

データサイエンスを通して学ぶ課題解決に有効な能力の育成

～決定木 AI を用いた仮説生成の試み～

後藤貴裕

東京学芸大学附属国際中等教育学校

gotoh@u-gakugei.ac.jp

データサイエンスに関する学習を通して得られた、分類（決定木）AI アルゴリズムの手法およびその考え方は、大量の統計データから特定の問題（課題）に強く影響を及ぼしている要因を見出すことを容易にし、さまざまな統計データから問題発見（仮説生成）するための有効なツールとなりうることに期待される。データサイエンスに関する学びが、さまざまな課題解決に有効であることに気づかせる教材としてその具体を提案し、教材研究に関する議論を深めたい。

1. はじめに

AI が普通に使える今日の社会では、さまざまなデータから価値を生み出す方法としてデータサイエンスの需要が高まっている。日本学術会議数理科学委員会数学教育分科会（2020）は、氾濫する情報から、批判的にデータの質を評価したり、機械が行う判断の論理の基礎を数学的に理解したりしながら、適切に意思決定や行動に繋がれるようにすることが必要であると提言している。

また、初中等教育の現場では、総合的な探究の時間の探究活動や理数探究などの課題研究などで、課題（問題）解決のための生徒の探究活動が引き続きもとめられている。

本教材は、教科「情報」に関わるトレンドとしてデータサイエンスおよびAI をとりあげ、その手法や考え方を、さまざまな分野の課題解決で活用できる汎化スキルとして位置付け、AI を活用したプログラミングの実習を通して、問題解決のスタートとなる問題発見（課題研究にとっては大事な過程であるが、現場では苦心している）や仮説生成につながる展開となっている。

本教材は、企画・計画の段階であり、授業実践には至っていないが、その具体を提案し、教科「情報」の教材研究の議論の題材としたい。

2. 教材設計とその展開

2.1 教材の概要

データサイエンスにおける分類 AI を用いた機械学習アルゴリズムの学習（実習）を通して得た技能や考え方を転移させることで、統計データから分類（決定木）AI を用いて特徴量重要度を明らかにし、対象の状態（ここでは2値）を決定（予測）するために強く関与している説明変数を特定し、それを課題解決のための仮説生成に役立てようとするものである。

2.2 教育課程における位置付け

教科「情報」の学校設定科目「インフォマティクス」（高校2学年・2単位科目として開設）の教材として開発。学習指導要領においては、次の複数の科目の内容と深い関わりがある。

○情報「情報Ⅱ」:

（3）情報データサイエンス（多様かつ大量のデータを活用することの有用性に着目し、データサイエンスの手法によりデータを分析し、その結果を読み取り解釈する活動を行う）

○理数「理数探究基礎」:

様々な事象に関わり、理数的な見方・考え方をを用いて、探究の過程を通して、課題を解決するために必要な資質・能力を身につける。
（オ）事象を分析するための技能〈知識・技能〉
（イ）数学的な手法や科学的な手法などを用いて、探究過程を遂行する力〈思考力・判断力・表現力〉

2.3 教材のねらい

- 開いた探究や生徒の課題研究において、探究課題や問いの設定のためのツールとして、分類 AI（決定木）を用いて仮説生成・課題発見をおこなう。
- （オープン）データや統計データから課題を発見する。データを基にして結論（決定）と関係の強い説明変数を特定して、そこから仮説を生成する。その後、別途課題研究などにおいて仮説に基づく研究を展開したり、仮説検証を行ったりする。
- 分類 AI（決定木）を Python でプログラミングする。また分類 AI の過学習を回避し汎化性能を高めるため「枝刈り（Pruning）」を行うことを通し機械学習の特性を実践的に学ぶ。

<https://www.nstac.go.jp/use/literacy/ssdse/>