

学習者の能力の時系列変化を畳み込む Temporal Convolutional Network を組み込んだ Deep-IRT

西尾 徹朗^{†a)} 堤 瑛美子^{††b)} 植野 真臣^{†††c)}

Deep-IRT with a Temporal Convolutional Network to Convolute Student's Past Ability Change

Tetsurou NISHIO^{†a)}, Emiko TSUTSUMI^{††b)}, and Maomi UENO^{†††c)}

あらまし 近年、人工知能分野、教育工学分野では、教育ビッグデータを分析し、学習者個人の特性に合わせて最適な問題提供と学習支援を行うために、高精度に学習者の能力値と反応の予測を行う Knowledge Tracing が盛んに研究されている。近年ではパラメータの教育的解釈性と高精度な反応予測を両立するために、深層学習と項目反応理論を組み合わせた手法が提案されている。しかし、既存手法では、学習者の能力値を推定する際に最新の潜在能力パラメータのみを用いるため、過去の能力値を反映した能力推定が行われず、パラメータ解釈性と反応予測精度が低下している可能性があった。この問題を解決するために、本研究では、各時点での学習者の潜在能力パラメータを保存し、過去の多次元の能力状態を畳み込む Skill Convolutional Network を組み込むことで、長期の過去の能力値を反応予測に反映する新たな Knowledge Tracing 手法を提案する。評価実験では、提案手法が既存手法と比較して高い反応予測精度とパラメータの解釈性をもつことを示した。

キーワード Knowledge Tracing, Deep Learning, 項目反応理論, Temporal Convolutional Network

1. ま え が き

近年、人工知能分野、教育工学分野では、オンライン学習システムで収集した大量の教育ビッグデータを分析し、教育現場でどのように有効活用するかが課題となっている。特に、オンライン学習システムを用いて学習者の個々の特性に合わせて最適な問題提供と学習支援を行うアダプティブラーニングが注目されている[1]～[13]。また、人工知能分野では、学習効率を最大にする学習支援を行うために、高精度

な学習者の能力値推定と未知の項目への反応予測を行う Knowledge Tracing (KT) が盛んに研究されている[2], [3], [6], [10]～[20]。具体的には、学習者の項目への反応データ（正答・誤答）から習熟度を推定し、未知の項目に対する反応を予測することで、個々の学習者に適切な学習支援を行う。

KT 手法は一般に、確率モデルを用いた手法[2], [3], [10]～[14]と深層学習を用いた手法[6], [15]～[19]が知られている。確率モデルを用いた代表的な手法の一つに項目反応理論 (Item Response Theory; IRT) [21] がある。IRT は項目への反応履歴データから学習者の能力値と項目の難易度を推定し、各項目への学習者の正答確率を予測する数理モデルである。学習者の能力パラメータと項目の難易度パラメータの解釈性が非常に高いことが知られており、教育应用到いられている。通常の IRT モデルでは、学習者の能力値は学習過程で一定値であることが仮定されているため、時間変化する能力値を推定することはできない。そこで、隠れマルコフモデルを用いて時系列変化する能力値を表現した時系列 IRT が複数提案されてい

[†] エスエイティーティー株式会社 駿台人工知能研究所、東京都 Sundai Advanced Teaching Technology, 1-3-12 Kandamisaki-cho, Chiyoda-ku, Tokyo, 101-0061 Japan

^{††} 東京大学大学院情報理工学系研究科、東京都 Graduate school of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan

^{†††} 電気通信大学大学院情報理工学系研究科、調布市 Graduate school of Informatics and Engineering, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, 182-8585 Japan

a) E-mail: nishio@ai.lab.uec.ac.jp

b) E-mail: tsutsumi@ai.lab.uec.ac.jp

c) E-mail: ueno@ai.lab.uec.ac.jp

DOI:10.14923/transinfj.2023JDP7018

る [12], [13], [22], [23]. しかし, これらの確率モデルで推定される能力値は一次元であり, 多次元のスキルの関係性を考慮した能力値推定はできなかった.

近年では, 複数スキルの関係性を考慮した多次元の能力値を推定することにより, 高精度に反応予測を行う深層学習を用いた手法が研究されている. 最新の研究では, 一般に自然言語処理の分野で用いられる Transformer を用いた手法として, Self Attentive Knowledge Tracing (SAKT) [24] が開発されている. SAKT は過去の全ての反応履歴データから複数のスキルと項目の関係性を推定しながら学習者の反応予測を行う. 更に, SAKT の構造に過去の反応履歴データの忘却機能を組み込んだ Attentive Knowledge Tracing (AKT) [17] が提案されている. AKT は反応予測に必要な反応データを Attention を用いて抽出し, 更に直近の反応データをより強く反映することで, 高い反応予測精度を示す手法として知られている. また, Zhang ら [16] は, Recurrent Neural Network (RNN) の一種である memory augmented neural network [25] に Attention 機構を組み合わせた Dynamic Key-Value Memory Network (DKVMN) を開発した. DKVMN は時系列変化する多次元スキルに対する能力状態を保存するための Memory Network をモデルに組み込むことで高精度な反応予測を行うことが報告されている. しかし, これらの深層学習手法は学習者の能力変化を表すパラメータをもたないために教育応用ができない問題があった.

