

卒業論文

IP ランドスケープによる 経営戦略支援ための 共起語ネットワーク作成

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

富山県立大学 工学部 情報システム工学科

2020032 平井遥斗

指導教員 奥原 浩之 教授

提出年月: 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	1
§ 1.3 本論文の概要	2
第2章 知的財産戦略と特許情報	4
§ 2.1 知的財産戦略	4
§ 2.2 特許情報処理	6
§ 2.3 テキストマイニングと自然言語処理	7
第3章 特許情報の可視化	10
§ 3.1 特許情報のベクトル化	10
§ 3.2 次元削減手法とクラスタリング手法	12
§ 3.3 単語間のつながりと共起ネットワーク	13
第4章 提案手法	14
§ 4.1 Google Patents からの取得, 分類, 抽出	14
§ 4.2 トピックからの 3D グラフによる可視化	17
§ 4.3 IPL(Intellectual Property Landscape) への活用	20
第5章 数値実験並びに考察	24
§ 5.1 数値実験の概要	24
§ 5.2 実験結果と考察	26
第6章 おわりに	28
謝辞	29
参考文献	30

図一覧

2.1	IP ランドスケープの概要 [27]	6
4.1	csv のフォーマット	16
4.2	システムのフロントページ	16
4.3	因果探索の結果とアウトプットの例	17
4.4	DEA 部分のフロー	18
4.5	DEA のアウトプットの例	20
4.6	EBPM-GIS の実装	22
4.7	提案システム	22
5.1	射水市の推計人口 [23]	25
5.2	年少・老年人口割合推移 [23]	25
5.3	入力データと出力データ	27

表一覧

4.1	地理情報を持たない数値データ	15
4.2	地理情報を持つ数値データ	15
4.3	位置データ	15
4.4	各都道府県の DMU 内訳	18
5.1	「年少人口 [割合]」に対する Direct-LiNGAM の結果	26
5.2	DEA の入力・出力一覧	27
5.3	元データと改善案の比較	27
5.4	参照の市区町村とウェイト	27

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 i 番目の観測変数	x_i
LiNGAM における j 番目の観測変数から i 番目の観測変数へのパス係数	b_{ij}
LiNGAM における i 番目の観測変数に対する誤差 (非観測変数)	e_i
主問題における各入力に対する重み	v^T
主問題における各出力に対する重み	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 DMU _o の入力	x_o
CCR モデルにおける DMU _o の出力	y_o
CCR モデルにおける DMU の入力	X
CCR モデルにおける DMU の出力	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 i 番目の入力に対する改善案	$\hat{x}_i$
入力指向モデルにおける参照集合内の k 番目の DMU の i 番目の入力	x_{ik}
入力指向モデルにおける参照集合内の k 番目の DMU に対する重み	λ
出力指向モデルにおける対象 DMU の j 番目の出力に対する改善案	$\hat{y}_j$
出力指向モデルにおける参照集合内の k 番目の DMU の j 番目の出力	y_j
出力指向モデルにおける参照集合内の k 番目の DMU に対する重み	μ
提案手法における d 番目の市区町村の i 番目の入力	x_{id}
提案手法における d 番目の市区町村の i 番目の出力	y_{id}
提案手法における d 番目の市区町村に対する重み	λ_d
<i>robust Z-score</i> における正規化後の値	ι
<i>robust Z-score</i> を用いて正規化するデータ集合内の値	x
<i>robust Z-score</i> を用いて正規化するデータ集合	X
<i>robust Z-score</i> を用いて正規化するデータ集合の中央値	$median(x)$
<i>robust Z-score</i> を用いて正規化するデータ集合の正規四分位範囲	$NIQR$
0~1 変換の結果の値	ι'
0~1 変換を行うデータ集合内の値の最大値	$max \iota $

はじめに

§ 1.1 本研究の背景

近年、コロナウィルスの影響やグローバル化、インターネット技術や AI, IoT 等のデジタル技術の進展、顧客のニーズの多様化や社会環境などの急速な変化など、さまざまな要素が絡みあうことにより、将来を予測することが難しくなっている。急激な変化と不確実性が高まる社会に対応するためには、企業が保持しているコア技術を強化して差別化を行い、優位性を確立することが重要である。また多角的な視点から経営戦略を策定することが不可欠である [25]

特許庁の調査によれば、IP ランドスケープが必要であると回答した者は約 8 割であったのに対し、IP ランドスケープを十分に実施できていると回答したものは約 1 割であった [?]. 現在、必要性は理解しているがまだ実施に至っている企業が少ないという状態である。

§ 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）[5] 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）[?, 国土] 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．

知的財産戦略と特許情報

§ 2.1 知的財産戦略

知的財産戦略とは企業が保有する知的財産を経営戦略の一環として取り入れ、企業の競争力を高め、事業目標を達成することを目的とする戦略である。知的財産には、特許、商標、意匠、著作権、ノウハウなど、さまざまな種類があり、これらの知的財産をどのように活用すれば、企業の価値を最大化できるのかを考えることが重要である [26]。

知的財産戦略は、経営戦略と密接に関係しており、企業全体の戦略において各部門や機能の方向性を決定する重要な役割を果たしている。日本において、知的財産戦略は特許などの知的財産（Intellectual Property：IP）と景観や風景を意味する「Landscape」を組み合わせた造語で「IP ランドスケープ」と呼ばれることが多い。

知的財産戦略の目的は、以下の3つにまとめることができる。

企業の競争力を高める

知的財産は、企業の競争力を高める上で重要な役割を果たす。具体的には、競合他社との差別化、新製品・サービスの開発促進、収益向上とブランド力強化の3つの点で、企業の競争力を高める。

例えば、スマートフォンメーカーは、独自の技術やデザインを特許や商標で保護することで、競合他社との差別化を図り、消費者の購買意欲を高めている。また、自動車メーカーは、独自のエンジン技術や安全技術の特許で保護することで、新車の開発につなげ、他社に先駆けて市場シェアを拡大している。さらに、ソフトウェア会社は、自社製品の技術特許をライセンス提供することで、収益を向上させている。

このように、知的財産を戦略的に活用することで、企業は競争力を高め、企業の成長と発展を実現することができる。

事業目標を達成する

知的財産戦略は、企業が掲げる事業目標の達成にも貢献する。具体的には、新市場の開拓、新製品・サービスの開発、ブランド力強化の3つの点で、企業の事業目標の達成を後押しすることができる。

例えば、自動車メーカーが成長戦略を掲げている場合、新興国市場向けに独自の技術を開発することで、新市場の開拓を図り、新車の販売を拡大することができる。また、製薬会社が新薬の開発に特許を取得することで、他社に先駆けて新薬を市場に投入し、収益を拡大することができる。さらに、ファッションブランドが独自のデザインやロ

ゴを特許や商標で保護することで、ブランド価値を高め、消費者からの信頼を得ることができる。

このように、知的財産戦略を活用することで、企業は事業目標の達成を後押しし、企業の成長と発展を実現することができる。

企業価値を向上させる

知的財産は、企業の独自性や将来性を示すものであり、企業の価値を高める上で重要な要素である。企業は、自社の知的財産を適切に管理・活用することで、企業価値を高め、企業の成長と発展を実現することができる。

例えば、Apple の iPhone は、独自のデザインや操作性によって、他社製品と差別化を図り、高いブランド価値を築いている。また、製薬会社が新薬の特許を取得した場合、その新薬の市場規模や将来の収益性を評価することで、企業の将来性を評価することができる。

