

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

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

October 18, 2022

- 1. はじめに
- 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
- New. 修論に向
けて

1.1 本研究の背景

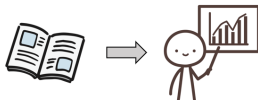
2/16

- はじめに
- データ間の因果性と改善値の導出
- 提案手法
- 数値実験並びに考察
- New. 修論に向けて

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。

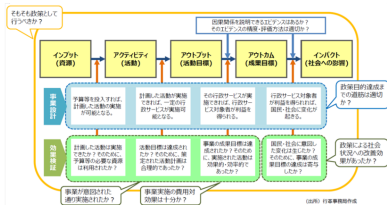


図1 EBPMのロジックモデル

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。

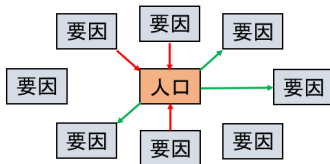
1.2 本研究の目的

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
 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
- New. 修論に向けて

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。

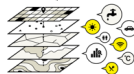


図5 GISのイメージ

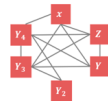


図3 因果グラフ



図4 データ分析

政策決定支援

3.1 因果探索によるデータ間の関係性

4/16

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。

- はじめに
- データ間の因果性と改善値の導出
- 提案手法
- 数値実験並びに考察
- New. 修論に向けて

LiNGAM

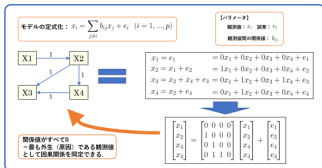


図15 LiNGAMのアルゴリズム

LiNGAMの制約

- 内生変数と外生変数をつなぐ関数は線形。
- 外生変数の分布は非ガウス連続分布。
- 因果グラフは非巡回
- 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

5/16

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。

<CCR モデルの主問題>

$$\begin{aligned} &\text{maximize} && \frac{u^T y_o}{v^T x_o} = z \\ &\text{subject to} && -v^T X + u^T Y \leq 0 \\ &&& u \geq 0 \quad v \geq 0 \end{aligned}$$

対象の効率性 = $\frac{\text{対象の仮想出力}}{\text{対象の仮想入力}}$

を可能な限り最大化する重みを定める。

他のDMUに対しても、効率値は1を超えないように制約。

重みが負の値をとらないように制約。

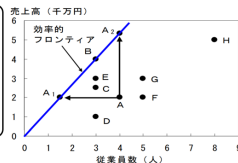


図16 1入力・1出力のDMU群

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

6/16

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

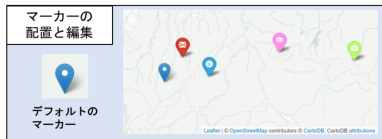


図19 各種マーカーの例

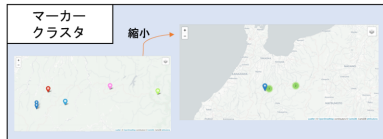


図20 マーカークラスターの例

1. はじめに
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
- New. 修論に向けて

4.1 データベース作成と因果探索によるデータの選定

7/16

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 畧 / 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

