

- LSI -

LLMと潜在空間を用いた
市区町村単位での持続性評価

Eye
of the
Storm

神戸大学
衣川凌太 橋本大和

Table of Contents

01 エグゼクティブ
サマリー

02 背景理解

03 分析手順・手法

04 分析結果

05 ダッシュボード

06 指標の利用例

Executive Summary

もうそろそろ SDGs 終わるけど...

達成を感じない！

エグゼクティブサマリー

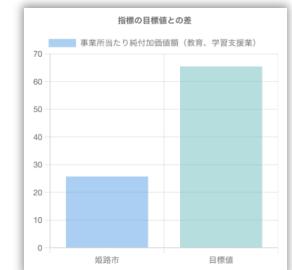
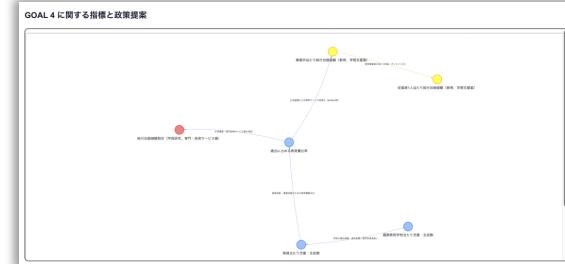
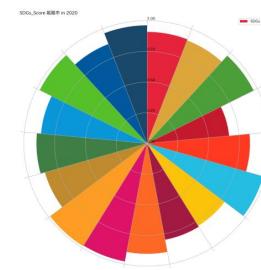
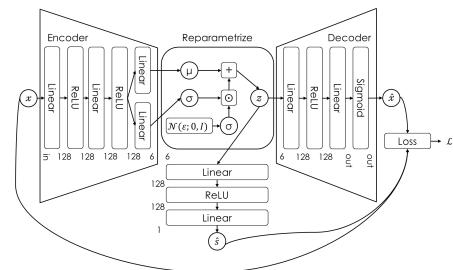
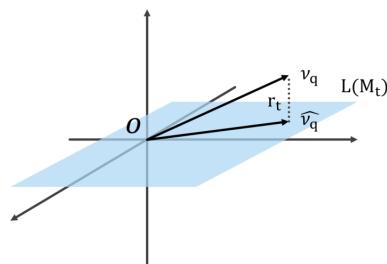
Executive Summary

SDGs進捗度を測る新たな指標

- ・持続的かつ統一的に収集可能なデータ
- ・指標間のトレードオフと地域差への配慮
- ・サブ指標ごとに目標値を生成

▶ **LSI**

AI/データを最大限有効活用した**持続可能な街づくり**のための**データ分析基盤**を提供する！



#

もくじ

Table of Contents

01 エグゼクティブ
サマリー

02 背景理解

03 分析手順・手法

04 分析結果

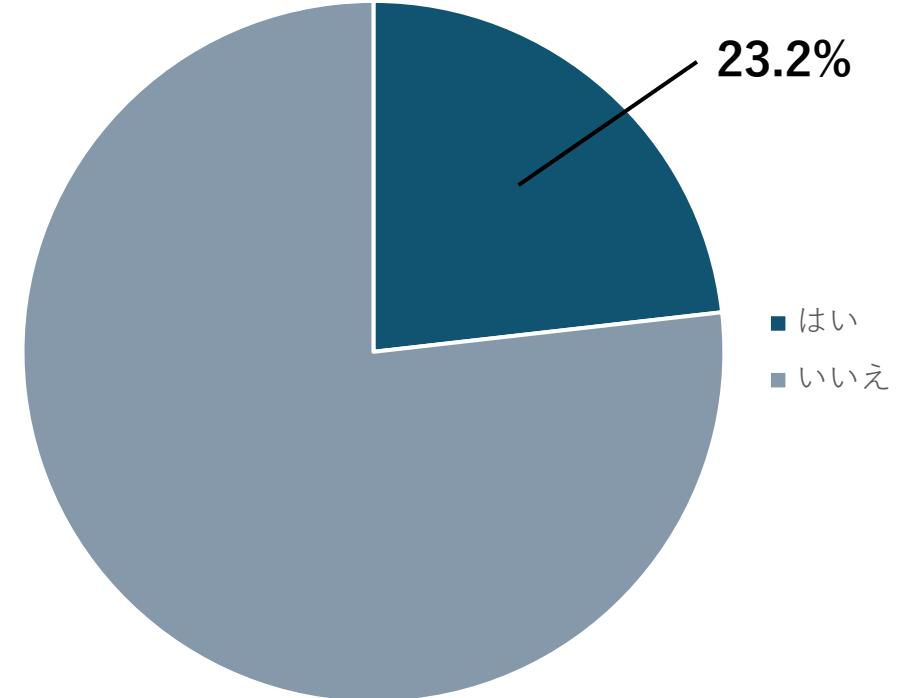
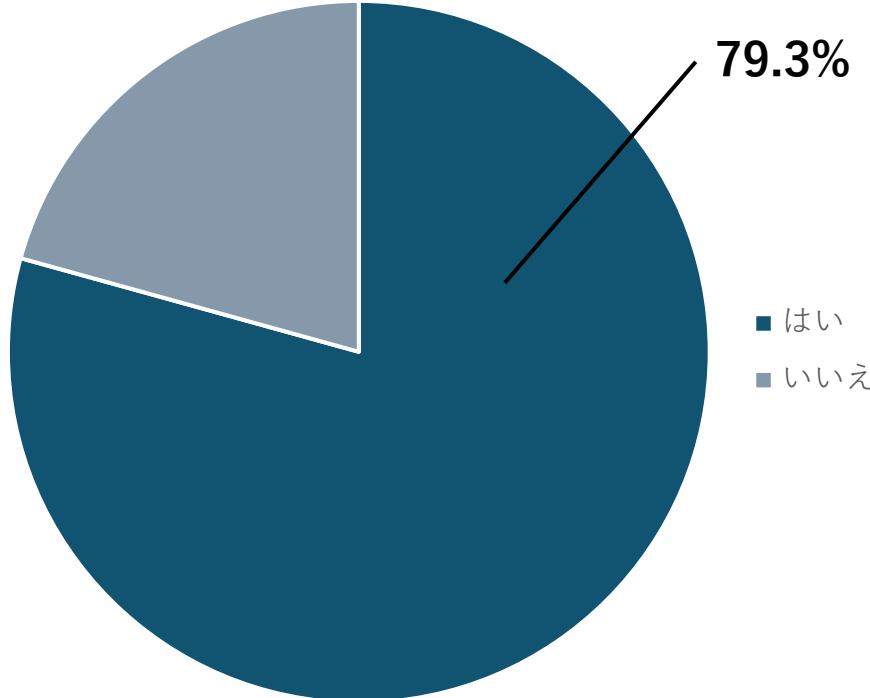
05 ダッシュボード

