

数理・データサイエンス・AI教育の紹介

滋賀大学データサイエンス学部の企業連携と数理・データサイエンス・AI教育プログラム(応用基礎レベル+)について

滋賀大学
データサイエンス学部副学部長

佐藤 健一

滋賀大学
データサイエンス学部教授

市川 治



(左から 佐藤、市川)

1. はじめに

本学では、わが国初となるデータサイエンス学部を2017年4月に開設し2019年度にデータサイエンス研究科修士課程を、2020年度には博士後期課程を開設しました。学生の定員は学部が100名、博士前期課程が40名、後期課程は3名です。本学部には情報学・統計学の教員をはじめ、医学、生物学、社会学、心理学等諸分野で活躍してきた教員、公的機関や企業での豊富な実務経験者など2023年7月現在28名の専任教員が在籍しています。

データサイエンス学部開設の1年前からデータサイエンス教育研究センターを開設し、データサイエンスに関する先端的な教育研究活動を行うとともに、企業や自治体との連携、多様な大学間連携に取り組んできました。2022年度からは全学体制のデータサイエンス・AIイノベーション研究推進センター(以下、センター)に発展改組され、併任しているデータサイエンス学部教員に加えて、21名の専任教員が所属し、学部講義なども担当しています。このように、総勢49名の専任教員が大学教育や企業連携に関わっており、西日本最大規模のデータサイエンティスト集団を形成しています。また、2023年度にはそれを支えるURA(ユニバーシティ・リサーチ・アドミニストレーター)を高度専門職員として2名雇用し、支援部門も大幅に強化されました。

企業連携の収益によるセンター教員の雇用拡充に加えて、本年7月には、大学院データサイエンス研究科博士前期課程の定員倍増(令和7年度から定員80名)とデータサイエンス学部定員増(令和7年度から定員150名)などを内容とする

提案が、独立行政法人大学改革支援・学位授与機構の大学・高専機能強化支援事業(高度情報専門人材の確保に向けた機能強化に係る支援(支援2))のハイレベル枠に採択され、今後も学生および教員の拡充が予定されています。

2. 企業連携について

本学データサイエンス学部の強みは企業との幅広い連携にあります。教員も実際のデータや課題に触れることで、データサイエンスの重要性・有用性を認識し、それが講義やゼミでの指導という形で還元されています。

現在継続中の企業連携は60件ほどあり、これまでに300を超える企業案件を担当しています。ここでは、共同研究の具体的な内容について、センターが発行する2022年度の年報「Data Science View」から5件を抜粋して紹介したいと思います。

- 大阪ガス株式会社：一般家庭の電力使用量を予測する手法について研究を行っています。電力の安定供給のために、電力会社には顧客の電力使用量を予測して需要と供給を均衡させることが求められます。しかしながら、特に一般家庭の電力使用量については、その変動の不規則性ゆえ、汎化性能の高い予測モデルを家庭ごとに作ることは困難です。そこで本研究では、新たなモデルベース時系列クラスタリング手法を開発し、使用パターンの近い顧客グループに分割しつつ、それぞれのグループに対して予測モデルを推定するアプローチによって高精度な予測を目指しています。2022年度の研究成果をまとめた論文は、機械学習・パターン認識に関するトップジャーナルに掲載されています。

ーナルの一つであるPattern Recognitionに掲載されました。

- 株式会社日立建機ティエラ：IoTを活用して収集した工場内のデータを分析し、生産設備の最適稼働を目的とした研究に取り組んでいます。工場では、建設機械が多種混流生産されており、生産状況・生産設備の見える化を推進し、生産にかかわる各種データの収集が行われています。本共同研究は2021年10月から開始しており、収集・蓄積した工場の稼働状況に関するデータを用いて部品投入の最適化を目指しています。2022年度は、塗装工程について構築したシミュレーションモデルを実際の稼働状況のデータと比較することで再現性の検証をしてきました。また、塗装工程後の組み立てに関する仕掛り管理についてもシミュレーションモデルの構築に取り組んでいます。大物塗装工程のシミュレーション結果と連結させることで、建設機械の組立計画、塗装設備の進捗、仕掛り数を管理しながら、最適な部品投入の計画をシミュレーション結果から検討できる技術の開発を進めています。
- トヨタ自動車株式会社との共同研究～その1：車載カメラ画像を利用した道路のオルソ画像生成手法の開発を行っています。道路のオルソ画像（真上から見た画像）にはナビゲーション用地図の作成・自動運転のための路面情報の取得など様々な用途があります。従来オルソ画像の作成にはコストのかかる飛行機による空撮などが必要でした。一方で路面の状態や標識は常に変化しており、その情報をリアルタイムで更新する方法が必要とされています。この共同研究では車載カメラで撮影された動画から道路の3次元復元を行い、そこから自由視点画像を生成するという方法でオルソ画像を生成する方法を開発しています。現在は特に複雑な路面形状に対応するため複数の車載カメラを使った画像生成を試みています。このような技術は将来自動車に搭載されている車載カメラを利用し道路の情報を常に更新し続けるシステムの開発に利用されることが期待されます。
- トヨタ自動車株式会社との共同研究～その

