

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
(の今後)

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

長瀬 永遠 (Towa Nagase)
u255013@st.pu-toyama.ac.jp

富山県立大学 工学研究科 電子・情報工学専攻
情報基盤工学講座

09:30-12:00 Tuesday, April 19, 2022.

1. はじめに
2. EBPM と ICT の有用性
3. データ間の因果性と改善値の導出
4. 提案手法
5. 数値実験並びに考察
6. おわりに

1.1 本研究の背景

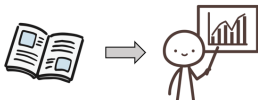
2/16

1. はじめに
2. EBPM と ICT の有用性
3. データ間の因果性と改善値の導出
4. 提案手法
5. 数値実験並びに考察
6. おわりに

近年、世界各国で証拠に基づく政策立案（Evidence-Based Policy Making: EBPM）に注目が集まっている。日本においても例外ではなく、研究機関でも取り上げられている。政府を中心に推進されている EBPM であるが、地方自治体ではいまだにエピソードベースの意思決定が根強い。

エビデンスベース

政策によって改善したい対象を明確化したうえでデータを収集し意思決定。



エピソードベース

住民によって役場に持ち込まれた問題に対して対面処理的に意思決定。

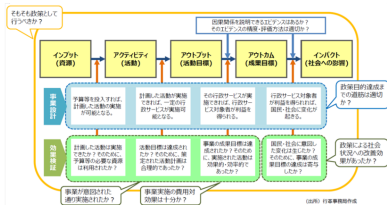


図1 EBPMのロジックモデル

EBPMの利点

- ・データに基づくため予測を行うことができ、問題が顕在化する前に対策を打つことが可能。
- ・その場限りのエピソードによるものではなく、明確な根拠があるため、住民の理解が得やすい。
- ・政策による効果が事前に逆算できるため、状況に応じた微調整が可能。

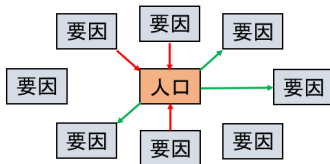
1.2 本研究の目的

3/16

政策決定における対象の複雑性という課題に対して、多種多様なデータから対象と因果関係のあるもののみを抜き出して分析し、GISを用いて結果を提示することで政策決定の支援を行う手法を提案する。

1. はじめに
2. EBPM と ICT の有用性
3. データ間の因果性と改善値の導出
4. 提案手法
5. 数値実験並びに考察
6. おわりに

人口を増やしたい。



問題が複雑すぎる。
全貌が把握できない。



図2 ビッグデータ

地形情報・施設分布
などと重ね合わせて
新知見を発見。

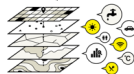


図5 GISのイメージ

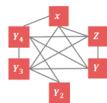


図3 因果グラフ



図4 データ分析

政策決定支援

2.1 ICT とデータの利活用による行政施策

4/16

1. はじめに
2. EBPM と ICT の有用性
3. データ間の因果性と改善値の導出
4. 提案手法
5. 数値実験並びに考察
6. おわりに

EBPM を実行するためには膨大なデータの収集・分析が必要になる。EBPM を広範囲に的確かつ効率的に取り入れるには ICT の有効活用がもとめられる。このことから、内閣府を中心に地方自治体における EBPM を支援するためのシステムが複数提供されている。

EBPM推進の向けたシステム

- ・ **内閣エビデンスシステム**
研究機関の運営評価に特化。研究機関における「研究力」、「教育力」、「資金獲得力」などを見る化。
- ・ **地域経済分析システム**
経済に関するデータを中心に幅広いデータを扱う分析が可能。
- ・ **V-RESAS**
COVID-19による経済の変化を分析可能。
- ・ **内閣府経済社会総合研究所**
内閣府を支えるシンクタンク。政策研究を担う人材育成も行う。



図7 地域経済分析システム



図6 内閣エビデンスシステム

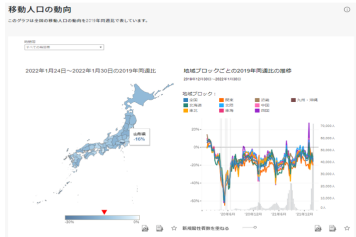


図8 V-RESAS

2.2 GIS による地域的な政策決定支援

5/16

GIS とは、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て視覚化できる技術。近年、多くの自治体で運用されている。