06 指標の利用例

背景理解

Background Knowledge

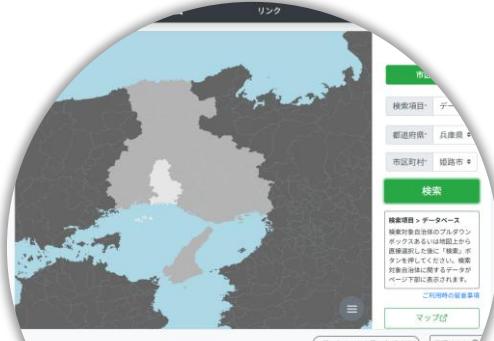
激減！



2

背景理解

Background Knowledge



自治体SDGs推進評価
調査検討会
「地方創生SDGsローカル指標」



日本経済新聞社
「全国市区SDGs先進度調査」



自治体SDGs達成度評価
「地域SDGs達成度評価指標」

対象

1741市区町村

式

$$S_{ij} = \frac{x_{ij} - \min(X_j)}{\max(X_j) - \min(X_j)}$$

公開

都道府県/市区町村単位で公開

815市区

- ・アンケート調査の定量化
- ・経済・環境・社会の3観点
- ・100点満点

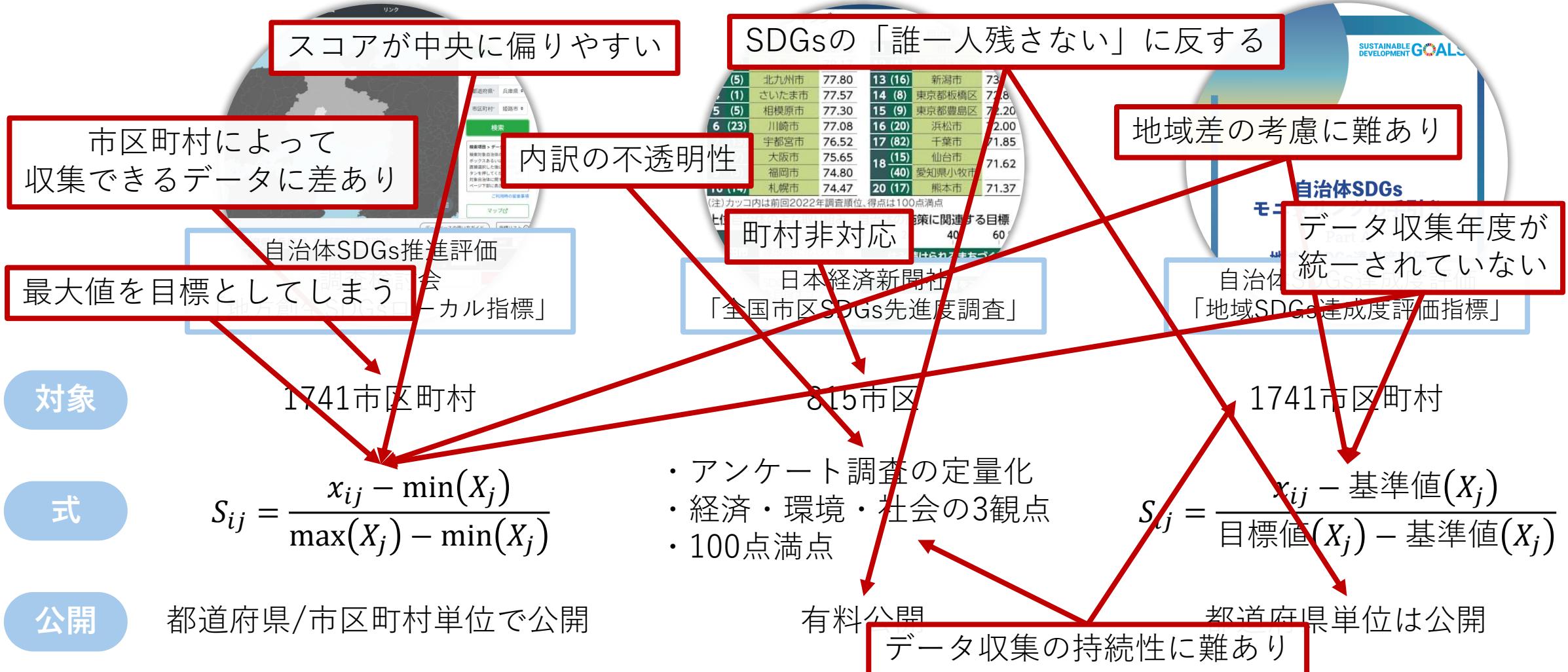
1741市区町村

$$S_{ij} = \frac{x_{ij} - \text{基準値}(X_j)}{\text{目標値}(X_j) - \text{基準値}(X_j)}$$

有料公開

都道府県単位は公開

Background Knowledge



Background Knowledge

データ

- ・データ収集の難しさ
 - 欠損値の存在
 - 持続的な収集可能性
 - 選択するデータの妥当性

指標式

- ・スコア計算方法の是非
 - 良くある流れ「正規化→平均」
 - 全てのサブ指標で最大値を目指すことに
- ・統一的な指標の是非
 - 地域差の考慮
 - データ間の多重共線性・トレードオフ

公開

- ・限定アクセス
 - SDGsの「誰一人残さない」の理念
- ・指標公開だけで終わる
 - 分析基盤としての利用促進

「国勢調査」「社会・人口統計系」
SDGsのゴールを
測るのに適切なデータを選択

サブ指標間のトレードオフと
地域差を考慮した
達成度評価と**目標提示**

誰でも使えるEBPM基盤として
可視化・分析ダッシュボード
の開発

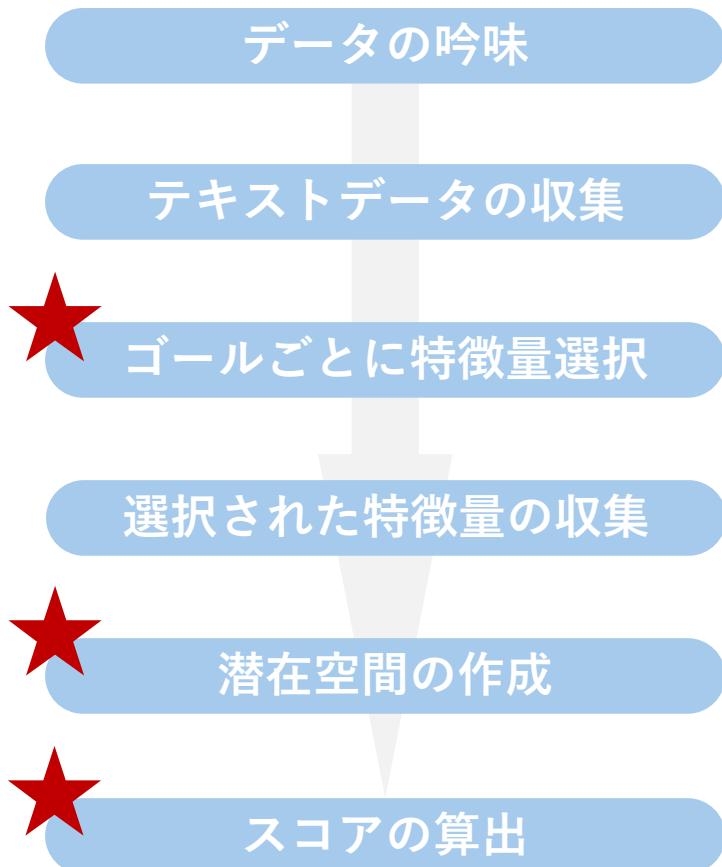
#

もくじ

Table of Contents

- | | | | |
|----|-----------------|----|---------|
| 01 | エグゼクティブ
サマリー | 04 | 分析結果 |
| 02 | 背景理解 | 05 | ダッシュボード |
| 03 | 分析手順・手法 | 06 | 指標の利用例 |

Analysis Method



データの収集元：国勢調査（e-Stat）
社会・人口統計系（e-Stat）

- ・市区町村単位で漏れなく収集可能であること
- ・5年単位で持続的に収集されていること



3

分析手順・手法

Analysis Method

通常ならこの時点でデータを収集するが...

それが無いのが現状

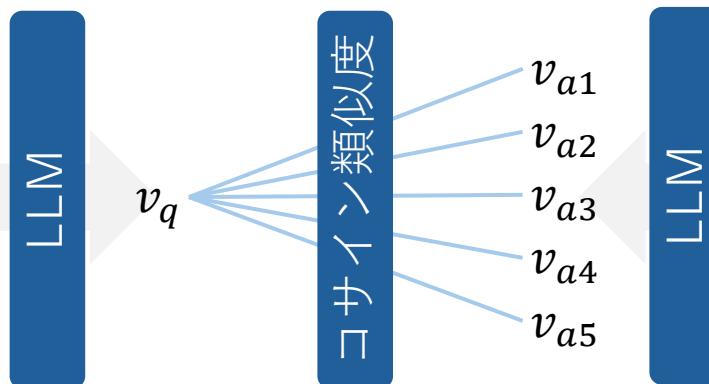
指標に適切なデータを選択したい！ \leftrightarrow 適切な目的変数が必要！

データの代わりに、データ名を使えないか...？

応用できそう！

LLMの埋め込みベクトルを用いたQAタスク

兵庫県の県庁所在地は



大阪市
神戸市
京都市
岡山市
姫路市

```
query = '兵庫県の県庁所在地は'
answers = ['大阪市', '神戸市', '京都市', '岡山市', '姫路市']

for i, answer in enumerate(answers):
    with torch.no_grad():
        answer_output = model.encode(answer)

    with torch.no_grad():
        query_output = model.encode(query)

    print(f'{answers[i]} : {cos_sim(query_output, answer_output)}')

大阪市 : 0.80019426
神戸市 : 0.8550547
京都市 : 0.775703
岡山市 : 0.74139714
姫路市 : 0.82805794
```

分析手順・手法

Analysis Method

目的変数のデータ名

説明変数のデータ名

「SDGsの各ゴールの文章」

「国勢調査/社会・人口統計系のデータ名」

「あらゆる場所のあらゆる形態の貧困を終わらせる」
を説明することができる特徴量は？

これに応答する
特徴量を選択する！

- 「ごみ最終処分率」
- 「パート・アルバイト男女比」
- 「世帯数割合（父子世帯）」
- 「事業所割合（情報通信業）」
- 「事業所当たり売上金額（複合サービス業）」
- 「事業所当たり純付加価値額（金融業、保険業）」
- 「付加価値率（卸売業、小売業）」
- 「付加価値率（第2次産業）」
- 「保護施設数密度」
- 「公民館密度（人口ベース）」
- 「合計特殊出生率」
- 「実質公債費比率」
- 「就業者数に占める正規雇用者数比率（男）」
- 「市街化調整区域面積割合」

⋮

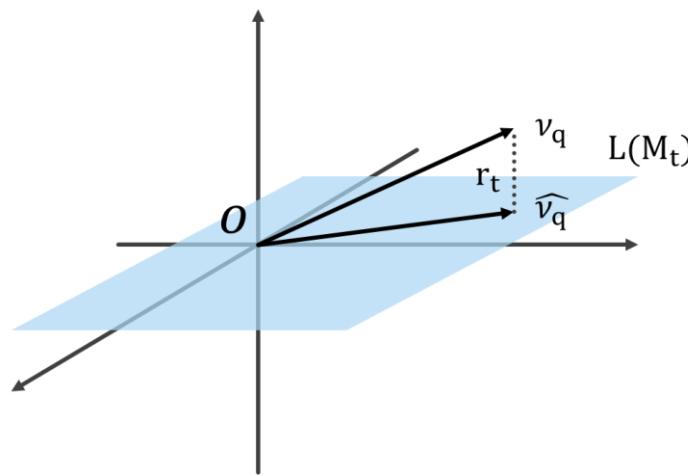
分析手順・手法

Analysis Method

これに応答する
特徴量を選択する！

残差ベクトル

説明変数	$: X = x_i \in \mathbb{R}^N$
クエリベクトル（目的変数）	$: v_q$
キーべクトル（説明変数）	$: V = v_{x_i} \in \mathbb{R}^N$
選択セット	$: S_t = j \in \mathbb{R}^T$
残差	$: r_t = v_q - \widehat{v_q}^{(t)}$ (N = 説明変数の数, T = 選択された説明変数の数)



$$k^* = \operatorname{argmax}_{k \notin S_t} \frac{r_{t-1} \cdot v_k}{\|r_{t-1}\| \|v_k\|}$$

$$S_t = S_{t-1} \cup \{k^*\}$$

$$M_t = \{v_{x_j} \mid j \in S_t\}$$

$$\widehat{v_q}^{(t)} = M_t (M_t^T M_t)^{-1} M_t^T v_q$$

$$r_t = v_q - \widehat{v_q}^{(t)}$$

$$\text{cov}_t = 1 - \frac{\|r_t\|^2}{\|v_q\|^2}$$

要約

クエリ（目的変数）により応答する
キー（説明変数）から選択していくが…

新しいキーを選択するごとに
まだ説明できていない部分（= 残差）を求める、
それを一番説明できるキーを
その次に選ぶ！



クエリを隅から隅まで
説明できるキーを選択できる！

基準化

- 全てのサブ指標で、値の上昇（下降）と意味のポジ（ネガ）が単純に結びついているわけでは無い
例) 転入超過率 … 高すぎても低すぎてもダメ
- ▶ 基準を定め、基準までの差をとって基準化することで、値と意味の関係を整理

標準化

- 次以降のステップのカギとなる

意味の方向の統一

- 「値の**上昇** → **ポジティブ**」と定める
- ▶ 「値の**上昇** → **ネガティブ**」のサブ指標には、**-1**をかける

なぜ、**潜在空間**が必要か。

スコア推移前				スコア推移後			
サブ指標A	サブ指標B	サブ指標C	サブ指標D	サブ指標A	サブ指標B	サブ指標C	サブ指標D
0.30	0.70	0.40	0.60	0.70	?	?	?

