

数理・データサイエンス・AI教育の紹介

オープンエデュケーション & オープンサイエンス with オープンマインド： 九州大学におけるデータサイエンス教育

九州大学数理・データサイエンス教育研究センター長 システム情報科学研究院・教授 **内田 誠一**



1. はじめに

本学においては「オープンエデュケーション&オープンサイエンス with オープンマインド」をスローガンとして「全分野横断・全学年縦断データサイエンス（以下、DS）教育」を2015年度から推進しています。図1に示すように、その構成は重層的です。すなわち全分野向けのリテラシーレベルから応用基礎レベル、そして数理・情報向けのエキスパートレベルまでと非常に多様です。

本稿では、特に「非」エキスパート向けのDS教育事業群について、実施開始の時系列に沿って、それぞれ概要を説明いたします。その中で、それら事業を効率的に推進できたポイントとして、上記の「オープン」ポリシーがあることを説明いたします。

的かつ再現性のある分析が必須となっています。その一方で、データ分析を習っていない、もしくはあまり積極的に学んでこなかった学生諸君の中には、卒業研究等で急にデータ分析が必要になった者がいるはずで。そこで、こうした学生の「駆け込み寺」が必要だと考えました。このモチベーションのもと、2015年より準備を開始、2017年度より開講し、現在まで毎年開講しています。

DS実践特別講座は、DS概論・DS実践（講義・演習）・DS発展の4科目で構成され、現在情報系教員4名により運営されています。DS概論は座学講義です。本講座を受講する学生は、そもそも数学やDS苦手意識があるはずなので、それを取り除くべく、図や例を豊富にし、なるべく用語も平易にしつつ、丁寧すぎるぐらい丁寧に進めています。次ページ図2はDS概論のスライドの抜粋です。例えば正規分布については、いきなりexpが入った式を見せるのではなく、「本質的には2次関数を変形したもの」といった説明を入れて恐怖心を下げるようにしています。

DS概論として教授すべき範囲は非常に広いために、丁寧さと範囲のバランスには注意しました。現状では、「データとデータ分析」「データのベクトル表現と集合」「平均・分散・相関」「データ間の距離と類似度」「クラスタリングと異常検出」「線形代数に基づくデータ解析の基礎」「主成分分析」「予測と回帰分析」「可視化」「確率と確率分布」「信頼区間と統計的検定」「非構造化データ解析」「パターン認識と分類」「データ収集とバイアス」「人工知能入門」という単元で構成されています。ご覧のように、線形代数から多変量解析、確率論や検定、さらには人工知能的な話

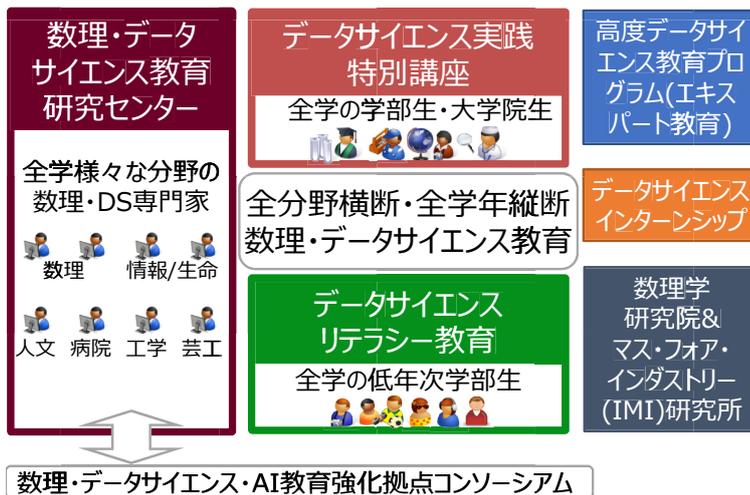


図1 九州大学におけるDS教育関連事業の全体像

2. データサイエンス実践特別講座

DS実践特別講座は、文理を含む全分野の高年次学部生・大学院生のうち、自身の研究でデータ分析が必要となった学生のために準備されました。いまや全分野の研究でデータに基づいた客観

