

数理・データサイエンス・AI教育の紹介

東北大学のAIMD応用基礎教育

東北大学
データ駆動科学AI教育研究センター
センター長

東北大学大学院
情報科学研究科教授

東北大学大学院
情報科学研究科准教授

早川 美徳

篠原 歩

鈴木 顕



(左から 早川、篠原、鈴木)

1. はじめに

何年も前の話になりますが、部局からの代表が多く参加する教務関係の会議の場で、学生の英語の実力を調査するために毎年実施している学カテストの結果を巡って、資料の点数やグラフを見ながら意見を交わしていました。こうしたデータの解釈はともすると主観的・独善的になりがちなので、そんな様子を見かねたのか、数学がご専門の先生が『私は統計関係の授業を長らく担当しているが、普段から学生にデータの平均しか見ないのは小学生並み、と指導しています…』と発言されたことを、今でも思い出すことがあります。

それから時を経て、データやAIに関係するリテラシーの重要性を誰もが認識するようになり、有償・無償を問わず、教材や教育プログラムは量・質ともに格段に充実してきました。

筆者らが所属する本学においても、ここ数年、AIMD（本学では、AI、Math. & Data scienceの頭文字をとって、AIMDと呼称しています）¹⁾教育体制の整備に取り組んできましたが、本稿では、その中でも全学部に向けた応用基礎レベルでの事例について紹介いたします。

2. 東北大学挑創カレッジ

本学では、現在の執行部体制がスタートして間もない2018年に「東北大学ビジョン2030」を公表し、「社会の転換期を生きる学生の創造力を伸ばす教育の展開」を重点戦略のひとつに掲げました。具体的な施策として、未来社会に立ち向かうための基盤となる学士課程教育を実現すべく、数理・データサイエンス・AI教育をはじめ、グロー

バルリーダー育成、アントレプレナーシップ養成等の現代的なリベラルアーツを、学部・学科の枠組みを超えて学修できるカリキュラムの実現を目標に設定しました。この「東北大学ビジョン2030」は、その後さらに更新され、2020年に「コネクテッドユニバーシティ戦略」を策定し、デジタル技術とデータを活用した大学の変革に一層力点が置かれるようになっていきます。

こうした目標を具現化すべく、2019年度から、新入生全員に基本的なAIMD教育を実施することに加え、意欲的な学生に向けた学修機会を提供するため、「東北大学挑創（ちょうそう）カレッジ」を創設しました。これは、欧米の大学で用いられる意味でのカレッジではなく、リベラルアーツ教育プログラムの枠組みの中で科目の新設や内容の見直しを行いつつ、定められた科目群から決められた単位を取得した者を認定することにより、課題に挑戦するマインドの醸成と、現代的な素養を涵養する機会を提供しようとするものです。学部や学年に依らず、全ての学生に対して挑創カレッジの科目群は開放されています。

当初、東北大学挑創カレッジは、グローバルリーダー育成プログラム（TGLプログラム）、企業家リーダー育成プログラム（TELプログラム）、そしてAIMDをテーマとするコンピューショナル・データサイエンス・プログラム（CDSプログラム）の3つのプログラムからスタートしました。2022年度からは、東北大学SDGsプログラム（SDGsプログラム）と東北大学プルリリングル・スタディーズ・プログラム（TU PlusSプログラム）も新たに加わり展開しています。

3. CDSプログラムの実施体制

挑創カレッジCDSプログラムは、発足当初から、意欲的な学生の好奇心と将来に向けての目標達成に応えられるようにと、従来は学部専門教育で実施されていた高度な内容を含む科目や、今後一層重要性を増すであろうAIに関する科目を新設し、これらに既存の基礎科目を組み合わせることで構成されました。結果として、数理・データサイエンス・AI（応用基礎レベル）モデルカリキュラムが正式に公表される以前から、同カリキュラムをほぼ包含する内容を提供してきました。