具体的には、知的財産を活用した新製品・サービスの開発、知的財産の譲渡・ライセンス、知的財産の活用による業務効率化などの方法で、企業価値を高めることができる。企業は、自社の事業目標や状況に応じて、これらの方法を組み合わせて、知的財産を活用して企業価値を高めることが重要である。

経営戦略

「経営戦略」とは、企業が競争環境の中で自らの経営目的・経営目標を達成するための方針や計画全般を意味する。どれほど巨大な企業であっても、保有する経営資源（ヒト、モノ、カネ）は有限だ。企業が掲げる目標や目的に応じて、選択し分配していく必要がある。そうした企業活動の基軸となる指針や指標、また方策を実現するための体制づくりなども「経営戦略」には含まれる。企業はすべてのリソースを有しているわけではない。当然ながら、強みや弱み、特性がある。それらを経営者は理解・把握した上で、組織改革や事業の方向性を決定していかなければならない。やるべきこと、やらなければいけないことは数多くある。それらにどのような優先順位を付けて実行していくかを明確に打ち出していくためにも、戦い方の根幹がどうしても必要になってくる。それが、「経営戦略」を策定する目的と言って良いだろう。

グローバル化の進展やIT・AIの普及、ニーズの多様化、競争環境の激化など、現代は変化のスピードがますます増している。こうした時代において、企業は10年、20年先の生き残りに向けてどのような成長シナリオを描いていくかが、一段と問われるようになってきている。そのためにも、経営者は自社の強みや特性を把握・理解し、組織改革や事業の方向性をスピーディ、かつダイナミックに決定していかなければならない。「経営戦略」の必要性が高まっているのもそのためである。

経営戦略のフレームワーク

経営戦略の策定では自社を取り巻く外部の環境要因に打ついて分析する外部環境分析や、自社内の環境を分析する内部環境分析を踏まえ自社の強みや弱み、機械や脅威を把握することで、戦略オプションを立案して最適な戦略を選択することが大切である。それらを行

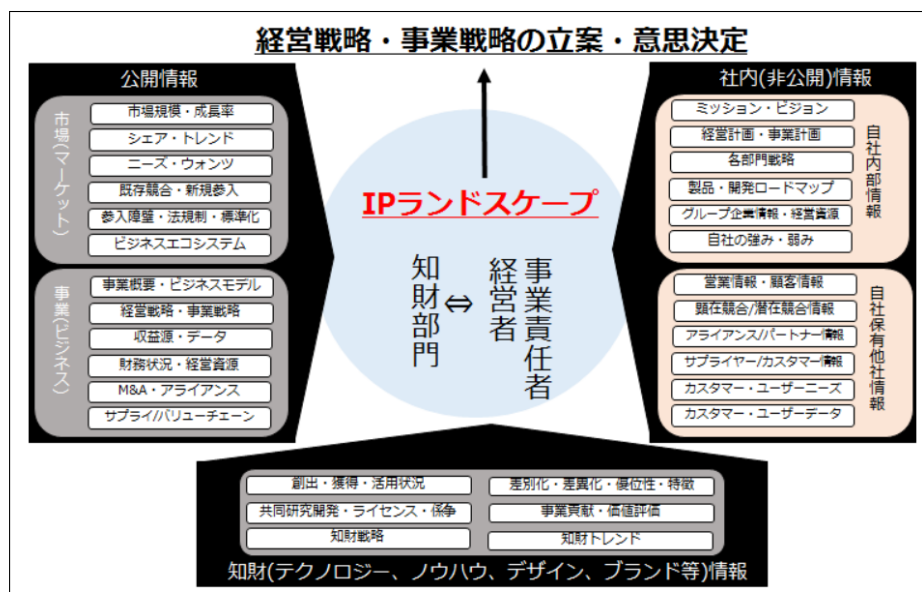


図 2.1: IP ランドスケープの概要 [27]

うために役立つ代表的なフレームワークとして、PEST 分析ファイブフォース分析、3C 分析、VRIO 分析、SWOT 分析、STP 分析、4P 分析などが挙げられる [28].

内閣府や特許庁による IP ランドスケープの積極的な推進に代表されるように、IP ランドスケープは研究機関においても積極的に検討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 を活用した取り組みが多数行われていることから、IP ランドスケープの効率的な実施には ICT の活用が不可欠であり、情報工学との親和性が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 IP ランドスケープの支援を目的とする。

§ 2.2 特許情報処理

特許情報とは、特許・実用新案・意匠・商標の出願や権利化に伴って生み出される情報である。この情報は、研究開発の重複防止、既存技術の活用、無用な紛争の回避などに役立つ。

特許情報は、研究開発の策定から商品化、更には他人の権利調査に至るまでの様々な事業活動において活用されている。

具体的な活用例は、以下のとおりである。[30].

特許情報の活用例

- 技術動向調査

将来性を見据えた研究テーマの選定や過去になされた研究との重複回避のために、特許情報を利用して技術動向調査が行われます。特定の技術分野における特許出願の動向や出願件数の推移を調査することにより、過去にどのような技術が存在したか、また、今後開発すべき技術分野の把握の参考になります。

- 出願前の先行技術調査

研究成果として発明がなされたとき、権利化するか否かの判断が必要となります。特許出願をする際に関連する分野の先行技術について調査することにより、権利として認められる見込みのない無駄な出願を未然に防止することができます。

- 権利調査

開発製品が他人の産業財産権を侵害すると、製造・販売の中止や製造品の廃棄、あるいは権利者への損害賠償にまで発展する恐れがあります。これらを未然に防止するために、設計から製造前段階にかけて、他人の権利範囲の調査を行います。

- 公知例調査

他の権利者から警告を受けた場合などの対抗手段として、自社の発明・考案を事業化する際に障害となる他人の特許権・実用新案権を無効にするため、その特許・実用新案登録の出願前の公知例を調査します。

- 公知例調査

事業を営む上で多くの場合には競合他社が存在しています。その競合他社がどのような戦略で事業を行っているか調査する上で、特許情報は貴重な情報源となります。競合他社の過去から現在に至るまでの出願動向を把握することにより、研究開発動向等を読み取ることが可能です。また、競合他社の出願動向を継続的に監視し、自社にとって障害となる出願等の早期発見に努めることも重要です。

§ 2.3 テキストマイニングと自然言語処理

テキストマイニングとは、定型化されていない文章から情報を抽出する技術です。SNSやアンケート、コールセンターの応対など、さまざまな場面で活用されています。テキストマイニングは、AI技術の進展により、より高度な分析が可能になりました。また、テキストデータの量も増加しており、テキストマイニングツールの種類も増えています。[31]

テキストマイニングを行うことで、企業は、顧客のニーズや市場動向を把握したり、新製品の開発やマーケティングの戦略を策定したりすることができる。

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。

形態素解析

形態素解析は、

分かち書き

自然言語処理（NLP）において、分かち書きは、テキストを単語や句などの意味的な単位に分割する処理である。分かち書きは、テキストの意味理解や解析の基礎となる重要な処理であり、多くのNLPタスクで必要となる。

分かち書きの目的は、テキストの意味を正確に理解するために、テキストを単語や句などの意味的な単位に分割することである。例えば、文の意味を理解するためには、文を主語、述語、目的語などの句に分割する必要がある。また、単語の意味を理解するためには、単語を品詞や語義などの単位に分割する必要がある。

分かち書きの処理方法は、大きく分けて以下の2つに分けられる [29]。

1. ルールベース型

ルールベース型分かち書きは、あらかじめ定義された対象となる言語の文法ルールに基づいて分かち書きを行う方法である。ルールベース型分かち書きは、人手でルールを定義するため、単純な分かち書きを行う場合は比較的容易に実装でき、調整も可能であるが、複雑な分かち書きを行う場合は、ルールを複雑にする必要があり、高度な専門知識が必要となり、誤りが生じやすくなる。