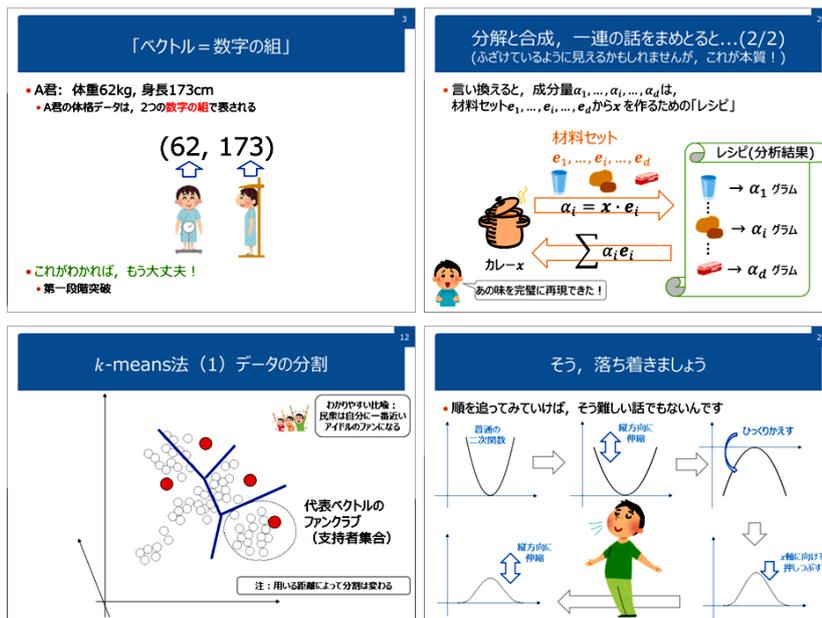


図2 DS概论で用いるスライド例

までと、それぞれで1科目になりそうな、非常に幅広い内容です。実は、開始当初は可視化や人工知能に関する話は入っていませんでした。その後受講者からのニーズに応える形で導入されました。

DS実践という科目では、概論で習った内容について、実際にPythonを使ったプログラミングを行いながら、その名の通り実践的な復習を行っています。多くはPython初心者であったり、それ以前にプログラミングの初心者であったり、さらには計算機の初心者だったりします。このため、まずはプログラミングの環境構築から始め、Pythonの文法を教え、その上でようやくデータ分析に取り掛かります。丁寧な支援のためには十分な数のTAも必要です。

DS実践の工夫の一つは、分析対象のデータに画像を多用している点です。表形式の数値データでもよいのですが、画像を用いることで結果が目に見えるというメリットがあります。例えば画像を対象としてクラスタリングすれば、どのような画像にグループされたか一目瞭然であり、それにより理解が深まる、といった具合です。また、全学様々な分野で画像を対象としたデータ分析が行われており、この点で実際の分析ニーズに即している、というメリットもあります。

最後のDS発展は、受講者がそれぞれ行っている研究課題について、データ分析の専門家が個別コンサルティングを行うというものです。1回

あたり45分程度のコンサルティングを、1か月程度の間を空けて合計3回行っています。

以上のDS実践特別講座の科目群について、ほとんどの学生は単位取得を希望していません。形式的には、同科目群は情報系大学院の講義と位置付けられていますので、学部生や情報系以外の学府の大学院生が受講したとしても、各自の部局で単位として認められないこともあるでしょう。(工学系大学院など、単位を認めるところもあります。)それ以前に、「寺小屋」である同講座は、「単位なんかとは無関係に、学びたい人が学べる場を提供したい」という考えが前提となっています。受講学生

もそれを理解した上で参加してくれているようです。なお、ご想像に難くないと思いますが、単位とは無関係に学びたいと思って参集してくれた学生諸君は非常に熱心です。教授する側としても教員冥利に尽きる感があります。

ここ数年、同講座の受講者は年々増え続けています。夏休みの集中講義形式にしたこと、自身の研究テーマをまだ持っていない低年次学部生も対象としたこと、さらにはDSの必要性を認識した学生諸君が増えたこと、などがその理由として考えられます。2021年9月の概論講義には全学から200名の申し込み(実参加は150名)がありました。図3は2021年度集中講義でのDS概论受講希望者の内訳です(150名登録段階)。期待した通り、受講者層は全学部・全学年に亘っていることがわかります。このうち実践まで受講するのが30~40名ぐらい、発展までは5~10名という状況です。