この課題を解決するために, DKVMN と項目反応理論を組み合わせた Deep-IRT [6], [18], [19] が提案されている. Yeung らの Deep-IRT [6] では, DKVMN において IRT と同様に学習者の能力パラメータと項目の難易度パラメータをもつように設計し, 解釈性のあるパラメータ推定を可能とした. しかし, Yeung らの Deep-IRT では能力値が項目の特性に依存するため, 双方の特徴を分離できず, 教育的応用のためのパラメータ解釈性には依然として課題があった. そこで, Tsutsumi らは学習者の能力パラメータと項目の難易度パラメータをそれぞれ独立したネットワークを用いて推定することで, パラメータの解釈性を向上させた Deep-IRT を提案している [18], [26]~[28]. この手法により, 学習者の項目への反応予測精度を保ちながら, 教育的解釈性のある多次元の能力パラメータ推定が可能となった. Tsutsumi らの Deep-IRT [18] は, 学習者の過去の反応データと潜在的な能力値を保存する

value memory パラメータをもち, DKVMN と同様の memory updating component と呼ばれる機構で前時点の value memory をある程度忘却し, 最新の反応データを用いて value memory を更新する. しかし, value memory を更新する際に最新の反応データのみを用いているために, 学習者の過去の反応データや潜在能力値を考慮できず, 反応予測精度が低下している可能性があった. そこで, Tsutsumi らは潜在変数 value memory を更新する際に, Hypernetwork 内で最新の反応データと直前の value memory の重みを最適化する Deep-IRT を開発し, 反応予測精度を大幅に向上させた [19], [20]. Tsutsumi らの Deep-IRT では, 未知の課題への正答確率を計算する際に使用する能力値パラメータを現時点での value memory のみから推定している. そのため, 過去の能力値を反映した能力パラメータの推定が行われず, 反応予測精度が低下している可能性があった.

近年, 時系列データ予測に用いる手法として Convolutional Neural Network (CNN) [29] と Transformer が注目されている [30]~[35]. CNN とは畳み込み層とプーリング層をもつニューラルネットワークであり, 画像などの 2 次元データにおいて局所的な要素の一つの特徴量として圧縮することで特徴抽出を行う手法である. 各要素の重みパラメータを共有することにより, 特定の位置における重要な特徴量を別の位置のデータでも検出できる. CNN を用いた時系列データ予測では, 1 次元の時系列データを入力として, Sliding Window 方式で各時点におけるデータの特徴量抽出を行う. 更に, 長期間における時系列データや時系列変化する潜在変数を多層で畳み込み, 広範囲の時系列データの特徴を反映させる手法として Temporal Convolutional Network (TCN) が開発されている [34]~[37]. TCN は LSTM [38] や GRU [39] などの RNN ベースのモデルよりも高い精度で時系列データを予測できることが報告されている [37], [40]~[43]. LSTM や GRU のように一時点前の状態のみを参照する再帰的なモデルと異なり, 複数の時系列データを畳み込む TCN は, より長期記憶性をもつ [37]. 一方, Transformer は入力されたベクトル要素間の相対距離情報と各要素間の関係性の重要度を計算することで, より長期のデータの情報を保持することができる手法である. 現在多くの分野で高精度なデータ予測が可能な手法として開発が進んでいる.

時系列予測における CNN と Transformer の比較を行った最新の研究 [31], [44]~[46] では, モデル学習の

ためのデータ数が多く、より長期の時系列データを予測する場合に Transformer の予測精度が優位であることが示されている。しかし、Transformer は少数またはスパースなデータセットにはモデルがオーバーフィッティングする問題が挙げられている。CNN は Transformer より若干予測精度が劣るものの、スパースデータにも対応可能でモデル学習時のメモリ効率が高いことが示されている。KT 分野においては、Transformer を用いた既存手法 ([17], [24]) は学習者の反応予測精度が高いことが示されているが、モデルで推定されるパラメータ解釈性が低く、教育応用には課題がある。以上より、スパースで広範囲な時系列データでもパラメータの解釈性を損なわずに推定精度を向上させるためには TCN を用いることが最適であると考えられる。

本研究では毎時点に value memory から出力された多次元的能力状態を保存し、TCN [36], [37] で過去的能力状態を畳み込むことにより、学習者の過去的能力変化を反応予測に反映する新たな Deep-IRT 手法を提案する。各時点での学習者の多次元の潜在能力を表す value memory を全結合層によって低次元の潜在変数にエンコードし、各時点で出力された潜在変数に対して畳み込みを行う。畳み込み時に最適化される重みパラメータによって、長期の過去的能力値を現時点での能力パラメータに反映することができる。評価実験では、提案手法と既存手法 (AKT [17], Tsutsumi らの Deep-IRT [18], Deep-IRT+Hypernetwork [19]) を用いて学習者の反応予測精度の比較を行い、提案手法の有効性を示す。更に、時系列 IRT モデルを用いてシミュレーションデータを作成し、提案手法の能力パラメータの推定値と真のパラメータを相関係数を用いて比較検証することで提案手法の有効性を示す。