2. 統計学習ベース型

統計学習ベース型分かち書きは、機械学習によって導き出されたルールに基づいて分かち書きを行う方法である。統計学習ベース型分かち書きは、複雑な分かち書きを行う場合でも、比較的正確に分かち書きを行うことができる。また、機械学習に大量のテキストデータが必要であり、計算量が大きいという問題もコンピュータの高速化と低価格化により解決に向かっている。

分かち書きの精度は、分かち書きの目的や、分かち書きを行うテキストの種類によって異なる。例えば、新聞記事などのフォーマルなテキストであれば、ルールベース型分かち書きでも比較的高い精度で分かち書きを行うことができる。一方、SNSの投稿などの非フォーマルなテキストであれば、統計学習ベース分かち書きの方が高い精度で分かち書きを行うことができる。

近年、NLP技術の進展により、分かち書きの精度も向上している。また、クラウドサービスやオープンソースソフトウェアの普及により、分かち書きの利用が容易になってきている。

また、現状の分かち書きには、以下の課題がある。

● 日本語の曖昧さ

ルールベース型分かち書きは、あらかじめ定義されたルールに基づいて分かち書きを行う方法である。ルールベース型分かち書きは、単純日本語は、英語と比べて曖昧な表現が多い言語である。例えば、「私は、彼に会いました。」という文は、文法的には「私は、彼に会いに行きました。」という意味にも解釈できる。このような曖昧な表現を正確に分かち書きすることは、困難である。

● 新語や流行語

常に新しい言葉や表現が生まれてくるため、分かち書きのルールや統計モデルを常に更新する必要がある。

● 誤ったデータの影響

分かち書きの精度は、分かち書きの対象となるデータの品質に大きく影響を受ける。誤ったデータが含まれていると、分かち書きの精度が低下する。

今回用いる特許の本文には、専門的な用語や複合語が多数含まれているため、それらを正しく抽出する必要がある。そのため、python のモジュール termextract を用いて専門用語や複合語の抽出を行い、それらを分かち書きの辞書に登録する。

< *termextract* >

今後、テキストマイニングと自然言語処理は、AI や ML 技術を活用することで、より高度な分析が可能となり、より幅広い分野で活用されるようになって考えられる。また、テキストデータの量の増加に対応するため、テキストマイニングと自然言語処理の自動化や、テキストデータの検索・分析・活用を効率化する技術の開発が進んでいくと考えられる..

特許情報の可視化

§ 3.1 特許情報のベクトル化

特許情報は、日々蓄積され、今では莫大な量となっており、それらの分析は困難を極める。そこで、特許情報を効率的に分析するためには、各特許をベクトル化して整理をおこない、全体を俯瞰できるように可視化する必要があると考える。本研究では、特許本文の文章を対象にベクトル化を行う。特許本文には、特許技術の内容が詳細に記載されているため、これらの情報をベクトル化することで、特許の技術分野や技術トレンドなどを把握することができると思う。

具体的には、特許本文を Sentence-Bidirectional Encoder Representation from Transform: Sentence-BERT を用いることで文章全体を単位にベクトル化を行う。Sentence-BERT は、Bidirectional Encoder Representations from Transformers: BERT をベースに開発されており、文章の単語の順序を考慮して、文章の意味を表現するベクトルを生成する。

sentence-BERT は、文章の意味を理解する能力に優れているため、自然言語処理の様々なタスクに活用されている。

Transformer

近年、翻訳などの入力文章を別の文章で出力するというモデルは、Attention を用いたエンコーダー、デコーダ形式の RNN や CNN が主流であった。しかし、Transformer は、RNN や CNN を用いず Attention のみを用いたモデルである。Transformer は、再帰も畳み込みも一切行わないので並列化が容易であり、他のタスクにも汎用性が高いという特徴がある。Transformer においては Attention を多数並列に配置した Multi-Head Attention が用いられ、一般的に以下の式 (3.2) のように定式化される [32]。

<Multi-Head Attention>

$$MultiHeadAttention(Q, K, V) = Concat(head_1, head_2, \dots, head_h)W_o \quad (3.1)$$

$$where head_i = ScaledDotProductAttention(QW_i^Q, KW_i^K, VW_i^V) \quad (3.2)$$

ここで、Scaled Dot-Product Attention では、内積を利用したベクトル間の類似性に基づく変換を行う。

<Scaled Dot-Product Attention>

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (3.3)$$

3.2では、学習パラメータを持っていない Scaled Dot-Product Attention の表現力を広げるために、入力直前に学習パラメータを持つ Linear 層の追加を行っている。これにより、入力されるベクトルの特徴空間に依存しない注意表現を学習することができる。Linear 層の追加を行った Scaled Dot-Product Attention を一般に Single-Head Attention を呼ぶ。

Scale Dot-Product Attention は、ある単語に対して、その単語が文章に含まれる単語とどれだけ類似しているのかを計算し、それらを確率的に表現したものである。Transformer における Attention の入力には主に以下の 2 種類の入力方法が用いられている。

1. Self-Attention (softmax に与える Query, Key, Value を同じ値にする)
2. Source-Target-Attention (Key, Value を同じ値にし、Query を異なる値にする)

Single-Head Attention では多種多様な意味や文法をもつ単語に対しても単一の注意表現が生成される。そこで、Single-Head Attention を多数並列に配置して Multi-Head にすることで、複数の特徴部分空間における注意表現の獲得をすることができる。

以上のことから、文章を行列で表せることが分かった。しかし、文章というのは、文字を読む方向が重要であり、行列として表され、かつ、一括で処理する場合、文字の順番の概念がなくなってしまう。これは、文章を正しく扱えなくなる可能性がある。そのため、Embedding 層からの行列に位置情報を含んだ行列を足し合わせることで、文字の順番の概念を扱えるようにする必要がある。これを可能にするのが Positional Encoding である。

<Positional Encoding>

$$PE_{(pos, 2i)} = \sin \left(\frac{pos}{10000^{\frac{2i}{d_{model}}}} \right) \quad (3.4)$$

$$PE_{(pos, 2i+1)} = \cos \left(\frac{pos}{10000^{\frac{2i}{d_{model}}}} \right) \quad (3.5)$$

入力文章の単語数が 50 個まで扱えて、Embedding 層の埋め込み次元数が 128 次元の場合、Positional Encoding が生成する行列は 128 次元の行ベクトルが縦に 50 個並んだ行列になる。この行列は、各行のベクトルが絶対に同じものにならないため、この行列から単語の位置情報を表すことができる。具体的には、行ベクトルの各次元は、単語の位置情報に応じて、異なる値が割り当てられている。例えば、最初の行ベクトルの最初の次元は、単語の位置が 0 であることを示し、最後の行のベクトルは、単語の位置が 49 であることを示す。このように、Positional Encoding は、単語の位置情報を行ベクトルに埋め込むことで、Transformer モデルが単語の順序情報を利用できるようにしている。

BERT

自然言語処理タスクにおいて、精度向上には言語モデルによる事前学習が有効である。この言語モデルによる事前学習には「特徴量ベース」と「ファインチューニング」の 2 つの方法がある。まず、「特徴量ベース」とは事前学習で得られた表現ベクトルを特徴量の 1 つとして用いるもので、タスクごとにアーキテクチャを定義する。ELMo[Peters, (2018)] がこ

の例である。また、「ファインチューニング」は事前学習によって得られたパラメータを重みの初期値として学習させるもので、タスクごとにパラメータを変える必要があまりない。例として OpenAI GPT[Radford, (2018)] がある。ただし、いずれもある問題がある。それは事前学習に用いる言語モデルの方向が1方向だけということだ

