

機械学習のデータ前処理において数理科学の知識を活用するシナリオ型学習教材の設計

Design of Scenario-Based Learning Materials for Using Mathematical Science Knowledge in Machine Learning Data Preprocessing

福田 美誉^{*1, *2} 久保田 真一郎^{*2} 戸田 真志^{*2} 喜多 敏博^{*2}
Miyo FUKUDA^{*1, *2} Shinichiro KUBOTA^{*2} Masashi TODA^{*2} Toshihiro KITA^{*2}

関西国際大学^{*1} 熊本大学^{*2}
Kansai University of International Studies^{*1}
Kumamoto University^{*2}

〈あらまし〉 本稿では企業調査に基づき、学習者が機械学習を実施する際のデータ前処理を、数理科学の知識を活用して理解するためのシナリオ型学習教材を設計した。業務フローに従い数理科学の学習単元を整理し、実際の機械学習の実施例、教材の構築方法を検討後、GBS 理論に基づき設計した。設計、構築方法ともに数式の解法や ICT の技能習得の内容を含めず、数値を読み取り判断する「知的技能」の学習を充実させる方針を採った。

〈キーワード〉機械学習, データ前処理, 数理科学, ゴールベースシナリオ (GBS) 理論

1. はじめに

筆者らは先進的に AI を活用する企業の調査結果に基づき、AI の実行前に数値の特徴を捉え適切に加工する「データ前処理」及びその際に必要となる数理科学の学習の必要性を報告した(福田・喜多 2025)。この学習部分は、現在の高等教育においては難解な数式を解く活動や、表計算ソフト、プログラミング言語の技能習得等が多く含まれ、数値の本質を読み取り判断する「知的技能」の学習を充実させることが難しい。

そこで本稿では、AI 技術の 1 つである機械学習の「データ前処理」の題材を収集し、福田(2024a)で取りまとめた数理科学の学習単元や、福田・喜多(2024b)で作成した GBS 理論のシナリオ設計を更新するとともに、AI を学習する者(主に非技術職を目指す大学生)が数理科学の知識を活用するための学習教材を設計した。

2. 数理科学の学習単元の整理

学習設計の前に、福田(2024a)や福田・喜多(2024b)で作成した課題分析図及び AI モデルの実行処理フローに企業調査から得た知見を反映し、必要な数理科学の学習単元を整理した。業務フローと数理科学の学習単元の対応を図 1 に示す。⑥の「生成 AI による言語化」は、調査先の企業現場の実態を反映した変更点である。機械学習を実行するストーリーと

しては①～⑥が対象となるが、今回はデータ前処理に相当する②～④の部分进行学习範囲とすることにした。また、※印の箇所は数理科学、AI モデル共に学習難易度が高く、実際の運用業務では自動処理されている部分のため、数理の学習単元とせず、あくまで教材のシナリオ文脈として取り扱う。

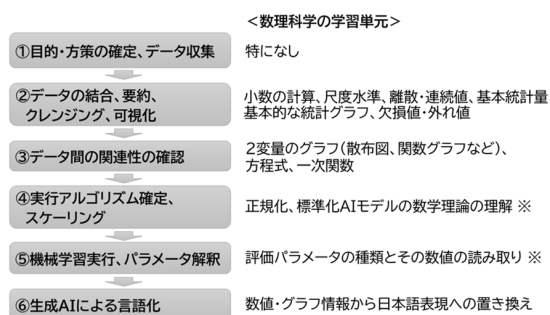


図 1 業務と数理科学の学習単元の対応

3. GBS 理論に基づく学習設計

3.1. 設計前の技術面の検討と方針の確定

次に、学習設計及び実装が可能な方法で教材化できるよう、AI 解析の題材と教材の構築方法を検討した。題材は AI の分野で世界的に有名なコミュニティプラットフォーム Kaggle において以下の条件を基に検索した。

