

論文

2100 年の岡山県の環境はどうなっているのか

PuLPを用いた未来予測

岩本 隆志¹⁾・神戸 康弘²⁾

キーワード：SDGs、データサイエンス、Python、環境問題、数理最適問題、線形計画問題、PuLP

1 はじめに

近年、地球環境の変化により、異常気象が各地で起こっている。国連では、このような状況を危惧し、SDGs が発表されている。筆者は現在岡山県在住であり、専門分野は経営工学であるが、ここ数年はAIやRPAの導入を研究している。筆者自身もSDGs への貢献として何ができるのかを考えたとき、データサイエンスを用いた予測が一番身近な取り組みであると考えられる。本研究では、岡山県の過去の気象データを元にして、Pythonの数理最適問題を取り扱うライブラリであるPuLPを用いて、2100年の岡山県の気象状況がどうなっていくのかを予測し、危機意識を高め、SDGs への貢献をめざすことを本研究の目的とする。

2 線形計画問題

線形計画問題とは、目的関数が線形で制約条件が線形等式や線形不等式で表される基本的な最適化問題であり、以下のように定式化されている。

$$\begin{aligned} & \text{minimize } c^T x \\ & \text{subject to } Ax \geq b, \\ & x \in \{0,1\}^n. \end{aligned}$$

2.1 PuLP

PuLP は COIN-OR プロジェクト[注 1]で開発されたモジュールである。COIN-OR Branch and Cutオープンソースのソルバーが含まれており、PuLP を利用することにより、比較的プログラム言語開発経験が少ない人でも、少ないソースコードを記述するだけ

1) 2) 山陽学園大学地域マネジメント学部地域マネジメント学科

で、数理最適化の問題を解くことが可能となる。PuLP が扱うことができるのは実数となる変数と整数となる変数が混在する問題であり、線形計画法や整数計画法で解くことを可能としている。PuLPを用いて、数理最適化問題を解くための流れを以下に示す。

- (1) モデラーで数理モデルを作成
 - (2) ソルバー [注 2] を呼び出し、解を得る
 - (3) ソルバーは、数理モデルを入力値とし、数理モデルを解いて、解を出力
- といった流れとなる [1-4]。

2.2 PuLPを用いた数理最適

ここでは、数理最適化問題として、線形計画問題を扱う。この線形計画問題の解法を利用して、回帰直線を求める流れを以下に示す。線形回帰は 2 つのデータ X と Y の関係を一次式

$$Y = a * X + b \quad (1)$$

で近似する。最終的に傾きの a と y 軸との切片 b を求める。計算対象の元データを X_i, Y_i とすると、初期段階の回帰直線から計算した値 $a * X_i + b$ と実際のデータ Y_i の誤差を Z_i とすると、 Z_i は、

$$|Y_i - (a * X_i) + b| = Z_i \quad (2)$$

で示される。次に目的関数は誤差の合計値を最小にすることとなるので、以下の式で示される。

目的関数:

$$\sum_{i=1}^n Z_i \rightarrow \text{minimize} \quad (3)$$

次に、制約条件はPuLP での制限で絶対値を求める関数 abs が利用不可のため、以下のよう
に+と-に分けて、以下のようになる。

制約条件:

$$Y_i - (a * X_i) \geq -Z_i \quad (4)$$

$$Y_i - (a * X_i) \leq Z_i \quad (5)$$

$$Z_i \geq 0 \quad (6)$$

2.3 相関係数

相関係数は -1 から 1 までの値を取る。相関係数がどの程度の値なら 2 変数のデータ間に相関があるのか、という統一的な基準は決まっていないが、表 1 に示した基準がよく用いられる。本研究では、Pythonのcorrメソッドを利用して、相関係数を計算しているため、ピアソンの積率相関係数を用いている。この係数は、2 つの量的変数間の直線的関連の程度を表す係数であり、一般的に利用されている相関係数のことである[5-6]。

表1 相関係数値と相関 [7]