[Sentence-BERT]

BERT では2つの文章を入力し、それらの類似度を測ることができる。しかし、複数の文章を入力する場合は BERT では容易ではない。そこで本研究では Sentence-BERT を用いる。BERT で求められた埋め込み表現を pooling し、それらを Softmax 関数を用いて、分類を行う。

$$O = \text{softmax}(W_t(u, v, \|u - v\|)) W_t \in R^{3n \times k} \quad (3.6)$$

事前学習モデルは Hugging Face や GitHub などのサイト公開されている。また東京大学や京都大学なども独自のモデルを公開している。本研究では Hugging Face に登録されている ”sonois/sentence-bert-base-ja-mean-tokens” を用いる。

§ 3.2 次元削減手法とクラスタリング手法

今回扱うデータは 768 次元と高次元であるためクラスタリングを行う際に次元の呪いが発生することが考えられるため、次元圧縮を行う。次元圧縮手法には Uniform Manifold Approximation and Projection of Dimension Reduction : UMAP を用いる。

UMAP

< 重み付き k 近傍 (高次元) >

$$v_{j|i} = \exp\left[\frac{-(r_{ij} - \rho_i)}{\sigma}\right] \quad (3.7)$$

$$\rho_i = \min_{j \in K} r_{ij} \quad (3.8)$$

< 対称化 >

$$v_{ij} = (v_{j|i} + v_{i|j}) - v_{j|i}v_{i|j} \quad (3.9)$$

< 重み付き k 近傍 (低次元) >

$$w_{ij} = \exp(-\max\{0, d_{ij} - \rho'\}) = \tilde{w}_{ij} \quad (3.10)$$

< 対称化 >

$$w_{ij} = \frac{1}{1 + a \cdot d_{ij}^{2 \cdot b}} \quad (3.11)$$

< トポロジカル表現の最適化 >

$$L = \sum [v_{ij} \log \frac{v_{ij}}{w_{ij}} + (1 - v_{ij}) \log \frac{1 - v_{ij}}{1 - w_{ij}}] \quad (3.12)$$

§ 3.3 単語間のつながりと共起ネットワーク

関連性の高い単語は、一緒に出現することが多いため、それらの単語の共起関係を調べることで、単語間の関係性を理解することができる。共起分析では単語同士の Jaccard 係数とい指標を用いて単語同士の共起度合いを比較し、共起関係にある単語と単語を線で結んで描かれる共起語ネットワークが利用される。このような共起語の分析を通じて、単語同士の意味的な特徴を理解することができる。本研究では、各クラスター内の単語にどのような関係があるのかを理解することを目的とする。

提案手法

§ 4.1 Google Patentsからの取得、分類、抽出

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??に示す。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

表 4.1: 地理情報を持たない数値データ

データ項目	単位	データ項目	単位
耕作放棄地率	%	経営耕地面積	1 畝 / 経営体
農業産出額	千万円	労働生産性	なし
企業数	社	従業員数	人
歳出決算額 [総務費]	%	農地平均取引価格	円 / m^2
歳出決算額 [民生費]	%	商業用地平均取引価格	円 / m^2
歳出決算額 [衛生費]	%	住宅用地平均取引価格	円 / m^2
歳出決算額 [農林水産業費]	%	林地平均取引価格	円 / m^2
歳出決算額 [商工費]	%	マンション等平均取引価格	円 / m^2
歳出決算額 [土木費]	%	1 人あたりの地方税	千円
歳出決算額 [警察費・消防費]	%	製造品出荷額	万円
歳出決算額 [教育費]	%	事業所数	事業所
歳出決算額 [公債費]	%	総人口	人
歳出決算額 [労務費]	%	老年人口	%
歳出決算額 [その他 (雑費)]	%	生産年齢	%
農業就業人口平均年齢	歳	年少人口	%
農業経営者平均年齢	歳	年間商品販売額	百万円
林作業請負収入	万円	海面漁獲物等販売額	万円
林産物販売金額	万円	付加価値額	万円
一人当たりの法人住民税	千円	1 人あたりの固定資産税	千円

表 4.2: 地理情報を持つ数値データ

データ項目	単位
施設数 [空港]	箇所
施設数 [工業団地]	箇所
施設数 [都市公園]	箇所
施設数 [道の駅]	箇所
施設数 [学校]	箇所

表 4.3: 位置データ

データ項目	単位
施設位置 [空港]	経度・緯度
施設位置 [工業団地]	経度・緯度
施設位置 [都市公園]	経度・緯度
施設位置 [道の駅]	経度・緯度
施設位置 [学校]	経度・緯度

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

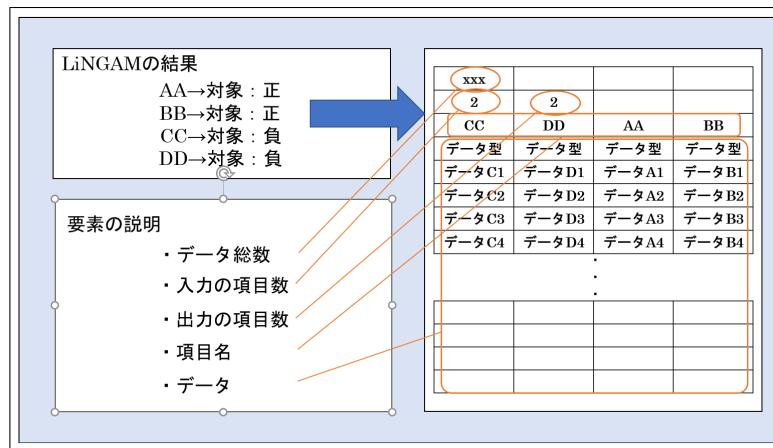


図 4.3: 因果探索の結果とアウトプットの例

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、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 および 4.2 に示した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 4.3 に示す。

§ 4.2 トピックからの 3D グラフによる可視化

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

表 4.4: 各都道府県の DMU 内訳

都道府県	市	区	町	村	都道府県	市	区	町	村	都道府県	市	区	町	村
北海道	35	0	130	20	石川県	11	0	8	0	岡山県	14	4	10	2
青森県	10	0	22	8	福井県	9	0	8	0	広島県	13	8	0	9
岩手県	14	0	15	4	山梨県	13	0	8	6	山口県	13	0	0	6
宮城県	13	5	18	1	長野県	19	0	23	35	徳島県	8	0	15	1
秋田県	13	0	9	3	岐阜県	21	0	19	2	香川県	8	0	9	0
山形県	13	0	19	3	静岡県	21	10	12	0	愛媛県	11	0	9	0
福島県	13	0	31	15	愛知県	37	16	14	2	高知県	11	0	17	6
茨城県	32	0	10	2	三重県	14	0	15	0	福岡県	27	14	29	2
栃木県	14	0	11	0	滋賀県	13	0	6	0	佐賀県	10	0	10	0
群馬県	12	0	15	8	京都府	14	11	10	1	長崎県	13	0	8	0
埼玉県	39	10	22	1	大阪府	31	31	9	1	熊本県	13	5	23	8
千葉県	36	6	16	1	兵庫県	28	9	12	0	大分県	14	0	3	1
東京都	26	23	5	8	奈良県	12	0	15	1	宮崎県	9	0	14	3
神奈川県	16	28	13	1	和歌山県	9	0	20	1	鹿児島県	19	0	20	4
新潟県	19	8	6	4	鳥取県	4	0	14	1	沖縄県	11	0	11	19
富山県	10	0	4	1	島根県	8	0	10	1	合計	733	188	727	203

