

ISSN 2758-7118

成城大学

データサイエンス 教育研究センター

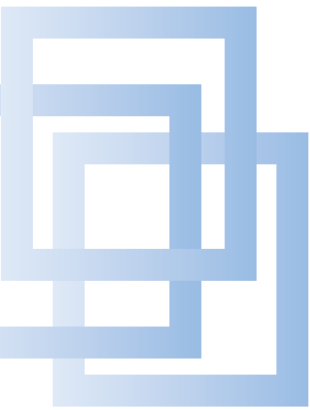
2022年度 年報

Education and Research
Center for
Data-driven
Social
Sciences & Humanities,
Seijo University



Seijo University
**Data
Science**
Since 2015

Annual Report 2022



成城大学データサイエンス教育研究センター
2022 年度 年報

目 次

巻頭挨拶

成城大学データサイエンス教育研究センター長／経済学部教授 小宮路 雅博 1

論文

「パターン認識におけるエラー測度について」

成城大学データサイエンス教育研究センター専任教員／経済学部専任講師 玉谷 充 3

講演録

データサイエンス研究 講演会 (2022 年 7 月 9 日開催) 18

成城大学におけるデータサイエンス教育 講演会 (2022 年 12 月 2 日開催) 37

センターだより

成城大学データサイエンス教育研究センター活動報告等 54

編集後記

成城大学データサイエンス教育研究センター専任教員／経済学部専任講師 玉谷 充 60

文系学生こそデータサイエンスを学ぼう

データサイエンス教育研究センター年報の創刊にあたって

小宮路 雅博

成城大学データサイエンス教育研究センター長／経済学部教授

文系4学部からなる成城大学では、全国の大学に先駆け2015年から全学的なデータサイエンス教育に取り組んできました。当初は共通教育研究センターの中の「データサイエンス科目群」として出発しましたが、2019年4月に専門の管轄部署として「データサイエンス教育研究センター」（以下、センター）が設置されました。

センターは、成城大学におけるデータサイエンス教育と研究を担うべく、その機能を充実させてきました（例えば、2019年11月には学外からも大きな注目を集めたシンポジウム「人文・社会科学系大学におけるデータサイエンス教育」が開催されました）。2020年4月にはそれまでの仮住まい的な4号館から大学新校舎9号館2階に移転し、センター事務室と「データサイエンスワークショップ」、専用教室「データサイエンススクエア」、センター特任教員・専任教員の研究室をそれぞれ設けることができました。

データサイエンスワークショップやデータサイエンススクエアを活用して、データサイエンス教育を展開しようとした矢先に世界的な「コロナ禍」となり、わが国におけるコロナ対応政策の推移に合わせて、本学においても2020年度、2021年度は全学的に遠隔授業となりました。遠隔授業の実施に取り組む中で、センターは2020年夏にはG検定^{*1}講座を遠隔方式で実施するなどして学生の自主的な学修を支援しました。

2021年度になると文部科学省の「数理・データサイエンス・AI教育プログラム認定制度」がスタートし、「リテラシーレベル」認定の申請公募が開始されました。センターも申請を行い、成城大学として同・認定を得ることができました。21年度はデータサイエンス研究講演会や第1回の学内「データサイエンス・コンテスト」も開催されました。また、21年4月からは、文系学生の「数学苦手意識」「統計苦手意識」といったマインドセットを変えるべく「文系学生こそデータサイエンスを学ぼう」の標語を掲げるようにしました。この標語は、新入生ガイダンス動画など学内広報で用いられ、他に、本学データサイエンス教育のブランディングの主要素の1つとして位置づけられて、オープンキャンパスなど入学広報においても様々な場面で一貫して用いられるようになりました。

2022年度は第2回の学内「データサイエンス・コンテスト」を実施すると共に、2回の講演会（それぞれデータサイエンス研究、データサイエンス教育をテーマとする講演会）、夏の体験型ワークショップ3種、授業内コンテストの「音楽プログラミング・コンテスト」が開催されました。G検定講座も21年に続き、3年目の開催となりました。

また、2022年度からデータサイエンスの新カリキュラムが開始されました（22年度新入生から適用）。新カリキュラムは、データサイエンス科目群をそれまでの6科目12単位から16科目32単位に拡大し、「データサイエンス概論」などこれまでの基幹科目に加え、データエンジニアリングや機械学習・AIに関する科目群^{*2}を設置し、データサイエンス教育の特に応用・実践分野への広がり意識して新たに「ワークフロー・プログ

ラム」や「特殊講義Ⅰ～Ⅳ」の科目が設けられました。また、センターの将来の展開を見据えて「数理科学」科目群も開設しました。

新カリキュラムの開始に合わせて、データサイエンス科目履修証明（ディプロマ）制度も変更となりました。データサイエンス科目群には、2015年度の発足時から履修証明制度があり、4科目8単位で「データサイエンス基礎力ディプロマ」、更に2科目4単位で「データサイエンスEMSディプロマ」を発行してきました。新カリキュラムでは、「リテラシーレベル」「応用基礎」「アドバンスド」の3段階の履修証明^{*3}を出すようになり、これに合わせて、2021年度後期分から発行していた（履修証明の）「オープンバッジ」もデザインを変更し、新デザインのオープンバッジが発行されるようになりました。

成城大学のデータサイエンスへの取り組みは元々、全学向けのデータサイエンス科目群の設置と運営という「教育の取り組み」としての展開をしてきました。センターが設立されてからも（上記、記してきたように）教育は依然として大きな任務ではありますが、「教育研究センター」としてデータサイエンス教育のみならず研究においてもその推進と貢献を行っていく必要があります。

センターは、これまで研究をテーマとする講演会を開催するなどして研究面での発信もしてきました。2022年4月に新・特任教授1名とセンター専任教員1名が着任し、23年4月にはもう1名の専任教員が増員されます。この体制整備を受けて、この度、『データサイエンス教育研究センター年報』を創刊することとなりました。

教育と研究とはセンターの両輪であり、両者は別々ではなく切り離すことができないものです。研究が教育実践に反映され、また、（とりわけ文系学生に対する／文系大学における）データサイエンス教育の実践を学生からのフィードバックを踏まえつつ研究のまなざしから捉えていくのも価値あることと考えます。毎年度発行される年報がセンターの当該年度の取り組みと諸活動を記録し、発信すると共に、成城大学におけるデータサイエンス研究の発表・発信の場としても大いに活用されていくことを期待いたします。

^{*1}G検定（DLA Deep Learning for GENERAL）は、一般社団法人「日本ディープラーニング協会」が実施する機械学習に関する検定試験。

^{*2}文部科学省「数理・データサイエンス・AI教育プログラム認定制度（応用基礎レベル）」に対応した科目群。

^{*3}「リテラシーレベル」は「データサイエンス概論」「データサイエンス基礎」の2科目4単位、「応用基礎」は「データアナリティクス基礎」「機械学習基礎」の2科目4単位、「アドバンスド」は「データサイエンス・アドバンスド・プログラム」など6科目から2科目4単位で認定・発行。

2023年3月吉日

パターン認識におけるエラー測度について

成城大学データサイエンス教育研究センター専任教員／経済学部専任講師 玉谷 充

1 はじめに

文字や音声を機械を通じて識別するにはどのようにすればできるのだろうか。例えば、ひらがなの「あ」という文字を認識するだけのために何千何万通りの「あ」の形状をデータとして記憶してしまうと膨大な量となってしまう、一番似ている文字を取り出すのに時間がかかってしまうので非常に効率の悪い認識システムとなってしまう。このような状況を改善するために、パターン認識の手法を取り上げる。分かりやすい例として、自動販売機の硬貨について考える。自動販売機の硬貨は10円から500円の硬貨を扱っているが、これらの硬貨はいずれも直径や重量が異なり、厚みや形状もまた異なっている。投入されたコインは、これらの情報に基づいていくらのコインなのかが判別される。これと同様に、文字や音声においても何らかの特徴というものがあり、その特徴に基づいて判別する手法がパターン認識の一般的な形式である。

パターン認識に関する応用例はいくつかある。一つは先に述べた音声に関する識別で、コンピュータのディスプレイ上に文字を入力する際に、キーボードを使わずに表示させる認識システムである。これは音声の周波数の形状を解析して得られたデータに基づき、文字を判別している。また、セキュリティシステムも応用例として大きく関連しており、個人の認証を指紋や目の虹彩などに基づいて識別する手法にまで現代では発展している。実際例として、島根県の宍道湖で採れるシジミ貝に対し、正常なシジミ貝と泥や水の詰まったシジミ貝を選別する装置が開発されている。そこでは、シジミ貝をコンクリート床にぶつけることにより発せられる音響信号に基づいて正常なシジミ貝とそうでないシジミ貝の判別が実行される。研究報告の詳細については堀内他(2006)を参照されたい。この自動選別装置の精度は約95%と報告されているが、より良い精度の自動選別装置を作るためには、更なる工夫が必要となるであろう。

では、どのようにすれば良い精度の判別ができるのだろうか。本論文では、Devroye et al. (1996)を参考にして基本的な数学モデルを構築し、理論的な観点で話を進めていく。特に、判別の誤りを確率論を用いて評価する手法により、最適な判別手法の構成について概観し、パターン認識に関する様々なエラー測度の関係をベイズルールで得られたエラー測度をもとにまとめることを目標とする。

2 パターン認識

パターン認識とは、クラスへの属性が既知の n 個のデータがあり、どのクラスに属しているかが未知の観測値が新たに得られたときに、 n 個のデータを利用してその新たな観測値のクラスへの属性を決定する行動を意味する。統計科学の枠組みでは、古くから判別分析と呼ばれてきた。

実際には、 M をクラスの数としたとき、クラスを与える方法として、関数

$$g: \mathbb{R}^d \rightarrow \{1, 2, \dots, M\}$$

を定義し、 \mathbb{R}^d における新しい観測値 x に対するクラス割り当てを行う。このような写像 g を“判別関数”
 といい、なるべく $g(x) \neq y$ とならないように g を構成する必要がある。

本章では2値判別における最良の判別関数 g^* の構築について述べていく。

2.1 パターン認識の枠組み

パターン認識における確率変数の組 (X, Y) は、直積集合 $\mathbb{R}^d \times \{0, 1\}$ に値を取るものとする。ここで、 x は
 観測値ベクトル (X の実現値)、 y は x の属すクラス (0 または 1) を表す。このとき、任意の観測値 $x \in \mathbb{R}^d$ に
 対して、

$$\eta(x) = P(Y=1|X=x)$$

と定義し、 $\eta(x)$ は一般に事後確率と呼ばれる。事後確率 $\eta(x)$ は観測値ベクトル $X=x$ が与えられたときにク
 ラスが1である条件付き確率のことであり、以後、観測値 $X=x$ が与えられた条件付き確率密度関数を

$$f_{Y|X}(y|x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)} & f_X(x) > 0, \\ 0 & f_X(x) = 0 \end{cases}$$

と定義する。このとき、条件付き期待値の定義より、 $\eta(x) = E[Y|X=x]$ とも書ける。ちなみに、 Y の取りう
 る値は“0”か“1”の2値であるため、

$$1 - \eta(x) = P(Y=0|X=x) = E[1 - Y|X=x]$$

と表わせる。一方、ボレル可測集合 $A \subseteq \mathbb{R}^d$ に対して、

$$\mu(A) = P(X \in A) = \int_A \mu(dx)$$

とする。また、集合 C を

$$C = C \cap (\mathbb{R}^d \times \{0, 1\}) \stackrel{\text{def}}{=} C_0 \times \{0\} \cup C_1 \times \{1\}$$

としたとき、

$$P((X, Y) \in C) = \int_{C_0} (1 - \eta(x)) \mu(dx) + \int_{C_1} \eta(x) \mu(dx)$$

と表わせる。これにより、確率変数の組 (X, Y) を (μ, η) で表現することができる。

3 バイズルール

3.1 バイズルールの構成

関数 $g: \mathbb{R}^d \rightarrow \{0, 1\}$ を、2値 (クラスが2つの) 判別関数とし、このときの g のエラー確率を

$$L(g) = P(g(X) \neq Y)$$

と表す。判別関数 g^* を事後確率 η を用いて

$$g^*(x) = \begin{cases} 1 & \text{if } \eta(x) > 1/2, \\ 0 & \text{if } \eta(x) \leq 1/2 \end{cases}$$

と定義する。この g^* をベイズ決定関数といい、事後確率 η に基づいてクラスを選択するルールをベイズルールという。この判別関数はエラー確率を最小にする特性をもっている。

定理 3.1. 任意の決定関数 $g: \mathbb{R}^d \rightarrow \{0, 1\}$ に対して、

$$P(g^*(X) \neq Y) \leq P(g(X) \neq Y).$$

また、 $f: \mathbb{R}^d \rightarrow \mathbb{R}$ とし、 Y を X の関数 $f(X)$ で予測するときの平均二乗誤差 $E[\{Y - f(X)\}^2]$ について考える。このとき、

$$E[\{\eta(X) - Y\}^2] \leq E[\{f(X) - Y\}^2]$$

となり、これは η が最良予測値であることを意味している。

以上より、 $\eta(x)$ は予測の誤差やエラー確率を最小にする性質を持っていることが分かる。以下から、ベイズ決定関数 g^* のエラー確率を

$$L^* = \inf_{g: \mathbb{R}^d \rightarrow \{0, 1\}} P(g(X) \neq Y)$$

とし、これをベイズリスク（ベイズエラー）と呼ぶことにする。

一般に、 g のエラー確率については、

$$L(g) = P(g(X) \neq Y) = 1 - E[I_{\{g(X)=1\}} \eta(X) + I_{\{g(X)=0\}} (1 - \eta(X))]$$

と書くことができるため、ベイズリスクもまた以下のように表せる：

$$L^* = 1 - E[I_{\{\eta(X) > 1/2\}} \eta(X) + I_{\{\eta(X) \leq 1/2\}} (1 - \eta(X))].$$

3.2 ベイズリスクにおける公式

ここではベイズリスクの別表現を扱っていく。ベイズリスクは、

$$L^* = E[\min\{\eta(X), 1 - \eta(X)\}] = \frac{1}{2} - \frac{1}{2} E[|2\eta(X) - 1|] \quad (1)$$

と表される。これはよく用いられる表現である。また、 $p = P(Y=1)$ をクラス確率とし、 f_i を X のクラス $Y=i$ での条件付き密度関数とすると、

$$L^* = \int_{\mathbb{R}^d} \min\{\eta(x), 1 - \eta(x)\} f_X(x) dx = \int_{\mathbb{R}^d} \min\{p f_1(x), (1-p) f_0(x)\} dx \quad (2)$$

を得る。このとき、 f_1 と f_0 が重複していない場合、すなわち、 $\int_{\mathbb{R}^d} f_1(x) f_0(x) dx = 0$ ならば $L^* = 0$ となる。さらに、 $p = 1/2$ と仮定すると、(2) 式は以下のように変形できる：

$$L^* = \frac{1}{2} - \frac{1}{4} \int_{\mathbb{R}^d} |f_1(x) - f_0(x)| dx = \frac{1}{2} \int_{\mathbb{R}^d} \{f_1(x) - (f_1(x) - f_0(x))_+\} dx.$$

ここで、 $g_+ = \max\{0, g\}$ とする。したがって、ベイズリスクはクラス密度間の距離に直接関連することが分かる。

3.3 バイズルールの例

3.3.1 学生の合否の予測 1

簡単な例として、学生の単位の予測を合格と不合格で与える。そこで、 $Y=1$ を合格、 $Y=0$ を不合格、そして X を 1 週間あたりの勉強時間とする。このとき、1 週間あたりの勉強時間が多ければ多いほど合格する確率も高くなると通常は考えられるため、回帰関数 $\eta(x) = P(Y=1|X=x)$ は x において単調増加であると言える。1 つの例として、もし $\eta(x) = x/(c+x)$, $c > 0$ であれば、関数 η は $x \geq 0$ に対し、単調増加関数となる。また、バイズ決定関数は

$$g^*(x) = \begin{cases} 1 & \text{if } x > c, \\ 0 & \text{if } x \leq c \end{cases}$$

と構成され、バイズリスクは (1) 式より

$$L^* = E \left[\frac{\min\{c, X\}}{c+X} \right]$$

と書ける。このように、 η からバイズ決定関数を導き出すことができるが、実際には L^* は X の分布がどのような分布であるのかが分からないと求めることができない。もし、1 週間決まって c 時間の勉強を強要させられている軍事学校のような場合、すなわち常に $X=c$ であれば、バイズリスクは $L^* = E[1/2] = 0.5$ となり、判別として非常に悪くなる。もし $X \sim U(0, 4c)$ 、すなわち 1 週間あたりの勉強時間が 0 時間から $4c$ 時間の間で不規則で勉強しているのであれば

$$L^* = \frac{1}{4} \log \frac{5e}{4} \approx 0.3058$$

となり、状況は改善する。

3.3.2 学生の合否の予測 2

先程と同様に、合格する学生を $Y=1$ とし、不合格する学生を $Y=0$ とする。また、勉強時間 X の代わりに新しく 3 つの確率変数を考える。ここでは、 T を学生がテレビを見る平均時間、 B を毎日ビールを消費した平均の数、そして E を怠惰、学習の困難（特別な負の要素）とする。そして Y を

$$Y = \begin{cases} 1 & \text{if } T+B+E < 7, \\ 0 & \text{otherwise} \end{cases}$$

とし、 T, B, E はそれぞれ数値を得ることができるとする。このとき、バイズ決定関数は

$$g^*(T, B, E) = \begin{cases} 1 & \text{if } T+B+E < 7, \\ 0 & \text{otherwise} \end{cases}$$

と構成され、エラー確率は $L^* = P(g^*(T, B, E) \neq Y) = 0$ である。

次に、 E の数値を得ることができない場合について考える。そこで、 T, B, E が *i. i. d.* であり、それぞれの確率密度関数は

$$f(u) = \begin{cases} e^{-u} & u \in [0, \infty), \\ 0 & \text{otherwise} \end{cases}$$

であると仮定する。ここで、 $P(Y=0|T, B)$ と $P(Y=1|T, B)$ を比較する。まず $P(Y=1|T, B)$ については、 $P(Y=1|T, B) = \max\{0, 1 - e^{-(7-T-B)}\}$ となる。一方、 $P(Y=0|T, B) = \min\{1, e^{-(7-T-B)}\}$ であるため、ベイズ決定関数 g^* を

$$g^*(T, B) = \begin{cases} 1 & \text{if } T+B < 7 - \log 2, \\ 0 & \text{otherwise} \end{cases}$$

と構成する。このとき、この判別関数は完全でなく、エラー確率は以下を得る：

$$P(g^*(T, B) \neq Y) = e^{-7} \left((7 - \log 2)^2 + 8 - 2 \log 2 - \frac{7^2}{4} \right) \approx 0.01996.$$

もし、 T (テレビを見る平均時間) だけしか得られなかった場合も、先程と同様に $P(Y=1|T)$ について求めると、

$$P(Y=1|T) = \begin{cases} 1 - (8-T)e^{-(7-T)} & T < 7, \\ 0 & T \geq 7 \end{cases}$$

を得る。そこで、 $1 - (8-T)e^{-(7-T)} = 1/2$ を満たす点を求めると、 $T=c \approx 5.322$ となる。よって、ベイズ決定関数 g^* を、

$$g^*(T) = \begin{cases} 1 & \text{if } T < c, \\ 0 & \text{otherwise} \end{cases}$$

と構成する。このときのエラー確率は、

$$L^* = e^{-7} \left(\frac{8^2-1}{2} - (8-c)^2 + e^{-(c-7)} \right) \approx 0.02707$$

となる。ベイズエラーはわずかに増加したが、多く増加したというわけではない。

最後に、全てが未知の場合を考える。このとき、ベイズ決定関数 g^* を構成することはできない。そこで、

$$P(Y=0) = P(T+B+E \geq 7) \approx 0.02964$$

を計算し、 $g \equiv 1$ と設定すると、エラー確率は $L(g) \approx 0.02964$ となる。

実際には、密度関数が未知であるとき、ベイズ決定関数もまた未知である。そこで、 (T, B) に基づいた判別関数を考えてみる。そこで挙げられるのが Rosenblatt (1958) のパーセプトロンである。この手法は

$$g(T, B) = \begin{cases} 1 & \text{if } aT + bB < c, \\ 0 & \text{otherwise} \end{cases}$$

により構成され、エラー確率の最も小さくなる重み付けの組 $(a, b, c) \in \mathbb{R}^3$ を選択する。この手法の利点として、多くのデータがある場合には、より最適な線形判別関数を選択するのが可能であるという点である。しかし、線形分離が可能ではない場合、例えば $Y = I_{\{T+B^2+E < 7\}}$ のとき、データの組がいくつあったとしても最も良い線形判別関数 g^* は準最適となる。

3.4 プラグイン決定関数

観測値 X から Y の最良の予測はベイズ決定関数であり、この関数は

$$g^*(x) = \begin{cases} 1 & \text{if } \eta(x) > 1/2, \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 1 & \text{if } \eta(x) > 1 - \eta(x), \\ 0 & \text{otherwise} \end{cases}$$

と書ける。しかしながら、関数 η は一般に未知の関数である。そこで、 $\eta(x)$, $1 - \eta(x)$ それぞれの近似非負関数 $\tilde{\eta}(x)$, $1 - \tilde{\eta}(x)$ が得られていると仮定する。このとき、ベイズ決定関数を近似するのに

$$\tilde{g}(x) = \begin{cases} 1 & \text{if } \tilde{\eta}(x) > 1/2, \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

とするのが自然である。この判別関数をプラグイン決定関数という。次のよく知られた定理は、真のエラー確率と似通っているほど、プラグイン決定関数 \tilde{g} のエラー確率は g^* の近くにあることを主張している。

定理 3.2. (3) 式で定義されたプラグイン決定関数 \tilde{g} のエラー確率に対して、

$$P(\tilde{g}(X) \neq Y) - L^* = 2 \int_{\mathbb{R}^d} |\eta(x) - 1/2| U_{\{\tilde{g}(x) \neq g^*(x)\}} \mu(dx)$$

および

$$P(\tilde{g}(X) \neq Y) - L^* \leq 2 \int_{\mathbb{R}^d} |\eta(x) - \tilde{\eta}(x)| \mu(dx) = 2E[|\eta(X) - \tilde{\eta}(X)|]$$