相関係数 r の値	相関
$-1 \leq r \leq -0.7$	強い負の相関
$-0.7 \leq r \leq -0.4$	負の相関
$-0.4 \leq r \leq -0.2$	弱い負の相関
$-0.2 \leq r \leq 0.2$	ほとんど相関がない
$0.2 \leq r \leq 0.4$	弱い正の相関
$0.4 \leq r \leq 0.7$	正の相関
$0.7 \leq r \leq 1$	強い正の相関

n 組のデータ

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

があり、それぞれの平均を \bar{x}, \bar{y} としたとき、ピアソンの積率相関係数 r_{xy} は、 x の標準偏差を s_x とし、 y の標準偏差を s_y とし、 x と y の共分散を s_{xy} とすると、以下の式で表される。

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{s_{xy}}{s_x s_y} \quad (7)$$

2.4 スピアマン順位相関係数

スピアマンの順位相関係数は、値の大小関係から計算される相関係数である。2群のデータにおける

$$(x_i, y_i) \quad (i = 1, 2, \dots, n) \quad (8)$$

に対し、順位

$$(x_i, y_i) \quad (i = 1, 2, \dots, n) \quad (9)$$

を求め、その順位をもとに相関係数 ρ_{xy} を計算する。ただし、

$$d_i = (x_i - y_i) \quad (i = 1, 2, \dots, n) \quad (10)$$

とすると、

$$\rho_{xy} = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (11)$$

となる [8-9]。

2.5 p値

統計的仮説検定において、帰無仮説[注3]の元で検定統計量はその値となる確率のことで

ある。P値が小さいほど、検定統計量がその値となることはほとんど起こりえないことを意味する。一般的にP値が5%または1%以下の場合に帰無仮説を偽として棄却し、対立仮説を採択する。P値は、帰無仮説が正しい場合に、実際に観察された、あるいはそれ以上の2群の差が観察される確率と言える。具体的には、p値はt値の絶対値と反比例の関係にある。つまり差分が大きいほど、p値は下がる [10-12]。

3 実験

本研究では、岡山県における環境変動を予測することとする。分析元データとして気象庁のホームページ [13] に掲載されている過去の気象データより岡山県のデータを抽出し、元データとし、3STEPで実験を進めた。

STEP1：散布図・回帰直線

雲量・年間日照時間合計・年間霧日数・年間降雪日数・風速・最大降水量・平均気温・海面平均気圧(hPa)・年間平均湿度・年間降水量について、岡山県の2100年の環境状況を散布図と回帰直線で示した。表2から表9に求められた回帰直線付散布図で示す。

STEP2：相関係数算出

次に年間平均気温・最大降水量・年間平均湿度・平均風速・年間降水量・年間日照時間・年間雪日数・年間霜日数について、相関係数を求め、どの様な相関関係があるのかについて計算した。計算結果をヒートマップとして纏めたものを表10に示す。

