

卒業論文

証拠に基づく政策立案のための オープンデータを利活用した 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	1
§ 1.3 本論文の概要	2
第2章 EBPMとICTの有用性	3
§ 2.1 ICTとデータの利活用による行政施策	3
§ 2.2 GISによる地域的な政策決定支援	4
§ 2.3 WebGISによるデータフュージョン	5
第3章 データ間の因果性と改善値の導出	6
§ 3.1 因果探索によるデータ間の関係性	6
§ 3.2 DEAによる効率値と入出力改善値の導出	6
§ 3.3 Foliumを用いたWebGISの開発	8
第4章 提案手法	9
§ 4.1 統計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	9
§ 4.2 選定されたデータに基づくDEA分析	9
§ 4.3 WebGISを用いたデータフュージョンのシステム開発	9
第5章 数値実験並びに考察	10
§ 5.1 数値実験の概要	10
§ 5.2 実験結果と考察	10
第6章 おわりに	11
謝辞	12
参考文献	13

図一覧

表一覽

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
インジケーターの計算に使用する期間	$n, N, m,$

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、これらをもとに大学等の研究機関でも○○や○○のようなEBPMに関する研究がなされている。

しかし、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。この方法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。

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思ったときに、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

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、サイバー空間上に存在する無数のデータに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] と包絡分析法（Data Envelopment Analysis: DEA）[4] を適用することで、取り扱いたい事柄に関係のあるものだけを抜き出し、それらを用いて分析を行うというアプローチでの解決策を提案する。

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、分析には直接使用していない地理情報等の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。

本研究では、まず、政策決定にデータマイニングや GIS が用いられた例について説明し、因果探索や DEA を用いたデータ分析手法を提案する。そして、提案手法による効率値の試算および WebGIS による結果の可視化とデータの重ね合わせ [3] を行う。

§ 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 WebGIS 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 WebGIS 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データの送受信を示しながら、本研究の提案手法の流れについて述べる

第5章

第6章

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、EBPMを推進することが重要視されている。

しかし、より効果的なEBPMの適用には、膨大かつ多種多様なデータを高速に高い信頼度を保って収集・分析する必要があり、人手のみに頼ると、担当者に対する大きな負担となる。そのため、EBPMを政策の広範囲に適用するためには、ICTを用いたデータ収集・分析が欠かせない。

本節では、はじめに○○における適用例を用いてEBPMの具体的なフローの一つを解説する。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。

EBPMのフロー

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。官民のビッグデータを集約し、可視化するシステムとして提供され、それを用いた地域課題の抽出、地域版総合戦略の立案といった活用が期待される。

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。

本節では、これらの情報が GIS によって分析・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にて、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。

被災初期

GIS によって、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に用いられた。

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と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。

○○

§ 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ものであるが、それらに加えて、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をいくつか挙げる。

都市計画と生態学

○○

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、医療をはじめとした幅広い分野での研究に用いられている。

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、

因果探索の手法として広く用いられる LiNGAM であるが、モデルを推測するための手法やモデルの拡張などによって現在までにいくつかのパターンが提唱されている。例として、Direct-LiNGAM, VAR-LiNGAM, ARMA-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。

§ 3.2 DEA による効率値と入出力改善値の導出

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事が挙げられる。

組織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組織との比較によって、その業績を評価する事が可能である。

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以下の理由が挙げられる。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における組織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。

「 $\Sigma(\text{各組織における出力} \times \text{出力のウェイト}) / \Sigma(\text{各組織における入力} \times \text{入力のウェイト})$ 」

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。

- すべての組織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、対象の組織の入力・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事である。以上をもとに、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し、その双対問題を考えると以下のようになる。

<入力指向モデル>

$$\text{Objective : } w = \theta \rightarrow \min$$

$$\text{Constrained : } Y\lambda \geq y_o$$

$$-X\lambda + x_o\theta \geq 0$$

$$\lambda \geq 0$$

<出力指向モデル>

$$\begin{aligned}Objective: \quad w = \eta &\rightarrow max \\Constrained: \quad X\mu \leq x_o \\&- Y\mu + y_o\eta \leq 0 \\&\mu \leq 0\end{aligned}$$

ここで、それぞれのモデルでは入力・出力の改善値を求めるために入力指向では（出力 × ウェイト）の総和を出力指向では（入力 × ウェイト）の総和を1とするという制約式を設けている。

§ 3.3 Folium を用いた WebGIS の開発

2.2節、2.3節では、すべてのGISを一般的なものとしてGISと呼称したが、GISは動作するプラットフォームによっていくつかの呼称に分類することができる。よって、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WebGISと呼称するものとする。

WebGIS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。

本節では、以上のような特徴を有するWebGIS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。

提案手法

§ 4.1 統計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

本研究では、Web上に公開されているオープンデータを用いて、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、効率値及び、改善値を算出する。
4. DEAの結果をWebGIS上に可視化し、

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.
- 2.
- 3.
- 4.

本節では、データベースに保存されているデータの種類とデータの収集方法と提案手法におけるLiNGAMによる因果探索を解説する。

§ 4.2 選定されたデータに基づくDEA分析

§ 4.3 WebGISを用いたデータフュージョンのシステム開発

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

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WebGIS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。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 價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（オアンダ 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] 刀根薰, “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.