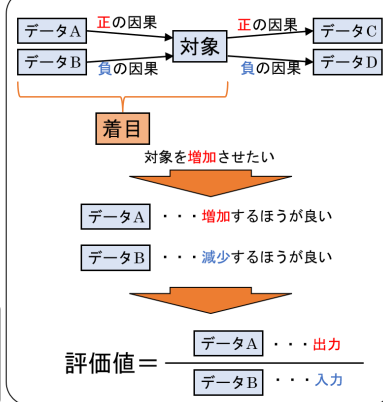
データの正規化

<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



1. はじめに

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

New. 修論に向
けて

4.2 選定されたデータに基づく DEA 分析

8/16

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。

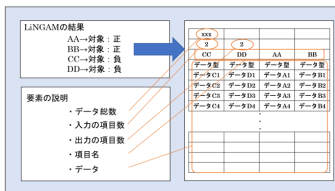


図21 DEAの入力・出力ファイル

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

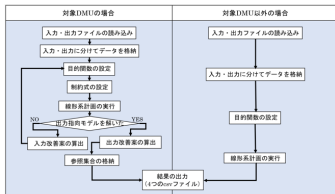


図22 DEA部分のフロー

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。

5.1 数値実験の概要

9/16

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

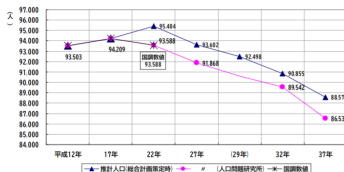


図24 射水市の推計人口

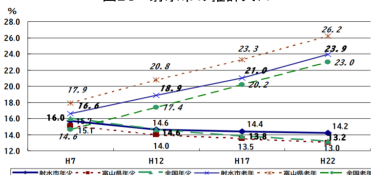


図25 年少・老年人口割合推移

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

10/16

システムの各部分における結果を示す.

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

1. はじめに
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

New. 修論に向けて

1. はじめに
2. データ間の因果性と改善値の導出
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

New. 修論に向けて

3つの課題

- 因果探索の結果は本当に妥当であるのか
- 因果探索の結果が妥当であるとして、その結果を政策に適用したら社会課題は解決するのか
- この研究自体に意義があるのか

解決策の案

現状の因果探索と DEA との間に因果推論を挟むことによって因果探索の結果を評価し、結果の精度を高める。 また、データベース内のデータを整理・拡充する必要がある。

因果推論

回帰分析

差分の差分

操作変数法

共分散構造分析

⋮

因果探索

ベイジアンネットワーク

LiNGAM

Figure 1: 因果推論と因果探索

解決の案

政策の適応範囲に対して数理モデルを構築し、シミュレーションによって効果を検証する。ただし、数理モデルに関しては恣意的ととられないように注意深く定式化する必要がある。また、それにあたって研究の対象とする社会課題に関しても再考の必要がある。

(例：将来の人口予想)

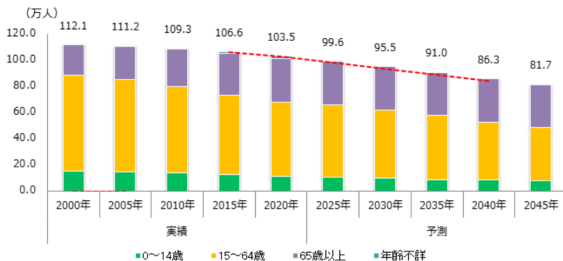


Figure 2: 富山県の人口推移と今後の予想

New4 この研究に意義があるか

14/16

1. はじめに
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

New. 修論に向けて

解決の案

正直，この懸念点があるために修論もこのテーマを継続するか迷っていた。ただ，日本では EBPM という分野自体がわりと新しいものであり，近年になって各研究機関をはじめ日立製作所など一般企業の研究所も研究を行っている分野であるので，その現状を上手くサーベイすれば価値がある研究だということは示せそう。

1. はじめに
 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
- New. 修論に向けて

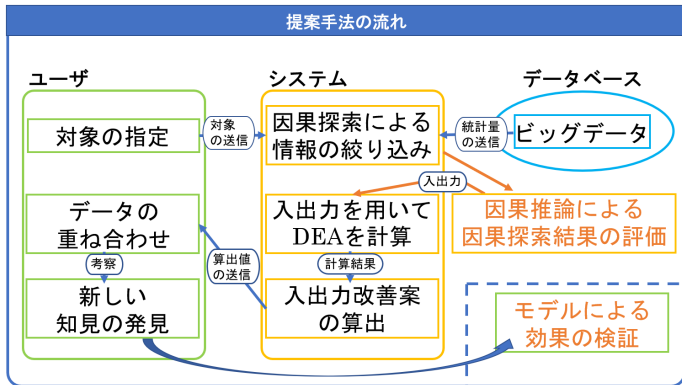


Figure 3: 修論での研究のフロー図

1. はじめに
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

New. 修論に向けて

(表)

- データベースの整理
- 因果推論の組み込み
- 因果推論込みの数値実験

(裏)

- モデル化の勉強