STEP3：p値算出・相関係数変更

特に、相関係数が大きい項目について、ピアソンの積率相関係数をスピアマン順位相関係数に変えて相関係数とp値を算出。

表2 岡山県における年間日照時間合計の散布図

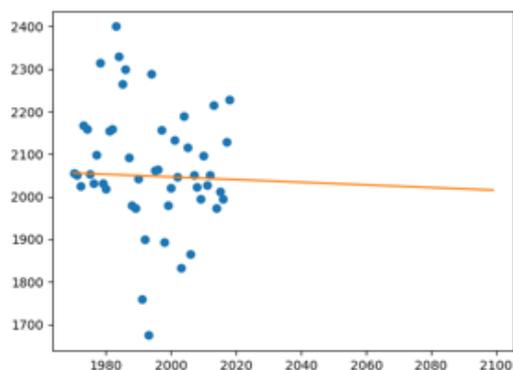


表3 岡山県における年間霧日数の散布図

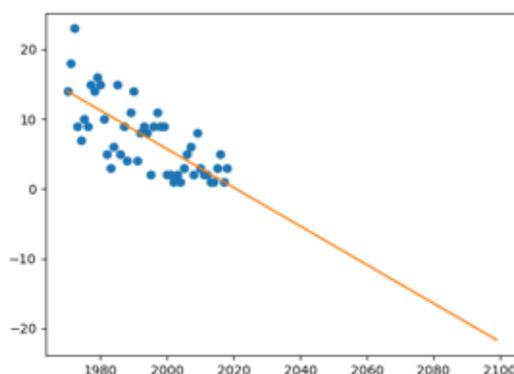


表 4 岡山県における年間降雪日数の
の散布図

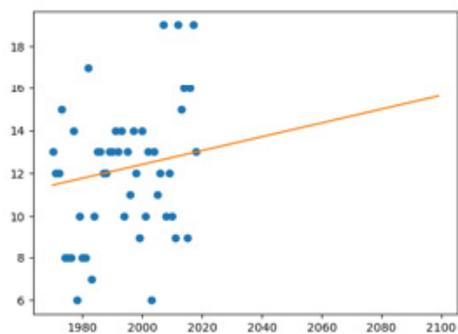


表 5 岡山県における風速の散布図

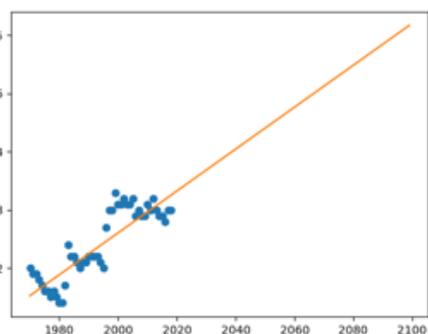


表 6 岡山県における最大降水量
(mm)/日の散布図

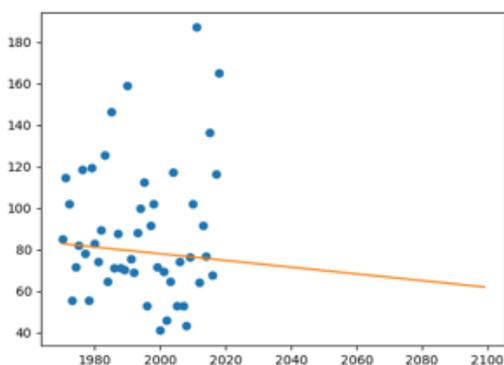


表 7 岡山県における平均気温の散布図

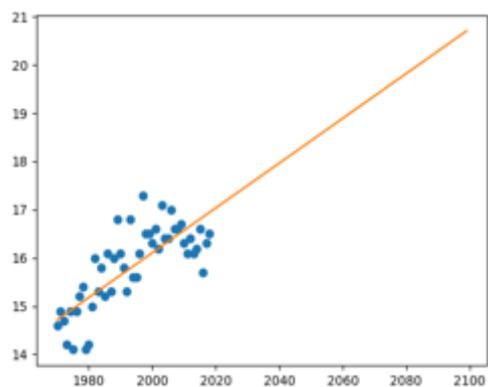


表 8 岡山県における年間平均湿度
の散布図

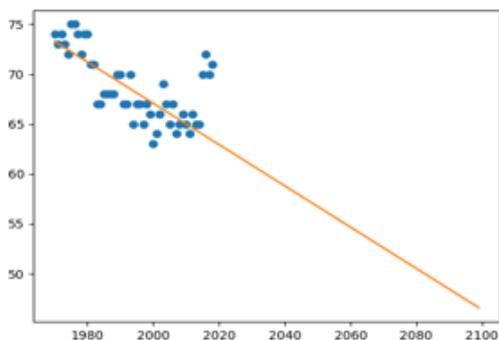
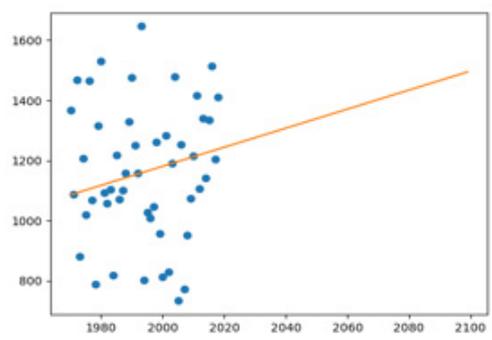