2. 関連研究

2.1 Item Response Theory

項目反応理論 (Item Response Theory; IRT) はテストのための測定モデルの一つであり、異なるテストを受けた学習者の能力を同一尺度上で評価することができる数理モデルである。近年では、Knowledge Tracing の分野においても時系列変化する学習者の能力値と項目の難易度を推定するために応用されている [6], [12], [13], [22]。ここでは、最も一般的な 2 パラメータロジスティックモデルについて説明する。2 パラメータロジスティックモデルでは、能力値 $\theta_i \in (-\infty, \infty)$ の学習者 i が項目 j に正答する確率を次のように表す。

$$P_j(\theta_i) = \frac{1}{1 + \exp(-a_j(\theta_i - b_j))} \quad (1)$$

ここで、 $a_j \in [0, \infty)$ は項目 j の識別力パラメータ、 $b_j \in (-\infty, \infty)$ は項目 j の難易度パラメータである。IRT では、能力値 θ は学習過程で一定であることが仮定されている。近年の研究では、学習者の能力値の時系列変化を隠れマルコフモデルで表現した時系列 IRT が複数提案されている [10], [12], [13], [22], [23], [47]。しかし、これらのモデルで推定される能力値は 1 次元であり、多次元のスキルに対する能力変化を考慮することはできない。

2.2 Attentive Knowledge Tracing

深層学習手法の中でも特に高精度な反応予測を行う手法として、自然言語処理の分野で広く用いられる Transformer とラッシュモデルを組み合わせた Attentive Knowledge Tracing (AKT) [17] が提案されている。AKT は学習者の最新の反応データと過去の反応データの関連を保持しながら、指数関数的に過去の反応データを忘却する忘却パラメータをもつ。更に、過去の全反応データから反応予測に必要なデータのみを抽出するように Attention を最適化することにより高精度な反応予測を達成している。AKT における Attention $\alpha_{t,\tau}$ は以下の式で求める。

$$\alpha_{t,\tau} = \frac{\exp(f_{t,\tau})}{\sum_{\tau'} \exp(f_{t,\tau'})}, \quad (2)$$

$$f_{t,\tau} = \frac{\exp(-\eta d(t,\tau)) \cdot \mathbf{q}_t^\top \mathbf{k}_\tau}{\sqrt{D_k}} \quad (3)$$

ここで、 $\mathbf{q}_t \in \mathbb{R}^{D_k}$ は時点 $t = 1$ から t までの、学習者の反応データを表し、 $\mathbf{k}_\tau \in \mathbb{R}^{D_k}$ は時点 τ でのキー行列、 D_k はキー行列 $\mathbf{k}_\tau$ の次元を表す。 $\eta > 0$ は過去データの忘却度を決定するハイパーパラメータである。 $d(t,\tau)$ は時点 τ に入力された過去の反応データと時点 t に入力された最新の反応データの関連性を表し、次式で求められる。

$$d(t,\lambda) = |t - \tau| \sum_{t'=\tau+1}^t \frac{\frac{\mathbf{q}_{t'}^\top \mathbf{k}_{t'}}{\sqrt{D_k}}}{\sum_{1 \leq \tau' \leq t'} \frac{\mathbf{q}_{t'}^\top \mathbf{k}_{\tau'}}{\sqrt{D_k}}}, \quad \forall t' \leq t \quad (4)$$

したがって、AKT では学習者の最新の反応データと過去の反応データの関連性を考慮しながら反応予測に必要なデータのみを抽出し、同時に、時点 τ から時

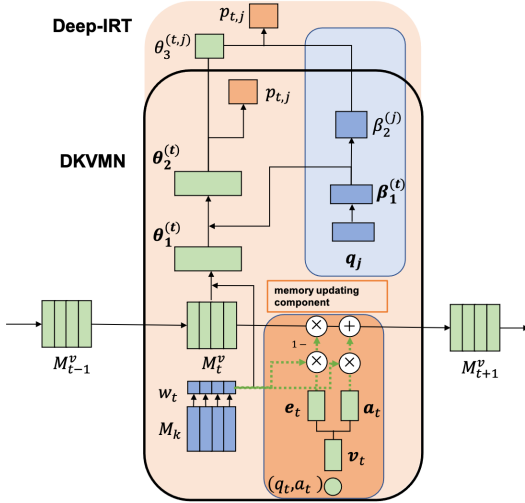


図1 DKVMN, Deep-IRT [6]

点 t までの経過時間 $|t - \tau|$ が大きくなるほど Attention $\alpha_{t,\tau}$ が減少する (忘却する) ように推定される。これらの構造により, AKT は現在最も高い反応予測精度を示す手法として知られている。ただし, AKT は解釈性のあるパラメータをもたないため, 学習過程での学習者の能力変化を表現することができない。

2.3 DKVMN・Deep-IRT

近年, 深層学習手法のパラメータ解釈性を向上させるために多くの手法が開発されている [48]~[54]。Zhang ら (2017) は, Recurrent Neural Network (RNN) の一種である memory augmented neural network [25] に Attention 機構を組み合わせた Dynamic Key-Value Memory Network (DKVMN) を開発した。DKVMN の構造を図 1 に示す。DKVMN は N 個の潜在スキルを仮定しており, 各項目とスキルの関係性を key memory $M_k \in \mathbb{R}^{N \times d_k}$, 時点 t の各潜在スキルに対する能力を value memory $M_t^v \in \mathbb{R}^{N \times d_v}$ にそれぞれ保存することで多次元の潜在スキルに対する能力変化を推定することができる。 d_k と d_v はチューニングパラメータである。