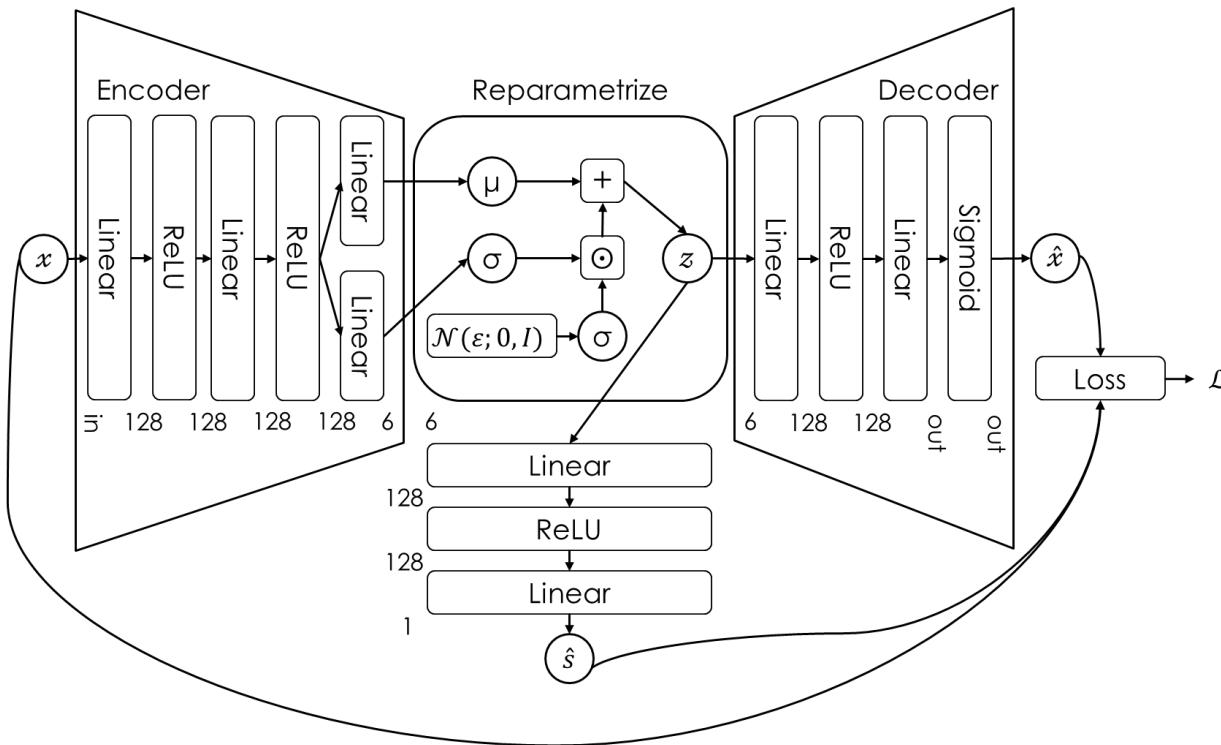
トレードオフ：あるサブ指標のスコアを動かすと、他のサブ指標も動く

地域差：サブ指標のとれる値の幅は地域によって異なる

潜在空間はこれらを反映したサブ指標の分布を求めることができる！

分析手順・手法

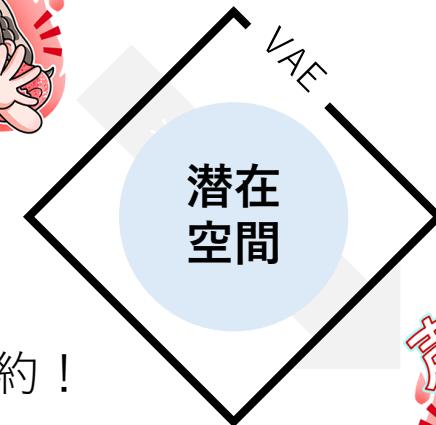
Analysis Method



VAEを用いてその分布を生成する



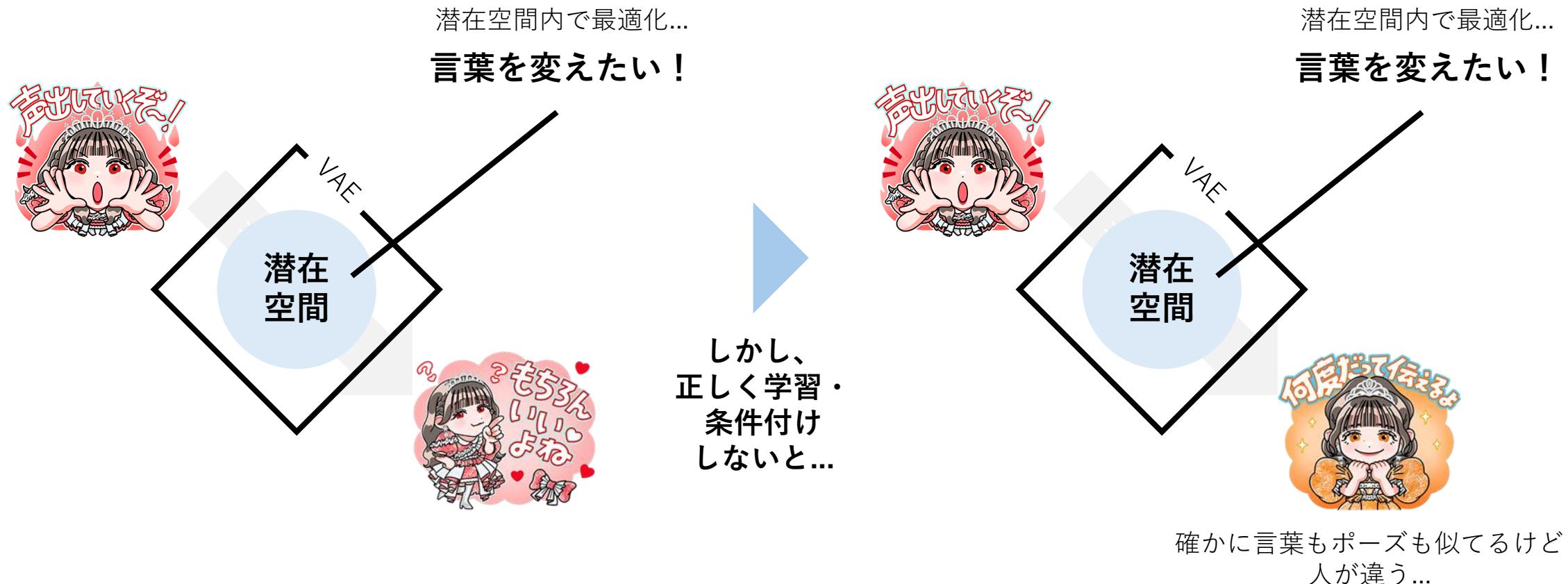
情報を
低次元に集約！



3

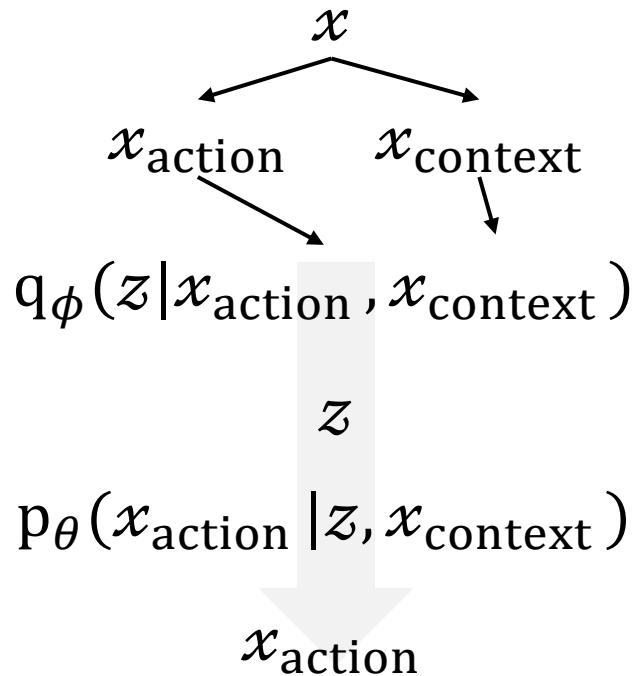
分析手順・手法

Analysis Method



Conditional VAE

地域特性を逃がさない！



損失関数の工夫

サブ指標の分布をうまく学習！

$$\begin{aligned}\hat{s} &= g_\psi(z) \\ \mathcal{L}_{\text{proto}} &= (s - \hat{s})^2 \\ \mathcal{L} &= \mathcal{L}_{\text{rec}} + \beta \mathcal{L}_{\text{KL}} + \gamma \mathcal{L}_{\text{proto}}\end{aligned}$$

