

RBSC-SubGenの概要

【RBSC-SubGen⁴の入力】

- 全体集合 U
- ハイパーパラメータ
 - ▶ 部分集合の大きさ S
 - ▶ 望ましいRBSC係数 p^*
 - ▶ 算出される実際のRBSC係数 p と望ましいRBSC係数 p^* との許容誤差 ϵ

【RBSC-SubGenにより得られる結果】

- RBSC係数が所望する値 p^* に十分近い2つの部分集合を構築する
- 反復回数が最大反復回数 M 以下である

⇒構築した2つの部分集合とRBSC係数の実測値 p を返す。

⁴Kohei Furuya et al. "Exploring the limits of an RBSC-based approach in solving the subset selection problem". In: *Proceedings of the International Congress on on Smart Computing and Artificial Intelligence*. Vol. 81. 2022, pp. 1-11. DOI: <https://doi.org/10.29007/11311>.

Webアプリケーションのアーキテクチャ

【使用言語及びツール】

- Python 3.11.1
- Streamlit 1.16.0



図: Webアプリケーションのアーキテクチャ

実行サーバ : Streamlit Cloud

処理 : 所望するRBSC係数 p^* に基づく2つの部分集合を生成する。


作成したWebアプリケーションのオープニングページ

RBSC-SubGen

[How to use?](#)

1. Data upload

Load a CSV data file

 Drag and drop file here
Limit 200MB per file • CSV

[Browse files](#)

2. Input parameters

[Subset size]

[RBSC coefficient]

[Tolerable error]

Insert subset size

Insert RBSC coefficient

Insert tolerable error

0.00 - +

0.00 - +

0.00 - +

⚠ The subset size must be smaller than the universal set size.

Your RBSC coefficient: 0.0

Your tolerable error: 0.0

👉 If you cannot create the expected subset, change Max. number of trials. ▾

3. Visualization parameters

[Number of histogram bins]

Insert number of histogram bins

0.00 - +

⚠ Number of bins must be greater than or equal to 1.

[Run](#)

Webアプリケーションの使用の流れ

入力のステップ

1. Data upload

データセット，
全体集合 U の設定

2. Input parameters

ハイパーパラメータ
の取得

3. Visualization parameter

出力されるヒストグラムの
ビン数の設定

出力

- 各部分集合の，選択した特徴の値の分布を示すヒストグラム
- 生成した部分集合のデータフレーム
- 各項目のRBSC係数の実測値 p
- 処理時間

図: 入力のステップと出力

Webアプリケーションの使用の流れ

1. Data upload

Load a CSV data file



Drag and drop file here

Limit 200MB per file • CSV

1.Data upload

- データセットのアップロード
- カラムの選択
- 全体集合の要素数Uが初期化

Browse files

☒: データセットのアップロード

Webアプリケーションの使用の流れ

Drag and drop file here
Limit 200MB per file • CSV Browse files

diabetes.csv 23.3KB ✕

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Ag
0	6	148	72	35	0	33.6000	0.6270	5
1	1	85	66	29	0	26.6000	0.3510	3
2	8	183	64	0	0	23.3000	0.6720	3
3	1	89	66	23	94	28.1000	0.1670	2
4	0	137	40	35	168	43.1000	2.2880	3
5	5	116	74	0	0	25.6000	0.2010	3
6	3	78	50	32	88	31.0000	0.2480	2
7	10	115	0	0	0	35.3000	0.1340	2
8	2	197	70	45	543	30.5000	0.1580	5
9	8	125	96	0			0.2320	5

Select the columns you want to apply to RBSC-SubGen.

Choose an option

Your number of data points: 768

- データセットのアップロード
- カラムの選択
- 全体集合の要素数Uが初期化

: データセットのアップロードの完了後

Webアプリケーションの使用の流れ



diabetes.csv 23.3KB



	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	6	148	72	35	0	33.6000	0.6270	5
1	1	85	66	29	0	26.6000	0.3510	3
2	8	183	64	0	0	23.3000	0.6720	3

Glucose

BloodPressure

SkinThickness

Insulin

BMI

DiabetesPedigreeFunction

Age

Outcome

Choose an option

1.Data uplod

データセットのアップロード

カラムの選択

全体集合の要素数Uが初期化

図: カラム選択

Webアプリケーションの使用の流れ

The screenshot displays a web application interface with the following elements:

- A blue box at the top right labeled "1.Data upload".
- Text: "Select the columns you want to apply to RBSC-SubGen."
- Two red buttons labeled "BMI ×" and "Glucose ×".
- A light blue box containing the text: "Your number of data points: 768".
- A list of three checked items with red checkmarks:
 - データセットのアップロード
 - カラムの選択
 - 全体集合の要素数 U が初期化
- A small grey box with a close button and a dropdown arrow on the right side.