各時点の潜在能力値 value memory $M_t^v \in \mathbb{R}^{N \times d_v}$ は memory updating component で更新される。このとき, それまでの value memory の値をどの程度保存し, 時点 t の反応データをどの程度反映するか調整する忘却パラメータを推定することで value memory の更新を行う。これらの更新・忘却機能により, DKVMN は先行研究に対して反応予測精度を大幅に向上したことが

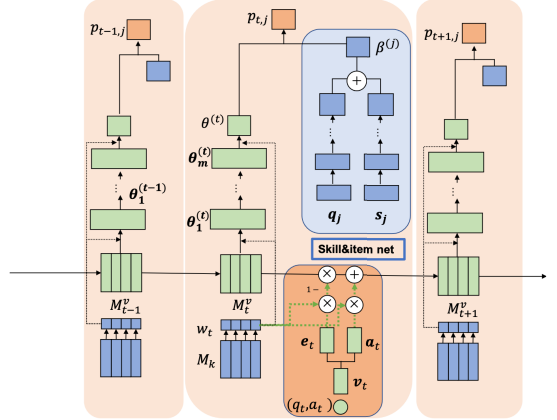


図2 独立なネットワークをもつ Deep-IRT

示されている。しかし, DKVMN は学習者の多次元の能力値が隠れ変数行列に圧縮されているため, 解釈可能な能力パラメータや難易度パラメータをもたないという問題があった。

そこで, Yeung らは DKVMN のパラメータ解釈性を向上させるための手法として, DKVMN と IRT を組み合わせた Deep-IRT [6] を提案した。Yeung らの Deep-IRT (図 1) は DKVMN において IRT と同様に学習者の能力パラメータと項目の難易度パラメータをもつように設計し, 解釈性のあるパラメータ推定を可能とした。しかし, Yeung らの Deep-IRT では能力値が項目の特性に依存するため, 双方の特徴を分離できず, 教育的応用のためのパラメータ解釈性には依然として課題があった。

2.4 独立な学習者・項目ネットワークをもつ Deep-IRT

Tsutsumi らは高精度な反応予測とパラメータ解釈性を両立させるために, 独立した学習者ネットワークと項目ネットワークを用いて学習者の能力パラメータと項目の難易度パラメータを推定する新たな Deep-IRT (図 2) を開発した [18], [28]。Tsutsumi らの Deep-IRT では学習者の能力値が項目の特性に依存せず, 多次元のスキルに対する能力変化を表現することができる。既存手法と同等の反応予測精度を達成しつつ, 学習者の能力推定精度を向上させた。一方, Tsutsumi ら [18] の手法では, DKVMN と同様の memory updating component を用いており, value memory を更新する際に最新の反応データのみを反映し, 学習者の過去の反応データや潜在能力値を考慮できず, 反応予測精度が低

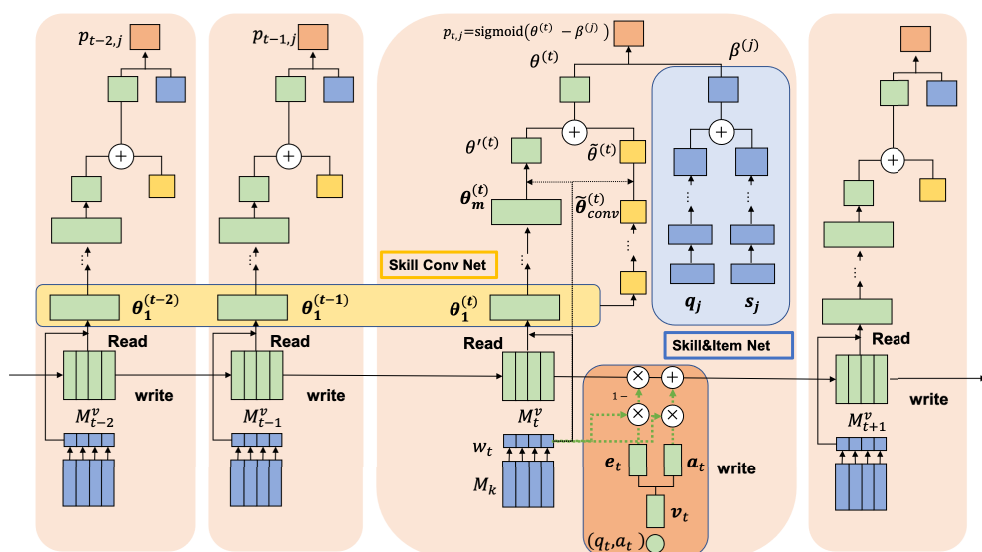


図3 提案手法の構造

下している可能性があった。

この問題を解決するために、Tsutsumi らは、Deep-IRT に Hypernetwork を組み込み、潜在変数 value memory を更新する際に、Hypernetwork 内で最新の反応データと直前の value memory の重みを最適化する新たな手法を開発した [19], [20]. 最新の反応データと直前の value memory のバランスを最適化することにより、memory updating component における忘却パラメータ (value memory の値をどの程度保存し、最新の反応データをどの程度反映するかを調整するパラメータ) を最適化することで、能力パラメータの推定精度と反応予測精度を向上させた。しかし、Tsutsumi らの Deep-IRT では、未知の課題への正答確率を計算する際に使用する能力値パラメータ $\theta^{(t)}$ を現時点での value memory $\mathbf{M}_t^v$ のみから推定している。そのため、過去の能力値を反映した過去能力パラメータの推定が行われず、反応予測精度が低下している可能性があった。