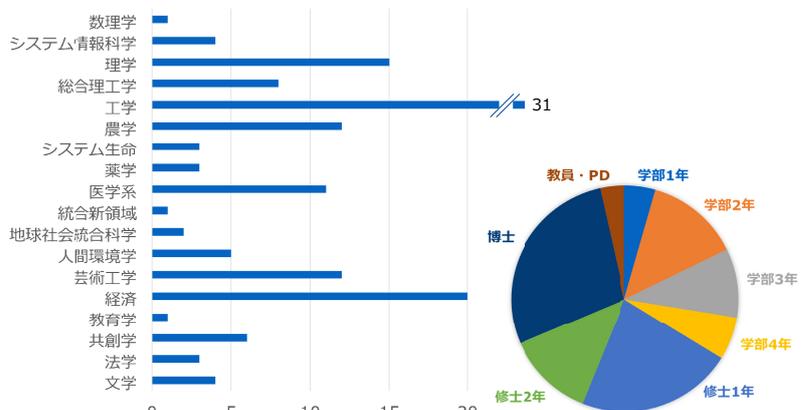


図3 2021年度DS実践特別講座 (DS概论) の受講希望者内訳

3. 低年次データサイエンスリテラシー教育

上述の通りDS実践特別講座には様々な部局からの参加があります。参加者に対する申込時アンケートを通して、所属部局ごとに必要としているデータ解析技術に差異があることがわかりました。さらにDS実践でのコンサルティングを通して、より具体的な研究内容についてもわかってきました。そうした情報をDS実践特別講座の担当教員だけで共有するだけでは「もったいない」ことに気づきました。すなわち、この情報を1年生向けのリテラシー教育に使えるか、と考えました。

もちろん本学では、他大学と同様、数学や情報学を含めて様々な基礎教育を低年次学生向けに数多く行っています。そうした教育と並行して、「君たちの部局の先輩には、実際にデータを使ってこのような研究をしている。だから、より興味を持って数学や情報学に取り組むべき」ことを伝えれば、それら数理・情報基礎教育に対する受講モチベーションがより上がるのではないかと考えました。

この考えに基づいて開始したのが、低年次データサイエンスリテラシー教育です。基礎教育の中に「情報科学」という情報系リテラシー科目があったので、同講義の2コマの内容を改変して、DSの内容を教えることにしました。新規の講義を立ち上げることも考えたのですが、入念に準備された基礎教育のカリキュラムを変更することは難しく、このような内容の一部改変からスタートしました。

2コマという限られた時間数で、データ分析の面白さや必要性を伝えるために、二つの工夫を行

いました。第一は、受講者の所属部局に応じて内容を変えることです。「情報科学」は、全学の1年生向けの基幹科目であり、同内容のクラスが並列的に開講されています。大雑把に言えば部局ごとに開講されているので、ある講義には理学部の学生だけが受講している、といった状況になります。この状況を「学術分野ごとに用いるデータ分析技術が異なる」ことに活かしました。すなわち、受講者の部局に必要な分析手法を中心に教授することにしました。図4はその状況です。各受講クラスに4つの分析技術をピックアップし、1コマで2技術を教えるということにしました。第二は、受講者の所属部局に即した実例を教えることです。本節冒頭でも述べたように、これは自分も将来データ分析をする必要があることを認識してもらうためです。講義資料はDS実践特別講座のものをベースとし、そこに実例スライドを追加したものとなっています。

学生からのフィードバックによる講義スライドの改善も行っています。本学では、非常に早い時期から、BYOD (Bring your own device)、すなわちPC必携化を始めており、低年次向けの講義の多くは、Moodleを介して実施されています。情報科学の講義スライドもMoodle上で閲覧されます。そこで、このMoodle等を管理する本学のラーニングアナリティクスセンターと協力して、学生が直接スライドの良し悪しをフィードバックできるようにしました。具体的にはスライドの横に「わかった」「わからない」ボタンを付け、自由にクリックできるようにしました。幸いにして「わからない」押下はあまり多くはありませんでしたが、学生が実例に対して好意的であるなどがわかりました。最近でも、各単元で「わかりにくいスライド上位5枚」調査を行い、例えば文字情報が多いスライドには抵抗を示しやすいことなどもわかりました。DSのリテラシー教育の使命として、DSそして数理・情報の「面白さ」を躊躇なく理解してもらうべきと考え、今後もこうした改善は続けていく予定です。

2021年度からは、2コマから7コマに拡張し、よりDSや人工知能に関して幅広く教授するようにしました。背景にはDS・AI教育が国策として推進されることになったことが挙げられます。現在、「様々なデ