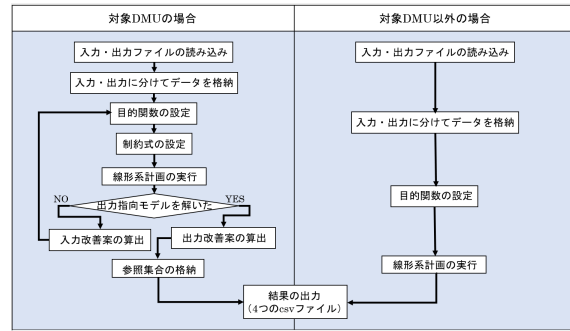


図 4.4: DEA 部分のフロー

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 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（??）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の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 773 個の市、188 個の区、727 個の町、203 個の村の合計 1891 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 4.4 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 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 >

$$\text{minimize} \quad \theta \quad (4.3)$$

$$\text{subject to} \quad \sum_{d=1} y_{id} \lambda_d \geq y_{io} \quad i = 1, 2, \dots, m \quad (4.4)$$

$$- \sum_{d=1} x_{id} \lambda_d + x_{io} \theta \geq 0 \quad i = 1, 2, \dots, m \quad (4.5)$$

$$\lambda \geq 0 \quad (4.6)$$

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．逆変換の式は以下ようになる．

< 正規化の逆変換 >

$$\iota = \iota' \times 2 \max |\iota| - \max |\iota| \quad (4.7)$$

< robust Z-score の逆変換 >

$$x = \iota \times NIQR + median(x) \quad (4.8)$$

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の扱う 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 4.4 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，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．

評価値	入力・出力改善案	参照集合・重み（入力）	参照集合・重み（出力）																																																																		
<table><tr><td colspan="2">cuty_codeeffic_value</td></tr><tr><td>1100</td><td>0.55679</td></tr><tr><td>1202</td><td>0.70558</td></tr><tr><td>1203</td><td>0.83029</td></tr><tr><td>⋮</td><td></td></tr><tr><td>⋮</td><td></td></tr><tr><td>⋮</td><td></td></tr><tr><td>47381</td><td>0.30172</td></tr><tr><td>47382</td><td>0.35638</td></tr></table>	cuty_codeeffic_value		1100	0.55679	1202	0.70558	1203	0.83029	⋮		⋮		⋮		47381	0.30172	47382	0.35638	<table><tr><td>imp_item</td><td>imp_value</td></tr><tr><td>AAA</td><td>0</td></tr><tr><td>BBB</td><td>13.244</td></tr><tr><td>CCC</td><td>0.86798</td></tr><tr><td>DDD</td><td>5.68598</td></tr><tr><td>EEE</td><td>3.69084</td></tr><tr><td>FFF</td><td>26.6893</td></tr><tr><td>GGG</td><td>6.20473</td></tr><tr><td>HHH</td><td>42026.3</td></tr></table>	imp_item	imp_value	AAA	0	BBB	13.244	CCC	0.86798	DDD	5.68598	EEE	3.69084	FFF	26.6893	GGG	6.20473	HHH	42026.3	<table><tr><td colspan="2"></td></tr><tr><td>city_code</td><td>weight_in</td></tr><tr><td>24303</td><td>0.0611</td></tr><tr><td>33663</td><td>0.17123</td></tr><tr><td>34203</td><td>0.01215</td></tr><tr><td>39424</td><td>0.02395</td></tr><tr><td>44462</td><td>0.32172</td></tr></table>			city_code	weight_in	24303	0.0611	33663	0.17123	34203	0.01215	39424	0.02395	44462	0.32172	<table><tr><td colspan="2"></td></tr><tr><td>city_code</td><td>weight_out</td></tr><tr><td>24303</td><td>0.13164</td></tr><tr><td>33663</td><td>0.3689</td></tr><tr><td>34203</td><td>0.02617</td></tr><tr><td>39424</td><td>0.05159</td></tr><tr><td>44462</td><td>0.69311</td></tr><tr><td colspan="2"></td></tr></table>			city_code	weight_out	24303	0.13164	33663	0.3689	34203	0.02617	39424	0.05159	44462	0.69311		
cuty_codeeffic_value																																																																					
1100	0.55679																																																																				
1202	0.70558																																																																				
1203	0.83029																																																																				
⋮																																																																					
⋮																																																																					
⋮																																																																					
47381	0.30172																																																																				
47382	0.35638																																																																				
imp_item	imp_value																																																																				
AAA	0																																																																				
BBB	13.244																																																																				
CCC	0.86798																																																																				
DDD	5.68598																																																																				
EEE	3.69084																																																																				
FFF	26.6893																																																																				
GGG	6.20473																																																																				
HHH	42026.3																																																																				
city_code	weight_in																																																																				
24303	0.0611																																																																				
33663	0.17123																																																																				
34203	0.01215																																																																				
39424	0.02395																																																																				
44462	0.32172																																																																				
city_code	weight_out																																																																				
24303	0.13164																																																																				
33663	0.3689																																																																				
34203	0.02617																																																																				
39424	0.05159																																																																				
44462	0.69311																																																																				

図 4.5: DEA のアウトプットの例

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 4.5 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。

§ 4.3 IPL(Intellectual Property Landscape) への活用

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 4.5 に示した 4 つ csv 内のデータをもとに Web-GIS による結果のフィードバックを行うアプリケーションである EBPM-GIS を作成した。各 csv ファイルが EBPM-GIS のどの部分の実装に使用されているかを図 4.6 に示す。

EBPM-GIS におけるベースマップは cartodbpositron および地理院タイルの 2 種類であり、cartodbpositron は白色でマーカーが視認しやすいという理由、地理院タイルは等高線や色分けによって地形が表されており、土地の地理的状況が視覚的に理解できるという理由で採用した。

各市区町村における評価値は矢印のアイコンを用いたマーカーで表示し、矢印の向きと色は三種類に分けて実装した。基準は評価値の大きさで、0.33 未満を青の下向き、0.33 以上 0.66 未満を黄の横向き、0.66 以上を赤の上向きとした。また、対象の市区町村および参照集合の市区町村に関してはアイコンの色も変更し、対象を赤、参照集合を青とした。

EBPM-GIS では、データ分析に用いた市区町村すべてに対してマーカーを立てるため、非常に多くのマーカーが画面上に表示される。そのため、視認性および処理速度の向上を目的として前述の矢印の種類ごとに別のレイヤに分けて実装した。各レイヤの切り替えは画面右上のレイヤコントロールで行うことができる。初期画面では全ての矢印が表示された状態となっている。

EBPM-GIS 上でマーカーをクリックするとポップアップが出現し、DEA における結果がテキストで表示される。対象の市区町村に関しては市区町村名、評価値、入力・出力の改善案、参照集合とその重みがポップアップ上に表示される。その他の市区町村においては市区町村名および評価値の表示のみである。

また、因果探索によって対象のデータと地理情報データの間に関係が示された場合、全国に分布するその施設の位置全てをマーカークラスタによって表示する。このマーカークラスタも評価値のマーカーと同様にベースマップから切り離された別のレイヤ上にプロットされており、レイヤコントロールを用いて表示・非表示の切り替えが可能である。

提案手法全体の流れ

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 4.7 に示す。

Step 1: 政策の対象決定・データ選択

フロントページにてサーバ上に存在する統計データ全ての項目名を公開し、全国の市区町村に対する項目ごとのデータを参照できるようにする。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。4.1 節の前半にて解説した部分がこれにあたる。