3. 提案手法

本研究では、Tsutsumi らの Deep-IRT [18], [27] に Skill Convolutional Network を組み込み、過去の能力値を反映する能力パラメータ推定を行うことで推定精度を向上させ、高い反応予測精度を達成する新たな Deep-IRT を提案する^(注1)。提案手法の詳細を図 3 に

示す。Skill Convolutional Network では、各時点の潜在能力値を表す value memory M_t^p から出力された多次元の能力状態 $\theta_1^{(t)}$ を保存し、Temporal Convolutional Network (TCN) [36], [37] で過去の能力状態を畳み込むことによって、長期の過去の能力値に依存した能力推定を行う。TCN は時系列データを予測する分野で使われているニューラルネットワークモデルであり、近年、LSTM [38] や GRU [39] などの RNN ベースのモデルよりも高い精度で特定の時系列データを予測することが知られている [37], [40]~[43]。LSTM や GRU のように一時点前の状態のみを参照する再帰的なモデルと異なり、複数の時系列データを畳み込む TCN では、より長期記憶性をもつことが示されている [37]。更に、TCN は長期間における時系列データや時系列変化する潜在変数を多層で畳み込むことで、より広範囲の時系列データの特徴を出力に反映させることができる。

3.1 Skill Convolutional Network

Skill Convolutional Network では、学習者の能力の潜在変数を時系列方向に畳み込むことで、学習者の能力変化を捉えるための最適な重みパラメータを推定する。Skill Convolutional Network の構造を図 4 に示す。Skill Convolutional Network の入力とは各時点の value memory M_t^v を全結合層でエンコードした値を時系列に並べた列 $\{\theta_1^{(t)}, \theta_1^{(t-1)}, \theta_1^{(t-2)}, \dots\}$ である。ここで $\theta_1^{(t)}$ は以下の数式で表される、 M_t^v と Attention Weight の内積である。

(注1) : https://github.com/UEC-Ueno-lab/TCN_Deep-IRT.git

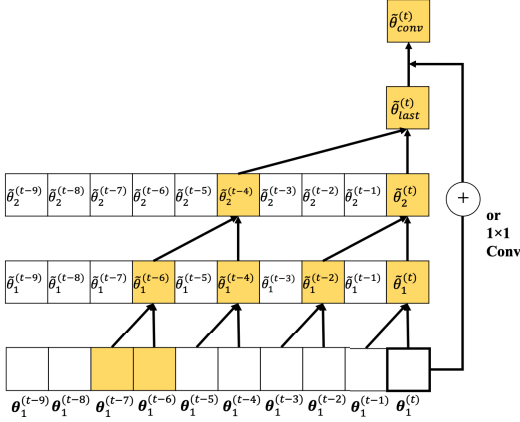


図4 Skill Convolutional Network

$$\theta_1^{(t)} = \sum_{l=1}^N w_{tl} (M_{tl}^v)^T \quad (5)$$

Skill Convolutional Network は Causal Dilated Convolution [36], [55], [56] と Residual Connection [57], [58] で構成される。Causal Dilated Convolution は入力値の各要素を間隔を空けて畳み込むことで、パラメータ数の増加を抑えながらより長い時系列データの畳み込みを行う方法である。畳み込む各要素の間隔 (dilation) は各層で指定することができる。ただし、dilation を 1 にしたときは通常の畳み込みと同様になる。また、Residual Connection は層の 1 層目を最後の出力に加算することにより、層が深い場合の勾配消失を防ぐ効果をもつ。上記の二つの方法を組み合わせることにより、潜在能力値に対する畳み込みを複数層にわたって行う。

提案手法では、時点 t における n 層での入力 $\{\theta_1^{(t)}, \theta_1^{(t-1)}, \theta_1^{(t-2)}, \dots\} \in \mathbb{R}^N$ は多次元であるが、簡単のため次のような 1 次元の列ベクトルを考える。

$$\{\theta_1^{(t)}, \theta_1^{(t-1)}, \theta_1^{(t-2)}, \dots\} \in \mathbb{R} \quad (6)$$

入力ベクトルに対して時点 t での 1 層目の畳み込みは次式で計算される。

$$\tilde{\theta}_1^{(t)} = \sum_{i=0}^{k-1} f_i^{(1)} \cdot \theta_1^{(t-d_1 \cdot i)} \quad (7)$$

同様に n 層目での畳み込みは

$$\tilde{\theta}_n^{(t)} = \sum_{i=0}^{k-1} f_i^{(n)} \cdot \tilde{\theta}_{n-1}^{(t-d_n \cdot i)} \quad (8)$$

と表される。ここで、 $f_i^{(n)}$ は畳み込みに用いる重みであり、 $d_n = \{1, 2, 4, \dots, 2^n\}$ は層ごとに畳み込む要素の間隔を決める dilation パラメータである。また k は畳み込みの際のカーネルサイズであり、以下のように設定する。

$$k = \begin{cases} k_{last} & (\text{最終層}) \\ 2 & (\text{それ以外}) \end{cases} \quad (9)$$