表 9 岡山県における年間降水量(mm)
の散布図



3.1 スピアマン順位相関係数

表 10 において、正の相関があると結果がでた年間降水量と平均湿度、最大降水量と年間降水量の相関係数をスピアマン順位相関係数とp値を用いて算出した結果を表 11 に示す。

有意水準を 0.05 とし、帰無仮説を「2 群間の平均値に差がないこと」、つまり、最大降水量と年間降水量の間とする。表 10 より、 $p > 0.05$ なので、帰無仮説は棄却されず、2 群間は等分散であることが示唆された。つまり、正規分布に従う為、スピアママン順位相関係数ではなくピアソン累積相関係数が有効であると考えられる。

3.2 考察

3 章のSTEP1 で作成した散布図・回帰直線についてそれぞれ見ていくことにする。表 2 から表 9 の散布図を元に 2100 年における岡山県の自然環境を表 12・13 に纏める。つまり、気温については、平均気温が 4.3°C 上昇し、「晴れの国」と呼ばれている岡山県であるが、年間日照時間合計は、44.2 時間減少する。雨については、年間降水量は、523.5 mm と大幅に増大し、最大降水量が 10 mm/ 日の微増であることから、雨量は増えるが、大雨が増えるわけではない。乾燥については、湿度が 26% 減少し、平均風速が 3.3(m/s) 増加し、年間霜日数は 0 になる。次に、積雪についてであるが、年間積雪日数は 6.6 日増えることとなっている。ここで得られた結果どおりになるとも限らないが、一般的に言われている温暖化が進む傾向は、間違いないであろう。更に、湿度が減る傾向にあるが、風速が増すことから、火事等の発生確率が上がる可能性が高い。次に、3 章のSTEP2 について考察する。表 10 のピアソン累積相関係数を各 8 項目について算出し、ヒートマップに纏めた。その結果から、年間平均湿度と年間降水量、及び、最大降水量と年間降水量には、正の相関があることが確認された。

最後に、3 章のSTEP3 について考察する。最大降水量と年間降水量は、当初の表 10 で示したPuLPのデフォルトであるピアソン累積相関係数で算出した。その結果として、正の相関があるとの結果が得られた。だが、表 5 の最大降水量と表 8 の年間降水量の散布図を見ると、ばらつきが激しく、回帰直線の正確性に疑いを持った。そのため、ピアソン累積相関係数では、正規分布を前提条件としているが、その正確性を検証する為、表 11 に示したように、スピアママン順位相関係数での計算を実施し、p値を求める事により、ピアソン累積相関係数で問題ないことを確かめた。

表 10 相関係数のヒートマップ

	年間平均気温 (°C)	最大降水量 (mm)/ 日	年間平均湿度 (%)	平均風速 (m/s)	年間降水量 (mm)	年間日照時間 合計 (h)	年間雪 日数	年間霜 日数
年間平均 気温(°C)	1	0.114354	0.336541	-0.165 466	0.195542	0.298938	0.126722	0.075052
最大降水 量(mm)/日	0.114354	1	0.331215	-0.213 556	0.684103	0.30011	-0.157 912	-0.06 4188
年間平均 湿度(%)	0.336541	0.331215	1	-0.343 213	0.543988	-0.073 664	0.034335	- 0.014515
平均風速 (m/s)	-0.165 466	-0.21 3556	-0.34 3213	1	-0.436 596	0.163756	-0.15554	- 0.076712
年間降水 量(mm)	0.195542	0.684103	0.543988	-0.436 596	1	0.117866	- 0.104251	- 0.153355
年間日照 時間合計 (h)	0.298938	0.30011	-0.073 664	0.163756	0.117866	1	0.316165	-0.237 621
年間雪 日数	0.126722	-0.157 912	0.034335	-0.155 54	-0.10 4251	0.316165	1	-0.059 542
年間霜 日数	0.075052	-0.064 188	-0.014 515	-0.076 712	-0.153 355	-0.237 621	-0.059 542	1