が成り立つ。

$\tilde{\eta}_1(x)$ を $\eta(x)$ の近似関数とし、 $\tilde{\eta}_0(x)$ を $1 - \eta(x)$ の近似関数とする。このとき、プラグイン決定関数 $\tilde{g}(x)$ を

$$\tilde{g}(x) = \begin{cases} 1 & \text{if } \tilde{\eta}_1(x) > \tilde{\eta}_0(x), \\ 0 & \text{otherwise} \end{cases}$$

とし、もし $\tilde{\eta}_0(x) + \tilde{\eta}_1(x)$ が必ずしも 1 と等しくなければ、状況は定理 3.2 の議論とは異なる。しかし、以下の不等式を得ることができる。

定理 3.3. 上で定義された決定のエラー確率は以下の不等式で与えられる：

$$P(\tilde{g}(X) \neq Y) - L^* \leq \int_{\mathbb{R}^d} |(1 - \eta(x)) - \tilde{\eta}_0(x)| \mu(dx) + \int_{\mathbb{R}^d} |\eta(x) - \tilde{\eta}_1(x)| \mu(dx).$$

さらに、クラス条件付きの密度 f_0 , f_1 が存在したとき、ベイズ決定関数は

$$g^*(x) = \begin{cases} 1 & \text{if } pf_1(x) > (1-p)f_0(x), \\ 0 & \text{otherwise} \end{cases}$$

と書ける。このとき、 $f_0(x)$ と $f_1(x)$ が密度 $\tilde{f}_0(x)$ と $\tilde{f}_1(x)$ により近似されているとし、クラス確率 $p = P(Y=1)$ と $1-p = P(Y=0)$ が \tilde{p}_1 と \tilde{p}_0 によってそれぞれ近似されているとする。その場合はプラグイン決定関数は

$$\tilde{g}(x) = \begin{cases} 1 & \text{if } \tilde{p}_1 \tilde{f}_1(x) > \tilde{p}_0 \tilde{f}_0(x), \\ 0 & \text{otherwise.} \end{cases}$$

となり，このときのエラー確率の誤差は以下で表される：

$$L(\tilde{g}) - L^* \leq \int_{\mathbb{R}^d} |(1-p)f_0(x) - \tilde{p}_0 \tilde{f}_0(x)| dx + \int_{\mathbb{R}^d} |pf_1(x) - \tilde{p}_1 \tilde{f}_1(x)| dx.$$

4 測度

4.1 エラー測度

測定には誤差というものが存在し，誤差は判別分析以外でも様々な分野で取り扱われる。前章においても判別の誤差，すなわち誤判別の確率をリスクとしてとらえ，ベイズルールが最小のリスクを与えることを示し，そのリスクをベイズリスクと呼んだのであった。本章では，ベイズリスクをもとに特色のある他の誤差の関係について見ていく。これにより，ベイズリスクとどのくらい離れているのかによって他の判別ルールの誤判別との関係も導くことができる。実際にはエラー確率全体を一般化することもでき，その概念も本章で扱っていく。一方，お互いの分布がどれだけの違いを持っているのかを表す尺度として「ダイバージェンス」というものがある。この尺度はエラー確率の一般形も含んでおり，その繋がりについても見ていく。

以下では，確率変数の組 (X, Y) の分布について特に仮定せずに話を進めていく。

4.1.1 ベイズエラー

2クラスの判別問題において，最小のエラー確率は (1) 式のベイズエラー

$$L^* = E[\min\{\eta(X), 1 - \eta(X)\}]$$

である。この量は前章でも紹介をしたが，エラー確率が最小であると同時に判別問題がどれくらい難しいかを測定する。 L^* はパターン認識のための尺度として役立つ。

4.1.2 コルモゴロフの変動距離

コルモゴロフの変動距離とは，

$$\delta_{KO} = \frac{1}{2} E[|P(Y=1|X) - P(Y=0|X)|] = \frac{1}{2} E[|2\eta(X) - 1|]$$

で定義され，これは2つのクラス間の距離を与え，分布間の距離に影響を与える。 L^* は δ_{KO} を用いて表記すると (1) 式より $L^* = 1/2 - \delta_{KO}$ と表せる。

4.1.3 NN法の漸近誤差

NN法の漸近誤差は

$$L_{NN} = E[2\eta(X)(1 - \eta(X))]$$

と表わせる。NN法とは Nearest Neighbor Rule の略であり，新しい観測値が得られたときにその近傍にある

観測値について、クラスの多数決をとってその新しい観測値のクラスを決定づける手法である¹⁾。一般に、 k 個のクラスから多数決をとる手法を k -Nearest Neighbor Rule という。この手法もまた、エラー確率は完全でなく、なおかつ、極限を用いて表現されるため“漸近誤差”と呼ばれる。一般に、 $2\max\{\eta, 1-\eta\} \geq 1$ として $\min\{\eta, 1-\eta\} \leq 2\eta(1-\eta)$ が成立するため、不等式

$$L^* \leq L_{NN} \leq 2L^*(1-L^*)$$

を得ることができる。このことから、 L_{NN} は L^* について関連していることが分かる。

4.1.4 バタチャリヤ距離

バタチャリヤ距離 (Bhattacharyya (1946)) は $-\log \rho$ であり、 ρ は

$$\rho = E \left[\sqrt{\eta(X)(1-\eta(X))} \right]$$

と書ける。これを Matushita エラーという。 ρ はパターン認識のための距離測度として Matushita (1956) によって定義された。また、数理統計学ではヘリンガー距離という形で現れ、こちらは LeCam (1970) や Beran (1977) を参照されたい。

$\rho=0$ と $P(\eta(X) \in \{0, 1\})=1$ 、すなわち、 $L^*=0$ が同値であり、さらに、 ρ が最大値 $1/2$ を取ることと $P(\eta(X)=1/2)=1$ は同値である。しかしながら、 ρ と L^* との関係は線形でなく、次の定理の不等式より、もし L^* の近似として利用されるならば、全ての分布において ρ より L_{NN} の方が役立つことが主張される：

定理 4.1. 任意の分布に対して、以下の不等式が成立する：

$$\frac{1}{2} - \frac{1}{2} \sqrt{1-4\rho^2} \leq \frac{1}{2} - \frac{1}{2} \sqrt{1-2L_{NN}} \leq L^* \leq L_{NN} \leq \rho.$$

4.1.5 エントロピー

離散的な確率分布 (p_1, p_2, \dots) のエントロピーは

$$H = H(p_1, p_2, \dots) = - \sum_{i=1}^{\infty} p_i \log p_i$$

によって定義される。また、ここでは $0 \log 0 = 0$ とする。エントロピーは Shannon (1948) によって定義された情報理論における基本的な量であり、多くの分野で有益な応用を持つことが知られている。エントロピーの主な特性は以下の3つにまとめられる：

- A. $H \geq 0$ ($p_i=1$ となる $i \in \mathbb{N}$ が存在するとき、等号成立)。
- B. $H(p_1, \dots, p_k) \leq \log k$ ($p_1=p_2=\dots=p_k=1/k$ のとき等号成立)。言い換えると、分布が一様であるとき、エントロピーは最大となる。
- C. ベルヌーイ分布 $B(1, p)$ による2進数のエントロピー

$$H(p, 1-p) = -p \log p - (1-p) \log(1-p)$$

は p において凹関数である。

1) 原著は Cover and Hart (1967) を参照されたい。

ここで1つの例を取り上げる。まず、 X を離散型の確率変数とする。ここで、任意の事象 A に対して、 $X \in A$ であるかどうかの問題について考える。このとき、 X を決定するための期待される最小の質問の回数が N とすると、 N はエントロピー H を用いて

$$\frac{H}{\log 2} \leq N < \frac{H}{\log 2} + 1$$

と表せる。これを可変長符号化定理と呼ぶ。

したがって、 H は X がどのように散らばっているのかを測定するだけでなく、あるアルゴリズムのために具体的なコンピュータの領域を我々に提供する。実際に、上の簡単な例では、 H が最も良いアルゴリズムの予想されたコンピュータの時間に比例している。また、もし観測値 $X=x$ を固定すれば、 Y の分布はベルヌーイ分布 $B(1, \eta(x))$ に従う。よって、観測値 $X=x$ が与えられた Y の条件付きエントロピーは

$$H(\eta(x), 1-\eta(x)) = -\eta(x) \log \eta(x) - (1-\eta(x)) \log(1-\eta(x))$$

である。ここでは期待値

$$\mathcal{E} = E[H(\eta(X), 1-\eta(X))] = -E[\eta(X) \log \eta(X) + (1-\eta(X)) \log(1-\eta(X))]$$

をとり、この量を条件付きエントロピーの期待値とする。このとき、 \mathcal{E} と L^* の関連性として、 $\mathcal{E}=0$ である必要十分条件は $P(\eta(X) \in \{0, 1\}) = 1$ である。

定理 4.2. 条件付きエントロピーの期待値として以下の不等式が成立する：

- (1) $\mathcal{E} \leq H(L^*, 1-L^*) = -L^* \log L^* - (1-L^*) \log(1-L^*)$.
- (2) $\mathcal{E} \geq -\log(1-L_{NN}) \geq -\log(1-L^*)$.
- (3) $\mathcal{E} \leq \log 2 - (1-2L_{NN})/2 \leq \log 2 - (1-2L^*)^2/2$.

4.1.6 ジェフリーダイバージェンス

ジェフリーダイバージェンス (Jeffreys (1998)) とは、カルバック・ライブラーダイバージェンス (Kullback and Leibler (1951))

$$\delta_{KL} = E \left[\eta(X) \log \frac{\eta(X)}{1-\eta(X)} \right]$$

の左右対称の形であり、

$$\mathcal{J} = E \left[(2\eta(X) - 1) \log \frac{\eta(X)}{1-\eta(X)} \right]$$

によって定義される。 \mathcal{J} は以下の性質をもっている：

- (1) 関数 $\varphi(\eta) = (2\eta - 1) \log(\eta/(1-\eta))$ が $1/2$ において左右対称である。
- (2) $\varphi(\eta)$ は凸関数であり、 $\eta = 1/2$ のとき最小値 0 を得る。
- (3) $\lim_{\eta \rightarrow 0} \varphi(\eta) = \infty$, $\lim_{\eta \rightarrow 1} \varphi(\eta) = \infty$ が成り立つ。
- (4) $P(\eta(X) \in \{0, 1\}) > 0$ ならば $\mathcal{J} = \infty$ である。

この性質 (3) より, \mathcal{J} は一般には上に有界でないことに注意しておく。そこで, L^* と \mathcal{J} の関連性について考えていく。このとき, 不等式として

$$\mathcal{J} \geq (1-2L^*) \log \left(\frac{1-L^*}{L^*} \right) \geq 2(1-2L^*)^2 \quad (4)$$

が成立する。これは, (X, Y) の分布を決め, L^* を固定したときに, 下界より上の \mathcal{J} の全ての値が存在することを意味している。そして, \mathcal{J} の定義から, $\mathcal{J}=0$ の必要十分条件は $P(\eta(X)=1/2)=1$, すなわち $L^*=1/2$ である。

また, (4) 式に関連した境界は,

$$\mathcal{J} \geq \sqrt{1-2L_{NN}} \log \left(\frac{1+\sqrt{1-2L_{NN}}}{1-\sqrt{1-2L_{NN}}} \right) \geq 2(1-2L_{NN}) \quad (5)$$

で与えられる。(5) 式の境界は, (4) 式で与えられた L^* の境界より厳密に良い。

4.2 F -Errors

今まで議論した各エラー測度は, $\eta(X)$ の凹関数の期待値に全て関係している。一般に, F が区間 $[0, 1]$ において凹関数であるならば, (X, Y) に対応する F -Error は

$$d_F(X, Y) = E[F(\eta(X))]$$

で定義される。 F -Error の例については, 以下の表 1 にまとめられている:

表 1: F -Error の例

エラー (ダイバージェンス)	記号: 式
(a) ベイズエラー	$L^* : F(x) = \min\{x, 1-x\}$
(b) NN 法の漸近誤差	$L_{NN} : F(x) = 2x(1-x)$
(c) Matushita エラー	$\rho : F(x) = \sqrt{x(1-x)}$
(d) 条件付きエントロピーの期待値	$\mathcal{E} : F(x) = -x \log x - (1-x) \log(1-x)$
(e) 負のジェフリーダイバージェンス	$-\mathcal{J} : F(x) = -(2x-1) \log \frac{x}{1-x}$

このとき, 誤差が最小という意味で良いエラーはベイズエラーであり, 任意の観測値 $x \in [0, 1]$ に対して $F(x) \geq \min\{x, 1-x\}$ とし, F -Error はベイズエラーにおいて上界であることを示唆している。つまり, $F(x)$ がベイズエラーの近くにあるほど $\min\{x, 1-x\}$ になり, より緊密となる。例えば, $F(x) = (1/2) \sin(\pi x) \leq 2x(1-x)$ は, L_{NN} より緊密な上界となっている。また, X が任意の関数によって変換されるならば, エラーの数値が増加する特性をもつ。

定理 4.3. $t: \mathbb{R}^d \rightarrow \mathbb{R}^k$ は任意の可測関数とする。このとき任意の確率変数の組 (X, Y) に対し,

$$d_F(X, Y) \leq d_F(t(X), Y)$$

が成立する。

要するに, F エラーは情報を失う (X を $t(X)$ に取り替える) ほど, そのエラー測度の量も増加することを

述べている。しかし、逆変換が存在するならば等号性を満たす：

補題 4.1. 任意の可測関数 $f: \mathbb{R}^d \rightarrow \mathbb{R}^k$, $g: \mathbb{R}^k \rightarrow \mathbb{R}^d$ に対して、以下が成り立つ：

$$d_F(X, Y) \leq d_F(f(X), Y) \leq d_F((g \circ f)(X), Y).$$

定理 4.4. $t: \mathbb{R}^d \rightarrow \mathbb{R}^d$ は可測関数とする。このとき、逆変換 $t^{-1}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ が存在するならば、

$$d_F(X, Y) = d_F(t(X), Y)$$

が成立する。

4.3 マハラノビス距離

2つの確率変数 X_0 と X_1 をそれぞれクラス 0 と 1 における確率変数とし、 $X_i (i=0, 1)$ の期待値を m_i 、共分散行列を Σ_i とする。このとき、 X_0 と X_1 の間のマハラノビス距離は

$$\Delta = \sqrt{(m_1 - m_0)^T \Sigma^{-1} (m_1 - m_0)}$$

によって与えられる。ここで、 $\Sigma = p\Sigma_1 + (1-p)\Sigma_0$ である。よって、 $\Sigma_1 = \Sigma_0 = \sigma^2 I$ ならば、

$$\Delta = \frac{\|m_1 - m_0\|}{\sigma}$$

と書ける。ただし、 $\|\cdot\|$ はノルムを表している。この形はその平均間の距離に分散を考慮した形であり、Mahalanobis (1936) によって提唱された距離である。そのため、分散が小さければ小さいほどその距離は大きくなることを意味している。また、 $\Sigma_i = \sigma_i^2 I (i=0, 1)$ であるならば、

$$\Delta = \frac{\|m_1 - m_0\|}{\sqrt{p\sigma_1^2 + (1-p)\sigma_0^2}}$$

となり、マハラノビス距離の取りうる値は p が 1 から 0 に変化するのにしたがって、 $\|m_1 - m_0\|/\sigma_1$ と $\|m_1 - m_0\|/\sigma_0$ の間の値を取りうることになる。

一般に、マハラノビス距離 Δ はバイズエラーに関連している。それを主張しているのが以下の定理である：

定理 4.5. $E[\|X\|^2] < \infty$ とし、全ての確率変数の組 (X, Y) に対し、

$$L^* \leq L_{NN} \leq \frac{2p(1-p)}{1+p(1-p)\Delta^2}$$

が成立する。

要するに、条件付きの分布間でのマハラノビス距離が大きいのであれば、 L^* はより小さくなることを表している。また、 X_1 と X_0 の分布が両方とも同じ共分散行列の正規分布である場合は、以下の定理が得られる。

定理 4.6. $Y=i$ が与えられたときの X を X_i とする。 X_i は期待値ベクトル m_i 、共分散行列 Σ の多次元正規分布に従っているとき、以下の等式が成り立つ：

$$\rho = E[\sqrt{\eta(X)(1-\eta(X))}] = \sqrt{p(1-p)} \exp\left(\frac{-\Delta^2}{8}\right).$$

もし、 $(x-m_1)^T \Sigma_1^{-1} (x-m_1)$ と $(x-m_0)^T \Sigma_0^{-1} (x-m_0)$ それぞれの関数として条件付き密度関数 f_1 と f_0 が書かれたなら、Matushita エラーは Δ で表すことができる。しかし、このような分布は一般的というよりむしろ例外である。一般には、 Δ が小さいとき、 L^* が小さいかどうかに関係なく、推論することは困難となる。

4.4 f -divergences

4.2 節では、凹関数 F に対して $F(\eta(X))$ の期待値をとり、これを一般のエラー測度と定義した。これにより、バイズエラー L^* と他のエラー確率の関連づけはより簡潔となった。ここでは、確率測度の間にある距離の古典的な統計理論を、より簡潔に関連づけていく。これらの距離測度に関する一般概念は、 f -divergence と呼ばれるもので与えられる。以前に定義された F -Error は、

$$\mu_i(A) = P(X \in A | Y=i) \quad (i=0, 1)$$

とし、クラスのもとでの X の条件付き分布 μ_0, μ_1 が既知であれば容易に計算される。そこで、クラス確率 p を固定したとき、2つの条件付き分布が互いに遠ければ、 F -Error は小さいことがいえる。この距離の定量化は、次のように定義される：

定義 4.1. $f: [0, \infty) \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$ を $f(1)=0$ となる凸関数とする。 \mathbb{R}^d 上の μ と ν の2つの確率測度の間にある f -divergence は

$$D_f(\mu, \nu) = \sup_{\mathcal{A}=\{A_j\}} \sum_j \nu(A_j) f\left(\frac{\mu(A_j)}{\nu(A_j)}\right)$$

と定義され、上限は全ての有限可測な \mathbb{R}^d の分割 \mathcal{A} でのとる。

もし μ と ν が共に λ に関して絶対連続であり、 $p=d\mu/d\lambda$ と $q=d\nu/d\lambda$ が密度関数であるとき、 f -divergence は

$$D_f(\mu, \nu) = \int q(x) f\left(\frac{p(x)}{q(x)}\right) \lambda(dx)$$

と表現される。また、 f -divergence は $D_f(\mu, \nu) \geq 0$, $D_f(\mu, \mu) = 0$ という性質を持っている。ここでは、 f -divergence の代表的な例をいくつか挙げておく。

4.4.1 全変動

全変動は $f(x) = |x-1|$ としたときであり、

$$V(\mu, \nu) = \sup_{\mathcal{A}=\{A_j\}} \sum_j |\mu(A_j) - \nu(A_j)|$$

で定義される。このダイバージェンスは、2つの確率分布が同じ事象に割り当てることができる確率の最も大きな違いを表している。これは密度関数 p, q によって表すことができる。その主張が次の定理である：

定理 4.7. (Scheffé(1947)) μ, ν を \mathbb{R}^d における絶対連続の確率測度であり, その密度関数を p, q とする。このとき,

$$V(\mu, \nu) = 2 \sup_A |\mu(A) - \nu(A)| = \int |p(x) - q(x)| \lambda(dx)$$

が成立する。ただし, 上限は \mathbb{R}^d 上の全てのボレル集合でとる。

4.4.2 ヘリンガー距離

ヘリンガー距離は $f(x) = (1 - \sqrt{x})^2$ としたときであり,

$$H_2(\mu, \nu) = \sup_{\mathcal{A}=\{A_j\}} 2 \left(1 - \sum_j \sqrt{\mu(A_j)\nu(A_j)} \right) = 2 \left(1 - \int \sqrt{p(x)q(x)} \lambda(dx) \right)$$

と定義される。これは, 分布間の変動距離の尺度を表している。また, Hellinger integral を $I_2(\mu, \nu) = \int \sqrt{p(x)q(x)} \lambda(dx)$ で表す。

4.4.3 カルバック・ライブラーダイバージェンス

カルバック・ライブラーダイバージェンスは $f(x) = x \log x$ としたとき, 以下の形式でまとめられる:

$$I(\mu, \nu) = \sup_{\mathcal{A}=\{A_j\}} \sum_j \mu(A_j) \log \left(\frac{\mu(A_j)}{\nu(A_j)} \right) = \int p(x) \log \left(\frac{p(x)}{q(x)} \right) \lambda(dx).$$

このダイバージェンスもヘリンガー距離と同様, 分布間の距離として情報理論においてよく用いられる量である。しかし, ヘリンガー距離とは違って対称性がないのがこのダイバージェンスの1つの特徴である。

4.4.4 χ^2 -ダイバージェンス

χ^2 -ダイバージェンスは $f(x) = (x-1)^2$ の場合であり,

$$\chi^2(\mu, \nu) = \sup_{\mathcal{A}=\{A_j\}} \sum_j \frac{(\mu(A_j) - \nu(A_j))^2}{\nu(A_j)} = \int \frac{p^2(x)}{q(x)} \lambda(dx) - 1$$

と定義される。

4.5 F -Errors と f -Divergences

最後に, F -Errors と f -divergence の間にある繋がりについてまとめる。まず, $\mu_i (i=0, 1)$ をクラス $Y=i$ が与えられた X の条件付き分布とする。ここで, クラス確率 p を仮定し, F が $[0, 1]$ に関して凹関数であれば, f を

$$f(x) = -F \left(\frac{px}{px + (1-p)} \right) (px + (1-p)) + F(p)$$

としたとき, $d_F(X, Y)$ は

$$d_F(X, Y) = F(p) - D_f(\mu_1, \mu_0)$$

と書ける。特に $p=1/2$ のときは, f を

$$f(x) = -\frac{1}{2} F\left(\frac{x}{1+x}\right) (1+x) + F\left(\frac{1}{2}\right)$$