最後に、最終層の出力に第一層の能力値 $\theta_1^{(t)}$ を加算する。

$$\tilde{\theta}_{conv}^{(t)} = \tilde{\theta}_{last}^{(t)} + \theta_1^{(t)} \quad (10)$$

以上の手順により、提案手法は学習者の過去の能力変化を考慮した能力値 $\tilde{\theta}_{conv}^{(t)}$ を推定する。

3.2 学習者ネットワーク・項目ネットワーク

本章では、学習者の能力パラメータと項目の難易度パラメータの推定方法について説明する。学習者の能力パラメータは、時点 t での value memory M_t^v から算出した潜在変数を m 層の全結合層で計算した列ベクトル $\theta_m^{(t)} = \{\theta_{m,1}^{(t)}, \theta_{m,2}^{(t)}, \dots, \theta_{m,N}^{(t)}\}$ と Skill Convolutional Network からの出力 $\tilde{\theta}_{conv}^{(t)} = \{\tilde{\theta}_{conv,1}^{(t)}, \tilde{\theta}_{conv,2}^{(t)}, \dots, \tilde{\theta}_{conv,N}^{(t)}\}$ のそれぞれに対して Attention Weight w_{tl} を用いて重み付き和を計算する。

$$\theta'^{(t)} = \sum_{l=1}^N \omega_{tl} \theta_{m,l}^{(t)} \quad (11)$$

$$\tilde{\theta}^{(t)} = \sum_{l=1}^N \omega_{tl} \tilde{\theta}_{conv,l}^{(t)} \quad (12)$$

列ベクトル $\theta_m^{(t)}$ は Tsutsumi ら [18] と同様の手順で求める。 $\theta'^{(t)}$ は時点 t に学習者が解答した項目に関連するスキルの能力値、 $\tilde{\theta}^{(t)}$ は過去の能力を反映した能力値を表しており、両者を加算することで学習者の能力値 $\theta^{(t)}$ を求める。

$$\theta^{(t)} = \theta'^{(t)} + \tilde{\theta}^{(t)} \quad (13)$$

次に、項目 j に関する難易度パラメータ $\beta^{(j)}$ を求める。はじめに、項目 j の項目 ID に対して埋め込み処理を行い、項目 j に関する難易度 $\beta_{item}^{(j)}$ を以下で求める。

$$\beta_1^{(j)} = GELU(W^{(\beta_1)} q_j + \tau^{(\beta_1)}) \quad (14)$$

表1 データセットの概要

Dataset	No. students	No. skills	No. Items	Rate Correct	Learning length
ASSISTments2009	4151	111	26684	63.6%	52.1
ASSISTments2017	1709	102	3162	39.0%	551.0
Statics2011	333	1223	N/A	79.8%	180.9
Junyi	48925	705	N/A	82.78%	345
Eedi	80000	1200	27613	64.25%	177

$$\beta_m^{(j)} = GELU \left(W^{(\beta_m)} \beta_{m-1}^{(j)} + \tau^{(\beta_m)} \right) \quad (15)$$

$$\beta_{item}^{(j)} = W^{(\beta_{item})} \beta_m^{(j)} + \tau^{(\beta_{item})} \quad (16)$$

ここでは活性化関数として GELU 関数 [59] を用いる。同様に、項目 j のスキル ID に対して埋め込み処理を行い、スキルに関する難易度 $\beta_{skill}^{(j)}$ を以下で求める。

$$\gamma_1^{(j)} = GELU \left(W^{(\gamma_1)} s_j + \tau^{(\gamma_1)} \right) \quad (17)$$

$$\gamma_m^{(j)} = GELU \left(W^{(\gamma_m)} \gamma_{m-1}^{(j)} + \tau^{(\gamma_m)} \right) \quad (18)$$

$$\beta_{skill}^{(j)} = W^{(\beta_{skill})} \gamma_m^{(j)} + \tau^{(\beta_{skill})} \quad (19)$$

次に、 $\beta_{item}^{(j)}$ と $\beta_{skill}^{(j)}$ から以下のように項目 j の難易度 $\beta^{(j)}$ を求める。

$$\beta^{(j)} = \tanh \left(\beta_{item}^{(j)} + \beta_{skill}^{(j)} \right) \quad (20)$$

各ネットワークの層数は $\{2, 3, \dots, m\}$ からデータに最適な値を決定する。

最後に、それぞれのネットワークからそれぞれ出力された、能力値と難易度を用いて時点 t での項目 j に対する正答確率を計算する。

$$p_{t,j} = \sigma \left(\theta^{(t)} - \beta^{(j)} \right) \quad (21)$$

深層学習手法では、一般的に誤差逆伝播法を用いて損失関数 l を最小化することにより、パラメータを学習する。提案手法では、損失関数にクロスエントロピーを用いて誤差逆伝播法で学習を行う。予測正答確率 $p_{t,j}$ と学習者の真の反応データ $y_t = \{1, 0\}$ から以下の式でクロスエントロピーを計算する。

$$l = - \sum_t (y_t \log p_{t,j} + (1 - y_t) \log(1 - p_{t,j})) \quad (22)$$

また、潜在能力値の更新には Tsutsumi ら [18], [27] と同様の memory updating component を用いた。

4. 評価実験

本章では、深層学習を用いた代表的な KT 手法である

AKT [17], Tsutsumi らの Deep-IRT (Tsutsumi-DI) [18], Tsutsumi らの Hypernetwork を組み込んだ Deep-IRT (Tsutsumi-HN) [19], [20] と提案手法を用いて学習者の反応予測精度比較実験を行う。