●AI 解析とデータの技術的条件

- ・データ前処理や探索的データ解析 (EDA) の実施例あり
- ・学習要素となるノイズがデータ内に存在
- ・機械学習での解析事例が複数あり

●学習教材としての条件

・若者が想像しやすい店舗や商品のデータ
 ・ライセンスで「改変禁止」の指定がない
 複数の候補のデータとコードの実行結果を確認し、Coursera コースの最終プロジェクトとして出題された「Predict Future Sales」の事例を活用することにした。これは DVD や書籍等を販売するロシアの大手企業 1c company の売上データである。

続けて今後の教材の実装方法を検討した。今回は技能の演習を含めず、数理科学に基づく判断に集中する形態を採ることから、物語を展開する動画の間に小テストを設け回答結果によりシナリオが分岐するよう、Moodle プラグインである H5P を使用することとした。

3.2. GBS 理論に基づくシナリオ設計

前述の検討を踏まえ、作成した GBS シナリオ設計を表 1 に示す。福田・喜多(2024b)からの変更点を下線に示している。題材と同じ店舗を舞台とすることで、教材としての実装を可能にしつつ、社会経験のない学生にも現場や学習の意義が伝わるよう工夫している。

表 1 GBS 理論に基づく学習設計

構成要素	教材の内容
使命	データと AI 技術を用いて、科学的に今後の売上予測をする。
シナリオ文脈	カバー ストーリー 小売店舗の販売数や来客数を予測する業務を担う。AI（機械学習）の実行結果から得られる販売アドバイスを「わかりやすい言葉」で他社員に送信する。画面上で数値データの欠測や不整合をクリアする必要がある。
役割	入社 3 年目の若手営業支援職
学習目標	データ前処理において、数理科学の知識や考え方を適用できる。
シナリオ操作	適切な数理判断や言語化ができていくかの問いを随時提示する。その問いに対する回答に応じて表示する売上データの処理や機械学習による実行結果により、販売予測の実行結果を得ることができる。
シナリオ構成	フィードバック 数理的判断に応じて正誤と解説が出る。どの様な問題が発生するのかを理解できる。
情報源	判断のシーンでヒントを提示する。

3.3. 実装方策の検討

学習設計を基にシナリオを作成した。その中に組み込まれる数理判断の質問項目を表 2

に示す。結果の数値を言葉で表現する問いが最後に用意されている。

表 2 シナリオ内の数理判断の質問項目

業務フロー	数理判断の質問項目の例
目的と方策の確定	不要なデータ列はどれですか
データの加工、要約	NaN は何を示していますか/クレンジングの結果として適切な出力はどれですか
データ間の関連性の確認	相関が見られる項目はどれとどれですか
分析方針の確定（特徴量エンジニアリング）	日付から加工して得られる情報はどれですか/正規化した数値はどれですか/目的変数となる項目はどれですか
機械学習モデル実行結果の解釈	パラメータの数値の解釈として正しいのはどれですか
生成 AI による提案	出力結果を適切に表現した文章はどれですか

4. まとめ

本稿では、機械学習におけるデータ前処理に必要な数理科学の学習教材を設計し、実装の採用技術を確定した。今後はコンテンツを実装し、専門家レビューや少人数の学習者を対象とした実験を通じて形成的評価を行い、教材の有効性を確認する予定である。

謝 辞

本研究は、JSPS 科研費 24K22713 の助成を受けたものです。

参考文献

- Kaggle, Predict Future Sales, <https://www.kaggle.com/competitions/competitive-data-science-predict-future-sales/> (参照日 2025.6.27)
 福田美誉 (2024a) AI 活用プロジェクトの計画に必要な数理学習単元の抽出. 日本教育工学会 2025 年春季全国大会講演論文集: 97-98
 福田美誉, 喜多敏博 (2024b) AI 活用実務に関連する数理学習単元表及び GBS 理論に基づく数理活用型問題解決学習の設計. 第 49 回教育システム情報学会全国大会講演集: 231-232
 福田美誉, 喜多敏博 (2025) 教育の視点でとらえた先進的 AI 活用企業の調査と分析. 日本教育工学会 2025 年春季全国大会講演論文集: 697-698