としたとき,

$$d_F(X, Y) = F\left(\frac{1}{2}\right) - D_f(\mu_1, \mu_0)$$

である。このとき, F が凹関数であるならば, f は凸関数である。これは f が凸関数であるならば, $-f$ は凹関数であることより成立する。また, $f(1)=0$ も満足していることが確認できる。これにより, F -Error はクラス確率 p や X の条件付き分布 μ_i が既知であれば, より早く計算されるものであることが分かる。

最後に, 実際に 4.2 節の F -Error で取り上げられた 5 つのエラー測度を上記の f -divergence に書き換えると, 表 2 を得ることができる :

表 2 : F -Error \rightarrow f -divergence 一覧

エラー測度	$F(x)$	$d_F(X, Y)$
L^*	$\min\{x, 1-x\}$	$\frac{1}{2}(1 - V(p\mu_1, (1-p)\mu_0))$
L_{NN}	$2x(1-x)$	$\int \frac{2p(1-p)p_1(x)p_0(x)}{pp_1(x) + (1-p)p_0(x)} \lambda(dx)$
ρ	$\sqrt{x(1-x)}$	$\sqrt{p(1-p)} I_2(\mu_1, \mu_0)$
\mathcal{E}	$-x \log x - (1-x) \log(1-x)$	$I(p\mu_1 + (1-p)\mu_0, \mu_0) - I(p\mu_1, \mu_0) - I((1-p)\mu_0, \mu_0)$
$-\mathcal{J}$	$-(2x-1) \log \frac{x}{1-x}$	$-I(p\mu_1, (1-p)\mu_0) - I((1-p)\mu_0, p\mu_1)$
エラー測度	$F(x)$	$d_F(X, Y) \quad (p=1/2)$
L^*	$\min\{x, 1-x\}$	$\frac{1}{2} \left(1 - \frac{1}{2} V(\mu_1, \mu_0)\right)$
L_{NN}	$2x(1-x)$	$\int \frac{p_1(x)p_0(x)}{p_1(x) + p_0(x)} \lambda(dx)$
ρ	$\sqrt{x(1-x)}$	$\frac{1}{2} I_2(\mu_1, \mu_0)$
\mathcal{E}	$-x \log x - (1-x) \log(1-x)$	$\frac{1}{2} (I(\mu_0 + \mu_1, \mu_0) - I(\mu_1, \mu_0))$
$-\mathcal{J}$	$(2x-1) \log \frac{x}{1-x}$	$-\frac{1}{2} (I(\mu_1, \mu_0) + I(\mu_0, \mu_1))$

参考文献

- Beran, Rudolf (1977) “Minimum Hellinger distance estimates for parametric models,” *The annals of Statistics*, pp. 445–463.
- Bhattacharyya, Anil (1946) “On a measure of divergence between two multinomial populations,” *Sankhyā: the indian journal of statistics*, pp. 401–406.
- Cover, Thomas and Peter Hart (1967) “Nearest neighbor pattern classification,” *IEEE transactions on information theory*, Vol. 13, No. 1, pp. 21–27.
- Devroye, L., L. Györfi, and G. Lugosi (1996) *A Probabilistic Theory of Pattern Recognition*. Springer.
- Jeffreys, Harold (1998) *The theory of probability*. OUP Oxford.
- Kullback, Solomon and Richard A Leibler (1951) “On information and sufficiency,” *The annals of mathematical statistics*, Vol. 22, No. 1, pp. 79–86.
- LeCam, Lucien (1970) “On the assumptions used to prove asymptotic normality of maximum likelihood estimates,” *The Annals of Mathematical Statistics*, Vol. 41, No. 3, pp. 802–828.
- Mahalanobis, Prasanta Chandra (1936) “On the generalized distance in statistics,” National Institute of Science of India.
- Matushita, K (1956) “Decision rule, based on distance, for the classification problem,” *Annals of the Institute of Statistical Mathematics*, Vol. 8, pp. 67–77.
- Rosenblatt, Frank (1958) “The perceptron: a probabilistic model for information storage and organization in the brain,” *Psychological review*, Vol. 65, No. 6, p. 386.
- Scheffé, Henry (1947) “A useful convergence theorem for probability distributions,” *The Annals of Mathematical Statistics*, Vol. 18, No. 3, pp. 434–438.
- Shannon, Claude Elwood (1948) “A mathematical theory of communication,” *The Bell system technical journal*, Vol. 27, No. 3, pp. 379–423.
- 堀内匡・別府俊幸・藤岡美博・原元司 (2006) 「音響信号による不良シジミ貝判別に関する基礎的検討」, 『電気学会論文誌 C (電子・情報・システム部門誌)』, 第 126 卷, 第 2 号, 297–298 頁.

講演録

成城大学データサイエンス教育研究センター講演会

データサイエンス研究 講演会

2022年7月9日(土) 13:00～14:40
オンライン開催 (Zoom)

Program

開会挨拶 小宮路 雅博 成城大学データサイエンス教育研究センター長／経済学部教授

講演

「ヘルスケアデバイスや質問票を活用した健康状態の可視化に関する研究事例」

森 由美 成城大学データサイエンス教育研究センター特別任用教授

「高次元小標本における判別手法について」

玉谷 充 成城大学データサイエンス教育研究センター専任教員／経済学部専任講師

閉会挨拶 増川 純一 成城大学経済学部長

開会挨拶

小宮路 雅博

成城大学データサイエンス教育研究センター長／経済学部教授

小宮路：皆さま、こんにちは。私は、データサイエンス教育研究センター長を務めております小宮路といたします。

成城大学は4つの学部がありますが、私はその中の経済学部の教員でもあります。本日は当センターの講演会にご参加くださり、誠にありがとうございます。成城大学は、2015年から共通教育の中にデータサイエンスの科目群を設けまして、本学には4学部11学科ありますが、その学生さんたちにデータサイエンス教育を提供してまいりました。その後、データサイエンス教育研究センターが立ち上がり、カリキュラムも変えていきまして、この4月から新しいカリキュラムということで、今年度入学者向けの新カリキュラムでの教育が始まったというところであります。最初は2015年からしばらくの間は、6科目12単位ということでデータサイエンス科目群がありましたが、現在の新カリキュラムでは、16科目32単位で、基礎・入門から割合と上級の、発展的な科目まで取りそろえて、学生の皆さんに提供しているということでございます。

本日の講演会は、本学データサイエンス教育研究センターに所属しております、お2人の教員から、ご自身の研究分野、その一端と言いますか、研究内容についてお話しをしていただくという趣旨でございます。

それでは、只今、画面に出ておりますが、本日の予定はご覧の通りになっております。最初に森先生、本センターの特別任用教授であります。13時5分、このあとから始めていただいて、終わりましたら、若干の質疑応答時間を設けたいと思います。次に目安として13時50分から玉谷先生、当センターの専任講師であります。また同じようにお話しいただいて、質疑応答をし、そして最後に閉会のご挨拶を本学の経済学部長であります増川先生からいただきたいと思っております。では、このようなタイムテーブルでやりたいと思っております。では、先ほど録画しますという表示も出ましたので、録画もしてということでございます。では、まず、森先生、早速よろしく願いいたします。

講演「ヘルスケアデバイスや質問票を活用した健康状態の可視化に関する研究事例」

森 由美

成城大学データサイエンス教育研究センター 特別任用教授

森：皆さま、こんにちは。データサイエンス教育研究センターの森と申します。短い時間ですが、どうぞよろしく願いいたします。では、画面を共有させていただきます。

本日は、「ヘルスケアデバイスや質問票を活用した健康状態の可視化に関する研究事例」ということをご紹介をさせていただきたいと考えております。こちら映っていますのは、成城大学のデータサイエンス教育研究センターにありますデータサイエンススクエアという教室になっておりまして、こちらで今後ともさまざま研究活動をやっていきたいと考えております。

本日の内容です。まず、「健康状態とは」について、今日のタイトルの中に健康状態という言葉がございますが、これについてお話しをさせていただきます。次に健康状態の測定手法の動向について、ヘルスケアデバイスの観点、質問票の観点からご紹介させていただきます。そして、健康状態の調査・可視化の事例として、少し過

去のことになります。私の研究事例についてお話しをさせていただき、最後に今後の展開としてどういった研究活動をしていくかということをお話しさせていただきたいと考えております。

まず、健康状態とは。これは、健康状態というふうに分けられたときに、どのように表現するかということについて考えてみました。まず、履歴書ではどのように書くかということで、転職支援サイトによりますと、履歴書では健康状態について聞かれると、まず、「良好」と書きましょと書かれています。そして、できれば、「極めて良好」と書きましょというふうに書いてあります。そしてさらに、これまで5年間、無遅刻・無欠勤ですという理由をつければ完璧ですという書き方がなされてきました。皆さまの中にも履歴書を書かれたことがある方がたくさんいらっしゃるかと思いますが、皆さまはどのように書かれましたでしょうか。ところで、この「良好」「極めて良好」「これまで5年間、無遅刻・無欠勤」という言葉は、実は危険な言葉になっています。なぜかといいますと、皆さまはこのプレゼンティーズムという言葉をご存じでいらっしゃると思います。最近よく使われるようになってまいりました。これは、健康問題をもちながら出勤している状態を言うということです。そして、最近ではテレワークが増えていまして、健康問題をもちながら勤務している状態、これのことをプレゼンティーズムというふうに使われています。

ところで、このプレゼンティーズムですが、労働生産性の損失という観点では、このアブセンティーズム、こちらは欠勤状態、具合が悪いので休んでいる状態。これに比べてこのプレゼンティーズムという状態は、労働生産性の損失が非常に大きいということで問題になっています。そして、この健康問題をもちながら出勤している、勤務しているという状態、これが健康問題をおろそかにしているという状態になりますが、これによってさらに健康を損なっていく。そして、労働生産性の損失をますます悪くしていくという悪い循環に行ってしまうということで問題になっています。

こちらは、みずほ銀行さまからの資料を少しこちらでお借りしたのになります。このプレゼンティーズムというのが、企業さまにおける従業員の健康関連コストという中でどれぐらいの割合を占めているかという、ほぼ78%が、このプレゼンティーズムが占めているそうです。医療費とか、傷病手当金、労災給付金、アブセンティーズムに比べて非常に大きくなっていることが分かると思います。ですので、履歴書に書くときには、「良好です」「極めて良好です」と書くことは問題ないかもしれないのですが、ご自身の生活の中で多少問題があっても、仕事をしようとか、そういったことではなく、もっと健康問題に向き合っていく必要があるというふうに使われております。

そこで今の健康状態ですけれど、健康状態は大きく分けて3つあるというふうに使われています。健康である状態、そして、病気である状態。それはそれぞれ自覚症状がなく、検査結果の異常もない場合を健康という。自覚症状もあって、検査結果の異常もある。これを病気と呼ぶ。それではということで、ここに「未病」という言葉があります。これは、日本未病学会さんが定義されているのですが、「自覚症状はない、けれども、検査結果の異常がある」または、「自覚症状がある、けれども、検査結果の異常はない」、こ



森 由美 データサイエンス教育研究センター特任教授による講演

ういった状態を未病と呼ぶというふうに定義されています。以前は、健康でもなく、病気でもない状態を未病と呼ぶというような言い方がなされていましたが、最近は、もう少し具体性をもって、こんな言葉で未病というのが説明されています。そこで、皆さまそれぞれご自身が健康か健康でないか、もしくは未病なのかを考えたときに、健康に関連する各種項目を数値化して可視化して、そして現在の状態や変化を把握していくということが今後ますます重要になってくるというふうに言われております。

そこで、その健康状態を知るためにどんな項目があるかということを考えてみましたときに、こちらみずほ銀行さまからの資料になりますが、こんな調査をされました。

医師の方、123人の方に通常の診療時以外にデータを活用するか、活用したことはないが今後活用したいかどうかを、そんな質問をされました。で、そのうちに、データを活用したことはないが、今後活用したいと回答したお医者さまは、約75%だったということです。医療機関で機器を使って検査して、得られたデータも重要だけれども、それ以外に普段の生活で知りたいデータがある。それを使うほうが、より良い診断、治療に向けていけることがあるということですね。そして、ここにさまざまな疾患に対して、どんなデータが重要なのかということが示されています。ここには、例えば血圧、血糖値、心拍数、心電図のように、医療機関に向いていかなないとなかなか測定できないようなデータだったり、または体重、歩数ですとか、食事ですとか、それぞれ個人が測定できそうなものもあつたりいたします。

日常のデータを取ろうとしますと、やはりそれぞれ個人が何かヘルスケアデバイスなどを活用してデータをストックしていくということが今後の流れになるかなと思います。そこで、広い意味でのヘルスケアデバイスについて、少し話をさせていただきます。ヘルスケアデバイス、広い意味でのヘルスケアデバイスを大きく分けると、医療機器、そして、医療機器ではないヘルスケア機器というものがあります。医療機器というのは、厚労省で認定されて、医療機器のクラス1、2、3、4というふうなかたちで、そのリスクの度合に応じて決められたものがございまして、それに当てはまらないもの、そういった認定を受けないものとしてヘルスケア機器というのがあります。最近はこのヘルスケア機器を使って、ご自分の健康を管理される、可視化するという方もどんどん増えてきているかと思えます。そして、そういったデータを取る際に、直接それに関連するデータを測定できるような機器の場合と、そして、直接のデータではないけれども、別のデータを測定して、それを解析して測定値を推定すると、こんな方法でデータを測定するというやり方がございまして。例えば、体温計ですとか、病院で使っているような心電図の測定機器などは、この直接測定に当たりますが、最近ではウェアラブルデバイス、時計型だったり、リストバンド式のウェアラブルデバイスでは心電図に代わるようなデータを取るということも最近ではできてきたりします。そんなふうに直接データを測定できるもの、そして、別のデータで測定値を推定するもの、このような機器がございまして。それから最近言われるようになりましたものとしては、そういったウェアラブル機器について侵襲性があるかないか、拘束性があるかないか、装着をしなければいけないかどうか、そういったことも最近よくテーマとして取り上げられるようになってまいりました。侵襲型か非侵襲型かと言いますと、例えば侵襲型として代表的なものとしましては、血糖値を測定するようなものがあります。ウェアラブル型で侵襲型のデバイスとして、少し針が付いていて、それを体に刺して使うようになる。こちらは侵襲型ということになります。それから、拘束型としては、心電図のデータを取るのに、専用のジェルを使ってびたっと体に張り付けて測定するというようなものもありますし、体のおなかの辺りにベルトで締め付けて、データを測定するというようなものもあります。こういったものは拘束型になるかなと思います。それから装着型。装着型は、そういったジェルを使って、体にびたっと密着させてデータを測定するようなもの。こういった装着型というものがあります。(拘束型よりは装着型と言う方が適切です。)

こういったものに対して、なるべく非侵襲で、非拘束で、非装着であるようなウェアラブルデバイスを活用したいというような要望が、今、どんどん増えてきているかと思えます。ウェアラブルなのに非装着というの

は少しおかしいんですけれども。例えば、靴の中に入れるようなインソール型のものですと、実際は装着しているけれども、装着感のないデバイスというふうに言えるかと思います。そういったものですと、あとはベッドのマットの下にシート型のデバイスを置いておいて、寝ている間中、寝ている間の睡眠状態を計測してくれるようなものも、非侵襲、非拘束、非装着のデバイスになるかと思います。それですとか、椅子のマットも、ベッドのシートと同じようなかたちで、そこで脈拍ですとか、心拍数ですとか、そういったものを測定できる、そんなデバイスもどんどん出てきております。こういった非侵襲、非拘束、非装着型のデバイスであれば、さりげなく使えるものとして、日常のデータ取得に今後ますます望まれていくのかなというふうに考えております。

そこで、このヘルスケアデバイスにつきまして、今、たくさんものが出ておりますので、その中でも最近私がちょっと興味を持ちましたものについて、皆さまにご紹介させていただきたいと思います。こちらは今年の2月のニュースに上がっておりましたものですが、スマホで撮影した顔の動画から、血圧、心拍数などを推定する。こういったバイタル測定サービスも商用提供が開始されたというニュースがございました。これは皆さまのお持ちのスマホで自分の顔の映像を撮影して……。しかも、30秒間撮影した顔の動画を皮膚下の血管内の血流量を特定し、その変動を解析することで、バイタルデータとして血圧、心拍数、ストレスレベルなど、こういった推定値を算出するというものになっております。これはアプリになりますけれども、こういったアプリのメリットとしましては、医療目的には使えないものの、専用デバイスを準備することなく、日々の心身の健康状態を推定できる。利用者の心理的・経済的負担が少ない、こういったメリットがございます。こちらのNTTデータさんも、こういったデバイスを世の中に広げていって、そして企業さんで働く従業員さんの健康を管理していく、そういったことに活用していきたいと、そんなふうに考えられています。この顔の動画からバイタルデータを推定するという手法なんですけれども、今のNTTさんのアプリではないんですけれども、最近映像から脈波情報を抽出するというアプリを使っていました。ちょっとその結果を皆さまにご紹介したいと思います。

これはつい数日前の昼間のデータなんですけれども。これはパソコン上にアプリを入れておいて、そしてパソコンのカメラを使って私の顔を撮影して。それも30秒間撮影して、そしてその映像から心拍数ですとか、それからここにLF/HFとありますが、これはストレス度合です。こういったものを算出するということをちょっと試しにやってみました。そうしますと、数日前は75.9という心拍数になりました。1分間の心拍数ですね。そして、このストレス度合が0.3ということで、この数値が小さいほどストレスが小さいということになります。ところが、つい先ほどの、30分ほど前のこの講演会が始まる直前にデータを取りましたところ、平均心拍数が96.4となっており、ストレス度合も1.577と大きくなっていることが分かります。こんなふうなかたちで、実際にデバイスを自分の体に着けておかなくても、普段使用してるパソコンにこのアプリを入れておくことで、自分の健康度合を可視化していく、測定していくということが可能になる1つの例かなというふうに思います。

この手法なんですけれども、映像脈波抽出技術と呼ばれていまして、こちらは東北大学さんが開発されたものになりますけれども。

血液に含まれるヘモグロビンは、可視光の中の、特に青色や緑色に見える比較的短い波長の光を多く吸収します。そこで、可視光からのこの反射光、これを測定することによって、反射光の中でも特に、緑色の光の成分を測定することによって、血管が収縮しているのか、拡張しているのか。拡張しているほど、このヘモグロビンが多くなりますので、反射光として返ってくる緑色の光は少なくなるんですね。そういったことを活用して、脈波として測定するという手法です。ただ、この手法としては、リストバンド型のデバイスの脈波測定と同じというかたちになっています。リストバンド型のデバイス、この場合は、緑色のLEDを血管に当てて、そして戻ってくる反射光を測定しているんですけれども。原理は同じなんですけど、こちらの映像脈波抽出技術

となりますと、緑色の光を当ててるのではなくて、周辺の光、普通の可視光を当てて、その反射を見るということで、こちらのほうがかなり、演算ですとかアルゴリズムが複雑になっているということが考えられます。また、測定する条件もかなり、実は絞られているのではないかなというふうに考えています。

実は、先ほどのこの結果を見ますと、信頼性指標といって0.03とありますが、実はこれ、ちょっと信頼度が低いということを示しています。恐らく明るさ環境光がちょっと暗かったので、信頼度が低いというふうな結果が出ているかと思いますが。測定するだけではなく、こういった信頼度合も一緒に求められることで、両方のデータを見て、そのデータの精度を考えていく。使ってよいものかどうなのかということを検討しながら、さらに測定するときの環境もちゃんと整えていくことが必要で、そういったことを考えながら測定を進めていくということが大事になってくるかなというふうに思います。

あと、最近気になったデバイスとしましては、これは香川大学の先生が開発されたものですが、爪の微小ひずみを利用した脈波計測というものを研究されています。これは付け爪の中にそのセンサーがもう入っているイメージになりますけれども。こちらは光を使わずに、圧電ピエゾフィルムというものを使われています。脈動するときに、その微妙なピエゾフィルムに掛かる圧力を測定して、そしてそこから脈波を測定するというものになっています。こんなふうに、こちらは装着するタイプのものですけれども、爪という、あまり気にならないところに貼るということで、簡単に生体データが測定されるということで、今後実用化と言いますか、社会にどんなかたちで広まっていくかというところを今後ともウォッチしていきたいというふうに考えております。

ところで、体調の変化に早く気付くためには、平常時の状態を知っておくことが大切ということで、今、ヘルスケアデバイスについてご紹介させていただきましたが、自宅でできる健康チェック方法として、特別なデバイスやアプリがなくても健康状態をチェックできる方法というのが、荒川区のホームページにございましたので、ここに少しご紹介させていただきます。

まず、健康状態を知るために大事なものとして、体温、呼吸数、呼吸の様子、脈拍、血圧、筋肉量というのが挙げられていました。それぞれ、体温は体温計を使って、自分の平熱を知っておくということが大事。

そして、呼吸数は、15秒間の呼吸数を数えて、その値を4倍にすることで1分間の呼吸数を求められます。脈拍も同じで、15秒間の脈拍の数を数えて4倍にすれば、求められます。血圧については、自宅でやはり血圧計を購入して、そして家庭血圧を測定しておきましょうということがよく言われています。あとは、筋肉量ですね。これについては、指輪っかテストというのがあります。自分の両手の親指と中指で輪っかを作って、ふくらはぎの一番太い部分を指でつくった輪っかで囲んだときに、囲めない、ちょうど囲める、隙間ができる、この隙間ができるという場合は、筋肉量が低下しているということで要注意ですよという。こんな簡単なことで、ご自分の健康状態をチェックできるということです。

こういった情報については、ご自分で日々そういったデータを残して行って、そして、何か変化があるかどうかを自分で管理していくということが必要になるかと思いますが、将来的にはこういったところを全てICTを活用して自動化していくことで、さりげない健康チェック、さりげない健康状態の管理ということが可能になっていくのかなというふうに考えます。ただ、デバイスがなくてもこういったことで健康チェックができますよということで、特にお一人住まいの学生さんなどは、こういったかたちで健康チェックをしていただきたいというふうに思います。

今日のお話の中に、ヘルスケアデバイスと、もう一つ質問票についてご紹介させていただくということで、最初にお話をさせていただきました。健康状態を知るためには、客観的にデバイスを使って、そのデータを測定するというやり方が一般的ですけれども、質問票を使うというやり方もあります。ただ、質問票で、何か質問されて、それに対して答えるということについては、その答えは主観的じゃないのでしょうか。それは客観