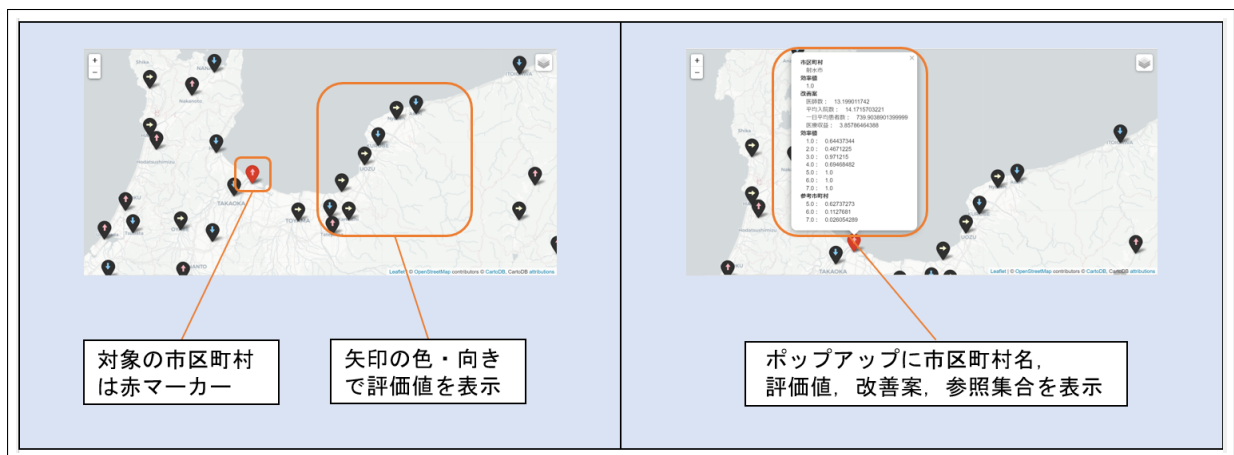


図 4.6: EBPM-GIS の実装

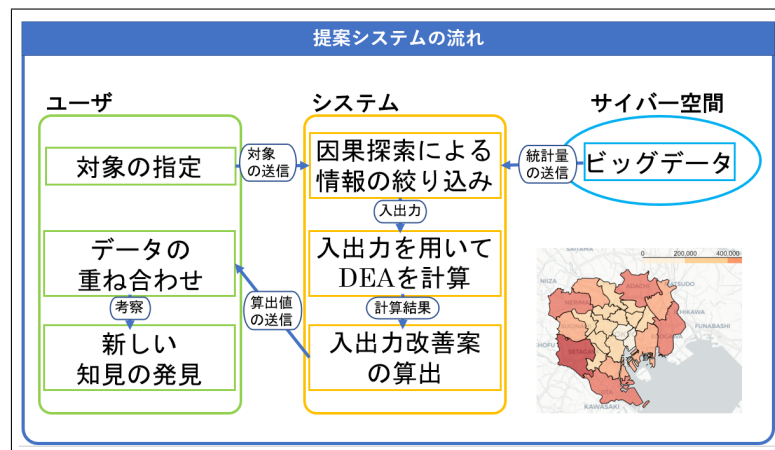


図 4.7: 提案システム

Step 2: 統計データの選別・入出力への振り分け

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。4.1 節の後半にて解説した部分がこれにあたる。

Step 3: DEA による評価値と入力・出力改善案の算出

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 (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その重み（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）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。4.2 節にて解説した部分がこれにあたる。

Step 4: EBPM-GIS の作成・データの重ね合わせ

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。4.3 節の前半にて解説した部分がこれにあたる。

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 EBPM-GIS が表示されるという画面遷移だけが提示される。

また、ユーザは出力装置に提示される EBPM-GIS を通して自身が対象としたいデータ項目に因果関係のあるデータにはどのようなものがあるか、それらを用いて DEA を行った結果の評価値や入出力の改善案、改善のために参考とすべき市区町村とその重みはどのようなものであるか、対象としたいデータと因果関係を持つ施設はどのように分布しているかを受け取ることができる。

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。

数値実験並びに考察

§ 5.1 数値実験の概要

本研究における提案手法は、ユーザサイドでシステムを見た場合、フロントページで対象のデータおよび市区町村を指定すると DEA の結果や地理情報が含まれた EBPM-GIS が表示されるという形になるが、サーバサイドでは 4.3 節で述べたように 4 つの段階を踏み、その各段階でデータの編集や分析を行うことによってユーザに分析結果を提示するというしくみになっている。

5 章では、本研究における提案手法の有効性を示すために現在とある市区町村が抱えると思われる問題を例として取り上げ、その問題に対して提案手法を適用する。そのうえで 4 つの段階それぞれにおいてどのようなデータの編集・分析が行われたのかを各段階の操作後に排出される結果を用いて示し、各結果に対する考察を行う。最後に、それらの結果および最終的な EBPM-GIS の表示内容をもとに取り上げた問題に対する解決策を考察することによって提案手法の有効性の検証とする。

取り上げるモデルケース

本実験では、問題を抱える地方自治体のモデルケースとして、富山県立大学工学キャンパスの所在である富山県射水市を取り上げる。当市は富山県の沿岸部中央付近に位置し、土地面積 109.43km^2 、人口約 90,000 人を有する市である。

また、現在の市域に至ったのは 2005 年 11 月で当時の新湊市、小杉町、大門町、大島町、下村の 1 市 3 町 1 村が合併することで射水市となった。そのため、富山県に占める土地面積は比較的広大で前述の人口は富山市、高岡市に次ぐ県内 3 位の値となっている。

以上のような特徴を持つ当市であるが、当市が抱える問題として最も深刻なものに人口減少問題が挙げられる [22]。当市の人口は 2005 年にピークを迎えて以来、年々減少しており、特に老年人口の増加に反比例して若年人口が減少することで少子高齢化に歯止めが利かない状況となっている。20 歳から 24 歳までの若い世代の人口が他市町村へ流出、とりわけ女性の転出の増加が原因の 1 つと考えられる。2010 年の国勢調査に基づく当市の推計人口の推移を図 5.1、年少（15 歳未満）・老年（65 歳以上）人口割合の推移を図 5.2 に示す。

国立社会保障・人口問題研究所が作成した当市における 2025 年の推計人口は 86,535 人であり、統計上最多である 2005 年の 94,209 人から約 10% 減少した値となっている。また、1995 年から 2010 年にかけて老年人口割合が約 8% 増加し、年少人口が約 3% 減少していることから前述の少子高齢化が統計からも読み取れる。

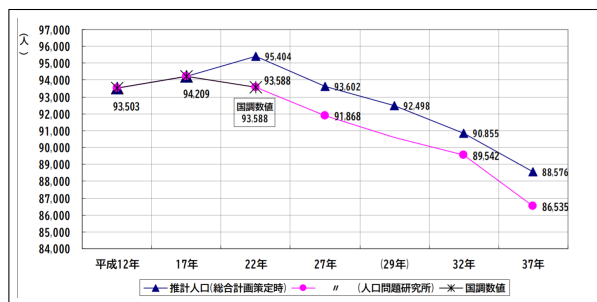


図 5.1: 射水市の推計人口 [23]

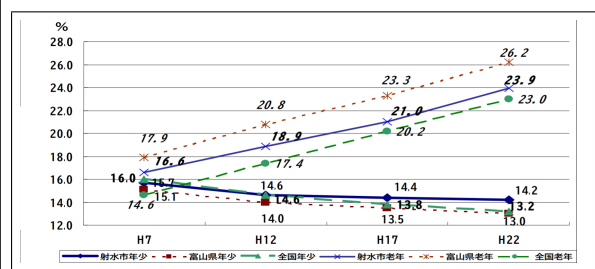


図 5.2: 年少・老年人口割合推移 [23]