様々なデ



図4 低年次DSリテラシー教育における受講者部局別教授内容

ータとデータ分析」「AIと機械学習」「非構造データとパターン認識」「データの収集」「ベクトル・距離・類似度」「データの可視化」を6コマでそれぞれ教授し、残り1コマでは以前のように「画像解析」「回帰分析と時系列分析」「相関と検定」「アンケート解析」の中から2技術を受講学生の所属部局に応じて選んで教授しています。

2021年からは、工学部全学科の1年生向けに、「データサイエンス序論」を新設・必修化しました。同学部の改組のタイミングだったため、カリキュラムへの導入は比較的容易でした。同講義では、半期15コマの授業すべてがすべてのDS・AIに関するもので、内容的には上記の情報科学の7コマをさらに拡張したものとなっています。同講義の教材も、やはりDS実践特別講座のものがベースとなっています。工学部は全学学生定員の30%を占めるため、この科目導入により、さらにDS教育が全学的に拡大推進されたこととなります。

4. 数理・データサイエンス教育研究センター

高年次・低年次向け教育を準備・推進する中の2016年12月、本学は文部科学省の重点支援事業である「数理及びデータサイエンスに係る教育強化」に、九州ブロック校として採択されました。これを受け「数理・DS教育研究センター」が2017年度に設立されました。

センター設立に当たっては、部局すなわち学術分野の壁を越えた協力を得ました。具体的には、人文科学研究院・芸術工学研究院・工学研究院(防災分野)・病院・統合新領域学府ライブラリサイエンス専攻・システム生命科学府など、数理・情報系以外の分野の協力を得ました。それら部局には、それぞれ1名のセンター専任教員を配置しました。さらに、リテラシー教育の責任部局である基幹教育院やラーニングアナリティクスセンターも協力部局になっていただきました。

こうした分野横断型のセンターを作ることは、むしろ当然と考えました。なぜなら、第一には本事業の目的が「数理・DS教育の全分野および全国展開」であったためであり、第二にはデータ分析の方法に部局ごとの差異があることもわかってきたからです。上記部局にはそれぞれの分野でのデータ分析の専門家が所属しており、協力を頂くのは難しくはありませんでした。数理・統計の専門家を擁する数理学研究院やマスフォアインダ

ストリ研究所の協力も非常にありがたいことでした。

センターの活動は多岐に亘っており、その一つが教材のオープン化です。センター発足に伴い、DS実践特別講座や低年次DS教育といった事業も、同センターで統括することにしました。特に、センターのホームページ⁴⁾を整備し、そこからDS関連講義スライドをダウンロードできるようにしました。一部の講義資料については、センター教員の尽力により英語化が図られました。現在も40種類以上のPDFを無償公開しています。また希望に応じて編集可能なPowerPointのファイルは無償提供しています。著作権もCC-BYとし、各大学の特性等に応じて自由に改変していただけるようにしています。公開後これまでに、19国公立大学、21私立大学、3高専、6企業・自治体にPowerPointファイルを提供しています。

センターが教員すなわち各分野の専門家で構成される点は重要です。それまでの高年次・低年次向け教育で得られるのはあくまで受講者である学生からの情報やフィードバックです。これに対し、センター関連教員はその分野のデータ分析の専門家です。いわば多様な分野におけるDSの「アンテナ」と言えます。このアンテナが収集した情報をベースに、互いに交流ができれば、各分野のデータ分析に関するより深い理解が進み、教育へのフィードバックも可能と考えました。

センター関係者の交流の場として、毎年一泊二日の合宿を実施しています。文字通り泊まり込みで、各分野のデータ分析動向や自身の研究内容をディスカッションします。人文・芸術・工学・病院・ライブラリサイエンスなどの研究者が一堂に会して、DSを合言葉に議論する状況は、非常に珍しいのではないのでしょうか。コロナ禍によりここ2年はオンライン開催となってしまいましたが、夜を徹して議論するあの熱い合宿の再開を楽しみにしています。

センター専任教員の所属部局は前述の通りですが、本学には他にも様々な部局があり、そこでも様々なデータ分析が行われています。そこで、そうした部局の研究者に「臨時的アンテナ」になっていただくべく、「数理・データサイエンスに関する教育・研究支援プログラム」を毎年実施しています。具体的には、全学を対象に「データ分析関連研究」に関する公募を行い、センターにて書類選考を行います。例年20~25件ほどを採択します。採択者にはセンター運営経費から研究支援

