特集　数理・データサイエンス・AI教育

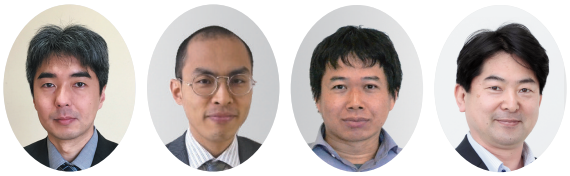
データサイエンス教育の先進的な取組み

滋賀大学
データサイエンス学部教授　佐藤　智和

滋賀大学
データサイエンス学部准教授　田中　琢真

滋賀大学
データサイエンス学部准教授　姫野　哲人

滋賀大学
データサイエンス教育研究センター教授　佐藤　健一



（左から佐藤智和、田中、 姫野、佐藤健一）

1．はじめに

現在日本ではデータサイエンティストの不足が深刻化しており、その育成が急務となっています。そのような状況に対応すべく、本学ではデータサイエンティスト育成を目的とする日本初の学部として2017年４月にデータサイエンス学部を設置しました。

図１は学部のカリキュラムマップです。本学部のカリキュラムは大きく分けて、データ駆動型PBL演習、価値創造基礎・応用科目、データエンジニアリング系科目、データアナリシス系科目、調査系科目に分類されています。これらの科目群は、データサイエンティストに必要とされる、情報系、統計系スキル、に加えて、多様な分野のドメイン知識、問題解決の経験およびコミュニケーション／プレゼンテーション能力を育成するためものです。学生は入学後すぐに広い分野の基本的な知識を身に付けることが要求され、１～２年生の講義には必修科目が多く配置されています。一方で、その後は学生の興味に応じて各分野のスキルを磨くことが可能となっています。また文理融合型カリキュラムとなっており、本学経済学部が開講している経済、経営等の文系科目も受講することが可能です。