GISの特徴

- ・ **データの可視化**
統計データ等を2Dや3D、アニメーションなど多様な表現方法で地図上にプロット。
- ・ **データ間の関係性把握**
複数データを同一画面に表示し比較。
- ・ **データの統合と分析**
位置情報をキーとして異なる特徴を持つデータを統合。
- ・ **データの作成と更新**
データ更新の負担を軽減し鮮度を保つ。

GISの利点

- ・ **業務効率化による情報管理のコスト削減**
電子化により作業時間や人員を縮小。
- ・ **最適な意思決定の促進**
同一画面上に複数データを表示。
- ・ **コミュニケーション性の向上**
視覚的な情報共有や議論が可能。



図9 被災下でのGIS



図10 地方自治体におけるGIS

1. はじめに
2. EBPM と ICT の有用性
3. データ間の因果性と改善値の導出
4. 提案手法
5. 数値実験並びに考察
6. おわりに

2.3 Web-GIS によるデータフュージョン

6/16

GIS は多様なデータに対して地理空間を介在することで同一プラットフォーム上に可視化できる。そのため、普段は別々に研究が行われている分野の研究結果を重ね合わせて新たな知見を得ることが可能となる¹。

都市計画と生態学

都市計画と生態学のそれぞれで作成されたマップを重ね合わせることで異分野の知見を融合。



図11 二次草原潜在生育地マップ



図12 都市化ポテンシャルマップ

二次草原が失われる可能性のある土地

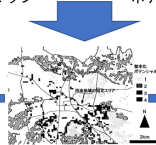


図13 重ね合わせマップ

類似な特徴を持つ土地を分析

身体情報に対する応用

GISにおける位置情報をキーにして情報を重ね合わせるという特徴を人体モデルに応用



図14 身体地図情報システムの構成

オーバーレイ機能

レイヤ同士を重ね合わせることで重畳部分のみの新たなレイヤを作成。部位別の受傷頻度等を算出。

受傷面積算出機能

傷害として塗りつぶされた範囲を自動算出し、体表の何%に当たるかを計算。熱傷の重症度の測定等を使用。

ICD-10コード変換機能

傷害情報と受傷位置それぞれの情報を持つレイヤを重ね合わせ、傷害情報をICD-10コードに自動変換。

¹ 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", 日本知能情報ファジィ学会誌, Vol. 20, No. 2, pp. 155-163, 2008.

3.1 因果探索によるデータ間の関係性

7/16

因果探索とは、観測データを用いて、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である。

1. はじめに
2. EBPM と ICT の有用性
3. データ間の因果性と改善値の導出
4. 提案手法
5. 数値実験並びに考察
6. おわりに

LiNGAM

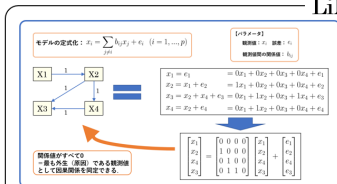


図15 LiNGAMのアルゴリズム

LiNGAMの制約

1. 内生変数と外生変数をつなぐ関数は線形。
2. 外生変数の分布は非ガウス連続分布。
3. 因果グラフは非巡回
4. 外線変数同士は互いに独立。

内生・・・観測済み 外生・・・未観測

Direct-LiNGAM

Direct-LiNGAMのアプローチ

・・・回帰分析を用いる手法

- ・ 内生変数群から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因果グラフの始点を探索。
- ・ その変数を内生変数群から除外し、残った変数のみで内生変数群を再形成。

適当な2変数

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

目的変数を x_1 、説明変数を x_2 としたときの回帰残差(r_2).

$$r_2 = \left\{ 1 - \frac{b_{21} \text{cov}(x_1, x_2)}{\text{var}(x_2)} \right\} e_1 - \frac{b_{21} \text{var}(x_1)}{\text{var}(x_2)} e_2$$

ダルモア・スキットピッチの定理

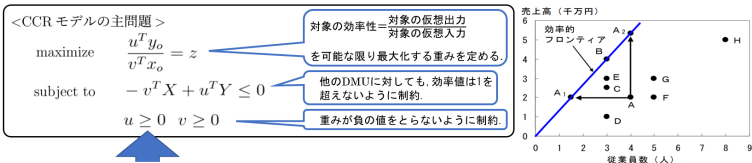
2つの確率変数 y_1, y_2 が互いに独立な確率変数 $s_i (i = 1, \dots, q)$ を用いて下記のように表されるとき、 y_1, y_2 が独立なら、 $\alpha_j, \beta_j \neq 0$ となるような変数 s_j はガウス分布に従う。

$$y_1 = \sum_{i=1}^q \alpha_i s_i \quad y_2 = \sum_{i=1}^q \beta_i s_i$$

3.2 DEA による効率値と入力・出力改善値の導出

8/16

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ノンパラメトリックなアプローチである。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）を指す。