的な答えが得られないんじゃないでしょうかということをよく聞かれるんですけども。質問票は、医療の分野で古くから使われてきたもので、自覚症状を調査するために非常に有用なものとなっています。この中で少し幾つかご紹介させていただきたいと思います。

まず、一番最初は、FSSG (F スケール) という問診票があります。これは、逆流性食道炎・胃腸の運動不全に関連する問診票となっています。これは2004年に群馬大学の先生が作成されたものとなっています。全部で12問ありまして、それぞれについて、例えば、「胸やけがしますか」について、「ない」「まれに」「時々」「しばしば」「いつも」を答える。そして、その項目別にその点数、スコアを足し合わせて、結果がどうなるかを見ることで、何か問題があるかどうかを確認するという、そういったものになっています。この中で、この1から12番まで全部のスコアを足し合わせると、逆流性食道炎の有無を推定することができると言われていいます。そして、この中でも、緑色の2, 3, 5, 8, 11の5つのスコア、これを足し合わせたときに、機能性ディスぺプシアの有無が推定できるというふうに言われています。この機能性ディスぺプシアについて、あまりなじみのない言葉かもしれませんが、またあとでご紹介したいと思います。

例えば、今、皆さんは、ここでパッと見て、例えばこの2, 3, 5, 8, 11だけでも、それぞれスコアが幾つになるか見ていただいて、それをちょっと足し合わせてみていただきたいと思います。ちょっと10秒ほどお時間を置きたいと思います。2, 3, 5, 8, 11ですね。これをそれぞれのスコアを足し合わせると、運動不全(もたれ)、機能性ディスぺプシアに関連する症状・有無を推定することができるというふうに言われています。

皆さんも足し合わせられましたでしょうか。対面の講演会ですとここで、何点以上の人と言って、手を上げていただくんですけども、それぞれ皆さんは心の中で、何点だったというふうに思っていただけだと思います。実はこれは、4点以上だと、この運動不全の可能性が出てくるというふうに言われています。(4点以上の場合に胃腸の運動不全である可能性が)60%ぐらいあるというふうに言われています。この運動不全ですけども、機能性ディスぺプシアというふうに言われています。これは上腹部を中心に、むかつき、胃痛、胃もたれ、早期満腹感などの症状が慢性的に続く疾患であって、日常生活に支障をきたしているにもかかわらず、内視鏡検査では明らかな異常所見が認められないので診断が難しい。これは消化管の運動機能異常、臓器過敏、ストレス負荷などが原因というふうに言われています。そして、日本人の10%から20%の方には、このような慢性的な不快な症状が経験されているというふうに言われています。皆さまの中で足し合わせて4点以上になった方はいらっしゃいますでしょうか。ただ、内視鏡を使っても異常は見られないんですけども、こういったことで機能性ディスぺプシアの可能性があるとしますと、今はいいお薬があるということですので、もし、悩んでいらっしゃる方がいらっしゃいましたら、消化器系の診療科を受診していただくのがいいかなというふうに思います。以前、こういった問診票の検査等を幾つかの企業さんでやらせていただいたときに、そのときは対面でしたので、何点以上の方、何点以上の方というふうに手を上げていただいたときに、十何点以上で、手を上げた方がいらっしゃったんですね。で、その中に実は、胃がんでしたという方がいらっしゃったので、やはりこういった問診票は普段からでも活用しながら自分の健康を見直していくということにはとてもいいツールかなというふうに思いました。

次に、SF8 質問票というのがあります。これは健康関連 QOL 尺度というふうに言われていまして、身体機能、そして日常役割機能、全体的健康感、活力、社会生活機能、日常役割機能、心の健康、そういった内容で8つの質問事項が含まれています。ただこれは、実は有料の質問紙になっていまして、でもこれは、医療機関ではよく使われる健康関連の QOL 尺度が分かる質問票というふうになっています。ここで求められるスコアとしては、身体的ストレス、精神的ストレスのスコアが求められまして、50が国民標準値として数値化されています。それと対応して、ご自分は今、その平均に対してどれぐらいなのかということを見ることができます。

例えば、このSF8の結果例としましては、横軸に身体的健康、そして縦軸に精神的健康の軸を持ってきて、

ここの50, 50のここが、日本人の平均となりますが、この4つの象限のご自分はどこにいるのかを見ることで、ご自分の健康の健康度合の弱いところなどが数値化されるということになります。また、企業さまのような組織のメンバーの皆さまにこれをしていただくことで、そしてこういった散布図をつくることで、こちらの組織はこの身体的健康度合、精神的健康度合がどの辺りにあるのかということが可視化できるということになります。

そして3つ目に、職業性ストレス簡易調査票というのがあります。これはお勤めの皆さまにはもうなじみの深いものかと思います。2015年12月よりストレスチェック制度が開始され、そこで使われている質問票になります。

この中に幾つか身体の不定愁訴に関する項目がありまして、これを活用したりすることがございます。内容としては、めまいがする、体のふしぶしが痛む、頭が重かったりと、こういった内容になりますけれども、こういったものについて定期的に答えていただいて、健康状態の度合がどういふふうに移りしていくのか、その組織の全体の様子はどうなっているのか。そして、その組織全体としてどういふふうな傾向に時間的推移しているのか、そういったことを見ることが出来ます。

例えば、ある企業さんで、外観検査をされている検査員さんに対して、こういったストレスチェックをしていただいたんですが。そうしますと、よく起こる問題としては、目の疲れ、首筋や肩のこり、腰の痛みというのがあります。4というのが、ほとんどいつもそれが起きるんですね。そして、しばしば、ときどきが黄色というふうになっています。こういったふうに見ますと、特に目の疲れというのについては、外観検査員さんの場合、80%から90%以上の方が、この目の疲れを感じているということが分かります。こういった問診票を使って、今の状態を知って、そして例えば、仕事の環境を変える、仕事のやり方を変える、そういったことで、どんなふうにも目の疲れが改善していくのかを問診票を活用して見ていくことが出来ます。

そこで、健康状態の調査・可視化の事例について、最後に少しご紹介させていただきます。今、外観検査員さんの調査というお話をしましたけれども、コロナ禍以前になんですけども、幾つかの企業さんで健康状態の評価というのをさせていただきました。手法としましては、短期的な調査としては、ウェアラブルセンサーを検査員さんに装着して、作業活動中の生体情報を計測する。それから、長期的なことにつきましては、健康改善効果について、数カ月ごとに問診票調査をするということを行いました。使った問診票は、こういったものになります。

ここで少しだけ皆さまにビデオを見ていただきたいと思いますが、今回、こちらの企業さんに対して行った調査としては、重心動揺、下肢の荷重バランスを測定するデバイスですとか、それから、心拍、脈拍、体動、体表温度が測定できるようなデバイスを活用して、それぞれにデータを取らせていただきました。こういったデータを使って、ストレスですとか作業負荷、衣服内の気候を測定することができます。そして、重心動揺の、この靴型のデバイスを使って、疲労度合ですとか、この検査のリズムを測定することができます。で、それぞれこんなようなデバイスを活用して測定を行いました。心電位センサーで得られるデータとしては、この心拍の間隔、これを活用してストレス度合を測定することができます。そして、衣服内気候の計測ということで、先ほど少しお見せしましたデバイスでは、衣服内の温度と、それから衣服内の湿度を測定することができます。作業中も両方のデータを測定して、こういった衣服内の快適域というものにそのデータを合わせますと、あるときの測定では、非常にやや不快から、不快の間にデータがあったというようなことが分かり、環境をどんなふうにも改善していったらいいかという指標を見つけることが出来ました。

そして、重心動揺の分析器、こういった靴型のデバイスを使いまして、静止をしているときに測定をすることで、疲労度合を測定することができます。20秒間、このデバイスを履いて、静止している間の重心の揺れ具合を測定します。そして、作業前に1回測定する。そして、作業後に1回測定するということで、今、こういったふうにも軌跡が見えていますが、重心が揺らいだ軌跡が見えていますが、これの矩形の面積を測定しまして、

この面積がどれぐらいに、何倍ぐらいになっているかを見ることで、疲労度合（の変化）を見ることができます。

例えばこの場合ですと、矩形面積が2.4倍になっているということで、非常に疲労度合が増しているということが分かったりします。あとは、作業中の測定ですね。このデバイスを履いてもらって作業をしていたんだけど、先ほど少し見ていただきましたビデオの中で、作業の方がすごく揺らぎながら、体を動かしながら、リズムを取りながら、検査をされているところが少し見えたかと思うんですけども、その状態をこんなかたちで測定することができます。非常に上級者の場合ですと、この波形が、左右の動揺の波形が非常にきれいで、そして高速にリズムを取りながら測定ができているということを確認することができました。こんなかたちで、ウェアラブルデバイスを活用して、作業者の方々の健康状態ですとか、それから作業のやり方とか、そういったことを確認することができます。

こちらの場合も、上下がそれぞれ同じ方なんですけれども。作業前と作業後でどんなふうに矩形面積の大きさが変わったかということで疲れ度合を見ることができます。

正しい検査方法をやっている方は、こんなような重心の揺れのグラフが得られるんですけども、まだ初心者の方でうまくできない場合は、こういったような重心動揺になっているということが分かりまして、そして、新しい検査手法の習熟度をこれによって測ることができるということも明らかになりました。

それから、三大身体愁訴、先ほどちょっとお話ししていました、目が疲れる、首筋や肩がこる、腰が痛い、この3つに関して、それぞれ相関があるということも分かりました。これはバブルチャートになっておりますけれども、円が大きいほど人数が多いということを示しています。

少し最後駆け足になりましたが、今後の展開としまして、できるだけ非侵襲で、非拘束で装着感のないデバイスというものを見つけだして、それを活用した日常の生体データの取得手法の確立を進めていきたいと考えております。また、そこから得られたデータを自動収集・解析して、データの利活用ができる仕組みを構築していきたいというふうに考えております。先ほどのように健康状態を可視化するという方法ですとか、作業の習熟度を数値化するということもできます。そして、さまざまなユースケースでの検証、実証実験を行っていききたいというふうに考えています。先ほどのような製造現場であったり、または、教育の現場であったり、それから、いろいろな職業一般にわたって、どんなふうに使えるか、そんな実証実験をやりたいというふうに考えています。こういったことで誰もが簡単に健康状態を把握、行動変容にうながせるような仕組みづくりをしつつ、健康リテラシー教育ということを幼稚園、小学生、中高生も含めたかたちで何か応用しているようなことを考えつつ、これから取り組んでいきたいというふうに考えております。少し時間が長くなってしまいました。ありがとうございました。

小宮路：はい。森先生、お話、ありがとうございました。

森：ありがとうございました。

小宮路：では、質疑の時間をちょっと持ちたいと思います。チャットが入っていますので、チャット、杉山さま、ありがとうございました。はい。スライドで出てきた資料類のURLなどですね。本日の講演会にご参加いただいた方、あと、ご参加を表明されていましたが、用務で来られなくなった方には、期間限定ではありますが、動画としてお知らせする予定ではありますので、そこで確認することができます。では、フロアからの質疑の時間を少し持ちたいと思います。森先生にご質問のある方は、挙手をお願いします。こちらで見つけて、ご指名したいと思いますけど、いかがでしょうか。

すぐにはないようですので、森先生、では、私から。この最後のところで、習熟してくると、体が自然に揺れるというか、リズムを取るというか。これって、日常的にも我々、いろいろなところで自覚してたり、自覚してなかったりしますけど。でも、そういうことってありますよね。

森：はい、ございますね。

小宮路:何かの作業なり, いろんな身体運動, あとは, 勉強したり, 研究したり, 何かに集中しているときも, 割とそういうことありますよね。

森: はい。

小宮路: これって, 森先生, 医学のほうもご専門なので, 伺いたいのですが, 幅広くこういう現象って認識されているというか, 測定されていたり, 研究されているものなのでしょうか。

森: そうですね。リズム運動とそのメンタルとの関係については, いろいろ研究結果が出ていまして, リズム運動, 自然なリズムを取ることで, セロトニンが分泌されるので。セロトニンというのは, ストレスに耐えられる強いようなホルモンですけれども。そういったセロトニンが分泌されることで, 例えばいらいらした気持ちが静まったりですとか, 例えば, 先ほどのような検査の作業っていうのは, ずっと同じことをしているので, 割と初めての方が見たら, 飽きてしまうんじゃないかと思うんですけども, ああいうリズム運動を取ることで, あの方々は楽しくされているんですね。そのセロトニンが分泌することで, そういうメンタルにいい効果があるということで。今後とも先ほどみたいなデバイスを活用して, いろいろなケースによって, そのリズムの取り方って変わってくると思うんですけども, そういったことも調べてみたいなというふうに考えております。

小宮路: 分かりました。何か身体運動と精神活動が連動しているというか, 割合そういう場面ってありますよね。こういうことをしていると, 体がこういうふうに動くとかです。はい。チャットが入りました。森先生, ご覧になれますか?

森: はい。

小宮路: ちょっとお答えいただけると。

森: ご質問いただきまして, ありがとうございます。身体計測などデータの取り込みをできるキットなど……。実は先ほど出てきた中に, もう販売していないものも含まれたりしてるんですけども。別途メール等でお知らせさせていただきたいと思っております。ご質問ありがとうございます。

小宮路: よろしいでしょうか。それでは, 森先生, ありがとうございます。

森: ありがとうございます。

講演「高次元小標本における判別手法について」

玉谷 充

成城大学データサイエンス教育研究センター専任教員／経済学部専任講師

小宮路: 皆さま, ありがとうございます。では, 次です。次は, 玉谷先生より, 「高次元小標本における判別手法」というタイトルでご講演をお願いいたします。では, 玉谷先生, お願いします。

玉谷: 成城大学データサイエンス教育研究センターの玉谷です。今年の4月からこちらのほうのデータサイエンス教育研究センターの専任教員としてやっているということで, よろしくをお願いいたします。本日なんですけれども, ちょっと理論的な部分が主となるんですけども, 「高次元小標本における判別手法」という内容についてお話をしていきたいというふうに考えていますので, よろしくをお願いいたします。画面, 見えていますか。大丈夫ですか。

小宮路: 見えています。

玉谷: ありがとうございます。では, よろしく申し上げます。今日の内容なんですけれども, いろいろと概念

図とかそういうのもちょっと踏まえたもとの、高次元小標本の話はどういうふうを持ちかけていくのかというところを見ていこうかなというふうに思います。

まず、パターン認識に関する内容なんですけれども。

従来パターン認識というのは、画像とか音声とか、心電図とか文字とか、そういったいろいろな信号とかそういったものをパターンとして読み取って、そこからいろいろとやっていくわけなんですけれども。実際には、そのパターンをまずはそのデータとして抽出をしてやって、そこから前処理だったりとか、特徴抽出だったりとか、そういったものを経て、1つのデータというものを得るわけです。で、この得られたデータというのを実際に判別というものを通じて、このデータはどのクラスなのかというのをやっていくような、そういった手法をここでは、この判別という枠組みがありますけれども、ここを中心に見ていこうかなというふうに考えています。

これはちょっと1つの具体的な例になりますけれども。例えば今、ここにデータがありますと。で、これはどういったデータがあるかと言ったら、Aのがんの人のデータとか、Bのがんの人のデータとか、5種類のがんのデータがあったとしましょう。そのときに、今、ここではこのデータを用いて判別関数というものをつくってやるということを考えてやります。ここにどのがんになってるか分からない人のデータがあったとしたときに、この人のそのデータを判別関数というものに放り込んでやったときに、そのときに5種類のがんのうちのいずれかは返すようになる。そういったものをここでは考えたいということになります。

そこで、本研究の目的なんですけれども、まず1つの目的としては、この x というのは、データということにしておいて。それで、このCl（シーエル）というのをクラスということ。例えば先ほどの例だったら、AからEのがんがありますけれども、その場合だったら、C1（シーワン）からC5をそれぞれAがんからEがんというようなかたちで与えるような、そういったようなかたちで考えたときに、判別関数として今、このデータというのを放り込んだときに、このC1からCkのいずれかを返すような、そういった判別関数を考えたいわけなんですけれども。理想的なものとしては、このgの精度をなるべく良くしたいというのが、これが1つの研究の目的として挙げられます。じゃあ、実際、その精度を良くしていくにはどうしたらいいのかということなんですけれども。ここも一つ、もうちょっと別の例で見ていきたいと思います。

例えば、こういったかたちでデータが与えられていると。今、ここでは、青と赤のデータが与えられていて、この白丸というのが、先ほどの例で言うところの、どのがんに属しているのか分からない人のデータというようなかたちで見てくれればよろしいかと思ます。このときに、ここから実際にその判別をするためにはどうすればいいかといいますと、ここでは、1つ、こういった線というものを引いて考えたりすることになります。

例えば、実線の場合とかだったら、これ、今、曲線で与えられていますけれども、曲線で与えることによって、その赤と青をうまく識別をするということも、これもできるわけですね。で、今、この実線のところで与えられているものなんですけれども、これについては、今、こういうように境界線を与えたら、この白

多クラス判別関数

$$\hat{y} = [y_1, \dots, y_{K+1}] = \hat{D}^{-1/2} (\mu_1, \dots, \mu_{K+1}) = \hat{D}^{-1/2} \hat{\mu}$$

Native Canonical Correlation に基づく多クラス判別関数

- $\hat{y}: \mathbb{R}^K \rightarrow \{1, \dots, K\}$
- $\hat{y}(x) = \arg \min_{1 \leq k \leq K} (x - \hat{\mu}_k)^T \hat{D}^{-1} (x - \hat{\mu}_k)$
- $K=2 \implies$ Native Fisher's Rule (Sample var: $\hat{E} \rightarrow \hat{D}$) に帰着

本講義の目標

- 高次元小標本漸近理論のもとで \hat{y} と \hat{D} の漸近挙動を導出する
- 高次元小標本漸近理論のもとで判別関数の漸近上界を導出する

本研究の目的

- $x \in \mathbb{R}^D$: 観測値ベクトル (データ)
- C_k : クラス k : $C_1 = "Aがん", \dots, C_K = "Eがん"$

判別関数 \hat{y}

$$\hat{y}: \mathbb{R}^D \rightarrow \{C_1, \dots, C_K\}$$

\hat{y} の精度を良くしたい!

玉谷 充 データサイエンス教育研究センター専任教員/経済学部専任講師による講演

丸というのは、この赤に属しているというふうなかたちで認識というか、クラスを与えることができるということが分かります。一方で、今、点線で与えていますけれども、この点線で与えるようなかたち、こういうのは直線で与えるというふうなかたちの方法もありまして、このような場合、今のこの例だと、今、この線の上に来ているということなので、この白丸は青で返されるというふうなかたちで考えることができるわけですが、けれども。

一般的な話で言うと、この曲線で判別の境界を引くという方法は、ちょっといろいろと複雑になってしまうということなので、ここではよく直線でその判別の境界線を引くということは、これはシンプルでやりやすいということで、ここではそういったかたちで直線で判別の境界線を引くというふうなかたちで、判別関数のその精度を上げるためにはどうすればいいのかということを考えていきたいと思います。

ただ、直線を引くと言っても、いろいろと線の引き方というのがあります、試しにちょっとここでは5本の線を引いていますけれども。その線の引き方によっては、その白丸が青に属するようなかたちで返されたりだとか、赤に返されたりとかするといったような、そういったかたちで線の引き方によって、そのクラスの返し方というのは異なってしまうということになるので、果たしてこの判別の境界線をどういうふうなかたちで決めていけばいいのかということここでは考えていかなければなりません。

そこで、まず、線形判別関数の定義ということで、少し数式的なものは出てきているわけですが、ここでは今、判別関数として g を、先ほどクラスの C_1 から C_k みたいな感じで書いてましたけれども。ここでは、ちょっと2クラスにおいて、それを見ていきたいと思えますけれども。要は、データを放り込んだときに、そのクラス1に属しているか、クラス2に属しているかみたいな、そういったものを判別関数として定義を与えてあげることになります。このときに、今、この判別関数として、今、1、2として、場合分けして返されるわけですが、今、このデータのその x に何らかの重み付けのベクトルというのを与えてやって、それでなおかつ、閾値の a_0 というのを与えてやって、で、この a_0 よりか大きければ1を返してやっています。その際に、先ほど挙げた部分を振り返ってみると、先ほどうまくその直線を引くためにはどうしたらいいのかということを考えてきましたけれども、この部分、クラス間をうまく分離するようなその重み付けのその a のベクトルと、それから閾値の a_0 、これを見つけないというふうなかたちで先ほどの問題は帰着させることができるということになります。

で、ここから、いろいろとちょっと仮定を踏まえたことで話をしますけれども。例えば、先ほどクラス1とクラス2で、それぞれ赤丸と青丸がありましたけれども。そちらのほうではそれぞれ、平均と分散というのをデータがあれば求めることはできますけれども。そのときの平均を μ_1 (ミューワン) と μ_2 (ミューツー) というふうなかたちで与えてやって、それで分散というのを Σ (シグマ) というふうなかたちで考えてやることにします。特にここでは、各クラスの分散というのが等しいというふうな仮定のもとでちょっと考えて見ていきたいと思えます。

そこで、実際に先ほどクラス間をうまく分けるような判別境界線の引き方っていうのは、実際には研究によって知られていて、どういうふうにしてやっていくかということ、こういうふうにしてやっていくことになります。まず、1つは、Fisherの評価基準というものがありまして、ここではクラス内の変動行列というのが、これは Σ として、要は分散というふうなかたちで与えていますけれども。それに加えて、クラス間の変動行列というものも考えています。これ、どういうものかと言ったら、要は、平均の値がありますけれども、その平均がどれだけ散らばっているのかを表したものを、これをクラス間の変動行列というふうなかたちで定義を与えます。そのときに、そのFisherの評価基準というものがありまして、これはどういうものかと言ったら、クラス内の変動行列でもありますけれども、それをなるべく小さくして、なおかつ、クラス間の変動行列をなるべく大きくするような、そういったような a というものを求めてやりたいということを考えてやります。

