

画像解析による表情推定に基づく In-process feedback の実装と試験的評価 Implementation and Preliminary Evaluation of In-process Feedback Based on Facial Expression via Image Analysis

中野 晴^{*1}, 山元 翔^{*2}, 田和辻 可昌^{*3}
Haru NAKANO^{*1}, Sho YAMAMOTO^{*2}, Yoshimasa TAWATSUJI^{*3}

^{*1}近畿大学理工学部情報学科

^{*1}Faculty of Science and Engineering, Department of Informatics, Kindai University

^{*2}近畿大学情報学部/情報学研究所

^{*2}Faculty of Informatics / Cyber Informatics Research Institute, Kindai University

^{*3}東京大学大学院工学系研究科

Graduate School of Engineering, The University of Tokyo

Email: 2110370175u@kindai.ac.jp

あらまし：システムのフィードバックを受けても演習が改善しない Wheel-Spinning 問題を解決するために、学習者の脳波データに基づいて行き詰まりのタイミングを検出し、その理由を記号モデルによって推定し、フィードバック生成する In-process feedback が提案されている。しかし、脳波計の活用は構築モデルの検証や現場利用における困難さが残る。本研究では、脳波の代わりに画像解析を用いることで、これらの問題を解決するシステムを提案する。また、このシステムの試験的検証についても報告する。

キーワード：Intelligent Tutoring System, Wheel-Spinning, In-process feedback, 画像解析, 機械学習

1. はじめに

個別適応学習は学習において有用な枠組みであり、ITS (Intelligent Tutoring system) のようなフレームワークにより様々なシステムが構築されている⁽¹⁾。ITS は学習者の回答などから学習者がどのような状態にあるかを推定し、学習者モデルを構築することで、適切なフィードバックの生成に繋げることができる。しかし、真に学習者の状態を推定することは難しく、生成されるフィードバックでは学習が改善せず、不正解のループに陥ってしまう Wheel-Spinning (WS) という問題が指摘されている。

筆者らはこの問題に対して、In-process feedback というフィードバック生成フレームワークを提案している⁽²⁾。先行研究では算数文章題を対象とした作問学習支援システム「モンサクン」を対象とし、脳波データに基づいて学習中の行き詰まりのタイミングを検出し、記号モデルに基づいて回答の様子から行き詰まりの理由を推定、フィードバックをリアルタイムで生成する。しかし、脳波で計測した結果を解釈することは困難であり、タイミング検出のモデルの妥当性を確認することが難しい。また、脳波計は学習者の偽装を防ぐことはできるが、一般的な機器ではないため、少なくとも現代においては、現場適用が難しいという課題もある。

一方で、画像からの状態推定を行う機械学習モデルはより高精度なものが提案されており、学校現場でも一人一台のパソコンが普及してきていることから、カメラによる画像に基づいた行き詰まり検出も可能ではないかと考えた。したがって本研究では、脳波計の代わりに PC の内蔵カメラを用いた In-process feedback の実装と検証を行い、In-process

feedback の汎用性と実用性について検証する。

2. 画像解析を用いた In-process feedback の構築

2.1 モンサクンと三文構成モデル⁽²⁾

今回対象とするモンサクンは加減算の算数文章題であり、図 1 に示すように、学習者は 3 つの単文カードを用いて課題に沿った問題を作問する。作成された問題は三文構成モデルという記号モデルに基づいた診断が行われ、制約違反について学習者にフィードバックを生成する。In-process feedback は右上に提示されているフィードバックで、学習者の行き詰まりを検出すると、リアルタイムで回答状況に合わせたフィードバックを生成する。

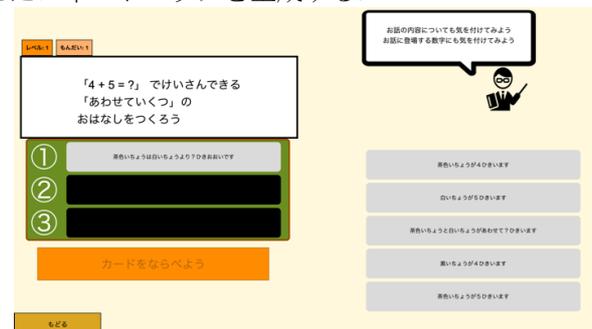


図 1 提案システムの演習画面

2.2 機械学習モデルの構築

本研究では Transformer をベースとした Swin-Transformer⁽³⁾を活用した。本研究の入力データである顔には多様な特徴量が含まれており、それらを考慮に入れることが必要である。そこで、(Transformer 自体エポックメイキングな機械学習