表 11 ピアソン累積相関係数とスピアママン順位相関係数とp値での比較

	最大降水量と年間降水量の相関係数
ピアソン累積相関係数	0.684103
スピアママン順位相関係数	0.564221
p 値	3.7043218

表 12 2019 年と 2100 年の平均気温・最大降水量・年間平均湿度・平均風速の差異

	年間平均気温 (°C)	最大降水量 (mm)/日	年間平均湿度 (%)	平均風速 (m/s)
2019 年	16.5	59	71	2.9
2100 年	20.8	69	45	6.2
差異	4.3	10	-26	3.3

表 13 2019年と2100年の年間降水量・年間日照時間・年間雪日数・年間霜日数の差異

	年間降水量 (mm)	年間日照時間 合計(h)	年間雪 日数	年間霜 日数
2019年	921.5	2065.7	9	2
2100年	1445	2021.5	15.6	0
差異	523.5	-44.2	6.6	-2

3.3 地球温暖化

IPCC [注4] 第5次評価報告書では、20世紀末頃（1986年～2005年）と比べて、有効な温暖化対策をとらなかった場合、21世紀末の世界の平均気温は、2.6～4.8℃上昇、厳しい温暖化対策をとった場合でも0.3～1.7℃上昇する可能性が高くなる。本研究での予測は、4.3℃上昇となっており、IPCCの予測の2.6～4.8℃上昇の範囲内にあり、地域差を考慮しても、大きなズレはないように思われる [14-15]。

3.4 地球乾燥化

この100年で深刻化してきた地球の乾燥化の主な原因は、人間の活動が引き起こした地球温暖化にあると考えられる。地球温暖化に伴う気候システムの変化は、大雨や猛暑日などの極端現象の増加や降雨量の変化といった形で人間社会に影響を与える。その中でも乾燥度の変化は、生態系や農業に大きな影響を与えると考えられており、温暖化の影響評価において極めて重要である。つまり、地球温暖化対策を怠ると、乾燥化が進むこととなり、食料といった生命維持に関わる部分にも大きな影響を及ぼす。地球乾燥化については、温暖化対策の成否により違ってくるため、そのシナリオはIPCCにより示されている。本研究で示した湿度26%減少と明確に比較する対象はないが、乾燥化が岡山県でも進むであろうことは間違いない [16-18]。

4 まとめ

本研究では、数理最適化の問題を解くことが可能となるPythonのモジュールであるPuLPを用いて、岡山県における2100年の環境変動状況を予測することを試みた。このようなAIが行う予測としては、機械学習が有名であるが、PuLPモジュールを用いる事により、数理最適化の問題として対応できることを示した。2100年における岡山県の環境を表12・13に示したが、よく世間で言われる地球温暖化が進むことが間違いないことが理解いただけるであろう。更に、注目すべきは、この試算が正しければではあるが、乾燥が進むということである。筆者も温暖化の影響を正確に理解している訳ではないが、温暖化が進み、雨量は増え、乾燥が進むということは、どういう事なのであろうか。2020年に発生した新型コロナウイルスの影響が今も続いているが、予想もできないようなところに影響がでるのではないであろうか。その為にも国連が提唱するSDGsを本気で皆が考え、大量生産大量消費の考えを正し、人としてのあるべき姿を模索することが求められる時代であると考え。

(注記)

1. 非営利の教育財団であるCOIN-OR Foundation、Inc. により管理され、モデル、アルゴリズム、最先端計算研究の研究開発と、その成果を効率的に展開するための、OR Softwareのオープンソースコミュニティ。