2：結晶構造から有用な物質特性を有する物質候補を予測する研究に取り組んでいます。結晶構造を取得する有用な手段である結晶構造データベースには、毎年数万件以上のペースで結晶構造が登録されており、誰でも自由にアクセスすることができます。その一方で、材料特性が登録されているデータベースの登録数は比較的少数に留まっているため、多くの結晶構造に対しては材料特性を得ることができません。材料特性を知るためには、実験を行う、または長時間かかるコンピュータシミュレーションを実行する必要があります。きわめて高い時間的、金銭的コストが必要になります。本研究では、機械学習を活用するマテリアルズ・インフォマティクス手法を用いて、有用な材料特性を示すことが期待される候補物質を効率的に探索する手法の構築を目標としています。これにより、材料特性が未知であるために使われていない物質の中から、優れた性質を示す素材を発見し、新製品開発などへと貢献することが期待されます。

- 日東電工株式会社との共同研究：粘着テープやフィルムを始めとした様々な製品を製造する日東電工、およびグループ企業で分析を専門とする日東分析センターと製品の設計・開発における課題解決のためのデータサイエンスに関する共同研究を行っています。材料の機能理解のためには、静的な評価による構造理解だけではなく、動的なその場測定による機能発現のメカニズム解明が必要となります。そのための世界最先端レベルの画像や動画解析の研究を共同で進めています。本研究では、画像のノイズ除去や高画質化、セグメンテーションなどの画像処理により材料の特徴量を抽出し、特性を評価する手法を開発しています。低コントラストでノイズの多い膨大なX線画像データに対して、スパースモデリング等を用いて高速に鮮明な4D-CT動画を取得し、形状解析を行う手法の開発に取り組んでいます。

また、企業連携だけでなく、地方自治体や地域経済の発展にも貢献しています。本学データサイエンス学部が位置する彦根市の彦根商工会議所と

は2020年11月に包括的連携協定を締結し、長期有給型インターンシップにより本学の学生が企業での実践力を習得するとともに、事業所の経営イノベーションを創出する人材育成や、データサイエンス教育プログラムの開発など、地域経済の活力創造に向けて連携を強化してきました。また彦根市は内閣府の地方創生テレワーク交付金制度に応募し、高い目標設定と先駆的な要素が必要な高水準タイプで採択されました。この制度はサテライトオフィス、シェアオフィス、コワーキングスペース等の整備・運営等により、地方への新たな人の流れを創出する地方公共団体の取組みを支援することを目的としています。この交付金を活用するために、彦根商工会議所、本学、地元金融機関および賛同企業等がコンソーシアムを組み、2021年6月に近江テック・アカデミー株式会社を設立しました。近江テック・アカデミーは商工会議所近くの彦根市中央町にサテライトオフィスなどの機能を持つ施設「INSPILAKE（インスピレイク）」を開業・運営しています。コンソーシアム機能を活かして、入居者間のビジネス交流、スタートアップ支援、ジュニアIT教育、社会人のリカレント教育、事業所へのDXサポート等の機能を備えるものです。近江テック・アカデミーではすでに本学のデータサイエンス学部生を採用して、ジュニアデータサイエンティスト育成スクールなど、様々な事業を開始しています。また、執行役員としてデータサイエンス研究科学生が就任しました。

このように、企業連携などを推進することで、連携先からバラエティに富んだゲスト講師を迎えることができ、また、卒業研究の一部としてデータの提供もされています。さらに、データサイエンス研究科においては、連携先から派遣社会人を受け入れており、連携先が持つデータを利用した研究指導が行われています。

3. データサイエンス学部のカリキュラムについて

本学部では、データサイエンスの専門知識やスキルといった理系的基礎に加えて、データ利活用の現場で相互補完的な専門性を有する仲間とコミュニケーションを図りながら、データから価値のある情報を取り出し、それを意思決定に活かす能

力を備えた文理融合型の人材の育成を目指しています。また、カリキュラムも文理融合型となっており、データを管理、加工、処理、分析するためのスキルは情報や統計のスキルなので理系的ですが、分析結果を価値創造に生かすためには、データの背景を十分に知る必要があり、多くの場合において文系的要素が必要となります。