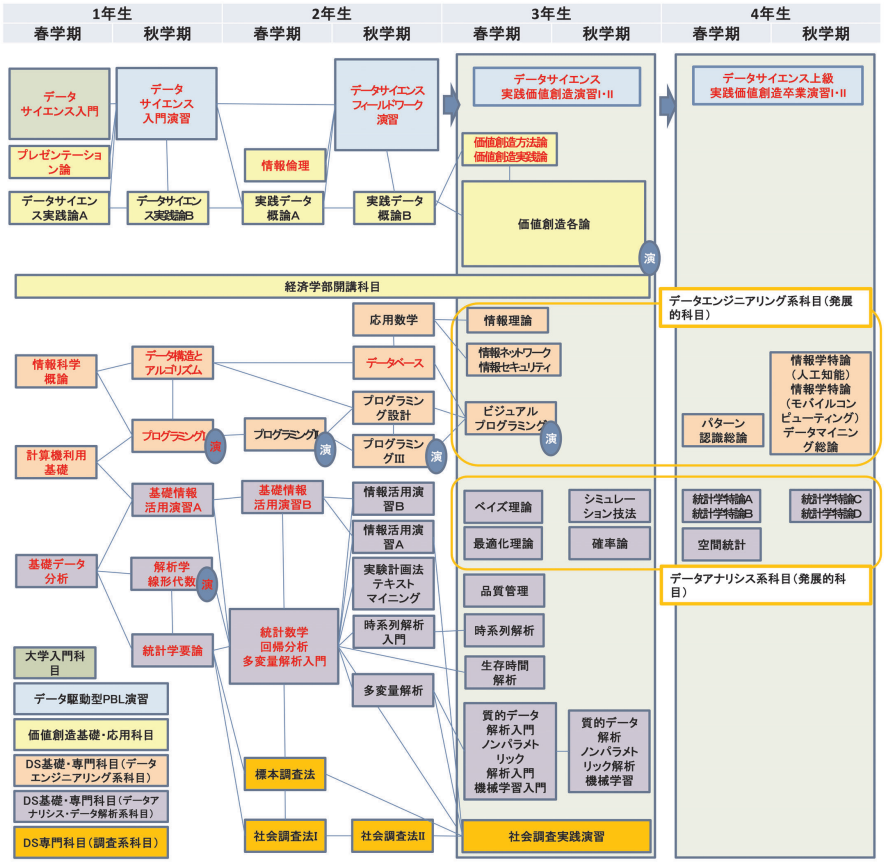


図１　学部のカリキュラムマップ（赤字は必修科目）

これらの科目群の中でも、データ駆動型PBL演習は全て必修科目となっており、ここでは実際のデータを用いた演習を行っています。以下では、本学データサイエンス学部におけるデータサイエンス教育のなかでも、特色のある先進的な取組みとして、PBL演習科目群の一部および企業のデータサイエンティストを招いて最先端の事例に触れるデータサイエンス実践論について、具体的な内容を紹介します。

2．データサイエンス入門演習

データサイエンス入門演習は、１年生春学期の入門的講義「データサイエンス入門」に対応する演習科目として、１年生秋学期に必修科目として設けられています。この科目はデータサイエンス学部生が実データの分析に取組む最初の科目です。問題解決型学修（PBL演習）によってデータ分析の「問題提起→データ収集→分析→発表」の全段階を体験することがこの科目の主目的です。また、グループワークやプレゼンテーションによってコミュニケーション能力を涵養することも目的です。

約100人のデータサイエンス学部生に加えて経済学部生も若干名を受け入れ、合同で演習を行います。学生を５クラス（20人あまり）に分け、さらに１クラスを５グループ（４～５人）に分けてグループワークを行います。

授業の最初の２回は合同授業でPPDACサイクル（Problem、Plan、Data、Analysis、Conclusion）に則った分析の進め方を学び、これ以降の回で実践します。２回から３回、グループワークの成果を発表する回があり、最終回は他のクラスの学生の前で発表する機会になっています。合同授業とクラス間の交流によって、５人の教員が演習を分担しながら、すべての学生に同一水準の技能修得の機会を提供しています。

なお、データが使われている現場について知るため、2019年度までは事業所見学も行っていましたが、2020年度は新型コロナウイルスの感染拡大を受けて中止となりました。しかし2020年度も、クラスによっては企業・官公庁の方をお招きしてデータ活用の実際について講義を行っていただいています。

演習ではクラスごとに異なるデータを使った分析課題に取組みます。2019年度は各クラスで以下のような演習を行いました。すべてに共通しているのは、統計学の教科書でよく使われるような整形されたデモデータではなく、様々な不規則性や欠測や異常値を含む実データを用いていることです。様々な前処理を必要とする実データに早い段階で触れることは、データサイエンティストとして必須の技能を身に付ける上で有効だからです。

テレビ字幕データ

ガラポン株式会社から提供を受けたテレビ番組の字幕テキストデータで演習を行いました。このテキストは社会の動きを反映していると考えられ、分析によって大きな価値が生み出される可能性があります。字幕テキストデータを時系列的にあるいは地域別に分析することで、社会のトレンドや人々の関心事を捕捉することを目指しました。

高齢者歩行データ

滋賀県米原市河内会館で計測した高齢者の歩行データを使用して演習を行いました。60歳代から80歳代の高齢者の歩行能力を複数の関節（脊椎・手・膝・踝・足）の３次元位置データで分析し、高齢者の歩行能力の問題点を発見するものです。与えられたデータに測定誤差がある前提で、データを補正し、被験者の歩行特徴をデータと映像を利用して分析しました。

大津市役所オープンデータ

大津市役所の公開しているオープンデータを用いて大津市の魅力を明らかにするというテーマで演習を行いました。まず、大津市役所を見学し、データに基づく施作づくり（EBPM）を学び学習のモチベーションを高めました。演習では大津市のデータを地域経済分析システム（RESAS）や政府統計（e-stat）のデータと組み合わせ、合わせて自分たちで探した大津市の魅力を同規模の他市のデータと比較し分析しました。さらに、大津市役所でのEBPM実務において学生の提案例を活用いただきました。