双対問題

< 入力指向モデル >

$$\begin{aligned} &\text{minimize} && w = \theta \\ &\text{subject to} && Y\lambda \geq y_o \\ &&& -X\lambda + x_o\theta \geq 0 \\ &&& \lambda \geq 0 \end{aligned}$$

< 出力指向モデル >

$$\begin{aligned} &\text{maximize} && w = \eta \\ &\text{subject to} && X\mu \leq x_o \\ &&& -Y\mu + y_o\eta \leq 0 \\ &&& \mu \geq 0 \end{aligned}$$

< 入力改善案 >

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

< 出力改善案 >

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

それぞれのDMUに対して各入力をどれだけ減少、各出力をどれだけ増加させれば評価値が1になるかが算出できる。また、その際に参考としたDMUもわかる。

3.3 Folium を用いた Web-GIS の開発

9/16

Web-GIS とは、html 形式で記述され、World Wide Web 上でアプリケーションとして動作する GIS の総称。また、Web-GIS における html は直接手書きで作成するほかプログラミング言語で自動的に生成も可能。Web-GIS を記述する Python のモジュールを Folium と呼ぶ。



図17 レイヤコントロールの例



図18 ヒートマップの例



図19 各種マーカーの例

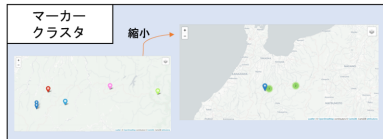


図20 マーカークラスターの例

1. はじめに
2. EBPM と ICT の有用性
3. データ間の因果性と改善値の導出
4. 提案手法
5. 数値実験並びに考察
6. おわりに

4.1 データベース作成と因果探索によるデータの選定

10/16

データベース内には地理情報を持たない数値データ、地理情報を持つ数値データ、施設等の位置データがある。データは RESAS API および国土交通省のウェブサイトを
用いて収集した。それらを因果探索で絞りこみ、DEA の入力・出力に振り分ける。

表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
歳出決算額 [総務費]	%	従業員数	人
歳出決算額 [民生費]	%	農地平均取引価格	円 / m ²
歳出決算額 [衛生費]	%	商業用地平均取引価格	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 [労働費]	%	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	%
林業業諸収入	万円	年間商品販売額	百万円
林産物販売金額	万円	海面漁獲物等販売額	万円
一人当たりの法人住民税	千円	付加価値額	万円
		1 人あたりの固定資産税	千円

表2 地理情報を持つ数値データ 表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

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

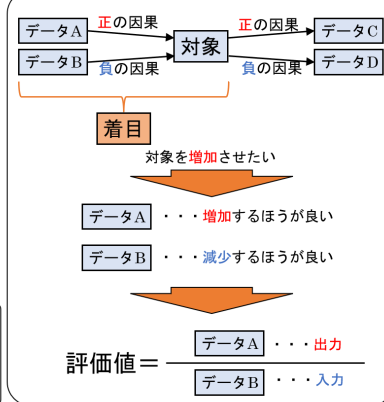
データの正規化

<robust Z-score>

$$x' = \frac{x - \text{median}(x)}{IQR}$$

median(x) ... データの四分位範囲 IQR ... データの中央値

入力出力への振り分け



4.2 選定されたデータに基づく DEA 分析

11/16

因果探索によって振り分けられた入力・出力を用いて DEA を行うことで評価値，入力・出力改善案，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算出する。種類ごとの計 4 つの csv ファイルで結果を出力する。

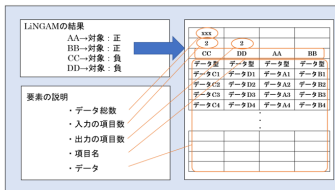


図21 DEAの入力・出力ファイル

表4 各都道府県におけるDMUの内訳

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

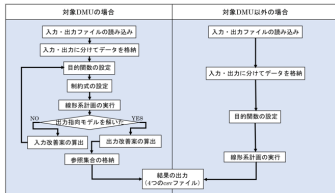


図22 DEA部分のフロー

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）
cuty_code effc_value	imp_item imp_value	city_code weight_in	city_code weight_out
1100 0.55679	AAA 0	24303 0.0611	24303 0.13164
1202 0.70558	BBB 13.244	33663 0.17123	33663 0.3689
1203 0.83029	CCC 0.86798	34203 0.01215	34203 0.02617
.	DDD 5.68598	39424 0.02395	39424 0.05159
.	EEE 3.69084	44462 0.32172	44462 0.69311
.	FFF 26.6893		
47381 0.30172	GGG 6.20473		
47382 0.35638	HHH 42026.3		