であるが、) 画像中のよりきめ細かな特徴量を取得可能な形で発展的に構築された機械学習である Swin-Transformer を採用した。

データセットの収集は 12 名の大学生を対象とした。これはシステムの検証には大学生を対象とするためである。収集には、通常のノート PC のフロントカメラを動作させながらモンサクンのレベル 1 からレベル 3 (各 10 問) の演習を行なってもらった。その際に別のカメラで学習者の様子と演習画面が映るように動画を撮影しており、各レベル終了後にその動画を確認しながら行き詰まりの箇所を報告してもらった。実験後、動画は 1 秒ごとの画像に変換し、その画像の時点で行き詰まっているかどうか (1/0) を記録することで教師データとした。結果、合計で 67887 個のデータを取得した。

学習に関しては、今回複数の Swin-Transformer ブロックを組み合わせることで機械学習を行い、Softmax 関数で 2 クラス分類としている。また、行き詰まりが多い 1 ユーザのデータを用いて交差検証を実施し、パラメータチューニングを行なった。結果、学習率を $1e-4$ 、バッチサイズを 64、エポック数を 100、ドロップアウト率を 0.03 に設定した。モデルの構築には 10 名のデータを用い、学習者ごとにモデルを構築し、合計 10 個のモデルを構築した。

残り 2 名のデータをテストデータとし、テストデータに対する 10 個の学習器の出力と正解ラベルを比較することでモデルの検証を行なった。システムの動作時には、画像を入力し、10 個の学習器の出力の平均値が 0.5 以上であれば行き詰まり、0.5 より小さいと行き詰まりなしと判断するように実装した。

3. 試験的評価

3.1 検証内容

5 名の大学生に、開発したモンサクンを利用してもらい、本フレームワークが問題なく動作するかを検証した。評価は、演習中に生成された In-process feedback のタイミングと内容の適切さ、および行き詰まりタイミング全てにおいて In-process feedback が生成されたかである。この検証により、画像解析に基づく In-process feedback の構築と実用性について確認する。

3.2 検証結果

表 1 は、検証実験中に表示された In-process feedback のタイミングと内容が適切であったかどうかの割合である。回答は 1 回カードを変更することに行ってもらった。タイミングが適切と回答されたものは 90% と高い結果となっており、概ね誤検出はしていないことが確認できた。内容は不適切という回答がみられたが、これはいずれもフィードバックの提示期間を 10 秒ほどにしており、学習者が次の思考ステップに進んでも前のフィードバックが提示されたままであったことが理由であり、記号モデルを活用することによる精度は高いことが確認できた。

表 2 は、行き詰まりの検出精度の評価を割合で示している。表の項目は(a)学習者が行き詰まっている時行き詰まりを検出してフィードバックを表示した、(b) 学習者は行き詰まっているがフィードバックを表示しなかった、(c)学習者は行き詰まっていないがフィードバックを表示した、である。結果として(b)の割合が多くなっており、行き詰まっていることを検出できないことが多い。これは今回のデータの少なさや、大学生ゆえに行き詰まりのデータが少なかったことが原因であると考えられる。

結果として、フィードバックの提示時間の考察やモデルの精度向上は課題として残ったものの、In-process feedback は画像でも適用可能であり、行き詰まり検出モデルの精度向上によって実用性もある枠組みであることが確認できた。

表 1 表示された In-process feedback の評価

| レベル | タイミング | | | | 内容 | |
|------|-------|------|------|------|------|------|
| | 適切 | 早い | 遅い | 不適切 | 適切 | 不適切 |
| 1 | 1.00 | 0.00 | 0.00 | 0.00 | 0.75 | 0.25 |
| 2 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 3 | 0.88 | 0.00 | 0.08 | 0.04 | 1.00 | 0.00 |
| 全レベル | 0.90 | 0.00 | 0.07 | 0.03 | 0.97 | 0.03 |

表 2 行き詰まり検出精度の評価

| レベル | (a) | (b) | (c) |
|------|------|------|------|
| 1 | 0.29 | 0.71 | 0.00 |
| 2 | 0.00 | 0.00 | 0.00 |
| 3 | 0.36 | 0.62 | 0.01 |
| 全レベル | 0.35 | 0.64 | 0.01 |

3.1 まとめ・今後の課題

本研究では、Swin-Transformer を用いた表情に基づく行き詰まり検出による In-process feedback を構築した。結果としてフィードバックの提示時間の検討や機械学習のモデル生成における課題は残ったものの、汎用的で実用性もある枠組みである可能性が示唆された。

参考文献

- (1) 林勇吾：“知的学習支援システム (Intelligent Tutoring Systems)”，人工知能学会誌，Vol.33, No.4, pp.527-530 (2018)
- (2) Yamamoto, S., Tobe, Y., Tawatsuji, Y. and Hirashima, T.: “In-process feedback by detecting deadlock based on EEG data in exercise of learning by problem-posing and its evaluation”, Research and Practice in Technology Enhanced Learning, Vol.18, No.028, pp.1-26 (2023)
- (3) Liu, Z., Lin, Y., Gao, Y., et al.: “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows”, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.10012-10022 (2021)