消費購買データ

株式会社マクロミルに提供していただいた消費購買データを用いて演習を行いました。このデータは全国のモニタを通して収集されたもので、日々の購買データが蓄積されています。まず、このデータの分析ツールQPR-TRACEの利用方法に習熟してから、新商品、新ブランドの提案を行いました。マクロミルの方にお越しいただき、マーケティングの専門家の観点から学生の提案に対するコメントを頂きました。

健康診断データ

本学保健管理センターから提供を受けた学生定期健康診断のデータ（本人提供承諾済み）で演習を行いました。定期健康診断の十年以上にわたるデータを利用し、厚労省や日本学生支援機構のデータと比較し本学学生の健康状態の傾向を調べました。各回にPythonを使った分析や回帰や検定などのテーマを設定し、分析技能の底上げを図りな

がら課題に取組みました。

3．データサイエンスフィールドワーク演習

データサイエンスフィールドワーク演習は２年生の秋学期に配置されているPBL演習です。ここではデータサイエンス入門演習に引き続き、実データを用いた問題解決型の学習を行います。データサイエンス入門演習と同様に、学生を５つのクラスに分け、さらに４～５名程度のグループでデータ解析を実施します。演習は各クラスを担当する５名の教員から、それぞれのクラスで扱うデータについての詳しい説明があり、学生はグループ内で話し合いをすることでどのテーマに取組みたいか希望を出します。その後、クラスに分かれた演習を実施し、最終回には全体での発表会を開催します。発表会ではグループで取組んだ内容をスライドにまとめてそれぞれ３分程度で発表します(写真１参照)。これにより、他のクラスでどのような内容に取組んだかについて共有するとともに、大人数の前でプレゼンテーションを行う機会を与えています。2019年度の演習において各クラスで取り扱ったデータと取組み内容は以下の通りです。

写真１　学生による全体発表会の様子

移動軌跡データ

本テーマでは、各自のスマートフォンで記録した移動履歴の分析を行いました。スマートフォンには所有者の移動軌跡を記録する機能が付いていて、集めた記録をライフログや健康のための運動増進のアシストなど、様々な用途に生かせます。演習ではデータを実際に収集し、経路検索結果の精度を高めることや、パラメータを班内で比較して経路による差や個人差があるかを調べることなど、各班で目標を設定し分析に取組みました。

防犯カメラ映像

防犯カメラ映像を解析することで、各駐車スペースに車両が駐車中かどうかを自動判定し、その結果を使って各駐車スペースの利用頻度を可視化する演習を行いました。本演習では少人数のグループに分かれ、それぞれのグループ内で、プログラミング、アイデア出し、プレゼンテーション等をチームとして相談しながら進めました。学生にとって画像処理は初体験であり、試行錯誤しながらデータの抽出から可視化までを一気に体験しました。

ソーシャルデータ

本テーマでは、SNS (Social Networking Services) においてユーザーから生成されたソーシャルデータを集めてテキストマイニングの演習を行いました。学生が決めたキーワードに基づき、IFTTTというツールを利用してストリームデータをダイナミックに収集することを体験しました。文章の分解、形態素解析、データクレンジングとデータ整形を含めた基本的なテキストマイニングの流れに従って、単語の出現頻度や関連性を分析した上で感情分析や社会調査等の課題に取組みました。最後はワードクラウドや共起ネットワーク等の手法を通じて分析結果を可視化することもできました。

Kaggleチャレンジ

Googleが運営しているデータサイエンティストのためのコミュニティサイトであるKaggle（カグル）のコンペティション機能を利用し、演習を行いました。Kaggleチュートリアルのタイタニックコンペで参加、解析、結果の提出法を学んだ後は、グループそれぞれが挑戦するコンペティションを選びました。コンペティション毎に設定された異なるデータ、目的に応じ、学生はこれまで学んできた分析手法を活用しながら予測モデルを構築しました。