したがって、本学部では、情報、統計関連科目ばかりではなく、経済、経営等の文系の授業も受けることができます。また、ビジネス分野の第一線で活躍をしている方々の話を多く聞くことができる授業もあり、幅広いスキルを身に付けることができます。なお、統計学の基礎である数学については、高校で数学Ⅱ・Bまでしか学習しなかった学生でも対応できるよう、数学Ⅲの内容から学習し、数学についての質問対応として、データサイエンス研究科の学生などによる数学サポートが毎週開催されており、予約なしで指導を受けることができます。

カリキュラムは学生の興味に応じた学修ができるように、大きく分けて、①データサイエンス科目、②価値創造科目、が用意されています。①はさらに、データエンジニアリング系（情報関連）科目、データアナリシス系（統計系）科目に分かれます。②では経済、教育、心理、医療など多分野における価値創造の実例紹介、価値創造の実践等を学びます。これらの授業から自分の興味に応じた授業を受講することで、情報のエキスパート、統計のエキスパート、価値創造のエキスパートになることも可能です。

さらに細分化すれば、データアナリシス系科目では統計学と、その基礎である数学について学びます。データエンジニアリング系科目では情報学と、プログラミングによる演習を行います。データ駆動型PBL演習では、データアナリシス系科目とデータエンジニアリング系科目で学んだ知識や技術を活かして、実際のデータを分析する経験を積みます。価値創造科目では、データサイエンスの応用事例などを学びます。さらに、社会調査士の資格を取得できる専門科目も揃っています。これらの科目群は、データサイエンティストに必要なとされる、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。

このように、本学部の教育課程では、統計や情報の基礎力を身に付けるだけでなく、実際にデータの解析結果を意思決定に活かして、価値創造できる力を高めることを目的としています。このような目的を達成するため、1、2年次には統計学と情報工学の基礎的内容を身に付け、様々な応用分野におけるデータ分析の実例を学びます。それらの基礎をもとに、3、4年次では各種領域科学におけるデータ分析手法を学び、実際のデータを使った演習を通して価値創造の実践経験を積み重ねます。それに加え、各自の興味に応じ、様々な統計手法の数理的内容をより深く学んだり、より高度な情報処理技術を身に付けたり、より多くの分野における問題解決スキルを磨いたりできるカリキュラムを用意しています。

4. 滋賀大数理・データサイエンス・AI教育プログラム

本学のプログラムは認定制度が始まった2021年度に、リテラシーレベルの中でも先導的で独自の工夫・特色を有するとして数理・データサイエンス・AI教育プログラムリテラシーレベル+（プラス）として選定されました。さらに2022年度には、応用基礎レベルに認定され、特に、データサイエンス学部のプログラムについては、認定された教育プログラムのうち、先導的で独自の工夫・特色を有するとして応用基礎レベル+に選定されました。

データサイエンス学部が取得した応用基礎レベル+は13科目（選択項目科目を含めると25科目）から構成されます。プログラムを構成する下記の基礎科目から14単位、専門科目から12単位、合計26単位を修得します。なお、1科目は2単位として換算され、専門科目群に含まれる「AI・機械学習入門」を除き、すべて必修科目となっており、データサイエンス学部の学生にとって履修しやすいプログラムとなっています。

- 基礎科目：線形代数への招待、解析学への招待、データ構造とアルゴリズム、プログラミング1、プログラミング2、統計数学、データベース
- 専門科目：AI概論、AI・情報倫理、AI・機械

学習入門、データサイエンス入門演習、データサイエンス応用演習、ビジネス価値創造論

- 選択科目：解析学、線形代数、マルチメディア処理入門、テキストマイニング、AI・機械学習、多変量解析入門、標本調査法、実験計画法、最適化理論、シミュレーション技法、データ研磨、情報セキュリティ

ここでは、専門科目群から価値創造科目に分類される「ビジネス価値創造論」のシラバスから原文を抜粋して紹介します。

- 授業の目的：データ分析の手法論をしっかり学んでも、それだけでは企業で活躍できない。企業で待ち受けているのは「このデータでこんな予測をしてね」という具体的なお題ではなく、「君のデータ分析力で売上げを伸ばして欲しい」という漠然としたお題である。方法論だけでは立ち往生するだけである。本講義では、そのようなお題にも成果を出せる力を授ける。授業の概要) 皆さんが学ぶ分析手法を活かすには、「解決したい課題」と「分析に用いるデータ」を用意しなければならない。本講義では、ビジネスにおいて適切な課題を設定する力と、分析に役立つデータを自ら考える力について、演習を通して伝える。数学やプログラムと異なり体系化された教科書はなく、世界で唯一のオリジナルな講義である。この分野で成功してかつ自らの経験を形式知として教える能力を持つ希少な人を講師としてアレンジした。
- 授業の到達目標：データ分析で解決したい課題を設定する型を習得する。データ分析に役立てるデータを考える型を習得する。製造業におけるデータ分析の活用方法を一例であるが習得する。