最適化の工夫

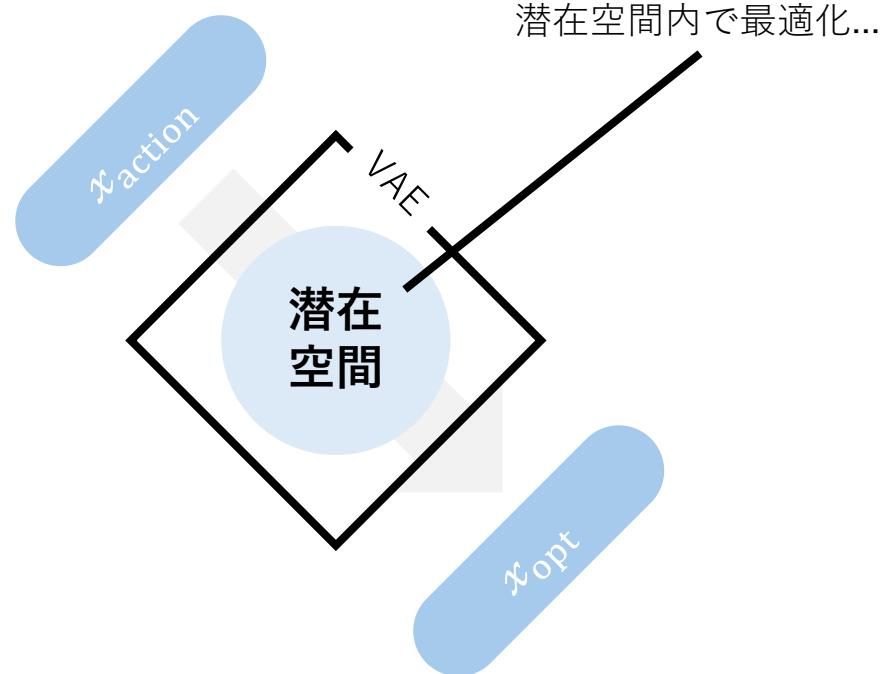
変な目標を生成させない！

$$\begin{aligned}\mathcal{L}_{\text{opt}} &= -g_\psi(z) + \lambda_{\text{L2}} \|z - z_0\|_2^2 \\ &\quad + \gamma \sum_{j \in A} \max(0, x_j^{(0)} - \hat{x}_j(z, x_{\text{context}}))^2 \\ z &\leftarrow z - \eta \nabla_z \mathcal{L}_{\text{opt}}(z)\end{aligned}$$



地域性を残しつつ、
トレードオフに基づいて、
伸ばせるサブ指標 (=目標)
を教えてくれる！

現在の値 (x_{action}) から **目標値** (x_{opt}) と **達成度** (Score) を求める！

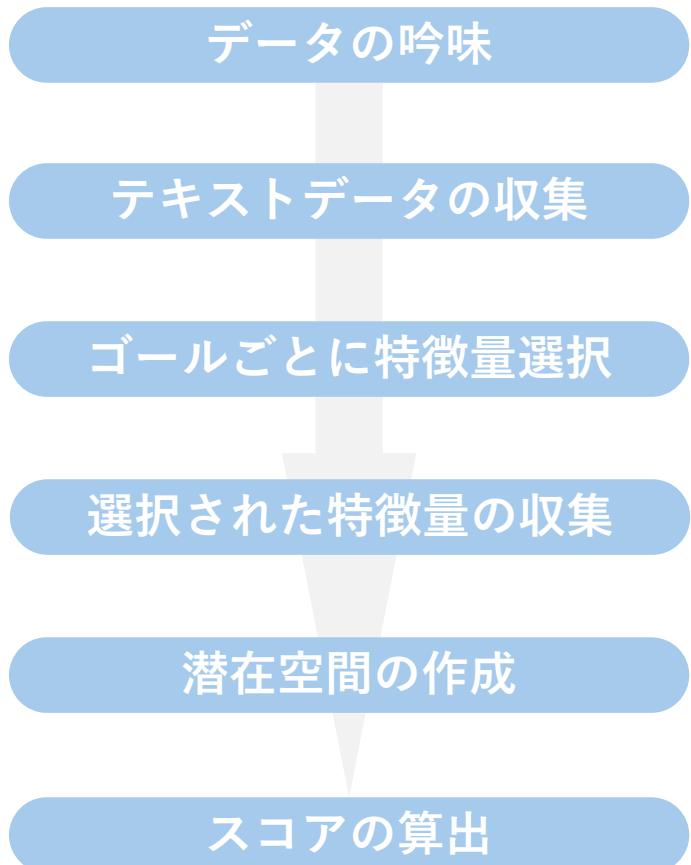


$$\begin{aligned} m_{\text{now}} &= \mathbb{E}[x_{\text{action}}] \\ m_{\text{opt}} &= \mathbb{E}[x_{\text{opt}}] \\ b &= -1.645 \end{aligned}$$

$$\text{Score} = \frac{m_{\text{now}} - b}{m_{\text{opt}} - b}$$

サブ指標は標準化しているので、
下位5%点をベース (b) とする
 $\rightarrow m_{\text{now}}$ が b より小さい場合は0点とする

Analysis Method



LSI
Local Sustainability Index

#

もくじ

Table of Contents

- | | |
|--------------------|------------|
| 01 エグゼクティブ
サマリー | 04 分析結果 |
| 02 背景理解 | 05 ダッシュボード |
| 03 分析手順・手法 | 06 指標の利用例 |