日本版総合的社会調査 JGSS

JGSSは、世界的社会調査General Social Surveyの日本版として大阪商業大学が実施している研究プロジェクトであり、各年度4000レコード超を収録するそのデータセットを教育目的のために東京大学SSJデータアーカイブから提供を受けました。学生は、年収、幸福度などに対して、就業、健康といったデータを組み合わせて、それまでに学んだ多重回帰やロジスティック回帰といった手

法を適用して分析に取組みました。

４．データサイエンス実践価値創造演習Ⅰ・Ⅱ

データサイエンス実践価値創造演習Ⅰ・Ⅱは、３年生向けのPBL演習科目です。この演習はいわゆるゼミであり、学生はどの教員の下での演習を行いたいか、２年生の秋学期に希望を出し、その後ゼミへの配属が決定されます。本演習で扱うデータ／テーマは担当教員の専門に依存しており、学生は幅広いテーマの中から専門的に学びたいテーマを選択します。この演習は、４年生向けにはデータサイエンス上級実践価値創造演習Ⅰ・Ⅱとして開講されており、４年生も３年生で選択したゼミのテーマに継続して取組みます。2019年度に実施したゼミのテーマ名は以下の通りです。

- 医療、防災、行政の実データの特徴を探索する
- 音声・テキストデータの分析と機械学習
- モバイルコンピューティング・センサーネットワーク
- 情報技術の基礎
- 元企業データサイエンティストによる企業に求められるデータサイエンティストを育てる
- 機械学習、ディープラーニングによる画像処理
- 画像解析・画像合成・AR／VR
- 統計的因果推論による価値創造
- ビッグデータマイニング
- 医学統計、ツリー構造機械学習、数理統計学の研究
- 社会調査を通じて社会・文化を読み解く
- 疫学・臨床研究のデータベース構築と統計解析手法開発
- 機械学習の数理的な基礎を学ぶ
- 統計的推測の経済・金融データへの応用
- 統計学・確率論の基礎

また、３、４年生向けのゼミとは別に、学生の多様な興味に応えるために、学年にかかわらず希望すれば参加できる自主ゼミを学期ごとに開催しています。2019年度の自主ゼミの開催実績は、春学期11テーマ、秋学期８テーマでした。

５．データサイエンス実践論Ａ・Ｂ

本学部のカリキュラムでは、初年度から様々な分野でのデータサイエンスの活用事例に触れることで、学生自身が本学部で学ぶモチベーションを高め、各自の目標を鮮明にすることを想定しています。データサイエンス実践論もその１つです。以下では、2019年度の授業内容について紹介します。

データサイエンス実践論Aでは、IT系企業を中心に様々な企業の現場経験者を講師として招き、全12回の講義をしていただきました。本講義では90分の講義に加え、各講師の方との20分の意見交換等を行うことで、データサイエンティストの方々がどのような仕事を行っているかを知る初めての機会となっています。講義テーマと担当者は以下の通りです。

- お客様の音声をビジネスに生かす音声認識ー音声ビッグデータの活用の広がり（市川治滋賀大学教授、元IBM研究員）
- AIブームの再燃：IBM Watsonの誕生、第３世代のコンピューティングWatsonの応用、さらに広がるWatsonとAIの世界（成城大学教授、元IBM TV・UR部長）
- 企業におけるサイバーセキュリティとデータ分析（外資系IT企業）
- データが繋ぐGEMBA～お客様を基点とするビジネスエコシステム～（株式会社小松製作所）
- テキストアナリティクス（Ⅰ）(外資系IT企業)
- テキストアナリティクス（Ⅱ）(外資系IT企業)
- パナソニックのデータ分析ってどんな仕事？（パナソニック株式会社　アプライアンス社）
- プロジェクトマネジメント概論～女性キャリアの一例として～（外資系IT企業）
- ネットワークデータサイエンス（Sansan株式会社）
- ビッグデータを保有している大企業に対してAI（機械学習）導入アプローチ実践（コグニロボ株式会社）