そうすると、そのときの評価基準を最大化にするような a と閾値 a_0 というのは、以下のように求めることができます。要はこの a というのは、このベクトルに比例したようなかたち、要は平行なかたちで得られますよということを表して、その得られたやつを使うと、その閾値 a_0 というのが次のようなかたちで得られると、そういったようなかたちで実際には求めることができます。ちょっとこれも数式がごちゃごちゃしてるので、難しいようなかたちにはなっているかもしれませんが、ちょっと概念的な図でまた表したいと思います。

これは先ほどの例で与えていた赤丸と青丸がありますけれども、今、ここでは各平均というのを求めてやると、バツ印のところはそれぞれの、青いところと赤いデータのところの平均であって、で、その中点が、今、この黒丸で与えているようなかたち、与えていますけれども、こういうようなかたちで与えています。そのときに、今、先ほどの Fisher の評価基準というのを求めてやると、そうすると先ほどの a のベクトルが、こういうふうなかたちで得られるということで。実際にはこの a というベクトルが得られたときに、この a のベクトルに向かって、そのまま垂直に下ろすようなかたちで、全部データをこの軸上に、垂直に落とし込んでやります。そうすると、こういうふうなかたちで。確かに、こういうふうな、射影というふうな言うんですけども、こういうふうなかたちで落とし込んでやると、データがちょっとうまく、青と赤で2つに分離されているような様子を見て取れることができるかと思えます。で、実際にはここから、このベクトルに垂直となるような線を引くことによって、この判別の境界線というのを与えてやるということによって、Fisher の評価基準を用いると得ることはできるという。これが実際の適用例ということになります。なので、今回の場合だったら、この白丸というのは、赤色のほうに属しているというふうなかたちで見てとらえることはできるかと思えます。

ちなみに、クラス内の変動行列を使わない場合、もし、そのクラス間の変動行列のみを用いた場合だったら、こういうふうなかたちになってしまって、ちょっとこの部分が、赤と青がかぶってしまうような部分も出てきているので、なのでクラス内の変動行列も使ったかたちで与えるというのも、これも1つの重要なポイントであるということが分かるかと思えます。

そういった Fisher の基準で用いられる判別基準を適用させた判別関数というのを、これは Fisher の線形判別関数というふうなかたちで、よく呼ばれています。特に、理論的に最適な関数であるというふうな言われているのは、今、ここでは正規分布というものをデータのところでは仮定をしておいて。正規分布というのは、いわゆる左右対称の山のかたちですけど。これを多次元化させたものを d 次元の正規分布というものになるんですけども。そのもとで判別関数を構成すると、結局 Fisher の線形判別関数はこういうふうなかたちで与えることはできるわけですけども。この式自体が理論上最適な判別関数であるというふうなかたちで、実際には言われていたりします。なので、これ以上良いような関数はほかにないよということを知らしめているのが、Fisher's Rule というふうな呼ばれているものになります。

先ほどはデータの仮定というのを、要は μ とか、平均とか分散というのを既知として与えていましたけれども、未知の場合でも推定を行うことによって得ることはできるということで。例えば、平均とか分散とかありますけれども、こちらのほうはデータに基づいて与えてやることによって、推定値として与えることはできるので、それを使って、実際には手元にあるデータから母集団の推定を行うことによって、判別関数というものをここでは求めることはできるということで。こういったかたちで、一般の判別関数を与える際には、推定値を用いて以下の式で構築してやって、それで実際にその x がデータなわけですけど、データを放り込んでやって、正か負によって、どちらかのクラスに属しているかということを与えるといったようなもの、これが一般的な流れというものになっています。

ただ、ちょっと題目にもあったとおり、その高次元小標本というふうな話になると、また話が別になっていて、今ここでは特に、高次元上の話をしたいと思うんですけども。今、ここでは、2つのクラスを仮定して、

それで仮にちょっとデータ数は、各クラス 500 個用意して、それで、ある条件のもとで次元を上げたような設定を考えてやります。そうすると、今、ここでは横が次元になっていますけれども、次元を上げていけば上げていくほど、誤判別の確率というものが 0 に近づいているような様子がここでは見て分かるかと思うんですけども。これはあくまでも理論的にはこういうふうなかたちで 0 に落ちていくというふうなかたちになっているわけですが、先ほどの、手元にあるデータから推定をしてやって、そこから判別関数を構築して、それで誤判別確率を求めてやるというふうなかたちで考えてやると、そうすると次のようなかたちで、その黒線のようなかたちで、なかなか 0 に落ちていかないような問題というものが生じてきます。これを「次元の呪い」というふうなかたちで呼んだりして。実際、理論上では 0 に落ちていくんだけど、実際手元にあるデータとかで構築していくと、うまく汎化性能を得られないというような、そういった問題が生じているというのが、1 つ、高次元上における問題になっています。

あと、もう一つ問題点というものがあって。今、例えば仮定として、次元と n をデータの個数というふうなかたちで仮定して、定義しておきますけれども。例えば、クラス 1 のデータの個数と、クラス 2 のデータの個数を足し合わせたものを n としたときに、この n というのは、次元よりも小さいような、そういったような条件で考えてやることにします。通常のデータとかは、この次元とデータ数の関係を考えていくと、データ数のほうが圧倒的に大きいような状況で考えているような場合がほとんどなんですけれども。ここでは、そのデータがその次元よりも小さいようなかたちのものを、ここでは仮定をして考えてやります。そうすると、先ほどちょっと、平均と分散を推定して考えていきますよみたいなことを言っていたと思うんですけども、実際その各推定量の構築というのは、式で表すとこういうふうなかたちで与えることはできるんですけども。ここで問題となるのが、この $\hat{\Sigma}$ (シグマハット) の分散の推定の部分が、1 つ問題となっていて。これはどういうものかというのと、この $\hat{\Sigma}$ というのは、実際にランクを計算していくと、フルランクになっていないということ。これは何を意味するかというのと、逆行列的なものは取れないという、そういった問題が生じてきます。つまり、これはどういうことかというのと、先ほどこういった式で与えていたかと思うんですけど、この式のところの分散の部分ですね。

分散はこの逆の……。逆に言うと、逆行列の部分が取ることではできないということは、この判別関数自体をつくることはできないといったような、そういった問題が生じてしまうということで、これが 1 つの大きな問題となっていきます。

ということで、ここからが高次元小標本におけるその問題点をちょっと整理をするわけですが、今ここでは、高次元小標本というふうなかたちで取っていますけれども、実際には漸近理論というふうな話を考えていくときには、その n と d の関係が、 n と d を比べると、 d のほうが圧倒的に大きいような状況を考えているんですけども。一応本研究の目標としては、この高次元小標本における判別手法としては、このような大小関係の枠組みの中で、理論的な考察を与えていきたいというのは、これは 1 つの目標であると。あともう一つは、これまで 2 クラスで説明していたと思うんですけども、これを多クラスへ拡張して、それで考えていきたいということで。これは正準相関に基づく判別手法になるんですけども。こちらのほうを適用して考えてやりたいということ。あとは、この $\hat{\Sigma}$ というものは先ほど、逆行列を得ることではできなかったというような話があったと思うんですけども。その特異性を避けるための新たな手法の提案、これを考えていきたいというのがここでの目標となります。

そこで、正準相関に基づく判別手法を構築していきましょうということで、先ほど 2 クラスで考えていたんですけども、これはまず多クラスに拡張して、多クラス判別に拡張して、それで考えてやることにします。先ほど、データ X というのがありましたけど、これは、いわゆる確率変数というふうなかたちでちょっと置き換えて、それで与えてやることにしますが。そのときに、 Y というのを K 次元のクラスベクトルというふ

うにしておきます。で、この K 次元のクラスベクトルはどのようなものかと言ったら、例えばクラス i に属していたら、その i 番目の成分がその 1 で、それ以外が 0 となっているような、そういったようなベクトルになっているというふうに押さえておけばよいと思います。で、この行列 \tilde{C} (シーチルダ) というのがありますけれども、こちらが正準相関行列と呼ばれるものになっていて、これを使って判別に必要なベクトルというのを求めてやるというようなものになっています。実際にこの \tilde{C} とその \tilde{C} を転置させたような行列を考えてやると、そうすると次のようなかたちで得ることはできて。この M というのは、先ほどもクラス間の変動行列というのがありましたけれども、その k クラスのバージョンのかたちで得ることはできるということで、ここではこの M というのが出てきているということをおそらくここで押さえておいてくれればと思います。

そして、 \tilde{C} 、正準相関行列がありますけれども、この固有値問題について考えてやります。固有値問題について考えてやることによって、先ほどの Fisher の判別基準がありましたけれども、それに帰着させることができ、ここから幾つか判別に必要なベクトルというのが得ることができるといことになります。なので、実際には行列のその話で、今、これまでは既知のような状態でやっていたけど、実際にはデータ、パラメータが未知の場合とかだったら、ここではまた推定を行わなければいけないということになりますので、それぞれ推定したものを考えていかなければならないと。

ただし、 Σ というものは、先ほど述べたとおり、逆行列を取ることができないというふうに述べていたので、ここでの \hat{C} (シーハット) で言うところの、この $\hat{\Sigma}$ (シグマハット) のマイナスの 2 分の 1 乗という部分がありますけれども、これも実際に高次元小標本の枠組みだったら、求めることはできないということになりますので、次のようなかたちで、ちょっと回避させてやるということを考えてやります。

その考え方としては、こういうふうな考え方ですね。先ほど Σ というのが与えられていましたけれども、これを D というふうなかたちで置いてやります。これはどのようなものかと言ったら、この D というのは、対角行列のみで、要は分散のみを成分に持ったような、そういった行列で与えられていると。要は、行分散は全部 0 であるというようなかたちで与えたような、そういったような行列で、ここでは与えているというものになります。そういうふうにして与えると、先ほどの固有値問題も、次のようなかたちの、 b というベクトル、判別方向ベクトルというのを求めることはできて、なおかつ、行列のその C の推定量においても、この D に関する推定は、分散のみの推定を求めてやって、それを対角成分に並べてやればいいので、このところの D の部分を実際には求めることができます。そういった手法をナイーブの正準相関に基づく判別手法というふうなかたちで、ここでは新たに提案をするといったようなものになります。特にここでは、固有値問題を考える際に、ランクとしては $K-1$ (K 引く 1) というふうなかたちで得られているので、いわゆる判別に必要なベクトルというのが、 $K-1$ (K 引く 1) 本得られるということで、こういったかたちで判別に必要なベクトルのその b_1 (ビーワン) から b の $K-1$ (K 引く 1) です。これは式を分解したら、こういうようなかたちで、正準相関のベクトルは出てくるんですけども。こういったかたちで判別に必要なベクトルというのをここでは得ることができるというふうなかたちになります。

実際にじゃあ、多クラスのその判別関数を構築するためにはどうすればいいかというふうな話なんですけれども。これについては、実際に判別関数としては、データから、1 (イチ) から K を返すような、今、ここでは多クラスを考えているので、1 から K のいずれかを返すような関数を考えてやるわけですけども。実際には先ほどの $K-1$ (K 引く 1) 本の得られたそのベクトルを並べたやつを \hat{B} (B ハット) というふうなかたちで与えたときに、以下のようなかたちで判別関数というのを与えてやることによって、判別関数というのを構築することができるということになります。要は、この ℓ (エル) を 1 (イチ) から K まで動かしたときに、この値が最小となっているようなクラスを返すような、そういった多クラスの判別関数というのをここでは与えてやるということになります。実際にはこれ、 K を今、一般化してまずけれども、 K を 2 というふうな

たちで与えてやると、最初のほうで述べた Fisher's Rule がありましたけども、そのその $\hat{\Sigma}$ (シグマハット) の部分を \hat{D} (ディーハット) に置き換えたやつに帰着するというのも確かめることはできます。で、あとここでは、いろいろ理論的な考察を行いたいというふうに考えてるんですけども。今回は、今ここでやられている \hat{P} とか \hat{B} というものがありますけれども、その漸近挙動、要は、次元を大きくしてやったときのその振る舞いというのを見ることをここでの目標ということにしておきたいと思います。ちょっと時間の都合上、その誤判別確率の漸近上界については省きたいと思いますが。要は、得られたその $K-1$ (K 引く 1) 本、その判別に必要なベクトルというものが、どういうふうな振る舞いを行っているのかということに次以降のスライドでは紹介をしていきたいと思います。

ここでは、挙動というものを評価をする上で、以下の定義を与えておきたいと思います。これはどういうものかと言ったら、いわゆるそのベクトルのなす角についての定義になります。今ここでは、 χ という、いわゆるデータの χ と、それから確率変数ベクトルの $\hat{\chi}$ (エックスハット) があつたとしたときに、それらをそれぞれ長さを 1 にしたようなやつをこの $\hat{\chi}^*$ (エックスハットスター) と、それから χ^* (エックススター) というようなかたちで与えてやることにします。このときに、この 2 つの内積を取ったときに、 1 に確率収束をするといったような場合は、この 2 つのなす角度は 0 であるということが実際にベクトルの考え方から成り立つことができ、これを $\hat{\chi}^*$ は一貫性を持つというふうなかたちで考えることができます。今ここでは、先ほどの得られていた正準相関のベクトルと、それから判別方向ベクトルというものがありますけれども、正準相関のそのベクトルの p に関しては、これは長さが 1 というふうなかたちで得られているので、こちらのほうは特に、基準化する必要、長さを 1 にする操作は必要ないんですけども、判別方向ベクトルに関しては、こちらのほうは、それぞれちょっと基準化をする必要があつて、次のようなかたちで基準化を行うというふうにして与えてやることにします。

もう一つのポイントとしては、その漸近理論の設定ということで、要は高次元小標本というかたちで今、枠組みを考えているわけですけども。実際には次元とそのデータのその数の飛ばし方というのをうまくコントロールさせないといけなくて。そこで必要となってくるのが、以下の 4 つの部分となります。1つ目は、今、 n と d を無限に飛ばしているような状況を考えていますけれども、要は、 2 つとも、データも、その次元も飛ぶような設定で考えてあります。ただし、次元自体は、データよりも速くなるような性質を持っているということで、2番目の性質としては、 n と d を比べると、 d のほうが速く飛ぶような性質を持っているというものを考えてやります。で、3番目の性質としては、この d というのを対数を取ってやると、そうすると n のほうが速くなるというような。要は、指数的な速さでは飛んでいないというような、そういったような条件になっています。で、4番目の部分が、これはちょっとテクニカルな部分なんですけども、例えば、今ここでは Cd というような正の数値を与えてやるとします。そのときに、今、このクラス i とクラス j の間のところの距離というのを考えたときに、その下界が Cd となっているようなかたちで与えてやることにします。つまり、この Cd というのは、速く無限大に行くような場合とかだったら、このクラス i とクラス j の間の距離が大きくなっているということになるので、この 2 つの間の距離が大きくなっているということは、これは判別しやすいということの意味しているので、そういったかたちでその Cd というのをコントロールさせることはできるといったような、そういったかたちで与えることはできます。これを今、それぞれ設定として、 d とその nCd が対等なかたちであるか。もしくは、 d と比べると、その nCd が早く飛ぶような状況、この 2 つの設定をここでは漸近枠組みとしては導入していきたいというふうに思います。

実際には、漸近挙動に関するこれは補題になるんですけども。今ここでは、その \hat{P} (P ハット) の漸近挙動としては、以下の枠組み、以下のかたちで成り立ちますということで。今、ここでは \hat{P} という正準相関のベクトルがありますけれども、これをこのあるベクトルを与えたときに、これがちょうど 0 に、 2 つのなす角

は0になるような、そういうふうなかたちで得られますよといったようなものが、この補題の1の主張になっています。

ちょっとこの γ とか、そういったいろいろな条件が必要にはなってはくんですけども、適当なモーメント条件を満たすことによって、この2つのベクトルのなす角は、1に近づくということを利用してやると、そうすると次のような定理を得ることができるということで。

これが正準相関のベクトルに関する漸近挙動になります。これは数式はまた、いろいろごちゃごちゃしてるんですけども、要はこの d と nCd が同等なような状況の場合とかだったら、この2つのベクトルのそのなす角というのは、ある角度へと確率収束をするということを理論的に示すことはできたというのは、これは1つのその定理の部分になります。特に、この d と nCd の部分、 nCd はより速く飛ぶような感じの状況の場合とかだったら、その場合だったら、その2つのなす角というのは、0に収束をする。で、それ以外の、 \hat{P}_1 （ピーワンハット）と、それからそれ以外のベクトルだったら、直交するような、そういうふうな性質が成り立っているのは、これが1つの性質というものになっています。

もう一つ、判別方向ベクトルとして B というものを与えていましたけれども、こちらのほうも同じようなかたちで成り立ちますということで、今、ここでは d と nCd の関係は、同等のようなかたちで得られた場合には、2つのなす角というのは、不等号で与えられていますけど、この2つの間の値のところにとまるかたちで与えることはできる、というようなかたちで導くことはできます。さらに、 d と nCd の間の関係が、 nCd が早く飛ぶような状況とかだったら、これは2つのなす角というのは、0に収束をするというような、そういうふうな性質を持っているということを実際に示すことはできます。

ということで、理論的な部分を若干駆け足で述べていったんですけども。一応このまとめというか、そういった部分を見ていくと、今回正準相関に基づいて判別関数というのを構築したわけですけども、この場合、その d が、次元がそのデータ数よりも大きいような場合だったら適用することができないということになっていて、それを回避するために、ナイーブ正準相関に基づいた判別関数というものを提案するといったようなかたちで、今回は提案をいたしました。その枠組みとしては、先ほど述べたとおり、正準相関のベクトルの p_1 から p_{K-1} （ p_K 引く1）に関する部分に関しては、これはある条件を満たす条件のもとだったら、 d と nCd が対等なかたちで得られているのであれば、特定の角度に確率収束をする。で、一方で、 nCd が速く飛ぶような設定だったら、各 p ハットの a というのは、一致性を持つというふうなかたちで結論づけることができるということになります。で、判別方向ベクトルに関しても、同様のよう結果が得られているということで、いろいろと理論的な考察も与えることはできるんですけど、ほかにも誤判別確率とかそういった部分も挙動を調べたりすることもできるんですけども、ある一定のその条件のもとだったら、得られているその判別の境界に必要となるベクトルというのものも、うまく母集団のほうに一致性を持っているというような、そういうような結果が得られているということになりますので、そういったかたちで今回はまとめというふうなかたちで入れさせていただきました。ということで以上で、発表のほう、終わりにさせていただきます。ありがとうございました。

小宮路：玉谷先生、ありがとうございました。

では、質疑の時間を少し持ちたいと思います。玉谷先生にご質問のある方、挙手をお願いします。早速、じゃあ、増川先生、お願いします。

増川：経済学部の増川です。玉谷先生は、分類ということに興味をお持ちということで。

玉谷：はい。

増川：説明変数が山ほどあるのに、データの数が少ないというシチュエーションは、実際のデータでは普通に生じることだと思います。回帰問題でも、正則化の手法として、リッジ（Ridge）とかラッソ（Lasso）とか

あるわけですがけれども。次元の呪いの状況としては、例えば分類の問題だと、グループ化したいのに、自分のまわりにはほとんど誰もいないという状況の中でグループ化ということを考えなきゃいけないという、そういう困難な問題を理論的に挑戦されているということだと思います。理論的な部分ってなかなか、1回聞いて理解できないので、こういう問題の応用としては、どんなことがあるのかということをお教えいただければと思います。

玉谷：今回応用的な部分については、紹介をしていなかったですが。実際にはその応用的な部分としては、膨大にある説明変数がある状況の中で、やっぱりその判別に必要なものと、そうでないものというか、判別に不要なもの。実際には、それを使ってしまうと、むしろ判別を逆効果にさせてしまうような、そういったノイズを含んだようなものがあったりするかと思います。それを取り除いて、それで有効な成分だけを使った判別手法として、特徴選択手法というのがあります。それを使ってやることによって、実際に判別の、実際説明変数がいっぱいあるような状況のもとで判別するよりは、その有効なやつを使ったほうが、判別の精度が上がります。それも実際に自分の研究では、提案しています。そういうかたちで与えることによって、その高次元小標本のようなデータとかも、判別に有効なものを使ってやることによって、シンプルに与えたりすることができるというふうな、そういった応用例もあります。

増川：最初、いろんながんの種類があって、それを同定するという話をされていましたが、その例は非常に分かりやすかったです。例えばいろんな種類のがんがあって、いろんな特徴量、つまり症状であるとか、病理検査などの膨大な結果があって、それらを使って、症例としては、非常に少ないような難病といえますか、そういうものの診断を行うのに使われるような理論というふうには思っていますか。

玉谷：そうですね。症例とかの場合とかだったら、またちょっといろいろ工夫が必要かもしれませんが。要は、説明変数の中には、いろいろと有効なものとかそうでないものがあったりするので、その有効なものだけを取り出して与えてやるという特徴選択手法というのがあります。それをやることで、精度を上げることができるのではないかと考えています。

増川：ありがとうございました。

小宮路：ありがとうございました。他にございますでしょうか。挙手でも、あと、チャットでも。はい。それでは、ありがとうございました。玉谷先生、ありがとうございました。

閉会挨拶

増川 純一
成城大学 経済学部長

小宮路：それでは、最後に本学経済学部長、増川先生より閉会のごあいさつをお願いいたします。

増川：印象をまずお話ししますと、理系の人たちが来たというのが、非常に強烈な印象であります。かく言う私も、実は、今、経済学部というところに所属しておりますけれども、理論物理で学位を取った人間でございます。理系の気持ちは非常によく分かると思っております。

森先生、ウェアラブルデバイスによるリアルタイムの疲労度とか、あるいは習熟度の測定の話、非常に面白かったですね。今、健康への関心が非常に高い中で、非常にタイムリーな研究をされているというふうには思いました。

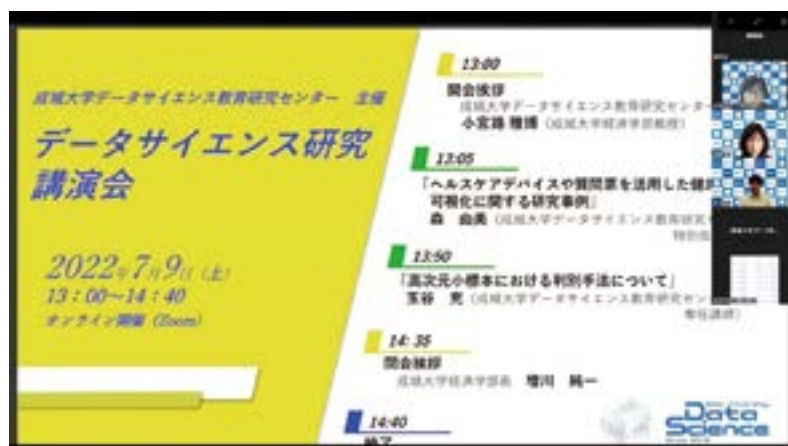
それから、玉谷先生。理学部ど真ん中の方なので、文系の視聴者の皆さんに、分かりやすいような話では、ひょっ