4.1 データセット

本実験には、データセット Statics2011, ASSISTments2009, ASSISTments2017, Eedi, Junyi を用いる。データセットの詳細を以下に示す。

(1) ASSISTments dataset (ASSISTments2009, ASSISTments2017)^(注2)：算数のオンライン学習システムで収集されたデータセット。

(2) Statics2011 dataset^(注3)：大学工学部の学生を対象に統計学の講義で収集されたデータセット。本研究では、先行研究 [17] で公開されたスキルタグのみをもつデータセットを用いる。

(3) Eedi Dataset^(注4)：数学の学習プラットフォーム Eedi で小学生から高校生を対象に収集されたデータセット。本研究では、各項目に設定された複数のスキルタグの組み合わせを一意的 ID に変換して使用した。

(4) Junyi Dataset Eedi^(注5)：E-Learning システム「Junyi Academy」で収集されたデータセット。

また、各データセットにおける学習者数、スキル数、項目数、正答率、平均解答項目数を表 1 に示す。平均解答項目数は学習者が解答した全項目数の平均値を表す。評価実験は学習者が解答した項目と付与されたスキルの両方の情報を入力データとして用いるが、Statics2011 と Junyi はスキルの情報のみ公開されており、項目の区別ができないためスキルタグのみを入力データとする。

本実験では、データセットを訓練データ、検証データ、評価データに分割し、5 分割交差検証を行う。まず、訓練データと検証データを用いてモデルの学習とパラメータチューニングを行う。各データセットにお

(注2) : <https://sites.google.com/site/assistmentsdata/home/assistment-2009-2010-data>

(注3) : <https://pslclatashop.web.cmu.edu/DatasetInfo?datasetId=507>

(注4) : <http://www.junyiacademy.org/>

(注5) : <https://eedi.com/projects/neurips-education-challenge>

表3 学習者の反応予測精度

Dataset	metrics	Tsutsumi-DI [18]	AKT [17]	Tsutsumi-HN [19]	Proposed
ASSISTments2009	AUC	82.09+/-0.28	82.20+/-0.25	81.98+/-0.54	82.95+/-0.30
	Accuracy	77.41+/-0.53	77.30+/-0.55	77.15+/-0.55	77.60+/-0.56
ASSISTments2017	AUC	73.56+/-0.27	74.54+/-0.21	75.13+/-0.20	75.49+/-0.36
	Accuracy	69.78+/-0.41	69.83+/-0.06	70.69+/-0.60	70.85+/-0.50
statics2011	AUC	81.15+/-0.37	82.15+/-0.35	81.57+/-0.50	82.02+/-0.39
	Accuracy	80.01+/-0.92	80.41+/-0.67	80.11+/-0.92	80.40+/-0.80
Eedi	AUC	78.93+/-0.12	77.58+/-0.21	78.97+/-0.10	79.14 +/-0.11
	Accuracy	73.38+/-0.17	72.35+/-0.21	73.38+/-0.13	73.55+/-0.13
Junyi	AUC	77.92+/-0.41	78.13+/-0.39	77.91+/-0.37	78.14+/-0.43
	Accuracy	86.79+/-0.15	86.79+/-0.17	86.65+/-0.15	86.85+/-0.14
Average	AUC	78.73	78.92	79.11	79.60
	Accuracy	77.47	77.54	77.60	77.81

表2 提案手法の条件

Data set	層数	k_{last}
ASSISTments2009	5	3
ASSISTments2017	8	2
Statics2011	3	7
Junyi	8	2
Eedi	8	2

ける最適な Skill Convolutional Network の層数と k_{last} の値は、それぞれの値を変化させて提案手法を学習したときに最も検証データの予測精度の高い条件を採用した(表2)。次に、評価データに対して学習者の未知の項目への反応予測を行い、実際の反応データとの比較を行う。反応予測精度の評価指標には Accuracy (正答・誤答の一致割合) と AUC スコアを用いた。

4.2 反応予測精度の評価実験

学習者の解答した項目に対する反応予測精度を表3に示す。表3より、提案手法は AUC, Accuracy とともに最も高い平均値を示した。提案手法は、既存手法の Tsutsumi-HN [19] を平均的に上回っており、Skill Convolutional Network を用いて学習者の過去の能力を反映した能力値推定を行ったことで、反応予測精度が向上したと考えられる。特に、ASSISTments2009, ASSISTments2017, Eedi, Junyi では AUC, Accuracy とともに全ての既存手法を上回る精度を達成した。これらのデータセットは学習者が1000人以上の大規模なデータセットであり、モデルの学習に十分な学習データがあったことも提案手法の精度向上につながったと考えられる。また、Statics2011においては、既存手法の AKT とほぼ同程度の予測精度を達成した。他のデータよりも精度向上が限定的になっているのは、各時点での学習者の能力値が過去の能力値に依存せずに変動しているために、Skill Convolutional Network が有効に働かなかったと考えられる。