2. 複数の変数を含む数式において、目標とする値を得るための、最適な変数の値を求めることができる機能のこと。
3. 統計的仮説検定の際に立てる仮説で、例えば、帰無仮説として「差がない」という仮説が立てられた場合、これが棄却されることにより、対立仮説の「差がある」が結論となる。
4. 国連気候変動に関する政府間パネル (Intergovernmental Panel on Climate Change) の略。

(引用文献)

- [1] デイップ株式会社：「AI研究は深層学習だけではない」、<https://ainow.ai/202019/03/29/166629/#AI> (2020/11/30閲覧)
- [2] Read the Docs, Inc & contributors、Eric：「PuLPによる最適化」、<https://coin-or.github.io/pulp/> (2020/12/23 閲覧)
- [3] Read the Docs, Inc & contributors、Eric Holscher、Bobby Grace、CharlesLeifer：「PuLP/PuLP dev」、<https://readthedocs.org/projects/pulp-deb/> (2020/12/23 閲覧)
- [4] Read the Docs, Inc & contributors、Eric Holscher、Bobby Grace、CharlesLeifer、Pulp Team.：「PuLP documentation」、<https://docs.pulpproject.org/pulpcore/> (2020/12/14 閲覧)
- [5] 株式会社ビープラウド：「数理最適化とは」、https://www.beproud.jp/business/mathematical_optimization/ (2020/12/15閲覧)
- [6] 日本ニューメリカルアルゴリズムズグループ株式会社：「最適化問題 (数理計画問題)」、<https://www.nag-j.co.jp/naglib/saitekika.htm> (2020/12/13閲覧)
- [7] 株式会社ダイナコム：「有意確率 (p-value)」、https://www.dynacom.co.jp/product_service/packages/snpyalyze/sa_t2_p-value.html (2020/12/11閲覧)
- [8] 株式会社 社会情報サービス：「統計用語集」、<https://bellcurve.jp/statistics/glossary/2172.html> (2020/12/13閲覧)
- [9] 株式会社システムインテグレータ：「P値とQ値 (Vo.11)」、<https://products.sint.co.jp/asia/blog/vol1-11> (2020/12/13閲覧)
- [10] 弘前大学：「-差の検定-」、<https://personal.hs.hirosaki-u.ac.jp/pteiki/research/stat/qa/qadiff.html> (2020/12/13閲覧)
- [11] 気象庁：「過去の気象データ検索」、<http://www.data.jma.go.jp/obd/stats/etrn/index.php> (2020/12/20閲覧)
- [12] 環境省：「地球温暖化の現状」、<https://ondankataisaku.env.go.jp/coolchoice/ondanka/> (2020/12/14閲覧)
- [13] 全国地球温暖化防止活動推進センター (JCCCA)：「温暖化とは？地球温暖化の原因と予測」、https://www.jccca.org/global_warming/knowledge/kno02.html (2020/12/14閲覧)
- [14] WWFジャパン：「地球温暖化が進むとどうなる？その影響は？」、<https://www.wwf.or.jp/activities/basicinfo/1028.html> (2020/12/14閲覧)
- [15] 日本放送協会：「1からわかる！地球温暖化 (3) 温暖化って本当？止められるの？」、[https://www3.nhk.or.jp/news/special/news_seminar/jiji/jiji59/\(2020/12/14閲覧\)](https://www3.nhk.or.jp/news/special/news_seminar/jiji/jiji59/(2020/12/14閲覧))
- [16] 国立環境研究所：「地球温暖化と「水」」、<https://www.cger.nies.go.jp/cgernews/20180>

9/333002.html (2020/12/14閲覧)

[17] 国立環境研究所：「温暖化による全球乾燥度の変化と人為起源の影響を分析～世界の平均気温の上昇を1.5℃に抑えることで、乾燥化を大幅に抑制可能～」、<https://www.nies.go.jp/whatsnew/20200917/20200917.html> (2020/12/14閲覧)

[18] 環境省：「2100年未来の天気予報」、<https://ondankataisaku.env.go.jp/coolchoice/2100weather/> (2020/12/14閲覧)