としたらなかったかも知れませんが、先ほども言ったように、例えば、難病、奇病の非常に患者数が少ないような病気の診断でありますとか、あるいは、遺伝子病の原因遺伝子の特定みたいな問題では、説明変数はいっぱいあるけれども、データが少ない。それからさっき、回帰という話をちょっとさせていただきましたが、最近日本も参加しているブラックホールの撮影をしたというプロジェクトがありましたよね。それって、地球規模でバーチャルに天体望遠鏡を作ったんですね。どのぐらいの視力の望遠鏡かというと、月の上のゴルフボールが認識できるぐらいの視力のある天体望遠鏡をつくったらしいです。そうすると、非常にコントロールが難しくって、コントロールするパラメーターが山ほどある中で、結果を画像として出さなきゃいけないわけですよね。そういうシチュエーションでも、玉谷先生に今日お話ししていただいたような分析の必要性が生じるということですね。なので、玉谷先生のご研究も、今やらなければいけない研究であるということは間違いなくって。そういう意味では2人とも非常にタイムリーな研究をされている方々だというふうに思いました。

で、我々、小宮路先生と一緒に成城大学のデータサイエンスの教育研究、主に教育面について、いろいろ企画してきたわけですけども。こういう新しい知と言いますか、理系の玉谷先生、森先生にご参画いただいて、今までの教育のフェーズからもう一つ先に進んで研究が発信できるようなフェーズにわれわれも移行できるような予感がした良い研究会だったと思います。以上です。

小宮路：ありがとうございます。では、以上をもちましてということですが、最後に皆さまにお願いがあります。チャットでアンケートのURLをお送りしますので、ご協力の方、よろしくお願いたします。お返しただく締め切りは、7月19日火曜日とさせていただきます。はい。今、ちょうどチャットで来ました。アンケートのURLをご確認いただいて、ダウンロードいただきましたら、本Zoomからご退出いただければと思います。

以上をもちまして、成城大学データサイエンス教育研究センター主催の講演会を終了いたします。皆さま、本日は誠にありがとうございました。



「データサイエンス研究 講演会」当日のタイムテーブル

講演録

成城大学データサイエンス教育研究センター講演会

成城大学におけるデータサイエンス教育

2022年12月2日(金) 18:00～19:10

オンライン開催 (Zoom)

Program

開会挨拶 小宮路 雅博 成城大学データサイエンス教育研究センター長／経済学部教授

講演

「データサイエンス科目でのSTEAM教育の実践」

森 由美 成城大学データサイエンス教育研究センター特別任用教授

座談会

「人とロボットの共生社会実現に向けて

一人型コミュニケーションロボットとの対話実験の試みー」

新垣 紀子 成城大学社会イノベーション学部教授

平野 なお佳 成城大学社会イノベーション学部4年生

森 由美 成城大学データサイエンス教育研究センター特別任用教授

質疑応答

閉会挨拶 小宮路 雅博 成城大学データサイエンス教育研究センター長／経済学部教授

開会挨拶

小宮路 雅博

成城大学データサイエンス教育研究センター長／経済学部教授

小宮路：皆さま、こんばんは。これより、成城大学データサイエンス教育研究センター主催の講演会を開始したいと思います。私は、データサイエンス教育研究センターのセンター長を務めております、小宮路と申します。よろしくお願いいたします。本日はお忙しい中、皆さまご参加くださり、誠にありがとうございます。

今日の講演は、「成城大学におけるデータサイエンス教育」ということで、こちらのセンターの教員による、データサイエンス教育の実践の一端といいますか、教育の現場からの視点を踏まえた講演をしてもらいます。その後、本学の学部の学生さんと、その指導教員の先生も含めて、座談会を行いたいと思います。

前期にはデータサイエンス研究ということで、講演会を開催しましたが、後期はデータサイエンス教育という、教育現場の視点を踏まえたお話をしていければと思います。

それではまず、当センターの特別任用教授であります森由美先生より「データサイエンス科目での STEAM 教育の実践」についてご講演をいただきたいと思います。

講演の後、引き続きまして、座談会を行いたいと思います。その後、最後の時間に、質疑の時間を設けたいと思います。それから記録のために、この今回の講演会、録画させていただいております。よろしくお願いいたします。

それでは森先生、よろしいでしょうか。

森：はい。

小宮路：はい。では、開始したいと思います。森先生、よろしくお願いいたします。

講演「データサイエンス科目での STEAM 教育の実践」

森 由美

成城大学データサイエンス教育研究センター 特別任用教授

森：よろしくお願いいたします。では、画面のほう、先に共有させていただきます。画面のほう、見えていますでしょうか。

小宮路：はい。見えております。

森：はい。ありがとうございます。では、始めさせていただきます。皆さま、こんばんは。成城大学データサイエンス教育研究センターの森と申します。今日はお忙しい中、お集まりいただきまして、ありがとうございます。私のほうからは、「データサイエンス科目での STEAM 教育の実践」ということで、お話をさせていただきます。

森：まず簡単に、自己紹介をさせていただきます。企業で研究開発に従事しておりまして、その後、横浜国立大学のほうで教員勤務を経て、4年前、2018年に成城学園に着任いたしました。成城学園は幼稚園から大学までワンキャンパスにある学校になっていまして、こちらでプログラミング教育支援ですとか、大学でのデータサイエンス科目を担当しております。

そして今年より、成城大学データサイエンス教育研究センターの教員として、引き続き、データサイエンス

科目を担当しております。興味のある研究活動としましては、健康推進ですとか、それから新しい医療機器、デバイスの開発提案、そういった活動を行っております。短い時間になりますが、どうぞよろしく願いいたします。

今日の内容はこちらになります。まず最初に「STEAM教育とは」ということで、お話しさせていただきたいと思います。

STEAM教育といいますと、もう多くの皆さまがご存知かと思いますが、ここで改めて、ご説明をさせていただきたいと思います。STEAM教育のSTEAMとは、科学、テクノロジー、工学、アート、数学、こういった言葉の英語の頭文字を取ってSTEAMと呼んで、その教育のことをSTEAM教育というふうに呼んでいます。

これらの科目を個別にはではなく融合的に統合して、応用して、新たな変化や新たなものを自ら生み出すような21世紀型スキルと呼ばれていますが、こういったスキルを育むこと、そしてまた、論理的思考力や問題解決能力を育んでいく、こういった教育のことをSTEAM教育というふうに呼んでいます。

これからはやはり、AIやロボットを活用した世の中になっていきます。そういった中で、人とAIとがうまく協調していくためには、人が得意なところを、やっぱりAIにできないところを伸ばしていくということが非常に重要になってくる。そういったことで、創造性をますます伸ばしていかなければいけないということでも、このSTEAM教育が重要だというふうに言われています。

また、社会が変わっていく中で、こういう、21世紀型スキルの論理的思考力や問題解決能力、そういった能力を持つ人材というものがますます必要とされていて、ただ、今、世の中では不足しております。そこで、こういった人材を育成していくことが急務ということで、STEAM教育を通して、そして、このSTEAM教育を行う実践の方法の1つとして、このプログラミング教育というものが、今、進められているという状況にあります。

そこで、私もデータサイエンス科目を担当している中で、このプログラミング教育といったものを、どうやったら学生さんたちに楽しく、モチベーションを高く進めてもらえるかということで、授業の中で実践してまいりました。それについてお話をしていきたいと思います。

データサイエンス科目として、その中で、「データサイエンス基礎」という科目を持っておりますが、こちらがそのシラバスになります。ここにありますが内容は、成城大学は文部科学省「数理・データサイエンス・AI教育プログラム」認定制度のリテラシーレベルの認定を取っておりますが、それに沿った内容として、このシラバスがございますけれども、この中で、プログラミング教育をうまく取り入れるということを行っております。

例えば、音楽演奏プログラムであったり、ロボットプログラミングであったり、そして、Pythonを活用した画像解析プログラミングであったり、また、別の科目では、自然言語処理のプログラミング、そういったことを行っております。これを少し、1つずつ、ご説明したいと思います。



森 由美 データサイエンス教育研究センター特任教授による講演

1つ目の音楽演奏プログラミング。これには「論理的思考力と情操を育む音楽演奏プログラミング」という名前を付けております。アメリカのマサチューセッツ工科大学で開発されたビジュアルプログラミング言語・Scratchという言語がありますが、ここに見ていただいているようなブロックを重ねていって、そして、それがプログラムになるという、そういったプログラミング言語を使います。とても扱いやすく、直感的で分かりやすいプログラミング言語になっています。これを使って音楽演奏プログラムを作成するという、そういった授業を行っています。イメージとしましては、こういった、楽譜があって、楽譜からプログラムに起こし直すという、そういった授業になります。

ただ、こういう音符を並べていって、プログラムにするというだけでなく、その音楽の中には、よく同じようなフレーズが使われているなどの規則性があります。そういった規則性を発見して、自ら新しい関数を作ったりですとか、それから、これはピアノの楽譜ですけども、ピアノだけではなくて、他の多数の楽器の音を重ね合わせて、同時に演奏をして、そしてその他に、音ですとかリズム、それ自体をアレンジしたり、音の強弱を工夫したり、そういったことをすることによって、「一人オーケストラ」という、そういうプログラムを作り出すということを行っています。

この、自分だけのハーモニーを作り出すというところで、それぞれの工夫を入れながらプログラミングしていくということで、このSTEAM教育に、まさにぴったりだというふうに考えております。

この授業は2年ほど前から始め出しましたけれども、だんだん学生さんが作ってくる作品はどんどんレベルがアップしているというふうに、今考えています。それを課題として提出してもらいますが、私1人で鑑賞するのは非常にもったいないと思っていて、コンテストという形で、学園内で共有できたらと思っておりました。それが今年の夏に実現しまして、この「成城大学 Scratch 音楽演奏プログラミングコンテスト」という形で実現することができました。

ここでは、8人の学生の作品が金賞、銀賞、銅賞という形で賞を付けることができて、こういったオープンバッジを発行して、それを渡すということもできました。

ここで、皆さまに少しだけ、この作品を聴いていただこうと思います。これは金賞を取りました『復活の祈り』という作品になります。途中までになりますが、少し聴いていただきたいと思います。〈音楽〉

はい。途中になりましたけれども、今ここに、画面上にたくさんの文字が出ていますが、この文字の1つ1つの中に、それぞれのプログラムがあって、そのプログラムが今、同時に並行して実行することで、今のような、オーケストラで演奏しているかのような音楽になっています。もう1つ聴いていただこうと思います。〈音楽〉

はい。聴いていただきました。こちらも、ドラムの音などを自分でアレンジして、そして、こういう音楽に仕上げています。他にも皆さまにご紹介したいものがたくさんありますけれども、お時間の関係で少しだけ聴いていただきました。

そして、今回のコンテストでは、この学園の教職員、それからこの授業を履修している学生が投票者として、コンテストを行いました。そして、この投票者の皆さんからのコメントもいただきましたので、ここで少しだけご紹介させていただきます。

「皆さんがたくさん試行錯誤され、苦労を重ねながら、でも楽しみながら作成している姿が想像できました」ということで、ほんとに作品を提出した学生さんは、とても楽しみながら、モチベーション高く、それぞれプログラミング的な試行錯誤も入れながら、作ってもらったと思います。

それから、「プログラミングというとデジタルで感情があまり感じられないイメージがありますが、道具を使いこなすと個性が出てくるものなんだなと思いました」、「自分もトライしてみたいと思いました」、それから、「他の人に聴いてもらう機会があると、ますます学びへのやる気につながり、とても良い企画だと思いました」、「学生さんの発表の場がこれからも保障されるといいなと思いました」、こんなコメントもいただいております。

次は2つ目として、ロボットプログラミングです。ロボットプログラミングのほうは、4年前から実践しております。こちらは、まだコロナ前の時の、対面で行っていましたが、ロボットを活用して、何か課題を解決するようなプログラミングを行ってありました。今はオンデマンドの授業になっていますので、実際にこのロボットを使って、目の前で動かすということはできませんが、プログラミングツールの中に、こういうバーチャルロボットというものがあって、このロボットを動かすということもできるようになっています。

〈映像開始〉

Pepper：では最初に、Pepperのプログラミングをするためには、Robo Blocksというプログラミングツールを使います。

〈映像終了〉

はい。こういったものも使いながら、先ほどのScratchも、こちらのロボットプログラミングも、プログラミングの基礎を一緒に学びながら、楽しく進めていける、そんな内容になっております。

こういったロボットプログラミングの授業を受けた学生の1人が、こちらは文芸学部の学生さんですけれども、自分のレポートを、自分の専門の授業の中で発表する際に、このロボットを活用して、さらにAIも活用して、授業の中で発表させたということがございました。この時には、担当の先生が「今まで、これまでの授業中でなかった、前代未聞のことだ。」と言って非常に驚かれていましたし、一緒にここに来ている他の学生さんたちもとても驚いて、そして、でも興味を持って、私もやってみたいと。その後、データサイエンス科目を受講した学生さんもございます。

こんなふうに、1人が何か行動を起こすことで、それがどんどん周りに波及していくということも、STEAM教育のいいところではないかなというふうに考えております。

そして3つ目に、Pythonを活用した画像解析・自然言語処理ということで、Pythonという、こちらは普通に企業の中でも使われているようなプログラミング言語になりますが、とても初心者でも分かりやすいプログラミング言語となっていて、今はプログラミングの人気ランキングでは1位になっているプログラミング言語です。

これを使って、画像解析であったり、自然言語処理という、とても最初は難しいように聞こえますが、もう数行のプログラムを使って、例えば、この猫の画像から、ここに映っているものが何なのかを検出するようなプログラムが簡単にできたり、それから、自然言語、文章ですね。その文章の中から特徴的な言葉を探し出してきて、こういう、ワードクラウドという、こういう絵柄のような、こういったものを作ったりですとか、そういった授業を行っています。

そうしますと、やっぱり学生さんも楽しく取り組んで、そこから自分でいろいろ工夫をしていくといったことがあります。例えば、ここの画像を大きくするとどうなるかとか、背景にいろいろなものが映っていた時には、どんなふうに検出されるんだろうかとか、それから、この形を変えるにはどうしたらいいんだろうかとか。そんなふうに、どんどん学生さんたちが自分で工夫して、新しいものを作っていかうとする姿が見えて、これもSTEAM教育としては、とてもいい題材ではないかなというふうに考えております。

プログラミング実習を受けた学生からの感想として、こんなものがあります。「簡単に機械学習を体験することができて、とても面白かったです」とか、「いい経験をしているなあ」、それから、「達成感を持てた」、こういったことがとても大事だと思います。一方で、やはり、簡単とはいっても「手間がかかる部分もあるということが体験できた」ですとか、それから、「画像認識を行う時にうまくいかなかったので、そこを何とか改善するように工夫をした」、そういったコメントもあつたりします。さらに、「もっと複雑なものにチャレンジしてみたい」、そんなコメントもございます。そして、この最後にありますのは、「今後も機械学習を利用して、

さまざまな業種との掛け合わせにより経済が活性化していくことを楽しみにすると同時に、私自身も関われるように日々自己研鑽を重ね、経済を後押ししたいです」, こんなコメントもございました。

私のほうではオンデマンド授業を行っています。オンデマンド授業のデメリットとしては、自己管理が難しいということがよく言われています。

この内容としては、分からないことをそのままにしてしまう可能性がある、学生自身のセルフモチベーションに委ねているところがある、こういったことが詳細な内容になりますが、この、分からないことをそのままにしてしまう可能性があるということについては、授業の後に、毎回感想文、それから質問を書いてもらって出してもらって、次の授業の時に、その振り返りということを必ず行うということで、分からないことをそのままにする可能性をなるべく少なくしているという工夫をしております。

また、学生自身のセルフモチベーションということにつきましては、先ほどのような、学生が興味を持てるような内容を題材にした授業にすることによって、面白く思ってもらえる、楽しく思ってもらえる、そこから自分で何かをしようと思う、そんなセルフモチベーションを高めるような授業を心掛けております。

今後の発展としましては、これまでの教育プログラムを継続していきたいと考えております。そして、学内コンテストを活性化していきたいと考えております。そして、データサイエンスの幅広い活用としまして、健康・ヘルスケアに関連した新たな教育プログラムの策定、それから、学内に幾つか学部がありますけれども、その学部との協業、データサイエンスの広い活用ということ、ぜひ進めていきたいなというふうに考えております。

以上で、私からのお話は終わりにさせていただきたいと思います。ご清聴ありがとうございました。

小宮路：はい。森先生、ありがとうございました。

森：はい。

小宮路：成城大学は、文系4学部だけがある大学ですがけれども、その中で、データサイエンス科目群は今、16科目32単位分の科目を取り揃えております。私たちのデータサイエンス教育の特徴というのは、どの学部どの学生さんにも履修してもらいたいのですが、必修科目とはしていないところです。卒業単位には組み込めるのですが、必修ではない。学生さん達からすれば、自動的に履修となる必修科目ではないので、彼らの自由意志、選択の判断の中で選んでもらう科目群という格好になっております。

一方で、これは文部科学省の認定制度との関係もあるのですが、データサイエンス教育研究センターとしては、学生さん全員、限りなく全員に履修してもらいたいというミッションがあります。その中で、森先生がいろいろ工夫していただいているように、学生の皆さんに興味を持ってもらったり、達成感をきちんと後付けしたりしながら、データサイエンス教育を進めているというのが、私たちの特徴かなというように、改めて森先生の講演を聴いて、私も思った次第です。

座談会「人とロボットの共生社会実現に向けて 一人型コミュニケーションロボットとの対話実験の試みー」

新垣 紀子

成城大学社会イノベーション学部教授

平野 なお佳

成城大学社会イノベーション学部4年生

森 由美

成城大学データサイエンス教育研究センター特別任用教授

小宮路：それでは、続きまして、データサイエンス教育の学部とのコラボレーションというか、融合的な部分ということで、座談会に移りたいと思います。森先生、引き続き、よろしくお願いいたします。

森：はい。ありがとうございます。これからは、座談会形式で成城大学の社会イノベーション学部4年生の平野なお佳さん、そして、社会イノベーション学部教授の新垣紀子先生をお招きしまして、「人とロボットの共生社会実現に向けて一人型コミュニケーションロボットとの対話実験の試みー」というテーマで進めていきたいと思います。どうぞよろしくお願いいたします。

平野：よろしくお願いいたします。

新垣：よろしくお願いいたします。

森：はい。よろしくお願いいたします。まずはこのタイトルに、「人とロボットの共生社会」という言葉がございます。少しだけご説明させていただきますと、

内閣府のムーンショット、内閣府が主導しておりますムーンショット型研究開発というのがございまして、その目標の3番目に「2050年までにAIとロボットとの共進化により、自ら学習・行動し人と共生するロボットを実現」という、こういった目標がございます。このように、これから人とAI、ロボット、それらが協調して社会を回していく、そういったことを国を挙げて押し進めていこうという、そういった目標になっています。

また、成城学園の幼稚園では、こういったような形でロボットを活用して、ロボット体験を行っていますし、また、世の中ではこういうロボットカフェみたいなものも最近出てきているというように、ロボットがかなり身近なものになってきています。

その中で、社会イノベーション学部の新垣先生のゼミでは、ここ2年、ここ3年ほど、ロボットを活用した授業を実践されていますし、また、平野さんにあたりましては、今度ご卒業ですけれども、卒業研究でロボットを活用するということがされていますので、今日はいろいろお話をお伺いしたいと思ひまして、ご参



座談会の様子

加いただくことになりました。どうぞよろしく願いいたします。

では最初に、新垣先生のほうから自己紹介をお願いしたいと思います。

新垣：はい。

森：よろしく願いいたします。

新垣：はい。よろしく願います。画面共有よろしいでしょうか。

森：はい。よろしく願います。

新垣：今回、私のゼミナールの平野さんの卒業研究における実験のお話をさせていただきますが、その前に自己紹介と社会イノベーション学部の紹介を簡単にお話しさせていただきます。どうぞよろしく願います。私も森先生と同じで、前職は企業の研究所で働いていました。2005年に成城大学の社会イノベーション学部ができた時から、成城大学に勤務しています。専門は認知科学という分野で、人のナビゲーション行動とか、人の周りの外界情報を用いた問題解決が研究のテーマです。

私たちの周りには、例えば、スマートフォンのように、非常に便利な道具があります。それを外界情報と呼んでいます。例えば、ナビゲーションの話で言うと、スマートフォンを使うと、屋外でも自分の位置を地図上に表示することができるようになりました。そういう新しい道具やシステムが現れてくると、人間の行動であるとか、認知や判断のプロセスとか、人と人のコミュニケーションとか、いろいろな人の認知活動が変わります。それらが人にどんな影響を与えるのかということが、研究テーマの1つです。

成城大学では、授業として、心理実験法と、ヒューマンインターフェイス論、こちらは、今の話と関係するようなテーマの授業です。それからゼミナールなどを持っています。

簡単に、社会イノベーション学部の紹介から入らせていただきたいと思います。社会イノベーション学部は、日本で最初にイノベーションという名前が学部名に付いた学部です。イノベーションは、すこし分かりにくいのですが、貧困であるとか、環境問題であるとか、世界にはさまざまな社会の問題があって、それを解決するためには、正解があるわけではない、解決方法に答えがあるわけではないような問題に取り組む必要があります。そういうことを解決するために、イノベーション、ここではイノベーションを技術革新ということ意味ではなくて、新しいアイデアなどの工夫によって、これまでにない価値を社会にもたらすことをイノベーションと定義していますが、私たちの学部では、そのような課題を解決するための学びとして、何かの学問体系を体系的に学ぶという形ではなく、課題志向型で、学問分野の横断的な学びを、カリキュラムとして取り入れているという特徴があります。