本プログラムによって、学生は自らの生活にも数理・データサイエンス・AIが深く関与し、産業界などでもデータにもとづいた合理的判断を求められている現状を把握するものと考えます。そして、データサイエンスの応用事例を理解することで、実データ、実課題（学術研究データ等も含む）を用いた演習についても、社会での実例として取り組むことができるようになり、その結果、現実

社会の課題に対する基本的な活用法を習得することが期待されます。

最後に、本プログラムデータサイエンス学部ならびにセンターが推進する企業連携で培った実践的なデータサイエンスの講義科目が含まれており、そのような点から応用基礎レベル+として評価されたのではないかと考えております。

5. 生成AIへの対応

本学部ではこれまでもAIの教育に力を入れてきました。科目名としては「AI・機械学習入門」「AI・機械学習」「人工知能論・演習」「画像処理・演習」「音声データと対話システム論・演習」があげられます。これらを履修すれば、CNNやLSTMを用いた分類、敵対的生成ネットワーク(GAN)による画像生成、変分オートエンコーダ(VAE)による潜在表現の獲得、Transformerによる系列変換などのAI技術を習得することができるカリキュラムになっていました。

しかし、最近になってChatGPTやStable Diffusionに代表される生成AIが急激に注目を浴びるようになり、カリキュラムとしても急遽対応を迫られることになりました。

生成AIとは端的に言ってしまえば、データを入力して、テキスト・画像・音声データを出力するものです。これまでの講義でもGANや系列変換器といった生成能力を持った機械学習モデルは教えてきましたが、最近の生成AIはそれよりも格段にレベルが上がっています。具体的には、自己教師あり学習や対照学習を用いることによって、ネット上で集めた超巨大なデータを（人手によって正解ラベルを付与する必要なしに）そのまま、巨大なモデルの学習に利用し、人間が持つような常識や感性に相当する潜在表現をモデルの内部に獲得させることができるようになりました。さらに、その際にテキスト・画像・音声データを融合させることができるようになりました。

この生成AIは、CoPilotのような製品群としてITベンダーから提供され、まずはオフィスワーカーの生産性向上に寄与していくと考えられます。今後は、音楽、絵画、映画といったアートやエンターテインメントの領域に広がっていくと見られており、データサイエンスを志向する学生のタイプも今後さらに多様化が進むと考えられます。

表1 生成AIに関する講義の各回構成

第1回	ガイダンスと生成AIの活用事例紹介
第2回	生成AIを支える理論(1) PyTorchを用いた深層学習
第3回	生成AIを支える理論(2) 敵対的生成ネットワーク
第4回	生成AIを支える理論(3) 自己符号化器 (AutoEncoder 他), 潜在表現, self-attention (BERT, ViT)
第5回	生成AIを支える理論(4) 系列変換器 (Encoder-Decoder), cross-attention, モダリティ変換, CLIP
第6回	生成AIを支える理論(5) 拡散モデル
第7回	生成AIを支える理論(6) 大規模言語モデル
第8回	生成AIの活用実習(1) 生成AIを使用したレポート作成
第9回	生成AIの活用実習(2) 生成AIを使用したレポート作成
第10回	生成AIの活用実習(3) 生成AIの創作応用
第11回	生成AIの活用実習(4) 生成AIのプログラミング応用
第12回	成果物発表会
第13回	成果物発表会
第14回	まとめの会 (またはゲスト講師による講義)
第15回	生成AIにまつわる諸問題: 生成AIに関する諸外国における議論の動向 (政府、企業)、生成AIに関する国内の議論の動向 (主に、政府)、生成AIと著作権に関する議論

以上の背景を踏まえて、生成AIを教える講義を「データサイエンス特論B (生成AIの理論と活用)」として、令和5年度の秋学期から開講します。

この講義では、前半を理論編とし生成AIの背後にある理論的な枠組みを学ぶとともに、後半を活用編とし生成AIを使用して実際の課題を解く演習を行います。最後に、著作権やプライバシーといった生成AIにまつわる諸問題について学びます。

生成AIに関する講義の各回の構成を表1に示します。今年度は音声生成やアート(感性)の領域には踏み込んでいません。まずは上記の安定的な内容で開講し、それらについては今後検討していくことになるでしょう。