- ▶ 特徴量選択法の妥当性
- ▶ VAEの妥当性
- ▶ LSIに基づく市区町村単位でのSDGs達成度

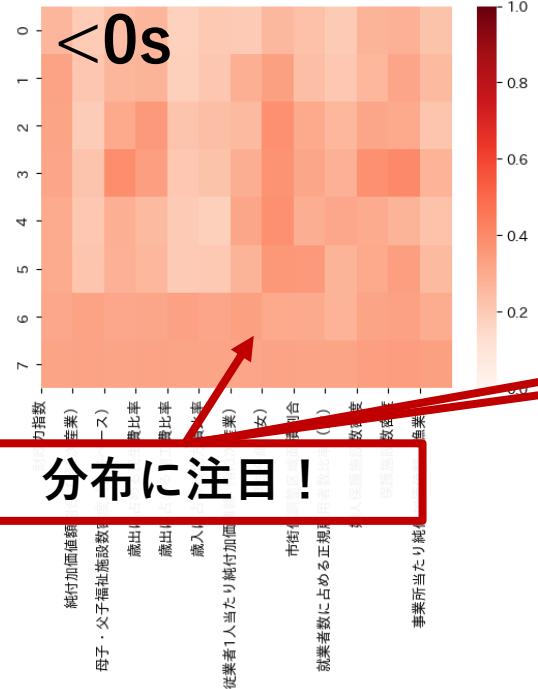
分析結果

特徴量選択法の妥当性

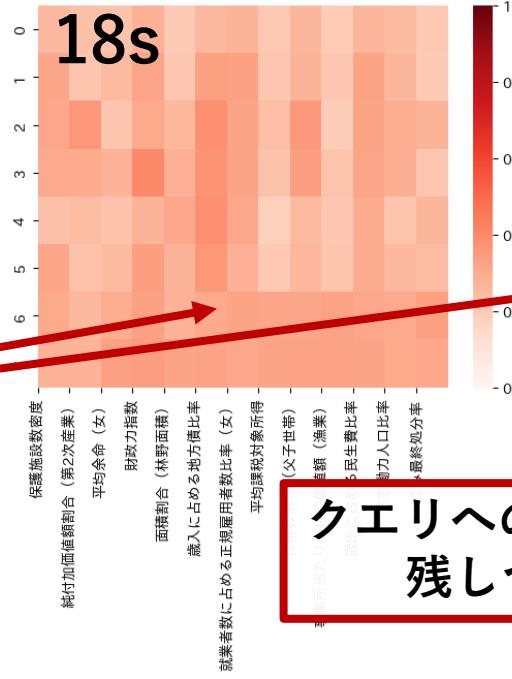
Analysis Results

Mean: クエリへの応答性, Std: 選択された特徴量の多様性

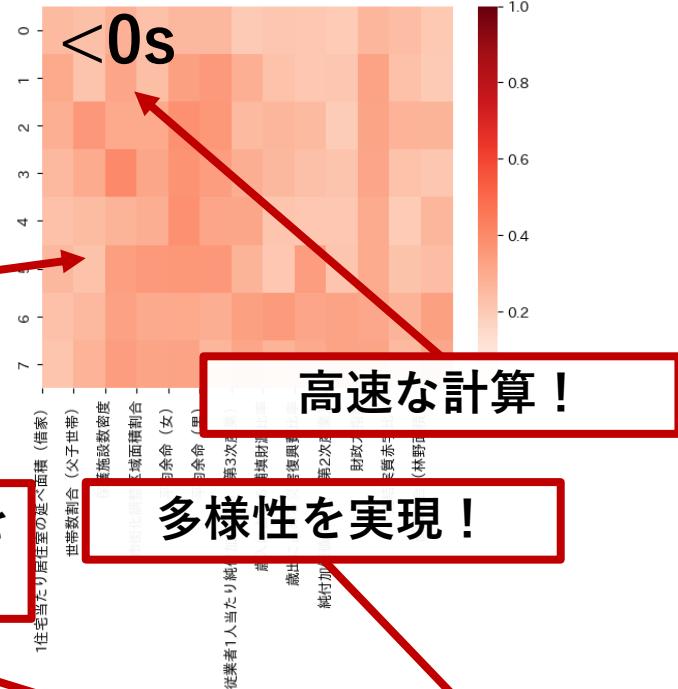
コサイン類似度降順



MMR



残差ベクトル

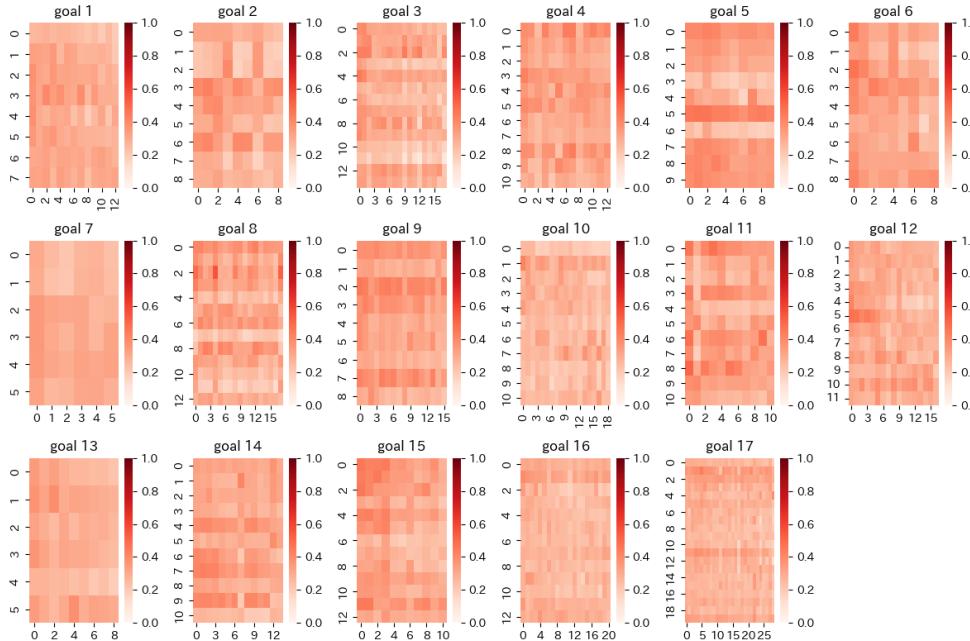


分析結果

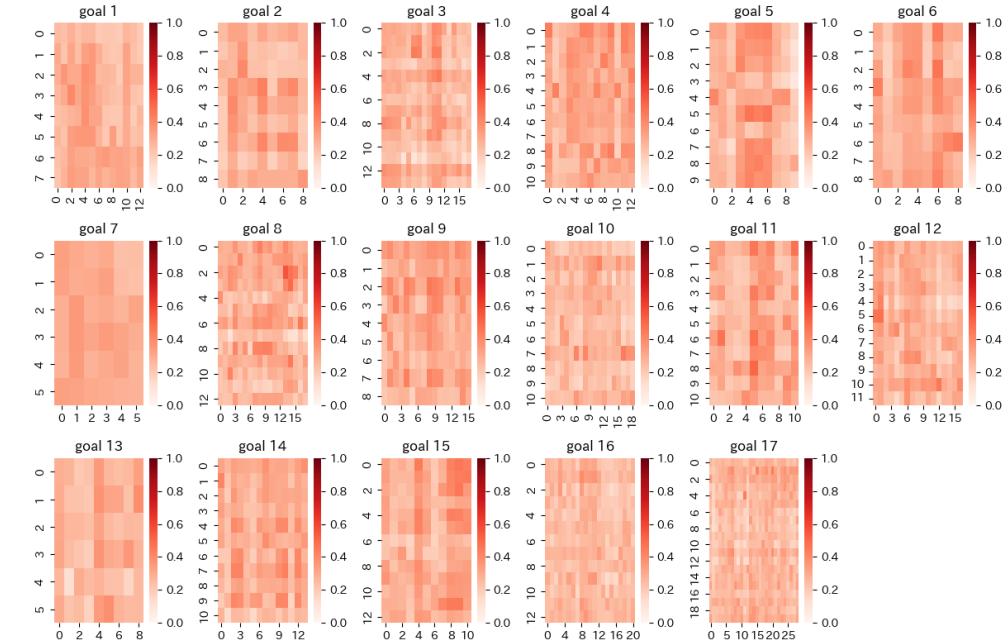
特徴量選択法の妥当性

Analysis Results

コサイン類似度昇順



残差ベクトル



残差ベクトルの方が、より多様なサブ指標を採用できることが分かる！

Analysis Results

ゴール別 選択特徴量数

G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17
13	9	18	13	10	9	6	18	16	19	11	17	9	14	11	21	29

ゴール4 選択特徴量

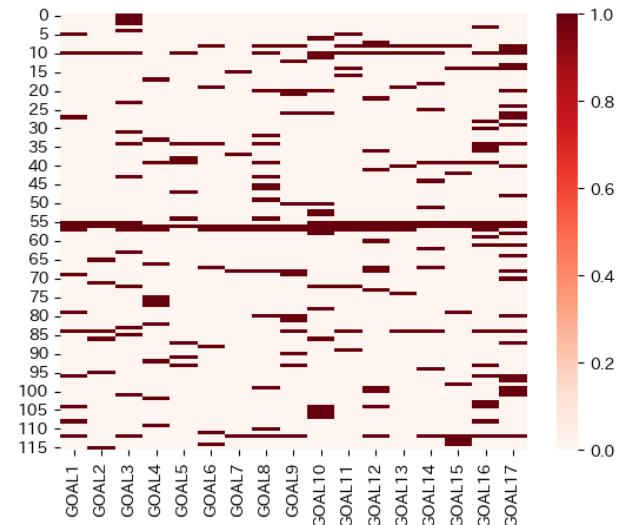
「質の高い教育をみんなに」

 x_{action}

- 事業所当たり純付加価値額（**教育、学習支援業**）
- 婦人保護施設数密度
- 従業者1人当たり純付加価値額（**教育、学習支援業**）
- 教員当たり**児童・生徒数**
- 最終学歴人口比率（**大学等**）
- 未就学児施設密度（人口ベース）
- 歳出に占める**教育費比率**
- 純付加価値額割合（**学術研究、専門・技術サービス業**）
- 義務教育学校当たり**児童・生徒数**

 x_{opt}

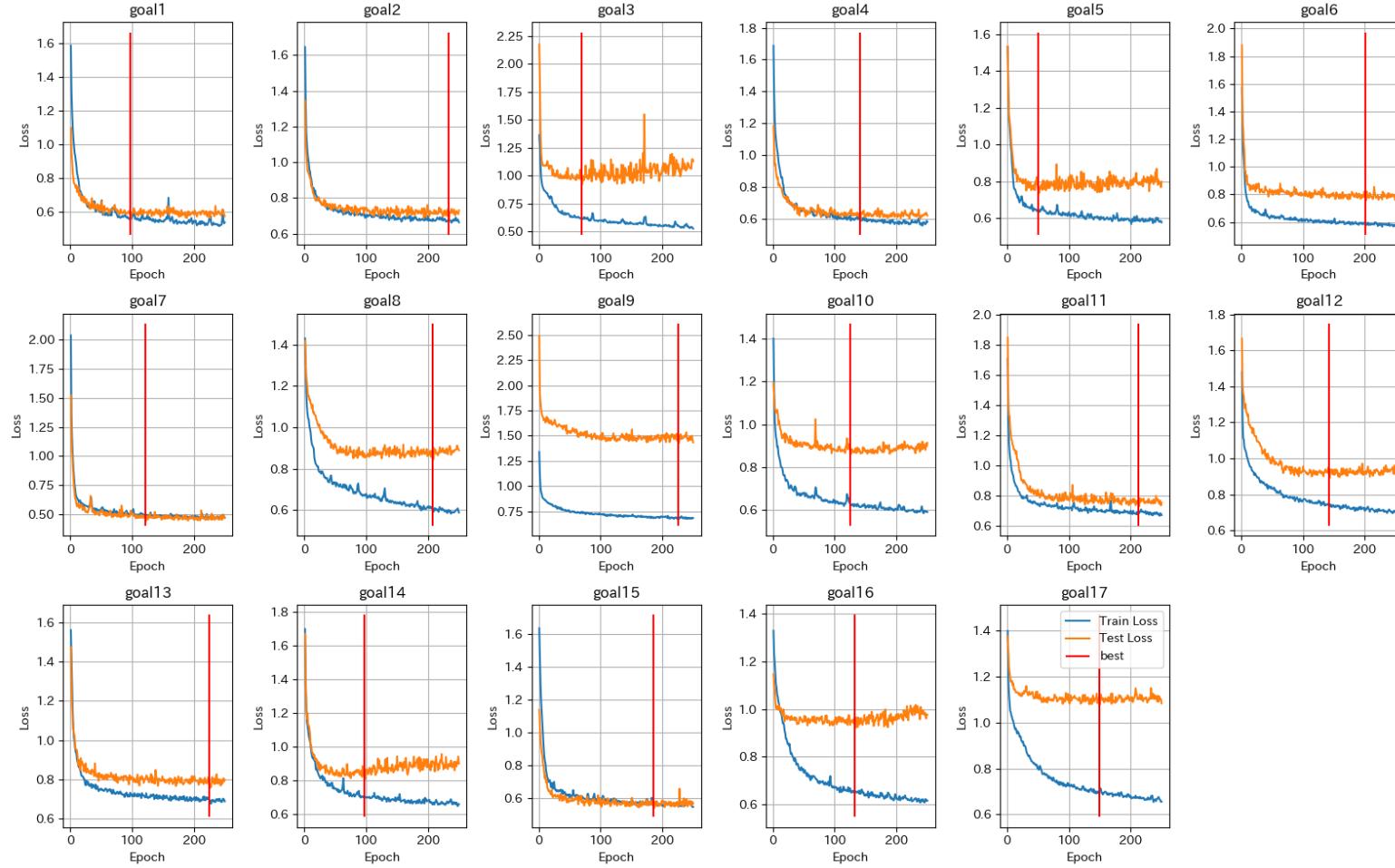
- 労働力人口比率
- 平均余命（女）
- 平均余命（男）
- 非労働力人口（通学）（女）比率