例えば、ここ2年ほどコロナウイルスが出てきて、私たちの生活様式は非常に変わりました。2年前には、大学は突然対面で授業を行うことが難しくなったわけですが、そのために行ったオンライン授業も1つのイノベーションの事例です。先ほど申しましたように、社会イノベーション学部は、例えば、国の政策や企業の戦略という、広い視点、マクロな視点とミクロな視点と説明することも多いのですが、こういった分野や、心理と行動や社会と文化のような、さまざまな視点を学ぶことにより、多様な視点から問題解決ができるような人になることを目指した、そういう人材を育成することがカリキュラムの目的になっています。

私は心理社会学科に所属しています。今日ご紹介するのは、心理社会学科の4年生の平野さんです。

ちょうど2年、3年ぐらい前に、新型コロナウイルスのために大学の授業がオンラインになった頃に、学内での情報共有のために、Slackを使っていろいろ情報交換をしていました。その時森先生と知り合いました。学内にロボットがいるのは知っていたのですが、森先生がそのロボットをお持ちだということを知りました。コミュニケーションロボットは、これから世の中で重要な役割をすると思うのですが、でも、実際は、世の中では、まだうまく使われているとは言えない側面があります。

それで、森先生にご相談して、3年生のゼミ研究のテーマで使わせていただくことになりました。それで、

初期の頃はこのような感じで、心理実験をロボットを使って行っているのですけれども、初歩的なのというか、人とロボット1対1の研究方法でしたが、最近、ロボットを2台使って実験を行う形になりました。

ロボットを2台使うというのは、今日、お話ししていただく平野さんの活躍でできるようになりました。2台でも実験できるようになると、研究の幅が広がります。

それから、ロボットを使った実験ですが、このロボットは開発環境がすごく整っていて、プログラミングを今まで経験したことがなくても、人との対話を簡単に作ることができます。その点がよいと思います。あと、森先生との関係では、3年生のゼミ生が、大学のゼミ活動の中で、一つテーマを決めて、グループ研究を行っているのですけれども。

この写真の階段ですが、今年、授業が対面式になって、学生が非常にたくさん学校に来ることになって、この階段が混雑して、危ないのではないかということになりました。結構人が多いので、人の流れを変えることができないう、階段に線を引いて、人の流れが変わるか、ということ进行分析しようとしていました。ビデオを見ながら、1人ずつがどのように動くか、という分析をしていたのですが、森先生にご相談をすると、こういうのはDeepSORTという機械学習の手法を用いて、人をトラッキング。

〈映像開始〉

トラッキングすることができて、こうやって追跡すると。

〈映像終了〉

X座標とかY座標とか、1人ずつの位置座標を出すことで、人の移動の定量的な分析が、できることが分かりました。このように森先生にご相談させていただいて、とてもゼミでの研究の幅が広がりました。

今日は、ロボットを使った実験の研究で、ご紹介する平野さんは、3年生の時に、このゼミ研究で、ロボットを使った実験をして、それを卒業研究でもぜひ行いたいということで、この1年というか、この夏から頑張って実験を進めました。その内容をご紹介するということになっております。

平野：ありがとうございます。私も画面共有させていただきます。

森：はい。

平野：共有できていますか。

森：はい。できてます。

平野：ありがとうございます。先ほど、新垣先生からご紹介にあずかった平野なお佳です。

私は新垣先生のゼミに2年生の頃から所属していました。2年生では認知心理学とか認知科学っていうものを勉強していたのですが、3年生になって、Pepperを使った実験をやりました。その時に、ロボットが教師の代わりになれるのかっていうテーマで実験を行いました。その時の実験で、結構ロボットも人と同じように理解度を高めることができたり、逆に、人よりも、人が教えるよりも、共感度とかそういったところではロボットのほうが上手に伝えられるみたいな、物語とかの共感度とかを上手に伝えられるっていうのが分かったので、ロボットの可能性がまだまだあるなというふうに思って、自分の卒業研究ではロボットを使った実験をしようと決めて、卒業研究でテーマにしました。

いろいろロボットを使った実験で、どういうことができるんだろうと思って先行研究を調べていくと、大体円滑なコミュニケーションを築くためにはというテーマで、ジェスチャーであったりとか、相槌であったりとか、あと、パーソナルスペースとかに注目した先行研究が多くて。そういった研究も面白いなと思ったのですが、大体そういう研究って、一対一の実験がほとんどで、人とロボットが一対一の実験で、うまくコミュニケーションを取れるか、みたいな実験がほとんどだったのですけど。

よく考えた時に、人間って集団で生活、行動する生き物なので、ロボットと人の共生というところに目を向けた時に、ロボットも集団となって、人と複数台のロボットと複数人の人の生活が今後活用されてくると思う

ので、そういった集団内でのロボットをどうやって活用していくかというところに目を向けるべきだなというふう思ったので、今回の卒業研究では、2台のロボットを使った実験をしようと思いました。

人は人間社会を生きる上で、暗黙のルールとか常識とかいうことを踏まえながら、いろいろ人と接しているなというふう感じたので、今後、複数台ロボットを研究するに当たって、ロボットの集団規範に目を向けて、実験をする必要があるなと思いました。

なので、私の卒業研究では「ロボットとの共生社会実現に向けて一人とロボットとのコミュニケーションと集団規範との関係性―」というテーマで進めています。では、ざっくりとした、自分の卒業研究の内容ですけど。2台のロボットが築く集団規範に人は気付いて、人もロボットの回答に影響されるかということをやっています。

それとプラスして、あらかじめ2台のロボットと交流をすることで、親密度を高めることで、集団規範に、より親密度が高くなることで影響されるんじゃないかっていう2本立てでやっているんですけど。ちょっと動画を見ていただきたいくて。これが2台のロボットとの対話をしている映像ですけど、聞こえますか。

〈映像開始〉

Pepper (タロウ)：こんにちは。僕はタロウだよ。

Pepper (ハナコ)：こんにちは。私はハナコよ。あなたのお名前はなんですか？

参加者 (サトウ)：私の名前はサトウです。こんにちは。

Pepper (タロウ)：サトウさん、今日はよろしくね。

参加者 (サトウ)：よろしくお願いします。こちらこそ。

Pepper (ハナコ)：サトウさん、お話できてうれしい。

参加者 (サトウ)：こちらこそ、お話できてうれしいです。

Pepper (タロウ)：ねえねえ、サトウさん。

参加者 (サトウ)：はい。

Pepper (タロウ)：今日はどんな天気？

参加者 (サトウ)：今日はすごくいい天気です。

Pepper (タロウ)：教えて欲しいな～

参加者 (サトウ)：今日はすごくいい天気です。暖かいです。

Pepper (タロウ)：晴れなんだね。天気いいとピクニックしたくなる～

参加者 (サトウ)：そうですね。ピクニックしたくなるね。お出かけしたくなりますね。

Pepper (タロウ)：でも、最近季節関係なく、暑いよね。

Pepper (ハナコ)：うん、暑いね。

参加者 (サトウ)：うん。でも、あしたから寒くなるみたいだから。今日はほんとにいい天気ですね。

Pepper (タロウ)：暑いと汗が出ちゃう。

参加者 (サトウ)：汗が出ちゃう。もう汗。走って来たから、汗が出ちゃう。

Pepper (ハナコ)：何を言ってるの！ 私たちは汗をかかないよ。

参加者 (サトウ)：あ、そっか～。羨ましいな～。僕は汗っかきだから、すぐ汗かいちゃう。

Pepper (タロウ)：サトウさん今笑いましたね～。

参加者 (サトウ)：ははっ。笑いました。はは～ (笑う)

Pepper (ハナコ)：タロウはほっといて、私とクイズで遊ぼうよ～

参加者 (サトウ)：やりましょ。でも、一緒にやりましょよ。(タロウに向かって)

Pepper (タロウ)：ハナコのいじわる～。僕もクイズやりたい～

参加者（サトウ）：一緒にやりましょう。

Pepper（ハナコ）：分かったよ。それじゃ私が問題を出すね。

参加者（サトウ）：はい。お願いします。

〈映像終了〉

ちょっといったんストップするんですけど。こうやって、2台のロボットと、まず対話をしてもらって親密度を築いた後で、2台のロボットと人が3人で、クイズに答えてもらう時に、ロボットの回答に影響されるかということをやりました。グループを2つに分けて、1つのグループは、今みたいに対話をしてもらいます。もう1つのグループは、ロボットとの対話はなしで、そのままクイズに答えてもらうということをやりました。

クイズの映像もちょっとあって、これがそうなんですけど。私がやったのが光点点減クイズっていうのをやって。見てもらったほうが早いかもしれないんですけど。

〈映像開始〉

今、画面に丸が出て、その後にもう一個丸が出たと思うんですけど、ここの丸の距離を大体何センチだと思うかを答えてもらうっていうクイズをやりました。明確な答えがあるクイズではなくて、曖昧なクイズのほうに影響が出るかなと思ったので、このクイズにしました。

実験者：タロウさん、お願いします。

Pepper（タロウ）：やっぱり僕は57センチだと思うよ。

平野：で、ロボットが次に答えて。

実験者：それでは、ハナコさん、お願いします。

Pepper（ハナコ）：さっきは私は今回は62センチ。

平野：で、2台目のロボットも答えてもらうんですけど。

〈映像終了〉

このクイズを5回やるんですけど。この2台のロボットが、それぞれ、自分たちの距離は、最初はすごい、20センチかな、離れているんですけど。それを徐々に、お互いの回答を近づけてって集団規範を築くというロボットがいて、それに人が気付いて、自分もロボットに合わせた回答をするかというのを見ました。

実験の結果としては、コミュニケーション、つまり2台のロボットの対話ありなしにかかわらず、ロボットの集団規範に影響されて、自分もロボットの回答に近づけるっていう人が多かったのです。やっぱりコミュニケーションを取ったグループのほうが、ロボットに対して、「もっと会話したいよ」とか、親しみやすさを感じた人も結構いて。集団規範も、交流なしのグループは、全然ロボットに影響されない人たちも、ある程度はいたんですけど、交流ありのグループは、大体みんなが影響されている結果が出て。

やっぱりロボットと会話をして、親しみやすさを与えることっていうのが大事なんだなっていうふうに思いました。

で、実験を通してなんですけど、私が3年生の時に実験した時は、まだまだロボットに対して、不安とか怖いっていうふうに思っている人たちも結構いて、自分の体感的には、肯定的な意見の人とロボットに対して否定的な意見の人が5対5ぐらいの感じだったんですけど、今回の実験では、意外と「ロボットかわいい」とか、「Pepperかわいい」と言ってくれる人たちのほうが多くて、自分の感覚的には、肯定的な人が8で、まだやっぱり怖いって思う人たちが2ぐらいいる印象で、何でこんなに違うんだろうと思いました。

実験参加者の人とインタビューとか感想とか聞いている中で、すごく思ったのが、みんな最初はロボットに対して、どういう生き物なのか分からない、何をされるのかが分からないっていう不安感が勝って、怖いっていう印象が強いけど、今回の実験は集団規範を見る実験だったので、ロボットも自分たちと同じように、集団規範、相手に合わせるとか、そういった暗黙のルールとかっていうのがロボットにもあるんだっていうふうに

感じた人は、ロボットに対して、結構親しみやすさを持っている人が多くて。

なので、ロボットと人の共生社会実現ということを踏まえると、ロボットにも暗黙のルールが分かる力とか、空気を読める力があるよというのを人間に気付かせることが、人とロボットの共生社会をする上で、すごく大切なんじゃないかなってというのが、今回の実験で分かりました。長々話しちゃったんですけど。

森：はい。ありがとうございました。とても興味深いお話、ありがとうございました。

平野：ありがとうございます。

森：新垣先生のほうは、平野さんをご指導されるに当たって、どういった感触を持ってらっしゃったんですか。

新垣：平野さんは、すごく行動力のある人で、ロボットを使うだけでも割と抵抗、やっぱりどうしても抵抗を感じる学生さんが多いのです。何か新しいことをするのは、大変なので。それだけではなくて、ロボットを2台使うというのが、2台を本当に使うのだと思って。それを実現する行動力があるところが、こういう実験の実施につながったのかなと思います。ちょっと Zoom の音声の関係で、ロボットの声が多分ノイズキャンセルみたいなのに引っかかったのかなと思うんですけど。ちょっと Pepper の声、聞こえなかったかもかもしれない。

平野：ほんとですか。

新垣：はい。ちょっと残念だったんですけど。

森：そうですね。

新垣：はい。Zoom の設定かな、と思います。

森：ちょっと笑えるところが、ほんとはあったんですね。

新垣：ほんとは。

平野：そうだったんですね。

新垣：難しいです。

平野：そうだったんですか。ノイズキャンセルされちゃうんですね。

森：すいません。ちょっと残念ですけど。でも、雰囲気はすごくよく伝わってきたと思います。ありがとうございます。

じゃあ、残り、あと 15 分ぐらいになるんですけども。今、平野さんのほうからご研究について説明していただいたところで、これからの人とロボットの共生社会として、今後さらにどんな課題があるのか、どんなことが必要になっていくと思われますでしょうか。

新垣：これは誰が。平野さん。

平野：はい。人とロボットの今回の実験の私のテーマも、人とロボットの共生社会の実現っていうテーマなんですけれども。主に3つ、思ったことがあって。

1つ目が、ロボットと共生社会を築く上で、やっぱりロボットに対して人間らしさとかを付与するのが大切だなんていうふうに思いました。結構、感想でも「人間みたいでかわいい」とかっていう感想が多くて。相槌を打ってくれたり、あとは人の顔を追跡できる機能もあるんですけど、Pepper に。そういった機能を付けることで、目を合わせてくれるとか、あとは一人称が、「僕」とか「私」とか付けることで、「人みたいだった」という感想がすごく多くて。そういった人たちは親しみやすさを結構高く評価してくれていたんで、そういう人間らしさというのも大切だなと思いました。

2つ目が、さっきも少し言っちゃったんですけど。「ロボットも空気を読めるよ」っていうのを与えることがすごく大切なと思って。ロボットって何をするか分からないから、結構人のほうが下手に出る、みたいな。接する時にロボットの意見をくみ取るように頑張ろう、頑張ろうとして疲れちゃうっていうのが先行研究でもあったんですけども。

そういった疲れを感じやすいので、ロボットも人間と同じような暗黙のルールが分かったりとか、そういう

た、常識って言うと、ちょっと言葉が違うのですが、そういった「空気を読めるよ」みたいな、これも人間らしさにつながると思うんですけど、そういうのを実装したロボットが、共生社会で大切かなと思いました。

3つ目が、こういったロボットの実験をする上で、すごく大切だなと思ったのが、人をよく知ることだなというふうに思っただけ。私も今回、ロボットをどのように使うかを考えた時に、結構、人の認知心理学、先生のゼミで2年生の時にやった勉強とかが結構大切になるんだなって、先行研究を読んでもすごく思ったので、ロボットを知るには、まず人を知ることがすごく大切なのかなと思いました。

森：はい。ありがとうございます。ロボットにも人間らしさを与えるということと、そして、人をよく知ることが大事だということで、とてもすてきなご意見、ありがとうございます。

平野：ありがとうございます。

森：はい。新垣先生のほうでは、学生さんのこういった実験をご覧になっていて、どういうことに注意してこられたかとか、今後どういうふうに展開していきたいとか、ございますでしょうか。

新垣：ロボットと人の対話実験は敷居が高いものですが、こういうふうにちょっとやってみればロボットを使えるのだとか、自分でもできるんだ、ということを実感します。そうすれば、研究の幅がすごく広がりますよね。

ロボットも、先ほどの階級の分析もそうですけど、技術を持ってなくても、いろいろと使えるツールが、世の中に出てきているので、それを使って文系の学生であっても研究の幅を広げることができます。ロボットと共生社会といっても、実験でできることは非常に狭い範囲の条件の違いしか見ることにはできないのですけれども。それでも、こういう技術を活用して、積極的に研究してもらえるといいなと思います。文系ならではの視点で問題をとらえて分析して、研究が進んでいくとよいなというふうに思っています。

森：はい。ありがとうございます。新垣先生にもおまとめいただきまして、ありがとうございます。今回、コロナ禍で新垣先生とお知り合いになって、そしてこういった形で、いろいろ協業っていうのがうまくできてきて、私もとてもいい学びができたと思っています。今後とも、ぜひ、こういった協業をさせていただければなというふうに考えておりますので、どうぞよろしくお願いします。

新垣：よろしく申し上げますというか、先生とはこういう形の授業ができるといいですね。

森：そうですね。

新垣：こういう形というか。データサイエンスの技術を使うと、いろいろな問題解決をするためのツールとかもあるということにも気が付きますし、そういうものを学生でも使えることができます。平野さんは、実際に実験をするのはすごく大変だったと思いますけれども。どうでしたか。何かもし、実験する上で大変だった感想とかあれば。

森：お願いします。

平野：すごい結構大変だったんですけど。Pepperにプログラミングするのとかも、一から全然分かんなかったんで、一から大変だったんですけど。結構、森先生も新垣先生も親身になって教えてくださっていたから。それで、大人2人が親身になってやってくれているのだから、もう頑張んなきゃ、みたいな感じが結構強かった感じです。

夏休みの時とかも、お二人とも時間削って、私のゼミ生の、もう1人仲間もいるんですけども、全然自分の卒業研究に関係ないのに、来てくれたりして。もう、みんながすごく手伝ってくれたので、優しい環境に恵まれ、できました。

新垣：ありがとうございます。

森：ありがとうございます。

新垣：どうぞ。

森：いいえ。やっぱり平野さんの頑張りモチベーションの高さ、そういったことが今回の成功につながったと思います。私も勉強になりました。ありがとうございます。

平野：いえ。こちらこそ、ありがとうございます。

森：はい。あと、お二人のほうから何か、ここで話ししておきたいこととかございませんでしょうか。

新垣：大丈夫です。

質疑応答

森：ありがとうございます。先ほど、おまとめいただきまして、ありがとうございます。こういったことで座談会のほう、終わりにさせていただきたいと思いますが、何か皆さまのほうからご質問等ございましたら。小宮路先生、まだお時間は、もう少しありそうでしょうか。

小宮路：はい、どうぞ。チャットでご質問があるようです。森先生、お願いします。

森：はい。チャットありがとうございます。何もしないロボット、弱いロボット。ロボットの人権についてコメントお願いしますということで、ありがとうございます。何もしないロボットというのは、何もできないけれども、そこにいるだけで和ませているようなロボットのことを、もしかしたら、おっしゃってますでしょうか。こういったロボットも人の心理的なメンタルに関連するロボットとして、とても重要だと思うんですけども。新垣先生、いかがでしょうか。

新垣：もしかしたら、豊橋技術科学大学の岡田先生の研究のことかと思いますけど。

平野さんの研究では、平野さんの意見では、ロボットに、もっと人間味を持たせようという形ですけども、ロボットが実際に有能とは限らなくて。私たち人間であっても、会話が最初から、ちょっとうまく言えないのですけれども、ロボットができないことを見せた時に、周りの人が助ける。ごみを拾えないごみ箱ロボットがいたら人間のほうがごみを拾って、入れてあげるというような関係、ロボットが自分が何ができるのかということを開示することで、そこで人間とロボットの関係ができていくような、ごみを捨てるということを互いに補い合って実現する。

実際の、人と人のコミュニケーションにも、そういう面があると思います。そのような要素を人とロボットとの関係にも取り入れていくというような研究は、すごく大事なことだというふうに、私は思っています。

割とロボットって、もう万能で何でもできるっていうような前提で考えられているのですけれども。やっぱり、ロボットには何ができて何ができないのかっていうのを見せる、外に見せることで、人とのコミュニケーションは円滑になっていくのではないかなというふうに、私自身はロボットの専門家ではないんですけども、そういうふうに思います。

森：はい。ありがとうございます。

平野：私からもいいですか。

森：はい。お願いします。

平野：ありがとうございます。実は最初、私は弱いロボット派で、実験を進めようとしてたんですけど。卒業研究でも弱いロボットっていう、岡田美智男先生が研究されてる分野で、そこにすごく共感を持って、弱いロボットみたいなのを卒業研究でやろうと思ってたんですけど。

いろいろあって、私は集団規範にしたんですけど。結構、集団規範も、人とロボットが相互に影響を与え合うことが大切だになっていうふうに感じました。この弱いロボットも、ロボット自体は弱い、何もできないんですけど、それに気付いて人間がやってあげるっていうので、人とロボットが相互に影響を与え合っているとい

う点では、全く違うところなんですけど、同じような観点を言ってるなと思っていて。

私は集団規範っていう、人とロボットの共生の、人間らしさっていうところに注目して卒業研究を書いたんですけど、また違う分野で、それこそペットとか。最近、ペットでロボットっていうのがあると思うんですけど、そういった点では弱いロボットもすごくいいなと思ったので、うまくまとめられないんですけど、人とロボットが相互に影響を与え合うっていう点では、大きなくくりでは、同じような感じの研究なのかなとも思いました。

森：はい。ありがとうございます。私のほうでは、ちょっと違う観点でお話ししてしまいましたが、弱いロボットということで、できないことを開示して、その上で、人とロボットがうまくコミュニケーション取れる、相互に影響し合えることができるロボットということで。ありがとうございます。

もう1つ、ご質問がありまして。ロボットの人権について、コメントお願いしますってことなんですけども。お二人のほうから何かございますでしょうか。

新垣：平野さん。

平野：すごく難しい問題ですよ。ロボットの人権は多分、政府側も難しいなって言っていると思うんですけど。私は、これは個人、ほんとに個人的な感想なんですけど、やっぱりロボットと共生っていうふうに政府も掲げてる以上、そういったロボットに適用した法律とかも今後は作らないといけないんじゃないかなというふうに思っています。

ロボットが人間と対等とは言わないんですけど、同じような感じで、人と人が会話するような感じで人とロボットが会話するみたいな感じにどんどん近づいていく時に、やっぱり人権問題は大きく取り上げられると思うので。そこは、時代の状況に合わせて、法律とか人権問題とかも変えていく必要が、考えていく必要があると思うので。私的には今ロボットの性能とかが結構注目されているけど、社会性っていう意味で、ロボットの人権は何かの授業の中で取り上げたりとか、そういった授業科目を作るのも必要かなって思います。