図: 全体集合 U の初期化

Webアプリケーションの使用の流れ

2. Input parameters

[Subset size]

Insert subset size

25.00

- +

Your Subset size: 25

[RBSC coefficient]

Insert RBSC coefficient of: BMI

0.50

- +

Your RBSC coefficient of BMI:
0.5

Insert RBSC coefficient of: Glucose

0.75

- +

Your RBSC coefficient of
Glucose: 0.75

[Tolerable error]

Insert tolerable error


0.05

- +

Your tolerable error: 0.05

2. Input parameters

- 部分集合の要素数S
- 所望するRBSC係数 p^*
- 許容誤差 ϵ

: 2. Input Parameters

Webアプリケーションの使用の流れ

3. Visualization parameters

[Number of histogram bins]

Insert number of histogram bins

10.00

3. Visualization parameter

- +

ヒストグラムのビン数

Your number of histogram bins: 10

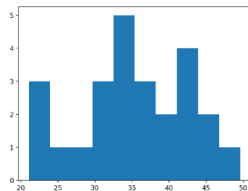
Run

 3. Visualization parameter

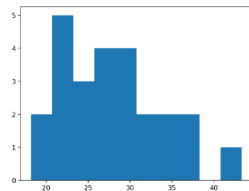
Webアプリケーションの使用の流れ

Histograms of distribution of [BMI]

a histogram of distribution of subset A values

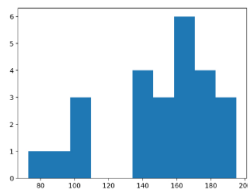


a histogram of distribution of subset B values



Histograms of distribution of [Glucose]

a histogram of distribution of subset A values



a histogram of distribution of subset B values

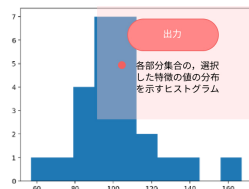


図: 出力されたヒストグラム

Webアプリケーションの使用の流れ

Subsets dataframe

	Pregnancies	Glucose	BloodPressure	SkinThick
0	7	133	84	
1	7	196	90	
2	5	117	92	
3	7	159	64	
4	1	189	60	
5	0	109	88	
6	0	100	88	
7	1	97	66	
8	0	118	84	
9	10	122	78	

Download data A as CSV

	Pregnancies	Glucose	BloodPressure	SkinThick
0	0	180	66	
1	1	95	66	
2	1	71	48	
3	1	85	66	
4	1	68	68	
5	2	101	36	
6	6	92	92	
7	7	105	0	
8	1	97	66	
9	0	101	65	

出力

- 各部分集合の、選択した特徴の値の分布を示すヒストグラム
- 生成した部分集合のデータフレーム

Download data B as CSV

図: 出力されたデータフレーム

Webアプリケーションの使用の流れ

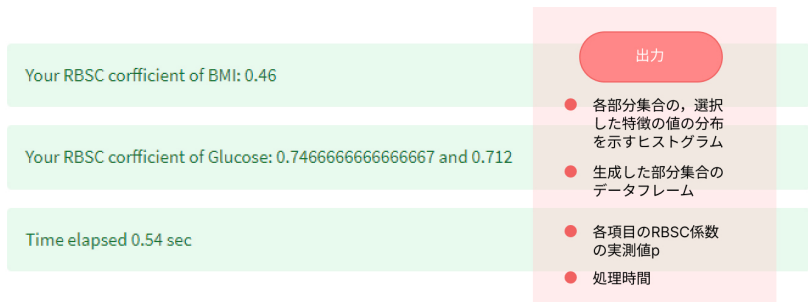


図: 出力されたRBSC係数 p と処理時間