CDSプログラムは、情報の基礎、統計の基礎、数学の基礎の3区分から科目群が構成されており、「実践的機械学習Ⅰ」および「機械学習アルゴリズム概論」（それぞれ2単位）を必修としています（表1）。これら必修科目については、次節以降で改めて、その様子を紹介いたします。さらに、必修2科目に加え、情報の基礎から4単位、統計の基礎、数学の基礎からそれぞれ2単位以上を取得することが修了の要件となっています。本プログラム修了者には修了認定証の他、2022年度からは、オープンバッジを発行しています^[11]。また、これらの科目の多くは、「学都仙台単位互換ネットワーク」に提供して学外に開放されています。

挑創カレッジの運営は、学務全般を所掌する学務審議会のもとに置かれた挑創カレッジ運営委員会が行い、CDS科目の運営や改善活動は、同じく学務審議会の各科目委員会（情報教育委員会、数学委員会）が担当しています。

各科目には、企画担当運営部局を定め、授業担当教員の推薦等について責任を担うかたちで、全学から授業担当教員が出動する体制をとっています。専門的かつ最新の内容を扱うことが求められ

るCDSの一部の科目では、担当者を見出すのは決して容易でない一方、受講者数は毎年順調に増加しているため、担当者を含めた時間割の調整が科目委員会の大きな任務になっています。

4. CDSプログラムの特徴と取組み

CDSプログラムは、全学教育の一環として実施されていることから、全学部の学生に開放されており、実際に文理を問わず多数が履修しています。2022年度については、CDS必修2科目の履修延べ人数は、人文社会系（文・教・法・経済学部）で169名、理工系（理・工学部）が855名、生命科学系（医・歯・薬・農学部）が63名でした。また、大学院生も12名が受講しており、CDSは学び直しの機会としても機能しているようです。データ科学やコンピュータ科学に関する科目の殆どは5時限目に設定されているため、大学院生が比較的受講しやすい時間帯であることも手伝っているのかもしれませんが。今後は分野に依らず、研究を進める上で機械学習等の知識やスキルが求められるケースが益々増えるでしょうから、高年次の学部学生や大学院生への教育ニーズにも応えていく必要があると考えています。

CDSプログラムでは、データ科学やAIの実社会での最新動向やユースケース、社会の変容等、多様な観点に接する機会を与えるべく、企業の実務家の協力を得ながら、教材開発や学修の場を設ける取組みを続けています。

リテラシーレベルの科目「情報とデータの基礎」で用いるオンデマンド型ビデオ教材を、地元のAI人材開発企業と共同で開発したことをはじめ、教材開発にあたっては直接・間接的に実務家の意見を取り入れてきました。また、大手eコマースから提供いただいた実データを、データ処理のサンプル用データとして、学内で公開しています。

「情報教育特別講義（AIをめぐる人間と社会の過去・現在・未来）」では、学内の教員に加え、大手IT企業、大手プラットフォームフォーマー、AIベンチャー等から講師を招き、AIと人間社会について現在進行形で考えるオムニバス形式の授業を行っています。

加えて、本学が包括連携協定を結んで

表1 全学教育のAIMD科目群（†CDS必修科目）

科目群	科目名
情報の基礎	情報とデータの基礎、実践的機械学習Ⅰ†、機械学習アルゴリズム概論†、実践的機械学習Ⅱ、データ科学・AI概論、情報教育特別講義（AIをめぐる人間と社会の過去・現在・未来）、情報教育特別講義（AI・データ科学研究の現場）
統計の基礎	統計学入門、数理統計学概論、数理統計学、情報教育特別講義（統計数理モデリング）
数学の基礎	解析学入門、線形代数学入門、解析学概論、線形代数学概論、解析学A、解析学B、線形代数学A、線形代数学B

いる企業の協力を得ての「リアルビジネスにおけるデータサイエンス/AIの活用」と題する連続セミナーや、特に文系の学部学生を想定した内容の「AI入門講座」シリーズ等、課外の学習イベントも企画・開催してきました。これらの講師は企業の第一線で活躍している方々で、学生の反応も良好です。