森：はい。ありがとうございます。

新垣：私はロボットの人権というのはあまり考えたことがないのですけれども。人は、こういう人工的なもの、ロボットに対してでも、すごく人間っぽいものを思い入れてしまう。人間じゃないけど、人のように感じてしまうことが知られています。ですので、ロボットが痛い思いをしたら、人もつらくなる、というような、そのような感覚を持つことが知られています。

だから、ロボットがどういう形であるべきか、人間っぽくあるべきなのか、ちょっと人権の話とは、ちょっとずれてしまうかもしれないんですけど。どういう形であるべきかというようなことは、本当は大切な研究テーマではないかと思います。倫理的な観点からも大切なテーマだと思っています。すいません。ちょっと回答にはなっていないと思うんですけど。

森：はい。ありがとうございます。そうですね。ロボットの人権というと、以前、ちょっと学生さんにもそういったことを聞いてみたことがあるんですけど、「ロボットは人ではないから人権ではない」というような答えも返ったようなこともありましたけれども。

確かに、ロボットがこれから人と共生していくっていうことを考えますと、ロボットがひどい目に遭うと、やはりそれを見ている人も傷ついてしまうっていうこともあるかもしれませんし。そういった影響も含めると、これからの時代に応じて、ロボットの人権、ロボット権といいますか、そういったことも考えていかないといけない時代が、これからやっぱり来るんだなというふうなことを少し、お二人のコメントを伺って感じました。ありがとうございます。

小宮路：はい。森先生、ありがとうございます。

森：はい。ありがとうございます。もう1つ、ご質問いただいているようですが。じゃあ、ちょっと読み上げ

させていただきます。まだお時間、よろしいですか。

小宮路：ご質問は、プログラミング学習において、学生差が大きいという点の工夫について、ということですね。これは、座談会というより、森先生が実際、授業担当されてということだと思いますけれども。

森：はい。プログラミング学習について、学生差が大きいことについて、どのような工夫を、ということ。

まず、オンデマンドで授業をしていますので、授業の中でこういうふうにはレコーディングをしていく話もやっていますけれども。対面ですと、それを聞き逃してしまって分からないとかいうのがありますけれど。まず、オンデマンドでしたら、映像をまた元に戻したり止めたりしながら、それぞれが何とか同じことをやっていけるということが、まずはできるというのがありますし。なるべく分かりやすい資料を一緒に提供するというところを行っています。

その中で、やはりデータサイエンス基礎、機械学習基礎ということで、あまり難しいプログラムにならないように、なるべく平易なプログラムで、短いプログラムで、こんなことができるっていうところをまず体験してもらおうということを大事にしていますので、そういった中で、「全く無理です」ということを言う学生さんは、今までは特にはなかったと思います。なるべく結果が分かりやすく、楽しく取り組めるもの、そういったことを題材に行っています。答えになってましたでしょうか。

閉会挨拶

小宮路 雅博

成城大学データサイエンス教育研究センター長／経済学部教授

小宮路：はい。どうもありがとうございます。大体少し時間が押してきましたので、私のほうで引き取ります。皆さん、どうもありがとうございました。

森：ありがとうございました。

小宮路：成城大学は、先ほども言いましたように、文系4学部での構成ですので、今まで学内で、こういった形で研究をしたり、あるいは学んだりという機会があまりなかったのですね。ちょっと縁遠かったということですね。

データサイエンス教育研究センターが出来上がって、データサイエンス教育の一環ですということで、予算を付けて、ロボット導入したりしまして、それから他のいろいろなものも、触って動かしたりするようなものも、少しずつ揃えたりしてきました。体験型ということで授業をしたり、それから、それを学部の教育というか、卒論でも良いのですが、いろいろ持ち帰ってもらって、そこから新しいテーマ性が生まれてくるようになりました。

工学部があって、ロボット研究の先生がいてというわけではありませんし、学生さんもまた、そういうことを学んでいるわけでもない。とは言っても、全然知らない世界ではなくて、触ってみたり動かしてみたり、試行錯誤しながら導入してみたりと

「データサイエンス教育 講演会」当日のタイムテーブル

というのが、成城大学らしい展開かなというように考えています。

これがデータサイエンス教育をやる波及効果なのかなと私は考えています。直接にはやはり、先ほども言ったように、なるべく多くの学生さんがデータサイエンスの科目群を履修してもらって、文部科学省さんが認定制度の中で想定しているような形で教育を受けていってもらえればというように考えております。

はい。それでは、どうでしょうか。ご質問のロボット三原則の話は、すみません。時間が来てしまいました。このご質問については、森先生のほうから、後でご返事いただくって形にいたします。すみませんが、よろしくお願いいたします。

森：はい。大丈夫です。お返事させていただきます。

小宮路：それでは皆さま、成城大学のデータサイエンス教育ということで、工学部があるわけでありませんので、最先端のことをやっているというわけではないですけども、社会に入っていく時は、最先端が入っていくとは限りませんので、今あるもので、その社会との、あるいは人間との関わり合いというところで、いろいろ取り組んでいけたらなというように思います。

それで、今日は私、個人的といったらあれですけども、とても感銘を受けたのは、ロボットの人権というご質問をいただいたことです。そうしますと確かに、本学には法学部もありますし、それから文芸学部で、ロボットの権利というようなテーマで、過去さまざまな作品がありますから、例えば、それは文芸学部的だなんていうように思いますし。それが社会とか、あるいは企業活動の中で、どのようにくみ取られるのかという観点からは、経済学部もまさしく該当しますので、とてもいいテーマをいただいたなというように思っています。

実現できるかどうか分かりませんが、法学部の先生、あるいは文芸学部の、そういうロボットや、あるいは人のように見えるけれども人でない存在が登場するような作品を研究している先生方と、データサイエンス教育研究センターが主催して、シンポジウム的なことを将来やればなというように思いました。

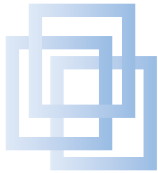
それでは時間が、ちょうど終了の時間になりました。今日は皆さま、ご参加いただきまして、ありがとうございました。チャットでアンケートの URL をお送りしますので、ご協力をお願いいたします。アンケートの URL ご確認いただきまして、ダウンロードいただけましたら、Zoom からご退室いただければと思います。

今日は皆さま、お忙しい中、ありがとうございました。以上をもちまして、成城大学データサイエンス教育研究センター主催の講演会を終了いたしたいと思います。皆さま、本日はありがとうございました。

森：ありがとうございました。

平野：ありがとうございました。

新垣：ありがとうございました。



センターだより

1. データサイエンス教育研究センター構成員（2022年度）

1) 成城大学データサイエンス教育研究センター委員会委員

委員長 小宮路 雅博 センター長／経済学部教授

センター員 森 由美 特別任用教授

センター員 玉谷 充 専任教員／経済学部専任講師

経済学部選出委員 渡邊 隼史 准教授

文芸学部選出委員 松村 一志 専任講師

法学部選出委員 村上 裕章 教授

社会イノベーション学部選出委員 積田 淳史 准教授

2) 成城大学データサイエンス教育研究センター自己点検・評価委員会委員

委員長 小宮路 雅博 センター長／経済学部教授

センター員 森 由美 特別任用教授

センター員 玉谷 充 専任教員／経済学部専任講師

大学事務局 大友 浩一 事務局長

大学事務局 中村 裕 総務課長

3) 外部アドバイザー委員会委員（50音順、敬称略）

大成 弘子 株式会社 Interbeing/CAO (Chief Analytics Officer)

標葉 隆馬 大阪大学 社会技術共創研究センター 准教授

田村 光太郎 株式会社ユーザベース UB Research チーフリサーチャー

三浦 航 楽天グループ株式会社 AIサービス統括部 アナリティクスデータエンジニアリング部サイエンスプロダクトグループ データサイエンティスト

山田 健太 琉球大学 国際地域創造学部 准教授

4) 事務局

中村 裕 総務課長

井上 貴博 総務課長補佐

2. 委員会活動

1) 成城大学データサイエンス教育研究センター委員会

2022年5月19日 第1回データサイエンス教育研究センター委員会 (Zoom)

2022年6月28日 第2回データサイエンス教育研究センター委員会 (Zoom)

2022年7月27日 第3回データサイエンス教育研究センター委員会 (持ち回り)

2022年9月20日 第4回データサイエンス教育研究センター委員会 (Zoom)

2022年11月7日 第5回データサイエンス教育研究センター委員会 (持ち回り)

2022年11月22日 第6回データサイエンス教育研究センター委員会 (Zoom)

2023年3月7日 第7回データサイエンス教育研究センター委員会（持ち回り）

2) 成城大学データサイエンス教育研究センター自己点検・評価委員会

2022年5月27日 第1回データサイエンス教育研究センター自己点検・評価委員会（持ち回り）

2022年7月18日 第2回データサイエンス教育研究センター自己点検・評価委員会（持ち回り）

3) 外部アドバイザー委員会

2023年2月17日 第1回外部アドバイザー委員会（Zoom）

3. データサイエンス科目群履修者数（2022年度）

1) データサイエンス概論 ※定員各80名

曜限	学期	経*	文	法	社	計
火3	前期	11	40	4	25	80
木3	前期	59	8	6	7	80
木4	前期	62	7	6	5	80
火2	後期	29	29	5	16	79
木2	後期	23	21	10	19	73
金5	後期	31	30	5	14	80
計		215	135	36	86	472

*経：経済学部、文：文芸学部、法：法学部、社：社会イノベーション学部、以下同じ。

2) データサイエンス基礎 ※定員各63名

曜限	学期	経	文	法	社	計
水2	前期	9	37	8	9	63
水4	前期	33	21	3	6	63
木1	前期	22	22	3	16	63
火3	後期	3	38	0	22	63
木3	後期	42	12	4	5	63
木4	後期	34	14	7	8	63
計		143	144	25	66	378

3) データアナリティクス基礎 ※定員各63名

曜限	学期	経	文	法	社	計
木1	前期	29	17	6	11	63
木1	後期	20	37	0	6	63
計		49	54	6	17	126

4) 機械学習基礎 ※定員各 63 名

曜限	学期	経	文	法	社	計
火 4	前期	25	21	2	14	62
木 1	後期	34	9	2	18	63
計		59	30	4	32	125

5) データアナリティクス応用 ※定員 40 名

曜限	学期	経	文	法	社	計
金 3	後期	5	0	2	2	9

6) データサイエンス・アドバンスド・プログラム ※定員 40 名

曜限	学期	経	文	法	社	計
金 4	後期	6	0	0	4	10

7) 数理科学基礎 a ※定員なし

曜限	学期	経	文	法	社	計
金 2	前期	8	4	0	3	15

8) 数理科学基礎 b ※定員なし

曜限	学期	経	文	法	社	計
金 2	後期	2	1	0	4	7

9) 合計 (延べ履修者数)

	経	文	法	社	計
計	487	368	73	214	1,142

4. 履修証明（ディプロマ）授与件数（2022年度前期履修分）

ディプロマ名称	授与件数	備考
リテラシーレベル・ディプロマ	11名（累計11名）	※2022年度以降入学者対象
基礎力ディプロマ	15名（累計73名）	※2021年度以前入学者対象

※2022年度後期履修分ディプロマ授与件数は2023年度年報に掲載予定。

5. 主なイベント

※2023年1月末までに行ったイベント

2022年6月28日 日本経済新聞出版ムック本

『大学選びを考えたらまず読む本 2023年度版』掲載

先方から依頼があり5月13日に取材を受ける。「デジタル化で大学が変わる」という特集ページの中で、成城大学が掲げている「文系学生こそデータサイエンスを」という考えと実践について詳しく掲載された。

<https://www.seijo.ac.jp/news/jtmo420000014npo.html>

2022年7月9日 「データサイエンス研究 講演会」オンライン開催（Zoom）

講師：森由美 特別任用教授、玉谷充 センター専任教員／経済学部専任講師

森由美 特別任用教授から「ヘルスケアデバイスや質問票を活用した健康状態の可視化に関する研究事例」と題し、健康状態の測定手法についての最新動向と、自身による「ウェアラブルデバイス等を活用した工場検査員を対象とした研究事例」についての紹介・説明が行われた。

玉谷充 センター専任教員／経済学部専任講師からは「高次元小標本における判別手法について」と題し、判別関数の精度を上げるための理論的研究の説明があり、講演の中では、理論上最適な判別関数と言われている Fisher's Rule, 「次元の呪い」といった高次元小標本における問題点等の説明とナイーブ正準相関に基づく判別関数の提案等が行われた。

参加者24名。

<https://www.seijo.ac.jp/news/jtmo420000014uyi.html>

2022年7月17日 オープンキャンパス

8月6日

8月7日

データサイエンススクエアにて、担当教員と話せる個別ブースや各種展示コーナーを設け、データサイエンスについての説明会を実施した。

2022年8月1日 ワークショップ「プログラミング入門—Telloを飛ばしてみよう—」

講師：増井敏克氏（増井技術士事務所代表）

ドローンの定義から、その市場規模、関連法規、産業用ドローンの技術向上、近年活用が進むドローン事例などを幅広く学んだ。アプリを使用し、小型ドローン「Tello」を操作し、離陸、写真撮影、着陸など基本的な動作を体験した。

主の目的であるプログラミングは「DroneBlocks」というScratchソフトを使用し、ドローンをさまざまな条件で自動飛行させるという練習課題に取り組んだ。

参加者10名。参加者にはオープンバッジが発行された。

<https://www.seijo.ac.jp/news/jtmo4200000153m1.html>

2022年8月2日 ワークショップ「統計分析ソフトR入門講座」

8月4日

講師：増井敏克氏（増井技術士事務所代表）

2日に渡り、統計分析ソフトRの入門講座を開催した。

Rの導入と基本操作からグラフの描画、偏差値の求め方など練習問題を解きながら基本的な機能を学習した後、回帰分析、主成分分析、クラスター分析といった高度な内容にも取り組んだ。

【8月2日】

2時限：Rの基本操作①、3時限：Rの基本操作②、4時限：各種グラフの描画

【8月4日】

2時限：回帰分析、3時限：主成分分析、4時限：クラスター分析

参加者13名。参加者にはオープンバッジが発行された。

<https://www.seijo.ac.jp/news/jtmo4200000155ok.html>

2022年8月3日 **ワークショップ「言語データと心のメカニズム」**

講師：新井学 経済学部准教授

言語データを用いて心のメカニズムを研究する方法を紹介した。

人間の高度な認知機能に欠かせない「予測」の働き、誤った判断・解釈に対する「抑制」の働きに焦点を当て、実際に眼球運動測定実験のデモを行い、得られたデータをどのように分析するのか解説が行われた。

2時限：(講義) 言語データを用いて心のメカニズムを研究する

3時限：(実践) 眼球運動を用いた心理言語実験

4時限：(演習) 実際に得られたデータを統計分析ソフト R で分析

参加者 17 名。参加者にはオープンバッジが発行された。

<https://www.seijo.ac.jp/news/jtmo4200000155uf.html>

2022年9月2日 **G 検定講習会オンライン開催 (Zoom)**

10月1日

10月15日

日本ディープラーニング協会主催「G (ジェネラリスト) 検定」受験対策講座として、初回の9月2日は事前数学講習会として、玉谷充 センター専任教員/経済学部専任講師よりG検定の問題に対応できる基礎知識を身に付ける為の講義が行われた。10月1日と15日は合同会社 Interbeing の CAO 大成弘子氏を講師として招き、更に理解を深める講義が行われた。

参加者 6 名。

<https://www.seijo.ac.jp/news/jtmo42000001577w.html>

2022年10月20日 **東洋経済新報社ムック本取材**

成城大学でのデータサイエンス教育の全体像に焦点をあて、文系大学でデータサイエンスを学ぶ意義や取り組み方を伝え、掲載された。併せて、特集ページ「変革の現場に迫る」へも、森由美 特別任用教授、玉谷充 センター専任教員/経済学部専任講師のインタビューそれぞれが掲載された。

2023年1月13日発行。

<https://www.seijo.ac.jp/news/jtmo420000017e0s.html>

2022年11月2日 **Scratch 音楽演奏プログラミング・コンテスト**

データサイエンス科目「データサイエンス基礎 (旧: データサイエンス入門 I)」の授業で、Scratch 言語プログラミングを体験した学生を対象とした、音楽演奏プログラミング・コンテストが行われた。9月に一次審査、10月に学内投票による二次審査を経て、金賞2件、銀賞2件、銅賞4件が決定した。

受賞者には賞状が授与され、オープンバッジが発行された。

<https://www.seijo.ac.jp/news/jtmo4200000167c2.html>

2022年11月12日 **データサイエンス・コンテスト 2022**

募集テーマを「オリンピック参加選手データ (1896年～2016年) を使って、自由に分析してください」とし開催。作品審査の結果、「チームエーケン」(提出作品「オリンピックから見える国家の経済力—金メダルとGDPの関係性—」)が「優秀賞」と「特別賞」のダブル受賞となった。

受賞チームには賞状と副賞のほか、各人に優秀賞 (1st Prize) のオープンバッジが発行された。

協賛：一般社団法人 ピープルアナリティクス & HR テクノロジー協会

https://www.seijo.ac.jp/news/jtmo420000016bav.html?fbclid=IwAR2k3d0Q9_XzBl8eJ8OAbWDkX-JsLT70tUDRrT3cEgqilURJq8aK5woX-rw

2022年12月2日 **講演会「成城大学におけるデータサイエンス教育」オンライン開催 (Zoom)**

講師：森由美 特別任用教授

ゲストスピーカー：新垣紀子 社会イノベーション学部教授、

社会イノベーション学部4年生 平野なお佳

森由美 特別任用教授が「データサイエンス科目でのSTEAM教育の実践」と題した講演を行った。社会イノベーション学部の新垣紀子 教授と同学部4年生の平野なお佳さんと3名で「人とロボットの共生社会実現に向けて—一人型コミュニケーションロボットとの対話実験の試み—」と題した座談会も併せて行われた。

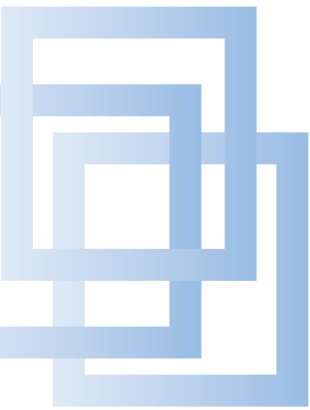
参加者 36 名。登壇した新垣先生と平野さんにはオープンバッジが発行された。

<https://www.seijo.ac.jp/news/jtmo420000016z9w.html>

6. 2022 年度に発行したオープンバッジ一覧

※ 2023 年 1 月末までに発行したオープンバッジ

発行日	バッジ名	バッジ	発行数
2022/8/1	成城大学データサイエンス・ワークショップ 2022 Summer 「Drone Programing」 参加証		11
2022/8/3	成城大学データサイエンス・ワークショップ 2022 Summer 「言語データと心のメカニズム」 参加証		17
2022/8/4	成城大学データサイエンス・ワークショップ 2022 Summer 「R 入門講座」 参加証		12
2022/9/20	成城大学データサイエンス基礎力育成・認定プログラム 「データサイエンス リテラシーレベル・ディプロマ (履修証明)」 2022 年度		11
2022/9/20	成城大学データサイエンス教育プログラム 「データサイエンス基礎力ディプロマ (履修証明)」 2022 年度		15
2022/11/2	成城大学データサイエンス 「Scratch 音楽演奏プログラミング・コンテスト (2022 前期)」 金賞		2
2022/11/2	成城大学データサイエンス 「Scratch 音楽演奏プログラミング・コンテスト (2022 前期)」 銀賞		2
2022/11/2	成城大学データサイエンス 「Scratch 音楽演奏プログラミング・コンテスト (2022 前期)」 銅賞		4
2022/11/12	成城大学データサイエンス・コンテスト 2022 優秀賞 (1st Prize)		4
2022/12/2	「成城大学におけるデータサイエンス教育」(2022/12/2 開催) 登壇者 (スピーカー) 参加証		3
2023/1/16	成城大学データサイエンス教育研究センター ティーチング・アシスタント 2022 年度 (活動証明)		3



編集後記

2022年度の成城大学データサイエンス教育研究センター年報を刊行いたしました。

本センターにおいては初めての刊行（創刊）であり、なおかつ、自身にとっては初めての編集長という役割をセンター長から任命され、一体どうなるのかと思いましたが無事に仕上げることができました。編集長という肩書きをいただいたとは言っても、特に大きな役割を担ったという訳ではなく、論文のページを年末年始に執筆する一方で、出来上がったページのチェックをするといった作業が主であり、事務室の井上さん並びに事務職員たちのお陰で表紙や講演録、センターだより等のページが仕上がったと言えます。この場を借りて、心から感謝申し上げます。

私は2022年度から本学のセンター専任教員として新たに着任しましたが、担当する「データサイエンス概論」のオンデマンド資料を作成したり、「数理科学基礎 a・b」の講義資料を作成したりと、年間を通じて教育業務に追われた1年間だったなと感じています。そして、2023年度から新たに1名のセンター専任教員が加入することになりますが、本学のデータサイエンス教育がうまく機能するようブラッシュアップしていきます。また、受講する学生にも「データサイエンス科目を履修して良かった」と思えるように努めていきたいと思いますので、2023年度も引き続きよろしくお願い申し上げます。

成城大学データサイエンス教育研究センター専任教員／経済学部専任講師 玉谷 充

成城大学データサイエンス教育研究センター
2022年度 年報

Education and Research Center for Data-driven Social Sciences & Humanities, Seijo University
Annual Report 2022

発行 成城大学データサイエンス教育研究センター
〒157-8511 東京都世田谷区成城 6-1-20
<https://www.seijo.ac.jp/education/support/cds3/>
発行日 2023年3月20日

※本年報の一部または全部を無断複写・転載・電子媒体等に加工することを禁じます。



The logo for Seijo University Data Science is centered on the page. It features the text "Seijo University" in a small, blue, sans-serif font at the top. Below it, the words "Data" and "Science" are stacked vertically in a large, bold, blue, sans-serif font. The "D" in "Data" is significantly larger than the other letters. At the bottom of the logo, the text "Since 2015" is written in a small, blue, sans-serif font. The background of the page is a light blue and green gradient with abstract, flowing shapes.