以上の問題のうち、少子化に対して2005年に「射水市次世代育成支援行動計画」を策定、2007年には「射水市子ども条例」を制定した。また、これらの理念に基づき、子ども医療費助成の中学校3年生までの拡大や第3子以降保育料の無料化、様々なニーズに応じた保育サービスの実施、放課後児童クラブの拡充といった福祉、保健サービスを始め、子どもの学び・育ちを支えるための学校教育の充実、子どもに関する相談体制の整備などの取り組みを行った。

これらの取り組みは2005年から2022年現在も継続して行われており2021年末における人口が91,783人と発表されていることから推計と比較して人口減少のペースは緩やかになっていると考えられる。ただ、いまだ人口が減少傾向であることは変わらず、2020年3月に策定され、2021年3月に改訂された「第2期射水市まち・ひと・しごと創生総合戦略」[]においても戦略全体の方向性として人口減少に対する対策が言及されている。このことから、本研究における提案手法の有効性評価として、「今後、射水市の人口減少を食い止めるためにはこういった政策を行うべきか」を対象とした数値実験を行うこととする。

数値実験のながれ

前述の問題に対してデータベース内のデータ項目のうち、最も関係があると考えられる「年少人口[割合]」をターゲットとして提案手法による分析を行う。また、対象とする市区町村には射水市を選択する。

はじめに、これらの情報をフロントページを用いてシステムに送信し、「年少人口[割合]」をターゲット、観測変数をデータベース上に存在する全てのデータとしてDirect-LiNGAMによる因果探索を行う。因果探索の結果は図4.3のようなcsv形式で示す。

次に、そのcsvに格納されているデータを入力・出力とし、対象DMUを射水市、DMUの集合をデータベース上に存在する全ての市区町村とするDEAを行う。分析結果については評価値、入力・出力の改善案、入力指向モデルおよび出力指向モデルにおける参照集合に属する市区町村名とそれらに対する重みを項目ごとに分けて4つのcsvで示す。

最後に、それらのcsvファイルをもとに作成されたEBPM-GISにおける射水市のポップアップおよび参照集合に属する市区町村、射水市周辺の市区町村のポップアップを示し、それらの結果を比較する。また、射水市と前述の市区町村の政策、特に因果探索によって因果関係が示されたデータに関するものを比べることで射水市が年少人口を増大させ、少子高齢化を食い止めていくためには今後どのような政策に取り組んでいくべきなのかを考察することとする。

表 5.1: 「年少人口 [割合]」に対する Direct-LiNGAM の結果

データ項目	パス係数	データ項目	パス係数
施設数 [空港]	0.059	企業数	-0.006
衛生費	-0.019	商工費	-0.024
警察・消防費	-0.038	教育費	0.017
住宅用地平均取引価格	-0.043	生産年齢人口	0.249
老年人口	-0.559		

§ 5.2 実験結果と考察

5.1 節で述べた通り、富山県射水市が抱える少子高齢化問題に対して今後どのような政策に取り組んでいけばよいかを模索するというモデルケースを想定して本研究の提案手法を用いた意思決定の支援を行うという数値実験を行った。その結果を 4.3 節で述べた 4 つの Step のうち、Step4 から Step4 における実行結果に基づいて示す。

統計データの選別・入出力への振り分け

因果探索の部分では、少子高齢化問題を解決するにあたってその値の増加が必要であると考えられるデータ項目である「年少人口 [割合]」をターゲットとしてデータベース内のデータ間における因果関係を Direct-LiNGAM を用いてもとめる。「年少人口 [割合]」とは、各市区町村における人口全体のうち、年少（15 歳未満）の人口が占める割合のことである。

Direct-LiNGAM によってもとめられたパス係数行列の中で「年少人口 [割合]」に対するパス係数が 0 ではないものの項目名とそのパス係数を表 5.1 に示す。また、結果に基づく DEA の入力・出力の項目を表 5.2 に示す。

入力・出力の項目数はパス係数の大きいものから老年人口、住宅用地平均取引価格、警察費・消防費を含む 6 つの入力と生産年齢人口、施設数 [空港]、教育費の 3 つの出力となった。本実験で用いたデータベースには老年人口、生産年齢人口、年少人口と 3 つの区分における人口割合が含まれる。これらのデータは割合のためお互いに作用しあうということから老年人口、生産年齢人口から因果関係が示されたのは妥当と考えることができる。

そのうえで、老年人口が入力、生産年齢人口が出力に振り分けられたのは年少人口の親世代が生産年齢人口にあたるためと考えられる。また、生産年齢人口を増やそうと考えた際に住宅用の土地の取引価格が低いほうがよいというのは容易に想像できる。教育費が出力に振り分けられたのも同様の理由が考えられ、安心して育児・教育に取り組める環境がある場合に年少人口が増加すると考えられる。

逆に空港の施設数や警察費・消防費から因果関係が示されたこと、企業数が入力に振り分けられたことなどは直接的な要因が想像しづらく、こういった項目が分析によって導き出されることで新たな発見につながるこれが本研究における有意性の 1 つといえる。

DEA による評価値と入力・出力改善案の算出

DEA の部分では、前述の因果探索における結果を入力・出力として DEA を行い、射水市における評価値、入力・出力改善案、参照集合に属する市区町村名とそれらにかかるウェ

住宅用土地平均取引価格 警察費・消防費 商工費 衛生費 企業数 老年人口	生産年齢人口 教育費 施設数 [空港]
---	---------------------------

15669									
6		3							
企業數	衛生費	商工費	警察費	土地	住老人口	施設数	教育費	生産年齢人口	／ 2015年
float	float	float	float	float	float	float	float	float	
0.682354	0.528872	0.580341	0.546548	0.523329	0.547238	0.75	0.586495	0.560284	
0.533241	0.513759	0.591818	0.560548	0.506019	0.507994	0.75	0.514985	0.507092	
0.512979	0.517407	0.53906	0.504677	0.501973	0.542878	0.5	0.574924	0.525709	
0.536653	0.521368	0.536283	0.53118	0.501208	0.500727	0.5	0.543816	0.500887	
0.508145	0.534188	0.517401	0.507127	0.500751	0.520349	0.5	0.515554	0.509752	
0.518957	0.509902	0.583488	0.510022	0.501485	0.507267	0.75	0.536798	0.516844	
0.520833	0.525976	0.601999	0.500445	0.504398	0.53561	0.75	0.548179	0.534344	
0.511683	0.513863	0.5087	0.515159	0.501375	0.508721	0.5	0.535281	0.516844	
0.502994	0.528455	0.533876	0.513586	0.504385	0.625727	0.5	0.648141	0.594858	
0.505912	0.501772	0.554054	0.523163	0.502149	0.509448	0.5	0.507398	0.501773	
0.501943	0.517094	0.600518	0.520713	0.501319	0.528343	0.5	0.530349	0.537234	
0.500091	0.571503	0.511477	0.527394	0.504385	0.518895	0.5	0.555008	0.506206	

企業数	3075	-27769.854 (社)
衛生費	10.08	-2.954 (%)
商工費	2.92	-1.647 (%)
警察費・消防費	2.88	-1.470 (%)
住宅土地平均取引価格	19507	-368560.385 (円)
老年人口	28.6	22.821 (%)
施設数 [空港]	0	0.143 (箇所)
教育費	12.5	32.502 (%)
生産年齢人口	57.5	66.692 (%)