データサイエンス実践論Bでは、データサイエンティスト協会の会員企業を中心に７つの企業から講師を招き、それぞれの企業の中でどのようにデータサイエンスが活用されているかを紹介いただきました。座学だけではなく様々な演習を取り入れてもらい（中には実データを扱わせていただくケースもあり）、より実践的な講義を実施していただきました。

- データサイエンスで実現ビジネス、顧客管理とデータサイエンス、人工知能とデータサイエンス、データサイエンティストに求められる力（株式会社 野村総合研究所）
- PropTech（不動産テック）におけるデータサイエンスの活用（株式会社GA technologies）
- ビヨンド・ザ・データサイエンティスト－社会を変えるデータサイエンティスト（PwCコンサルティング合同会社）
- 通信業界におけるデータサイエンス（株式会社シイエヌエス）
- デジタル広告におけるデータ解析（デジタル・アドバタイジング・コンソーシアム株式会社）
- データ分析&ダイバーシティー（株式会社SMBC信託銀行）
- マーケティング業界でのデータサイエンス実務とデータサイエンティストの生態（株式会社電通）

6．大学生向けのMOOC教材

本学部ではこれまで、データサイエンスに関するオンライン学習サービスMOOC（Massive Open Online Courses）向けの教材を作成しています。大学生向けの教材としては、「大学生のためのデータサイエンス（Ⅰ）」で、データサイエンス全般について概観し、現代社会におけるデータサイエンス、データ分析の基礎、コンピュータを用いたデータ分析、そして、その応用事例について解説しています。「大学生のためのデータサイエンス（Ⅱ）」では、技術的により進んだ内容として、機械学習の応用事例、分類問題および回帰問題を紹介し、さらに、近年、発展の著しいニューラルネットワークを取り上げました。これらの教材は一般的な配信とは別に、各大学のニーズ合わせたオンライン配信も行っており、高崎健康福祉大学、岡山大学、福井大学、金沢大学をはじめ、本学が主催する企業向けのセミナーの副教材としても利用されています。

また、大学生向けのMOOC教材の第３弾として、「大学生のためのデータサイエンス（Ⅲ）問題解決編」を制作しました（写真２参照）。こちらは2020年12月に開講されました。本講座では、データサイエンスの分析手法を使って実際の問題を解決することを目標に、必要な考え方、分析の進め方、問題設定や伝え方のスキルを習得します。今回の（Ⅲ）では、これまで学んだ様々なデータサイエンスの分析手法を使って実際の問題を解決することを目標に、必要な知識やスキルを具体的なデータとともに説明しています。この講座では初めに、PPDACサイクルを問題解決のための枠組みとして学びます。続いてデータ例として、１）自動車販売データ、２）地産地消データ、３）自由記述のテキストデータ、を扱い、どのような手順で分析が進められるかを示します。さらに、データサイエンスを推し進める上で重要となる問題設定のためのヒアリングや結果の伝え方にも言及します。より具体的な内容については以下のURLからご確認ください。

https://lms.gacco.org/courses/course-v1:gacco+pt067+2020_12/about

写真２　大学生のためのデータサイエンス（Ⅲ）

7．おわりに

本学データサイエンス学部は2020年度で第１期生が４年生となりました。これまで、複数の学生のデータ分析コンペティションへの入賞や、学生によるデータ分析に関する合同会社の立ち上げなど、学部教育の成果が出始めています。また一方で、現在のカリキュラムの課題も明らかとなってきました。例えば、文理融合型の現実として、学習すべき分野が広く学生にとって負担が大きいことや、数学やプログラミングに苦手意識を持つ学生が見られることがあげられます。本学部では、これらの問題に対応すべく、次年度からプログラミング講義の開講時期の変更や、必修科目の選択化などを含むカリキュラム改定を予定しています。

本学部では本稿で紹介した講義・演習以外にも、約50企業との連携により、実社会で扱われる生のデータを対象とした共同研究の推進や、企業向け講習会（道場と呼ばれています）の開催など、特色のある取組みを多数実施しています。詳しくは、学部Webページからも情報発信しておりますので、そちらもご参照ください。