分析結果

VAEの妥当性

Analysis Results

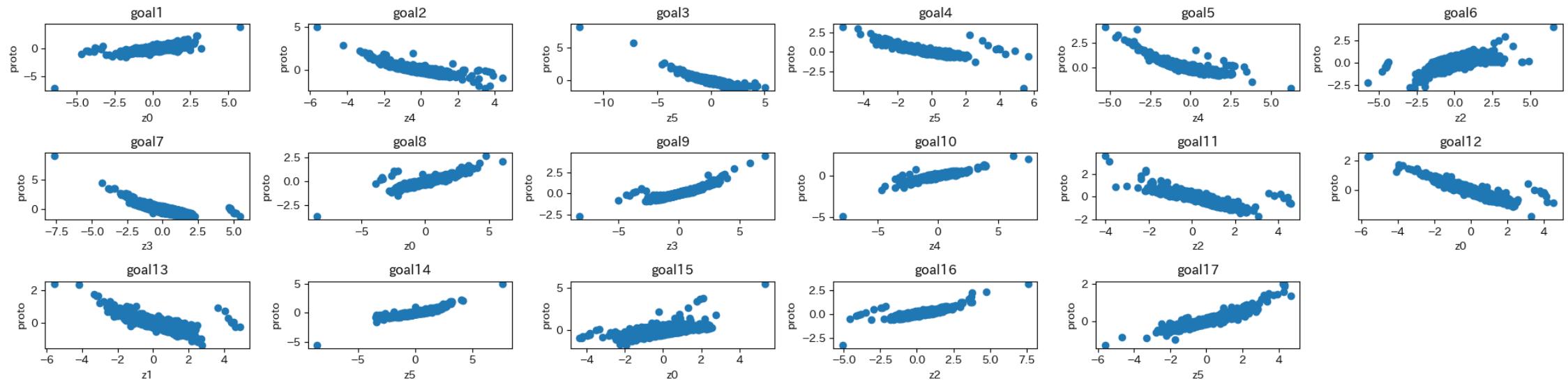


TarinとTestの誤差収束

安定している！

Analysis Results

各ゴールのサブ指標の平均値と潜在空間のある次元の値のプロット

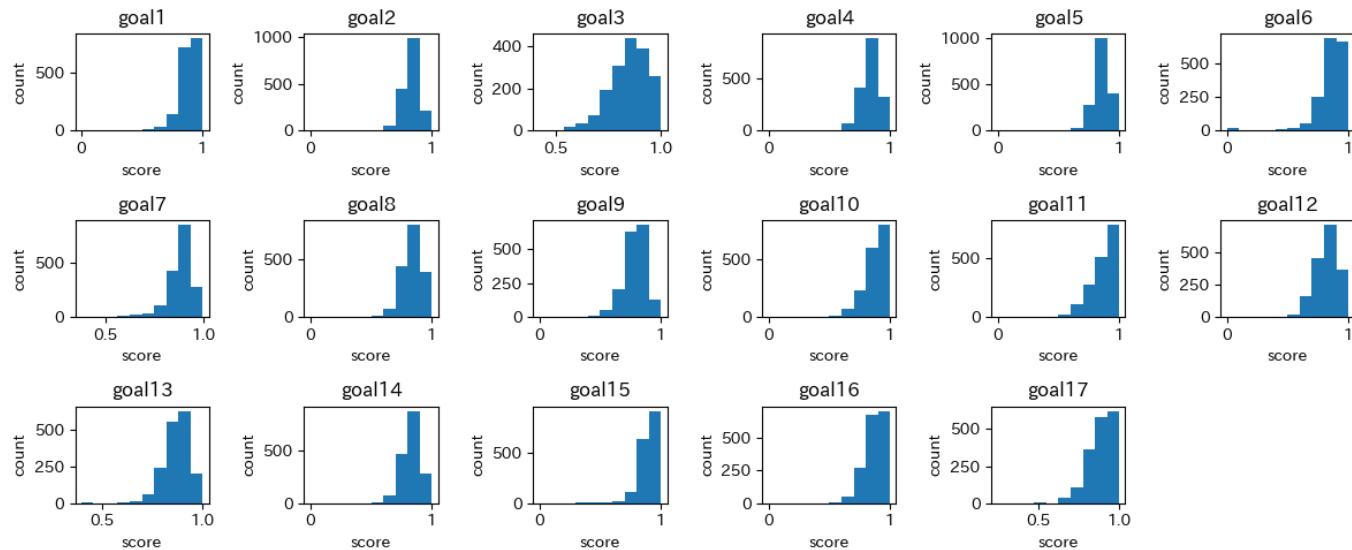


$\mathcal{L}_{proto} = (s - \hat{s})^2$ のおかげで、潜在空間内で少なくとも1つの次元がサブ指標の平均値と**高い相関**を持つ

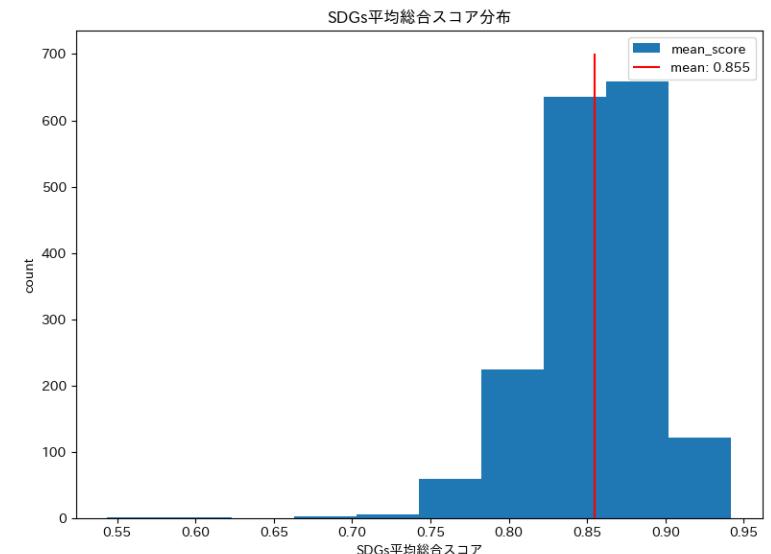
▶ 肌感にあった目標値と達成度の算出ができると期待できる！

Analysis Results

各ゴールのスコア分布



各ゴールの平均スコア分布



個別化した目標値と達成度の評価を行っているが、その分布は自然である

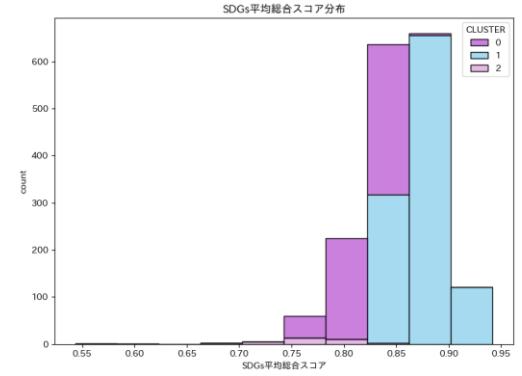
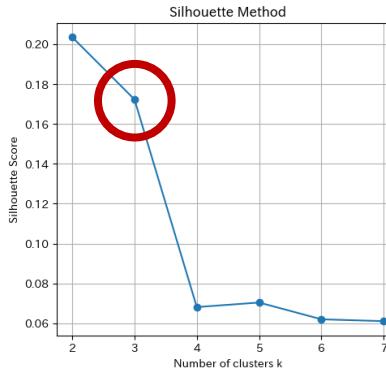
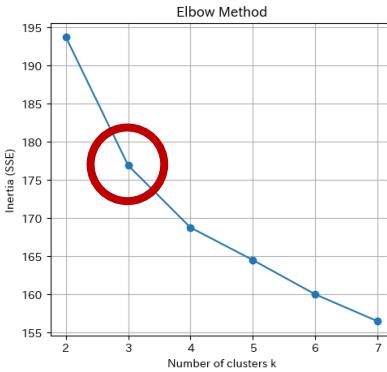
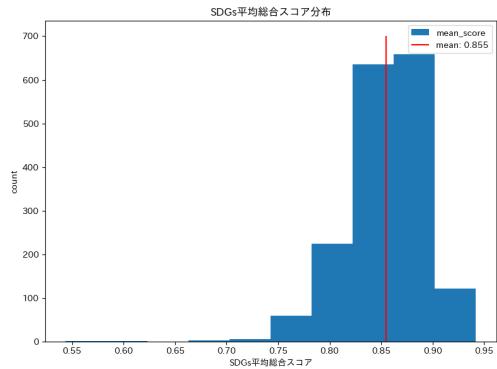


地域差を考慮すると、SDGsは達成に向かっていると言えるのでは？

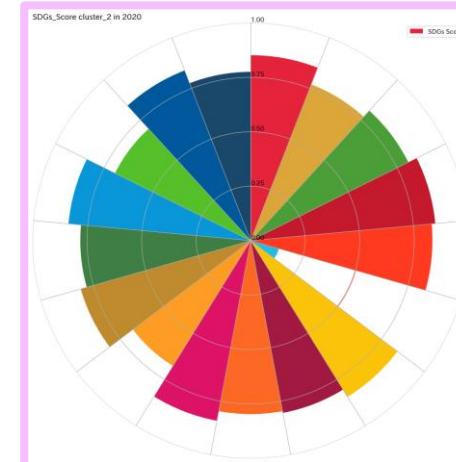
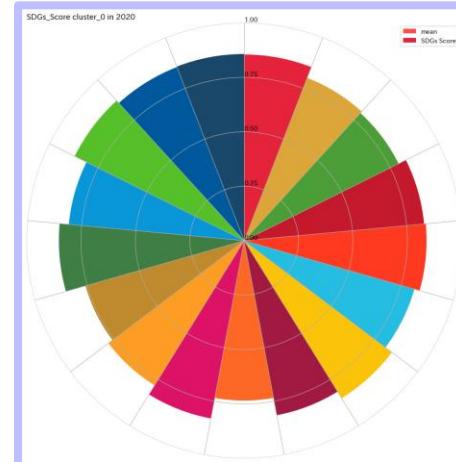
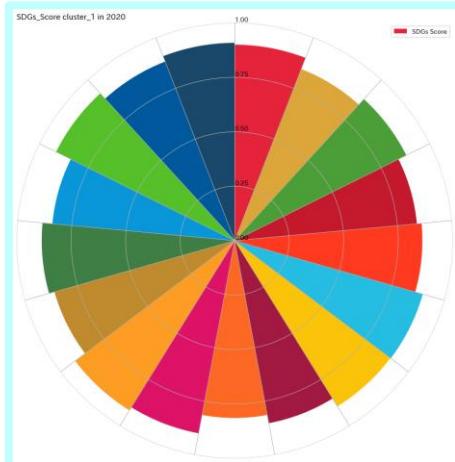
分析結果