図23 DEA部分のアウトプット

改善案は正規化されていた値を逆変換して表示。

4.3 Web-GIS を用いたデータフュージョンのシステム開発

12/16

動画

提案システムのながれを動画でお見せします。

1. はじめに
2. EBPM と ICT
の有用性
3. データ間の因果
性と改善値の導出
4. 提案手法
5. 数値実験並びに
考察
6. おわりに

5.1 数値実験の概要

13/16

富山県射水市における少子高齢化問題について、その解決に向けた政策を行うというモデルケースに提案手法を適用することによって本研究の有効性を検証・考察する。

射水市における人口問題

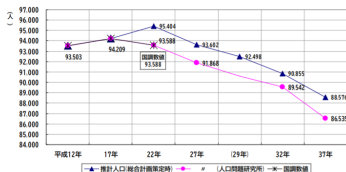


図24 射水市の推計人口

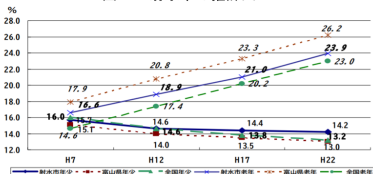


図25 年少・老年人口割合推移

数値実験のながれ

対象データを「年少人口[割合]」、対象市区町村を「射水市」と入力。

「年少人口[割合]」をターゲットとして因果探索。

結果をcsvファイル①として排出。

csvファイル①を入力・出力としてDEA分析

全市区町村の評価値をcsvファイル②
射水市の入力・出力改善案をcsvファイル③
参照集合と射水市に対する重みをcsvファイル④、⑤として排出。

csvファイル②～⑤を用いてWeb-GISを作成。

Web-GISをもとに考察。

5.2 数値実験の結果と考察

14/16

システムの各部分における結果を示す.

1. はじめに
2. EBPM と ICT の有用性
3. データ間の因果性と改善値の導出
4. 提案手法
5. 数値実験並びに考察
6. おわりに

表5 「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		

表6 DEAにおける入力・出力

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



図26 EBPM-GISの結果

表7 元データと改善案の比較

	元データ	改善案
企業数	3075	-27769.854
衛生費	10.08	-2.954
商工費	2.92	-1.647
警察費・消防費	2.88	-1.470
住宅用地平均取引価格	19507	-368560.385
老年人口	28.6	22.821
施設数 [空港]	0	0.143
教育費	12.5	32.502
生産年齢人口	57.5	66.692

入力：減少

出力：増加

表8 参照集合に属する市区町村とウェイト

入力指向モデル		出力指向モデル	
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	0.106	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

考察

- ・ 生産年齢人口が出力に振り分けられた。
 ・ ・ ・ 年少人口の親世代にあたるため妥当。
- ・ 警察費・消防費が入力に振り分けられた。
 ・ ・ ・ 意外性がある結果＝普通では気づかない関係性

まとめ

- データ間の因果関係に基づくデータ選択
 - 因果探索の結果をもとにデータの絞り込み，入力・出力を選定を行った。
- DEAを用いた評価と改善案の算出
 - 振り分けられたデータを用いてDEAを行った。
- GISによる結果の表示と重ね合わせによるデータフュージョン
 - 分析結果と地形データ，施設分布データを重ね合わせて分析が行われるようにした。

今後の課題

- データベース内の情報量の充実
 - 扱える問題の汎用性と分析結果の確からしさ両方の向上につながる。
- 因果探索における手法の洗練
 - 因果グラフにおけるパス係数をどの程度まで有意とするか検討する。
Direct-LiNGAMに限らず，より効果的な因果探索手法を模索する。
- DEAにおけるモデルの深化
 - 本研究のテーマによりフィットしたモデルを模索・提案する。

1. はじめに
2. EBPM と ICT
の有用性
3. データ間の因果
性と改善値の導出
4. 提案手法
5. 数値実験並びに
考察
6. おわりに

- 1 提案手法の洗練
- 2 テーマの再考