また、データ科学やAIのオンライン教育プラットフォームの開発・提供で実績のある地元企業の協力を得て、一般社団法人日本ディープラーニング協会が実施している「G検定」および「E資格」の受験に向けた学修支援を2020年度から実施してきました。毎年1回、CDSプログラムの履修者から希望を募り、資格取得に向けて企業が有償で提供しているオンラインコース（本学の教員が監修）の費用を大学側が負担する試みです（ただし、受験料は本人負担）。毎年、30名程度の学生が申し込み、年度中に受験し合格する者はその一部に留まっていますが、事後アンケートから1年以上経過した後に、「合格できたのは本プログラムを通してご提供いただいた教材のおかげです」といったメッセージとともに、合格の知らせが届くこともあり、意欲的な学生の背中を押していると強く実感しています。

こうしたCDSプログラムを含むAIMD人材育成の取組みについては、IT関連企業、大手シンクタンク、大手プラットフォーム、宮城県、仙台市、等の有識者から成るAIMDアドバイザリボードに意見や助言を求め、点検と改善に努めているところです。

次節以降は、CDSプログラムのコアとも言える2つの必修科目について、その内容と状況についてさらに詳しく紹介することにいたします。

5. 実践的機械学習

この授業は、人工知能を支える基盤技術である機械学習について、実際に手を動かしてプログラミングを行いながらそのエッセンスを学ぶことを目標とするもので、2018年度後期に試行的に実施した展開ゼミ「実践 機械学習」の内容を拡充し開講しています。受講者は年々増加し前期は約250名、後期は約100名となっており、大多数を学部1年生と2年生が占めていますが、大学院生や短期留学生もいます。学部別にみると工学部が

約4割、理学部が約3割、経済学部が約1割で、受講者の事前知識やプログラミング経験は様々です。当初からGoogle Colaboratoryを用いたハンズオン形式で実施しており、突然のコロナ禍による完全オンライン化や、その後の教室の収容人数制限への対応を経て、対面とオンラインを併用する現在の授業形態に落ち着きました。毎回、事前にColaboratoryファイルと説明動画を公開しておき、授業時にはハイフレックス型で演習を進め、受講生の質問に対して教室では対面で、オンラインではSlackを用いて、ティーチングアシスタントとともに応答しています。

前期の「実践的機械学習Ⅰ」では、Pythonの機械学習ライブラリscikit-learnを用いて教師あり学習で学びます。チュートリアル用のIrisデータとDigitsデータに対して、 k 近傍法や決定木、サポートベクトルマシン、ランダムフォレスト、ニューラルネットワークによる分類器を、それぞれまずは使ってみることを優先します。そして各種パラメータが分類結果にどう影響を与えるのかをMatplotlibやPlotlyによる2Dや3Dの可視化も使いながら観察し、アルゴリズムの挙動を理解します。さらに踏み込んで、 k 近傍法や決定木をPythonプログラムで実装することにも挑戦しています。もちろん決して容易な課題ではありませんが、まずその導入として、「丸暗記学習器」と呼ぶ簡単なプログラムから始めます。このとき、ちょっとした作法にしたがうことでscikit-learn準拠として既存の分類器と同様に使えるようになることを体験します。ライブラリの開発者側に立つこと自体は決して難しくないことを知るためです。この丸暗記学習器は、その名の通り訓練例を丸暗記するだけで、未知のデータに対してランダムな予測を行うという乱暴なものですが、それでも訓練例に対して100%の精度が出てしまうのが面白いところで、この経験から「汎化性能」の重要性を実感し、評価法を学ぶ動機となることを期待しています。

