, No. 1, pp. 1–13, 2013.
- [3] 片寄諒亮, 吉岡真治. “複数のテクニカル指標を用いた市場動向の予測”. 人工知能学会全国大会論文集 第 34 回全国大会 (2020), pp. 3Rin412–3Rin412. 一般社団法人 人工知能学会, 2020.
- [4] 長尾将宏, 長尾智晴ほか, “ファンダメンタル分析とテクニカル分析を組み合わせた投資戦略の提案”, 第 77 回全国大会講演論文集, Vol. 2015, No. 1, pp. 417–418, 2015.
- [5] “FX 初心者は必見!FX の始め方からトレードのやり方まで基礎を解説”, <https://www.randcins.jp/media/fx/beginner/>, 閲覧日 2020.12.25.
- [6] 徳岡聖二, 田中美栄子ほか, “進化計算による tick 価格変動のトレンド予測”, 情報処理学会論文誌数理モデル化と応用 (TOM), Vol. 48, No. SIG19 (TOM19), pp. 68–74, 2007.
- [7] 馬場則夫, 西田将人, 甲斐良隆, “ニューラルネットを活用した従来型テクニカル分析手法改善の試み”, 電気学会論文誌 C (電子・情報・システム部門誌), Vol. 126, No. 11, pp. 1324–1331, 2006.
- [8] “【FX 用語】バックテストとは?言葉の意味とその重要性について”, <https://fx-square.com/backtest/>, 閲覧日 2020.12.25.
- [9] “MT5 (メタトレーダー 5) のバックテストの方法 — OANDA FX/CFD Lab-education (オランダ ラボ)”, https://www.oanda.jp/lab-education/blog_mt5/auto_trading/mt5backtest/, 閲覧日 2020.12.25.
- [10] 平林明憲. “遺伝的アルゴリズムによる外国為替取引手法の最適化”. 人工知能学会全国大会論文集 第 22 回全国大会 (2008), pp. 282–282. 一般社団法人 人工知能学会, 2008.
- [11] 金井喜美雄, “ロバスト適応制御入門”. オーム社, 1989.
- [12] 木村英紀, “ロバスト制御”, 計測と制御, Vol. 22, No. 1, pp. 50–52, 1983.
- [13] 立林和夫, “実験計画法・タグチメソッドの活用”, 応用統計学, Vol. 42, No. 3, pp. 161–171, 2013.
- [14] 川岸裕之, 工藤一彦, “直交表による大域的最適解探索手法の開発”, 日本機械学会論文集 C 編, Vol. 72, No. 720, pp. 2344–2351, 2006.

- [15] 岩崎学, “統計的データ解析入門 実験計画法”. 東京図書株式会社, 2006.
- [16] 渕上美喜, 上田和明, 近藤宏, 高橋玲子, “Excel で学ぶ営業・企画・マーケティングのための実験計画法”. オーム社, 2006.
- [17] “Python 用 MetaTrader - 統合 - MQL5 リファレンス - MetaTrader 5 のためのアルゴリズムの/自動化されたトレーディング言語のリファレンス”, https://www.mql5.com/ja/docs/integration/python_metatrader5, 閲覧日 2020.12.25.
- [18] Abraham Charnes, William W Cooper, and Edwardo Rhodes, “Measuring the efficiency of decision making units”, *European journal of operational research*, Vol. 2, No. 6, pp. 429–444, 1978.
- [19] 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.
- [20] 杉山学, “データ包絡分析法による JR と大手私鉄の事業活動効率比較”, 群馬大学社会情報学部研究論集, Vol. 24, pp. 33–53, 2017.