

# カード操作によるプログラミング学習支援システムにおける 疑似ログ生成を用いた学習者行動の自動分類

## Automatic Classification of Learner Behavior Using Pseudo-Log Generation in a Card-Based Programming Learning Support System

矢上 果林<sup>\*1</sup>, 田辺 七海<sup>\*2</sup>, 松本 慎平<sup>\*1</sup>

Karin YAGAMI<sup>\*1</sup>, Natsumi TANABE<sup>\*2</sup>, Shimpei MATSUMOTO<sup>\*1</sup>,

<sup>\*1</sup> 広島工業大学情報学部

<sup>\*1</sup> Faculty of Applied Information Science, Hiroshima Institute of Technology

Email: {b22226, s.matsumoto.gk}@cc.it-hiroshima.ac.jp

<sup>\*2</sup> 広島工業大学大学院工学系研究科

<sup>\*2</sup> Graduate School of Science and Technology, Hiroshima Institute of Technology

Email: md24003@cc.it-hiroshima.ac.jp

**あらまし** : 本稿では, カード操作によるプログラミング学習支援システムを対象とし, 共起行列分析と疑似ログ生成を用いた学習者行動の自動分類手法について述べる. 本システムでは学習者の詳細な操作履歴が記録されるが, 非効率な学習行動の検出や, 分類器構築のための実データ不足が課題であった. そこで本研究では, 適切学習者の遷移構造を共起行列を用いてモデル化し, これに基づき大量の疑似ログを生成することでデータセットを構築した. 生成データにより学習させた識別器を用い, 実授業のログに対する評価実験を行った結果, 提案手法は極めて高い適合率で適切学習者を判別可能であり, 教育現場での自動介入に向けた有効性を示した.

**キーワード** : カード操作によるプログラミング学習支援システム, 機械学習

### 1. はじめに

分節化された意味のある部分を再構成する活動を通じプログラムの構想に関する学習支援を意図して, カード操作方式による学習支援システム(Card Operation-Based Programming Learning Support System, 以降, COPS) が開発されている<sup>(1)</sup>. COPS では詳細な操作履歴が記録されるため思考過程を分析可能だが, 総当たりの適切な学習活動の自動検出が課題である.

著者はこれまで, ダミーカード等を考慮した拡張レーベンシュタイン距離による可視化<sup>(2)</sup>を試みたが, 遷移特徴の欠落や学習データ不足により自動判別には至っていない. そこで本研究では, 共起行列を用いて適切学習者の遷移構造をモデル化する. さらに, これに基づく適切ログ生成と, 深さ優先探索 (DFS) による不適切ログ生成を組み合わせてデータセットを構築し, 学習過程に基づく自動判別手法の実現を目指す.

### 2. カード操作によるプログラミング学習支援システム

COPS は, ソースコードが記述されたカードを並び替えてプログラムを構築する環境である. 構文エラーを排除し論理構造の理解に注力できる点が特徴で, 選択肢にはダミーカードや空欄カードも含まれる.

システムは, カードの挿入・削除・置換や判定ボタン押下などの操作を詳細なログとして記録する. 本研究ではこの操作ログに着目し, 最終的な正誤だ

けでなく, 正解に至るまでの試行錯誤のプロセスを分析対象とする.

### 3. 提案法

本研究では, 実データの不足を補い, かつ分類根拠を明確にするため, 学習ログの分析および生成を行う. 具体的には, 過去の学習データから熟練者の遷移パターンを抽出して共起行列を構築する. 次に, この行列および探索アルゴリズムを用いて大量の疑似ログを生成し, これを機械学習モデルの訓練データとして利用する.

#### 3.1 共起行列の構築

まず, 収集された過去の実データから, 適切に学習を進めていると判断できるログを選別し, モデル構築の基礎データとする. 選別に当たっては, 課題ごとに正解に至るまでの手数や試行錯誤のプロセスを精査し, 各問題の難易度や構造に応じた基準を個別に設定して抽出を行った. 抽出された各学習者の状態系列から, あるカード配置から次の配置への遷移回数を集計し, これを正規化することで遷移確率行列すなわち共起行列を作成する. これにより, 熟練した学習者が辿りやすい論理的な思考経路を確率モデルとして再現する.

#### 3.2 疑似ログ生成

構築した共起行列および探索アルゴリズムを用い, 機械学習モデルの訓練データとなる疑似ログを生成する. 本研究では, 実際の学習環境における分布を考慮し, 適切ログと不適切ログを作成した.