費を配分します。採択者は同プログラムの成果発表会にて、自分の研究について発表する義務があります。この成果発表会は、上述の合宿に増して多様性に富んだものになります。文理を問わない20を超える分野からの発表がある場合も、やはり非常に珍しいものと考えます。同発表会は学内だけでも参加可能です。2021年度は23件の発表があり100名を超える参加者がありました。なお、採択者には、発表会で用いたスライドを、差支えない範囲でセンターに提供してもらっています。これまでに100を超えるスライドが集まりました。そして、それらの一部は、低年次DSリテラシー教育において「各分野におけるデータ分析の実例」として利用されています。すなわち、この支援プログラムは、教材収集の場でもあるわけです。

5. 数理・DS教育強化拠点コンソーシアム活動

本学センターは、以上の学内活動に加えて、全国の「数理・DS教育強化拠点コンソーシアム」の拠点校としての事業も様々に行ってきました。例えば、同コンソーシアムの教材WGの副主査として、スライド教材の提供や、教科書執筆、放送大学講師なども行っています。

同コンソーシアムのこれまでの活動は、我が国の様々な大学が、組織を越えて力とノウハウを結集してカリキュラムと教材を創り出し、それらをオープン化することで、「学びたい人が、何を学べばよいか分かり、その上で学びたいことを学びたいレベルで学べる」状況を創り出しました。これは大学教育の歴史においても極めて画期的なものと思われれます。その活動に本学が携われたことに感謝したいと思います。

本学はコンソーシアムにおける九州ブロック拠点校であり、九州・沖縄におけるDS教育を推進する責務があります。九州ブロック活動においてもオープンを重視し、過去3年間の全てのイベントでは、全参加校が発言・発表する場を設け、フラットでフランクな関係を築くことを心がけました。例えば2021年度9月に実施した九州・沖縄ブロック会議(オンライン)では、参加した16大学すべてに15分以上のプレゼンをしてもらい、さらに10分ごとにメンバーを入れ替える少人数ブレイクアウトルームを利用して、少しでも多くの参加者間で密な情報交換ができるようにしまし

た。

また2020年11月および2021年12月のブロック会議は、教材作成の実例や、(後述する)MDASH認定校によるDSカリキュラム構築の方法といった、ノウハウ交換の場にしました。特に前者については「データサイエンス教材バトル」と題し、模擬講義に対してフロアからダメ出しをしようというイベントにしました。

2021年12月に、本学が九州ブロック拠点校を2022年度以降も継続して務めることが決まりました。他ブロックの拠点校と協力しながら、九州ブロック内の大学・高専と、国立公立私立・文系理系・そして規模を問わず引き続きフラットな関係を保ち、効率的かつ効果的に「DSは面白い」と思ってくれる学生を増やせるよう、尽力する予定です。

6. DSインターンシップ

本学では「DSインターンシップ制度」を2020年度から開始しました。これは、図5に示すように、DS・AI応用分野の研究者(教員・博士・修士、文理を問わず全学から募集)が、情報系の研究室と中長期コラボレーションすることで、より実践的にDS・AIの専門家を増やすことを目指した事業です。先述の合宿の場でも出てきた意見が発端となって開始しました。

ホストとなる情報系の研究室を募り、各研究室で提供できる知識・技術を明確にした後に、そこへのインターンを希望する研究者を募り、最後にセンターにてマッチングを行います。2020年度は13件、2021年度は17件のインターンシップを行いました。

ホストとなってくれる情報系研究室があつて初めて成り立つ事業ですが、ありがたいことに当初予定よりも多くの研究室が参画してくれています。期間終了後も、共著論文や共同研究費申請など、より踏み込んだコラボレーションに至ったという報も聞いています。インターンに参加してく