後期の「実践的機械学習Ⅱ」では、教師なし学習として階層型クラスタリングや k 平均法に触れた後で、PyTorchを用いて深層学習に取り組みます。scikit-learnとの違いを緩和するライブラリskorchを介して、徐々に多層パーセプトロンやCNNの扱いに慣れていきます。ニューロン数や中間層の数が分類精度や実行時間にどのような影

響を与えるのかを観察し、特に訓練データが少ない場合には分類精度が大幅に落ちてしまうことなどを確かめます。またニューラルネットワークを構成する重み付き和や活性化関数について、それらの部品を単独で使ったり同等なプログラムを書いたりすることで理解を深めます。変分オートエンコーダや画像のスタイル変換などの応用例や、さらには強化学習による迷路探索、制御問題にも触れていきます。

このように盛り沢山な内容ですが、Python初心者向けの補助教材とその解説動画を併せて提供することで、多様な受講者が理解度に合わせて取り組めるよう配慮しています。毎回の講義では、動作の確認を主目的とする基本課題に加えて、難しめの自由課題もヒントとともに提供しています。これに対して興味を持った受講生がかなりの時間と労力をかけたレポートを提出してくれますので、その中から印象的なものを幾つか全体共有しています。他人の優れたレポートを閲覧するのはお互いに良い刺激になっているようです。

ブラックボックスにみえるAI・機械学習も、その中で行われているのはプログラムの実行によるデータ処理に他ならず、予測や分類には誤差が含まれていること、学習に用いるデータが異なると挙動が全く変わることを、したがってもしもデータの前処理を間違ってしまうとその結果は全く無意味であり、また実行結果だけからその誤りに気付くのは極めて困難であることなど、データサイエンスの基本的な事項を受講生が専門分野に進む前に体感できる機会になればと願いながら授業を進めています。



写真1 ハイフレックス型で実施中の「実践的機械学習Ⅰ」の授業風景

6. 機械学習アルゴリズム概論

「機械学習アルゴリズム概論」は2019年度に新設された授業科目です。この授業は、「人工知能や機械学習といった言葉を最近よく耳にするけれど、何のことだかよく分からない」という学生を対象として、機械学習の設計図ともいえる「アルゴリズム」に着目しながら、機械学習の世界に連れまわすことを目的としています。

初年度は50名程度しかいなかった受講者数も年々増加し、2023年度には1,100名を超えました(図1)。学部生向けの授業ですが、大学院生や短期留学生も受講しています。また、「学都仙台単位互換ネットワーク」や、宮城県の「高大連携事業」を通じて、他大学の大学生や県内の高校生も受講しています。さらに、東北地区+新潟の8国立大学からなる「東北創成国立大学アライアンス」にも本授業の教材提供が行われており、学内のみならず全国の数理データサイエンス教育に貢献しています。

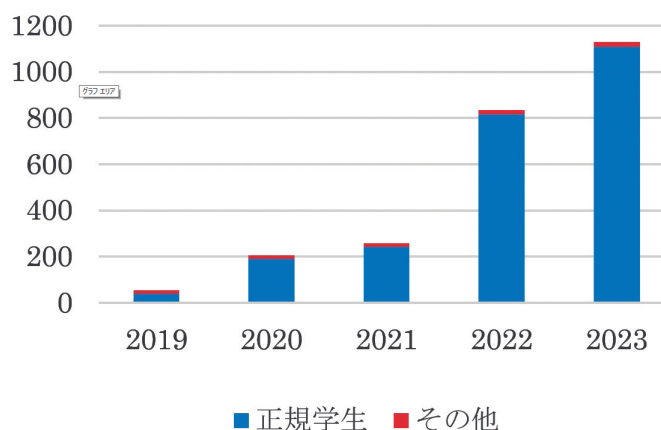


図1 「機械学習アルゴリズム概論」の受講者数の推移 (その他には本学の大学院生・短期留学生の他、他大学の学生や県内の高校生等が含まれます。)