適切ログの生成においては, 共起行列に記録され

た遷移確率に基づき、次の状態を確率的にサンプリングする。現在の状態から遷移先となる各状態への頻度を重みとして選択を行うことで、正答へと向かう大局的な流れと、微細な試行錯誤による揺らぎを併せ持つログを再現する。

一方、不適切ログの生成には深さ優先探索をベースとした総当たりの状態遷移アルゴリズムを用いる。ただし、実際の学習者が既知の誤配置を繰り返す可能性は低いという仮説に基づき、同一状態への再訪を禁止し、さらに一度誤判定されたダミーカードを以降の探索から除外する制約を設けた。これにより、論理的思考を欠いた機械的な試行錯誤による非効率な探索行動をシミュレートする。

#### 4. 分類モデルの構築

本研究では、生成した疑似ログおよび実ログから、学習過程の動的特徴を反映する5つの指標を特徴量として抽出する。

1つ目は傾きである。これは手数と正解距離であるLS距離の回帰直線の傾きであり、負に大きいほど効率的な接近を示す。2つ目は切片である。これは回帰直線の切片であり、初期状態における正解構造の把握度を反映する。3つ目は決定係数である。これは接近プロセスの直線性を示し、値が1に近いほど迷いのない一貫した操作であることを示す。4つ目は総手数であり、課題終了までに要した全操作回数を用いる。5つ目は提出回数であり、解答ボタンの押下回数を用いる。

これらの特徴量を用い、サポートベクターマシン、k近傍法、決定木の3つのモデルに対して分類器の構築を行った。評価には後述する実証実験のリアルログをテストデータとして用い、その判別性能を検証した。

#### 5. 実証実験

分類モデルの評価用データであるテストデータの収集を目的として、被験者による実証実験を行った。

##### 5.1 実験概要と対象

C言語の基本を理解している情報学専攻の学部3年次生と4年次生、および大学院生の計25名を対象とした。実験ではプログラミング初学者を対象とした難易度の課題を10問提示した。課題構成は、while文やfor文、1次元配列、二重ループ、if-else文を含む基礎的なものである。なお、システムエラーにより正常なログが取得できなかった2問を除外し、計8問を分析対象とした。

##### 5.2 被験者のグループ分け

被験者のプログラミング能力を事前に測定するプレテストを実施し、そのスコアに基づき訓練データと同じく8対2の比率となるよう、適切ロールである上位19名と不適切ロールである下位6名に振り分けた。そして、本実験により得られたラベル付きの実ログを、疑似ログで学習させたモデルの性能検証

用データとして使用した。

#### 6. 分類モデルの構築

各モデルの8問題平均の分類精度を表1に示す。

表1 分類モデルの全体平均性能

モデル	Accuracy	Precision	Recall	F1-Score
SVM_RBF	0.800	0.955	0.744	0.831
KNN	0.785	0.983	0.723	0.825
DT	0.665	0.876	0.670	0.753
全体平均	0.750	0.938	0.712	0.803

分析の結果、全体平均で適合率は0.938、特にk近傍法では0.983と極めて高い値を記録した。本手法が未知の実ログに対し高い汎化性能を示した要因は、共起行列を用いた疑似ログが、単なる最短経路ではない熟練者特有の試行錯誤の揺らぎを効果的に学習できたためである。これにより、適切に学習を進めている者の誤判定をほぼ完全に排除できており、教育現場での自動介入において、学習者の意欲を削ぐリスクの低い安全なモデルであると言える。

一方で、再現率は全体で0.712に留まり、多くの不適切学習者を見逃す結果となった。特に、一部の複雑な課題では再現率が著しく低下していた。詳細な分析の結果、見逃された学習者の多くは、初期段階で正解に必要なカード群を解答欄に揃えた後、正解スロット内のみでカードを入れ替え続ける機械的なシャッフル行動に移行していたことが判明した。また、頻繁な提出を繰り返すフィードバック依存の傾向も見られた。

このような限定されたスロット内での総当たり行動は、外部から観測される指標（傾きや決定係数）においては、部分構造を論理的に組み立てている過程と数学的に極めて酷似した挙動となる。その結果、特徴量空間上で適切群との重なりが生じ、見逃しの主因となった。結論として、本手法は典型的な非効率行動を高精度に特定可能であるが、複雑な課題における意図なき正解接近の識別が今後の課題である。

#### 謝辞

本研究は、独立行政法人日本学術振興会科学研究費助成事業(基盤研究(B)25K00844, 25K00843, 基盤研究(C)23K02697, 22K02815)の助成を受けて実施した成果の一部である。ここに記して謝意を表します。

#### 参考文献

- (1) 重松大志, 奥平泰基, 松本慎平, カード操作によるプログラミング学習支援システムにおけるコンパイラとの接続機能の実装, 情報処理学会論文誌, 65(6), pp.1102-1115(2024)
- (2) 矢上果林, 田辺七海, 松本慎平: “カード操作によるプログラミング学習支援システムの学習過程を分析するための評価手法の検討”, 教育システム情報学会2024年度学生研究発表会, A04(2024)