SDGs達成度

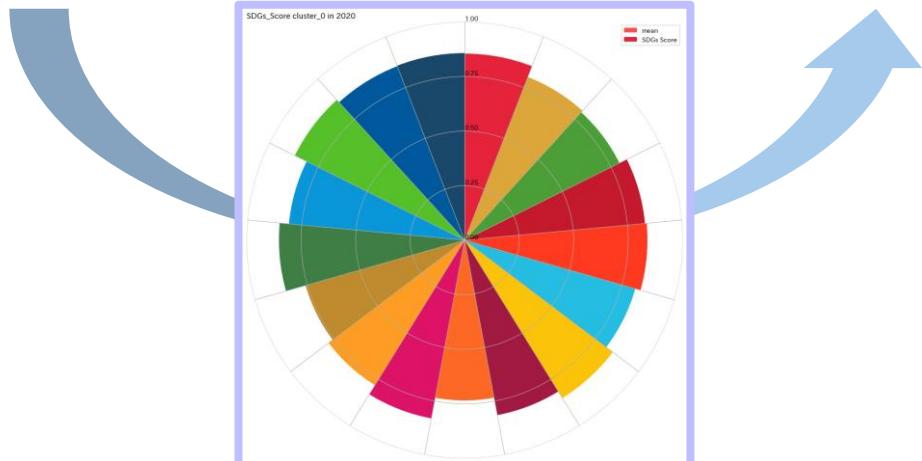
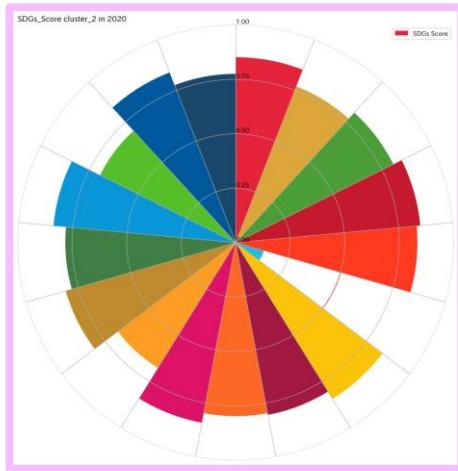
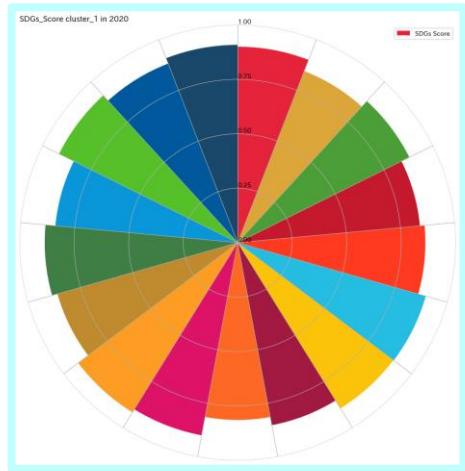
Analysis Results



3つにクラスタリング！



Analysis Results



地域差を考慮した評価
||
都市部と農村部の差が
明確に出ない！



SDGsの達成に
都市度合いは関係なく
個別に向き合う必要がある

#

もくじ

Table of Contents

01 エグゼクティブ
サマリー

02 背景理解

03 分析手順・手法

04 分析結果

05 ダッシュボード

06 指標の利用例

LSI Dashboard

- ✓ 好事例の普遍化を進めるためには、
まず**好事例を知る機会**、
そしてそれを学ぶことのできる**環境づくり**が不可欠である。

【地方創生2.0 2025】基本姿勢・視点より

- ✓ 全国で統一されたSDGs・EBPMの分析基盤の提供

課題把握、好事例の普及化を進めるための
ダッシュボード



lsi-dashboard.org で検索！

ダッシュボード



LSI Dashboard



可視化モード

- ・要因分解
→サブ指標を確認可能
- ・AIインサイト
→各市区町村のSDGs達成度について、
地域性とLSIに基づいて評価

要因分解

SDGsゴールをクリックして、スコア算出に使用された特微量を表示します

4 GOAL 4: 質の高い教育をみんなに
スコア算出に使用された特微量 (7項目)

- 事業所当たり純付加価値額 (教育、学習支援業)
- 婦人保護施設密度
- 従業員1人当たりの純付加価値額 (教育、学習支援業)
- 教員当たり児童・生徒数
- 歳出に占める教育費比率

AI インサイト

分析対象: 兵庫県 姫路市

AI分析結果

姫路市は、多くのゴールで0.8~0.9台と高い達成度を示す一方で、ゴール4「質の高い教育をみんなに」が0.67と17ゴール中で最も低く、全体のボトルネックとなっている市です。基礎的な生活環境や産業・インフラは比較的整っているものの、教育機会の格差や学びの質・継続性など、人への投資面に今後の重点課題が集中している自治体だといえます。

市区町村の説明

姫路市は兵庫県の南西部に位置する、人口約53万人の中核市です。世界遺産の姫路城を中心とした観光業が盛んである一方、瀬戸内海沿岸には鉄鋼や化学などの大規模工場が集まる播磨臨海工業地帯が広がり、製造業が市の経済を支えています。都市機能と豊かな自然、歴史文化が共存する多面的な特徴を持つ都市です。

最も高いゴール

ダッシュボード



LSI Dashboard

地域を選択 / 検索
比較したい市区町村を選択してください

選択 検索 地方: 近畿地方 都道府県: 兵庫県 市区町村: 姫路市

選択 検索 地方: 近畿地方 都道府県: 兵庫県 市区町村: 播磨町



比較モード

- ・任意の市区町村同士を比較
- ・どのゴールが
どれぐらい**強み**で
どれぐらい**弱み**なのかが
一目で分かる！





LSI Dashboard

➤ LSI Dashboardの強み



➤ 日本の全ての市区町村で利用可能



➤ いつでも誰でも利用可能



➤ 可視化・分析・比較に対応した多角的な分析ツール

指標の利用例



Example of Utilize

選択中: 兵庫県 姫路市
市区町村選択に戻る

GOAL 4に関する指標と政策提案

一覧に戻る

4 みんなに
指標の目標値との差

指標	目標値	姫路市
事業所当たり純付加価値額 (教育、学習支援業)	60	25

GOAL 4に関する指標一覧と政策立案

事業所当たり純付加価値額 (教育、学習支援業)

この指標の目標値との差を見る

現在のスコア
25.728
目標スコア
65.47539

政策の提案

- 民間教育事業の高度化支援**
ICT導入補助や教材開発支援により教育事業の付加価値を高める施策。
- 企業・大学連携プログラム創設**
大学と教育事業者が連携し新規教育サービスを開発するための枠組み整備。

参考自治体: 渋谷区 / 品川区
参考自治体: 仙台市 / 千葉市

婦人保護施設数密度

この指標の目標値との差を見る

現在のスコア
0.000167

政策の提案

政策立案モード

LSIが求めた目標に到達するために
・どのサブ指標を
・どの順番で
・どんな施策で
・どれくらい改善
 すれば良いかを提案するモード。

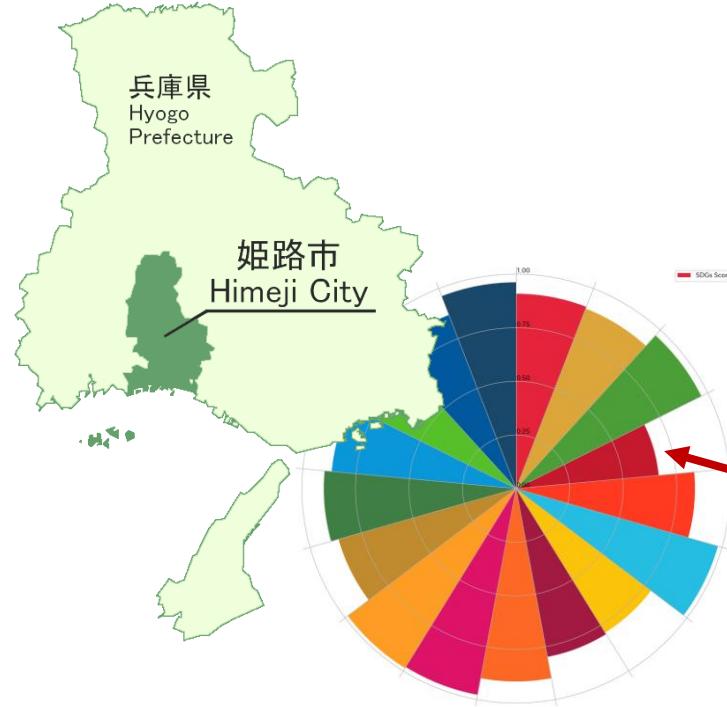
AIを利用した
誰でも使える、強力な政策立案補助ツール

6

指標の利用例



Example of Utilize



姫路市

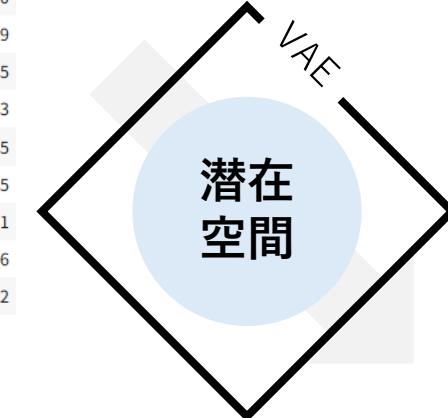
- ・兵庫県第二の人口を誇る中核市
- ・合併を経て都市部から農村部まで多彩な顔を持つ

x_{action}

姫路市	
事業所当たり純付加価値額（教育、学習支援業）	25.728000
婦人保護施設数密度	0.000019
従業者1人当たり純付加価値額（教育、学習支援業）	2.055515
教員当たり児童・生徒数	15.673653
最終学歴人口比率（大学等）	0.269635
未就学児施設密度（人口ベース）	0.005725
歳出に占める教育費比率	0.085771
純付加価値額割合（学術研究、専門・技術サービス業）	0.049566
義務教育学校当たり児童・生徒数	406.601942