参照市区町村	ウェイト	参照市区町村	ウェイト
山形県東置賜郡川西町	0.268	山形県東置賜郡川西町	0.315
長野県北佐久郡御代田町	0.197	長野県北佐久郡御代田町	0.231
滋賀県米原市	0.108	滋賀県米原市	0.127
滋賀県蒲生郡日野町	0.158	滋賀県蒲生郡日野町	0.186

また、参照集合には山形県東置賜郡川西町、長野県北佐久郡御代田町、滋賀県米原市、同県蒲生郡日野町が含まれており、川西町がとりわけ高くなっている。これらのことから、川西町をはじめとした参照集合に属する市町村での取り組みなどを参考にしつつ、目的別の歳出額の調整や住宅用の土地取引に関する補助金などのしくみを作っていくことが年少人口の増加につながると考察できる。

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、因果関係が示されたデータのうち、対象のデータに対してパスを持つデータをパスの正負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・出力の改善案を算出することによって、対象の自治体における現状を評価し、現状以上に優れた運営を行うためには入力・出力をどのように変化させればよいかを分析した。また、参照集合に属する市区町村とそれらにかかるウェイトを提示することで入力・出力を改善するうえで参考とすべき自治体を示した。

最後に、それらの結果に基づいてデータのフィードバックを行う EBPM-GIS を作成し、各自治体における評価値と対象の自治体に関しては、入力・出力の改善値、参照集合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マーカークラスターを用いて位置情報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。

今後の課題としては、データベース内のデータの拡充、因果探索および DEA におけるモデルや分析手法の深化などが挙げられる。本研究で取り上げた課題である政策決定における問題の複雑性の性質上、分析に用いるデータの集合は多種多様かつ多量であるほど課題解決に有意義な結果が得られることが考えられる。そのため、実際の政策決定で用いる際には本研究で扱ったデータベースからデータ量を大幅に増加させる必要があると考える。

また、本研究で用いた分析手法は因果探索や DEA における基本的なモデルであるため、多様な問題やデータに対応するためにはそれらによりフィットした分析手法やモデルについての議論が必要であると考えられる。特に、因果探索においてはどの程度のパス係数までを有意として因果関係を認めるかなど今後発展の余地が残されている。

今後の展望としては、本研究で扱ったような自治体における政策決定への適用だけでなく、社会活動を行う上で頻繁に遭遇する対象が複雑な問題やデータ分析に際してデータ群の全容が確定していない問題に対して本研究における提案手法を応用していくことも考えられる。

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電子・情報工学科３年生の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] 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.
- [2] 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.
- [3] 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.
- [4] Shohei Shimizu, Takanori Inazumi, Yasuhiro 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.
- [5] 末吉俊幸, "DEA-経営効率分析法-", 朝倉書店, 2001.
- [6] 国土交通省国土地理院, "GIS とは", 閲覧日 2022-02-08, <https://www.gsi.go.jp/GIS/whatisgis.html>.
- [7] 佐藤主光, "税財政分野における EBPM の基礎と活用", 閲覧日 2022-02-08, https://www.ipp.hit-u.ac.jp/satom/lecture/localfinance/2019_local_note07.
- [8] 内閣府, "内閣府における EBPM への取組", 閲覧日 2022-02-08, <https://www.cao.go.jp/others/kichou/ebpm/ebpm.html>.
- [9] esri ジャパン, "GIS (地理情報システム) とは", 閲覧日 2022-02-08, <https://www.esri.com/getting-started/what-is-gis/>.
- [10] 国土交通省国土地理院, "基盤地図情報の利活用事例集", 閲覧日 2022-02-08, <https://www.gsi.go.jp/common/000062939>.
- [11] esri ジャパン, "東日本大震災対応における政策形成支援に GIS を活用", 閲覧日 2022-02-08, <https://www.esri.com/industries/case-studies/35859/>.
- [12] 田中貴宏, 佐土原聡, "都市化ポテンシャルマップと二次草原潜在生育地マップの重ね合わせによる二次草原消失の危険性の評価: 一福島県旧原町市域を対象として", 環境情報科学論文集, Vol. 23, pp. 191-196, 2009.
- [13] 坪井利樹, 西田佳史, 持丸正明, 河内まき子, 山中龍宏, 溝口博, "身体地図情報システム", 日本知能情報フレンジー学会誌, Vol. 20, No. 2, pp. 155-163, 2008.
- [14] 杉原豪, 塚井誠人, "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", 土木学会論文集 D3 (土木計画学), Vol. 75, no.6, pp. 583-589, 2020.
- [15] Dentsu Digital Tech Blog, "Google Colab で統計的因果探索手法 LiNGAM を動かしてみた", 閲覧日 2022-02-08, https://note.com/dd_techblog/n/nc8302f55c775.

- [16] 藤井秀幸, 傅靖, 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.
- [17] 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.
- [18] 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.
- [19] 日本オペレーション・リサーチ, ”第 4 章 包絡分析-入力と出力と”, 閲覧日 2022-02-08, <http://www2.econ.tohoku.ac.jp/ksuzuki/teaching/2006/ch4>.
- [20] pork_steak, ”folium 事始め”, 閲覧日 2022-02-08, https://qiita.com/pork_steak/items/f551fa09794831100faa.
- [21] 保母敏行ほか, ”日本分析学会における標準物質の開発”, 日本分析化学会誌, vol. 57, No. 6, pp. 363-392, 2008.
- [22] 射水市役所, ”総合戦略-射水市”, 閲覧日 2022-02-08, <https://www.city.imizu.toyama.jp/appupload/EDIT/054/054185>.
- [23] 射水市役所, ”共通課題-射水市”, 閲覧日 2022-02-08, <https://www.city.imizu.toyama.jp/appupload/EDIT/024/024383>.
- [24] 特許庁, ”広報誌「とつきよ」”, 閲覧日 2024-02-04, <https://www.jpo.go.jp/news/koho/kohoshi/>.
- [25] 特許庁, ” “ 経営戦略に資する知財情報分析・活用に関する調査報告書”, 閲覧日 2024-02-04, <https://www.jpo.go.jp/support/general/document/chizaijobobunseki-report/chizai-jobobunseki-report.pdf>.
- [26] 東京知的財産総合センター, ”中小企業経営者のための知的財産戦略マニュアル”, 閲覧日 2024-02-04, https://www.tokyo-kosha.or.jp/chizai/manual/senryaku/rmepal000001vpyy-att/senryaku_all.vol.9.pdf.
- [27] 特許庁, ”「経営戦略に資する知財情報分析・活用に関する調査研究」について”, 閲覧日 2024-02-04, <https://www.jpo.go.jp/support/general/chizai-jobobunseki-report.html>.
- [28] 金融ナビ, ”経営戦略の策定に役立つフレームワーク 7 つ | 経営戦略の代表例も解説”, 閲覧日 2024-02-04, https://financenavi.jp/basic-knowledge/management_strategy_framework/#tag1.
- [29] gikyo.jp, ”Perl による自然言語処理入門”, 閲覧日 2024-02-04, <https://gihyo.jp/dev/serial/01/perl-hackers-hub/0031011>.

- [30] 特許庁, ”2019 年度_知的財産権制度入門”, 閲覧日 2024-02-04,
https://www.jpo.go.jp/news/shinchaku/event/seminer/text/document/2019_syosinsya/1_3.pdf.
- [31] 株式会社 日立ソリューションズ・クリエイト, ”テキストマイニングとは？ 手法や活
用法を解説”, 閲覧日 2024-02-04,
<https://www.hitachi-solutions-create.co.jp/column/technology/text-mining.html>.
- [32] AGIRobots Blog, ” 【Transformer の基礎】 Multi-Head Attention の仕組み”, 閲覧日
2024-02-04,
<https://developers.agirobots.com/jp/multi-head-attention/>.