授業を行う上で最も大切にしていることは、受講生との直接的なやり取りです。単に筆者が教壇でしゃべり続ける一方的な授業とならないよう、様々な工夫をしています。その取り組みのひとつとして、受講生には毎回授業の最後に、内容で気になった点、感想、疑問点などを自由に書いて提出してもらう時間を設けています。集まったコメントから、いくつかピックアップして、次の授業の際にフィードバックを行っています。受講生は自身のコメントが授業内で取り上げられると嬉しいようで、特に強制はしていないにも関わらず、毎

回8割以上の受講生からコメントが寄せられています。

本授業は受講生の増加に伴い、2023年度からはオンライン授業となりました。そのため、オンラインの特性を活かして、反転授業を導入しています。具体的には、受講生はオンデマンドで事前動画を視聴し、その後コメントを提出してもらい、コメントの内容に基づいてフィードバック形式で授業を行っています。また、授業がオンデマンドとなったことにより、受講する時間の自由度も増したため、以前は他の授業との兼ね合いで受講できなかった学生も受講しやすくなりました。

授業では実際のデータを用いた実践や、最近ニュース等で話題になった最新技術の解説など、受講生が意欲を持って取り組めるよう様々な工夫を行っています。特にソートアルゴリズムを解説する際には、コンピュータ内部と同様の動作を行う、手作りの実験装置を作成して、視覚的に理解を深めるための実演を行っています(写真2)。また、先述の通り、受講生は高校生から大学院生まで幅広く、持ち合わせている前提知識にも大きな差があるため、初めての人を置き去りにすることなく、一方では、ある程度知識を持つ人が退屈しないように、基礎から丁寧に解説しつつも、向上心がある人に向けた発展的な情報を発信する等、メリハリをつけた授業構成に尽力しています。



写真2 ソートアルゴリズムの実演の様子

7. まとめ

東北大学CDSプログラムは、2022年度に文部科学省数理・データサイエンス・AI教育プログラム(応用基礎レベル)に認定、加えて、同レベルのプラスにも選定いただくことができました。

最後に、その後の新しい取組みについても一部

ご紹介して、本稿を閉じたいと思います。

「数学の基礎」の科目群では、宿題等の時間外学習に、数学用eラーニングとして実績のあるWeBWorKを活用しています。全学のLMSとLTI(Learning Tools Interoperability)によって連携することで、全ての学生が簡単に利用できる環境を提供するとともに、標準的な問題について、大学院生が実際に解いてみせるショート動画を制作し、オープンコースウェアとして一般に公開しています²⁾。

最近、とみに生成AIが各方面で注目されるようになりましたが、データ駆動科学・AI教育研究センターの教員が中心となり、少人数制の課外講座「人工知能技術のためのプログラミング入門」を毎年開講しており、AIによる自然言語処理の進んだ話題にも触れる機会を提供しています。この講座の修了者には、2023年度からオープンバッジを発行し、習得したスキルを認定する予定です。

また、学生の多様な興味に応えるべく、2024年度からは「情報の基礎」科目群に「情報教育特別講義(実践的量子ソリューション創出論)」を加え、実データを用いつつ、量子コンピューティングの基礎から応用までを学部1年生の段階から学ぶことが可能となります。

以上のように、全ての学生に応用基礎レベルまでのAIMD教育を提供する試みの、当初の目標は概ね達成しましたが、限られたリソース(時間割や教員数)の中で、いかに高度な質を確保しつつ受講者数の増加に対応するかが、喫緊の課題となっています。コンソーシアム活動等を通じて、機関の取組みや課題を共有しながら、適応性と持続性を兼ね備えたAIMD人材育成を目指したいと考えているところです。

注

[注1] こうした取組みにより、本学は社団法人オープンバッジネットワークより「オープンバッジ大賞優秀賞(教育機関部門)」を受賞。

関連URL

[1] 東北大学AIMDウェブサイト:

<https://aimd.cds.tohoku.ac.jp/>

[2] CDSオープンコースウェア:

<https://ocw.cds.tohoku.ac.jp/>