ゴール4
「質の良い教育をみんなに」
のスコアが低い

政策立案モードで
改善策を考えよう！



x_{opt}

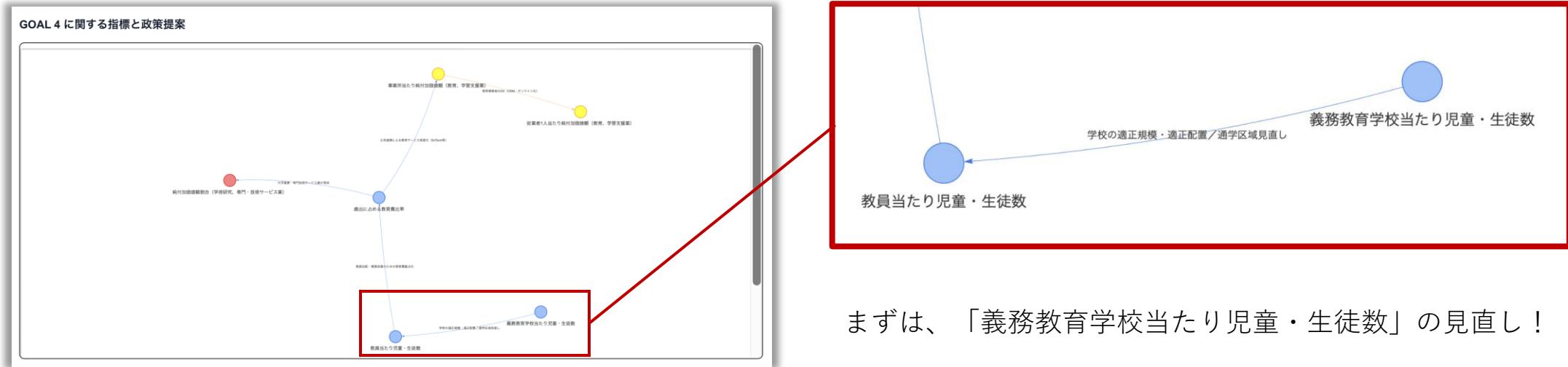
0	
事業所当たり純付加価値額（教育、学習支援業）	65.475388
婦人保護施設数密度	0.000019
従業者1人当たり純付加価値額（教育、学習支援業）	4.064561
教員当たり児童・生徒数	15.470827
最終学歴人口比率（大学等）	0.264218
未就学児施設密度（人口ベース）	0.007572
歳出に占める教育費比率	0.112240
純付加価値額割合（学術研究、専門・技術サービス業）	0.097587
義務教育学校当たり児童・生徒数	406.196158

6

指標の利用例



Example of Utilize



まずは、「義務教育学校当たり児童・生徒数」の見直し！

義務教育学校当たり児童・生徒数	この指標の目標値との差を見る
現在のスコア 406.6019	
目標スコア 406.1962	

目標値に近い市区町村を結び付け、実際の事例を検索し、具体的な政策の提案を行う

指標の利用例



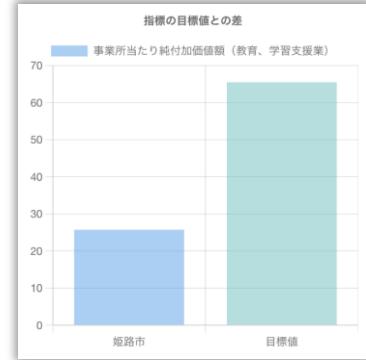
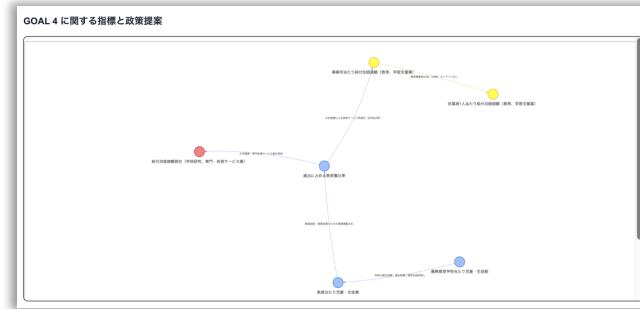
Example of Utilize

KPIの設定

ロジックツリーの作成

類似事例の検索

具体的な政策/施策の提案



義務教育学校当たり児童・生徒数

現在のスコア
406,6019

目標スコア
406,1962

この指標の目標値との差を見る

政策の提案

- ・適正増槽化と学校経営合意書
適大規模校の是正や競合により児童数を適正化し学校運営効率を改善。
- ・学区再編と進学環境整備
学区の見直しと進学支援整備で児童数の偏りを解消する取り組み。

参考自治体: 宍道市 / 千葉市
参考自治体: 平塚市 / 東松山市

データに基づく政策立案：EBPMが
市区町村と対象のゴールを選択するだけ、**2クリック&1ページ**ができる！

データを最大限に有効活用した**持続可能な地域づくりのサポート**を目指す

指標の利用例



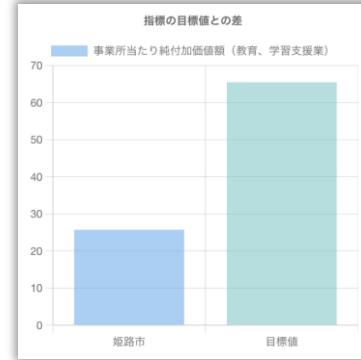
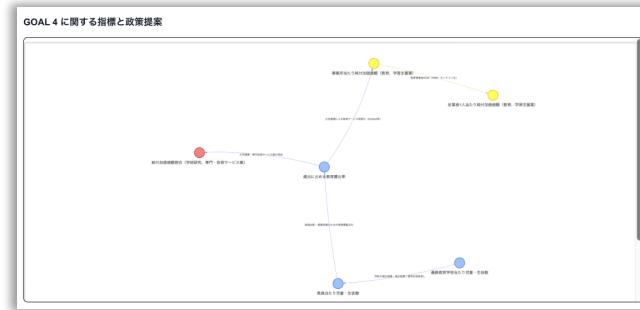
Example of Utilize

KPIの設定

ロジックツリーの作成

類似事例の検索

具体的な政策/施策の提案



義務教育学校当たり児童・生徒数

現在のスコア
406,6019
目標スコア
406,1962

この指標の目標値との差を見る

政策の提案

- ・適正増槽化と学校経営合意書
过大規模校の是正や競合により児童数を適正化し学校運営効率を改善。
- ・学区再編と進学環境整備
学区の見直しと進学支援整備で児童数の偏りを解消する取り組み。

参考自治体: 宮崎市 / 千葉市
参考自治体: 平塚市 / 東松山市

データに基づく政策立案：EBPMが
市区町村と対象のゴールを選択するだけ、**2クリック&1ページ**ができる！

データを最大限に有効活用した**持続可能な地域づくりのサポート**を目指す

今後の展望

References

- LLMの埋め込みベクトルが持つ「意味」の断片性吟味
 - 世の中の全ての「意味」を網羅しているわけでは無い
 - LLMを用いた特徴量選択においては、その限界を精査する必要がある
- VAEの生成する潜在空間の表現性吟味
- スコアの妥当性調査
 - 実際の肌感に合っているかどうか
- ダッシュボードの可用性調査



参考文献

References

- [1] 荒深凌馬. (2019) . ローカルSDGsの達成度を測る指標の開発. 法政大学大学院紀要デザイン工学研究科偏, 8. 1-7. <https://irdb.nii.ac.jp/en/01357/0004129223>
- [2] 自治体SDGs推進評価・調査検討会. (2020) . 令和2年度SDGsに関する全国アンケート調査結果. https://www.chisou.go.jp/tiiki/kankyo/kaigi/dai31/sdgs_hyoka31_shiryo3-1.pdf
- [3] 松本忠. (2019) . 持続可能な開発目標(SDGs)への地域的アプローチ-OECDによる自治体支援プロジェクトの紹介, 都市計画 110(7). 40-50.
- [4] 増原直樹. 岩見麻子. 松井孝典. (2019) . 地域におけるSDGs達成に向けた取組と課題：先進地域における目標・指標設定の傾向. 2019年度環境情報科学研究発表大会. 33. 43-48https://www.jstage.jst.go.jp/article/ceispapers/ceis33/0/ceis33_43/_article/-char/ja/
- [5] 自治体SDGs推進評価・調査検討会. (2024). 令和6年度 SDGsに関する全国アンケート調査結果. https://www.chisou.go.jp/tiiki/kankyo/kaigi/pdf/02_R6_kekka.pdf
- [6] 自治体SDGs推進評価・調査検討会. (2020). 地方創生SDGsローカル指標リスト. https://www.chisou.go.jp/tiiki/kankyo/kaigi/sonota/sdgs_shihyou_risuto_2.pdf
- [7] SUSTAINABLE DEVELOPMENT REPORT. (2024). Sustainable Development Report 2024. <https://s3.amazonaws.com/sustainabledevelopment.report/2024/sustainable-development-report-2024.pdf>
- [8] 西尾弘樹. 尾崎平. (2022). 自治体SDGsの評価指標の提案と適用 -大阪府下43市町村を対象として-. 環境共生. 38(1). 15-24. https://doi.org/10.32313/jahes.38.1_15
- [9] 高井亨. (2020). SDGsの到達度を測る -正射影ベクトルを用いた統合指標作成の試み-. 経済論叢. 194(1). 91-113. <https://doi.org/10.14989/262284>
- [10] 衣川凌太. (2024). 地方創生を推進する多角的複合指標の提案 —SVMに基づく主観的でない変数選択と重み付け手法の検討-. <https://www.nstac.go.jp/sys/sites/3/2024U5-shorei-02.pdf>
- [11] 田辺和俊. 鈴木孝弘. (2019). サポートベクター回帰における感度分析による変数選択の有効性の検証 -都道府県別全死因死亡率の影響要因の分析-. 統計数理(2020). 69(1). 175-192.
- [12] 外務省. (2023). 持続可能な開発目標(SDGs)実施指針改定版. https://www.mofa.go.jp/mofaj/gaiko/oda/sdgs/pdf/000101402_2.pdf
- [13] 総務省統計局. e-Stat. <https://www.e-stat.go.jp/>
- [14] 中部圏SDGs広域プラットフォーム. 地域で取り組むSDGs経営と効果的な情報発信. https://chubu-sdgs-platform.jp/information_dissemination/index.html