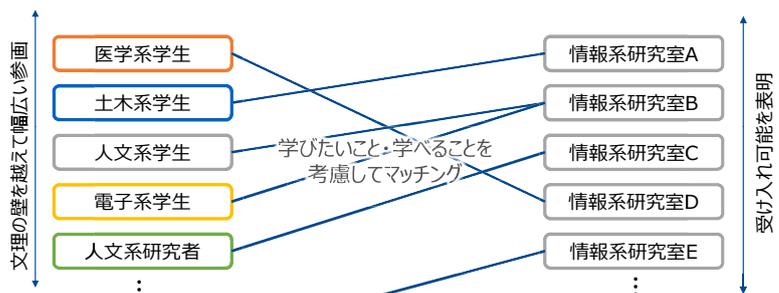


図5 DSインターンシップ

れた様々な分野の研究者が、ここで培ったノウハウを自部局に持ち帰り、その周りに伝えてくれることで、輪がさらに広がることを期待しています。

7. 数理・データサイエンス・AI教育プログラム認定制度（リテラシーレベル）

2021年度、本学の「低年次データサイエンス教育」プログラムは、標記認定制度（リテラシーレベル）においてMDASH+（プラス）に認定されました。同プログラムは、「情報科学」と「サイバーセキュリティ基礎論」の2科目から構成されています。前述の「情報科学」に加え、後者は1年次全学必修科目として数年前に導入されました。前者でDSの考え方や手法を教え、後者でデータリテラシーについて教える、という組み合わせで、リテラシーレベルに必要とされる内容をカバーしています。2020年度実績では、全学部1年生2,700名の約60%に相当する1,600名が同プログラムを受講しました。

「プラス」に認定いただいたのは、既述した以下の項目を評価いただいたためと思われます。すなわち、①授業内容については、「情報科学」において講義内容を受講部局ごとに変えている点、および「情報科学」において講義内容に当該部局でのデータ解析実例を導入している点、②学生への学生支援については、Moodleに独自機能として「わかった」「わからない」ボタンを付けている点、より発展的で網羅的かつ全学だれもが受講可能な「DS概論、実践、発展」科目やDSインターンシップが準備されている点、そして③その他として、九州ブロック拠点校としての貢献、スライド教材のオープン戦略、です。

8. おわりに

本稿では、本学のDS関連事業群のうち、特に「非」エキスパート向けの事業について説明しました。教材の公開や、誰でも受講できる講義、部局の壁を越えたセンター運営やインターンシップなど、様々な事業がオープンな考えに基づいて実施されていることを強調しました。

2015年度にDS実践特別講座のアイデアを思いついた頃には、ここまで急激に広がるとは全く思っていませんでした。その後には、全国コンソーシアムの発足や、AI戦略2019による（DSに加えての）AI教育の全国的推進、さらにはMDASH認定制度と、予期せぬ追い風（？）も色々ありました。今年度からはコンソーシアム事業の継続も

決まりました。DSやAIの基本的考え方が全分野に広がり、誰にとっても「当たり前」の知識になるまでは、まだやらなくてはならないことも多そうです。

最近、DX（Digital transformation）という用語を様々な分野で聞くようになりました。DSとDXには同じDがついているので混同されがちですが、DataとDigitalで全く違うものです。とはいえ、Digital化した結果は往々にしてDataになります。ですので、両者は無関係とは言ってられないようです。

本稿を結ぶにあたり、二点ほど付記します。第一点は、「本当は面白いデータ分析を、みんなにも面白いと思ってほしい」ということです。個人的な話で恐縮ですが、筆者自身は、画像などの実データを対象とした分野で研究活動しています。画像はデータとして身近であるだけでなく、非常に面白い性質を多数持っています。またその解析目的も多様なので、結果的に色々なデータ分析・AI系技術をフル活用しなくてはならず、その点でも飽きさせない面白さがあります。この生のデータを分析する面白さを、色々な方々、特に苦手意識を持った皆さんにも知ってほしいと願っています。

第二は、「データ分析は誰もがやっている」ということです。「線形代数、確率統計、微分積分などは苦手な全くなからぬ」「自分の人生には関係ない」と思っている人でも、無意識にデータ分析をやっています。本稿の文字を読む（認識・判別できる）のは、子供の頃に国語で叩き込まれたデータが脳内に残っていて、それを適切に分析して利活用しているからです。幼稚園児は「赤いアメはイチゴ味かも」と思うでしょう。これは彼らの経験で「アメの色が赤くなると、イチゴ味である確率が高くなる（色と味に相関がある）」ことを分析しているからです。幼稚園児は確率も画像が超高次元ベクトル表現されることも知りませんが、見事に分析をやっています。実際、推論や予測をやる時には、必ず経験すなわちデータを分析しています。類型化や類似性評価も同様です。このように、誰もが日々朝から晩までやっているデータ分析について、「無用な苦手意識」は是非忘れて、楽しく学んでほしいと思っています。

関連URL

[1] <http://mdsc.kyushu-u.ac.jp>