5. 能力パラメータの解釈性評価実験

5.1 能力パラメータ推定精度評価

本章では、提案手法における能力パラメータの推定精度を評価するために、Tsutsumi ら [20] と同様に、シミュレーションデータを用いて提案手法と既存手法 (Tsutsumi-DI [18], Tsutsumi-HN [19]) との推定能力値の比較実験を行う。シミュレーションデータは時系列 IRT の手法である TIRT を用いて生成する [22]。TIRT では、学習者 i が項目 j に時点 t で正答する確率を以下で計算する。

$$P_{ij}(x_{ij} = 1|\theta_{it}) = \frac{1}{1 + \exp(-\tilde{a}_{\Delta_t}\theta_{it} - b_j)} \quad (23)$$

$$\tilde{a}_{\Delta_t} = \frac{a_j}{\sqrt{1 + \sigma a_j^2 \Delta_t}} \quad (24)$$

ここで、 $\Delta_t = t - t_j$ であり、現在の時点 t と過去に項目 j を解いた時点の差分を表す。 $a_j \in (0, \infty)$ は項目 j の識別力パラメータ、 $b_j \in (-\infty, \infty)$ は項目 j の難易度パラメータである。項目の識別力パラメータ a と難易度パラメータ b の事前分布には、それぞれ $a \sim LN(0, 1)$, $b \sim N(0, 1)$ の対数正規分布と正規分布を用いた。また、 $\theta_{it} \in (-\infty, \infty)$ は時点 t での学習者 i の能力値を表す。 θ_{it} の事前分布は $\theta_{i0} \sim N(0, 1)$, $\theta_{it} \sim N(\theta_{it-1}, \sigma)$ の正規分布とした。 σ は過去の学習データの忘却度を決める忘却パラメータで、その値が大きいくほど現在の能力値が過去の能力値に依存しない。また、 σ は θ_{it} の分散を反映しており、 σ が大きく現在の能力値が過去の能力値に依存しないほど θ_{it} は急激に変動する。

本実験では、TIRT を用いて学習者2000人の50, 100, 200, 300項目への反応データをデータセットとして生成した。はじめに学習者1800人分の反応データを用いて項目の識別力パラメータ a と項目パラメータ b を

表4 推定能力値と真の能力値の相関係数

σ	No. items Method	50	100	200	300	50	100	200	300	50	100	200	300
		Pearson				Spearman				Kendall			
0.1	Tsutsumi-DI	0.732	0.771	0.856	0.815	0.745	0.786	0.882	0.863	0.548	0.595	0.701	0.686
	Tsutsumi-HN	0.728	0.765	0.85	0.821	0.739	0.784	0.880	0.867	0.542	0.592	0.698	0.691
	Proposed	0.859	0.896	0.913	0.892	0.866	0.909	0.943	0.939	0.678	0.741	0.789	0.788
0.3	Tsutsumi-DI	0.826	0.843	0.853	0.862	0.849	0.900	0.942	0.951	0.661	0.723	0.787	0.803
	Tsutsumi-HN	0.819	0.858	0.855	0.861	0.846	0.913	0.943	0.952	0.657	0.741	0.79	0.804
	Proposed	0.840	0.905	0.900	0.907	0.877	0.932	0.947	0.954	0.689	0.767	0.791	0.804
0.5	Tsutsumi-DI	0.826	0.843	0.853	0.862	0.849	0.900	0.942	0.951	0.661	0.723	0.787	0.803
	Tsutsumi-HN	0.849	0.839	0.833	0.818	0.895	0.927	0.939	0.951	0.714	0.759	0.780	0.804
	Proposed	0.846	0.839	0.820	0.810	0.894	0.916	0.917	0.892	0.710	0.732	0.730	0.702
1.0	Tsutsumi-DI	0.826	0.802	0.813	0.795	0.888	0.924	0.940	0.939	0.702	0.753	0.783	0.783
	Tsutsumi-HN	0.822	0.805	0.813	0.794	0.887	0.924	0.940	0.938	0.700	0.754	0.786	0.784
	Proposed	0.802	0.793	0.803	0.788	0.876	0.916	0.921	0.927	0.682	0.737	0.748	0.760

推定し、推定した a, b を所与として残りの 200 人の各時点での能力値を推定する。また、各データセットについて θ_{it} の分散 σ の値を $\sigma = \{0.1, 0.3, 0.5, 1.0\}$ に変化させて生成したデータセットを用いて実験を行った。 σ は学習者の能力値の変動を調整するパラメータであり、 σ の値が大きいくほど真の能力値は各時点で大きく変動するように推定される。本研究では、真のモデル (TIRT) から生成された真の能力値と提案手法、既存手法で推定された能力値を比較するために、先行研究と同様に、各時点での能力値について Pearson の積率相関係数、Spearman の順位相関係数、Kendall の順位相関係数を求めた [20]。母集団に分布が仮定されている Pearson の積率相関係数に対して、Spearman の順位相関係数は母集団に分布を仮定しないノンパラメトリックな指標である。Kendall の順位相関係数は異常値の頑健な推定について評価する指標である。TIRT によって生成した学習過程での能力値 $\theta_{it}, t \in \{1, 2, \dots, T\}$ と提案手法、既存手法で推定された能力値をそれぞれの相関指標を用いて算出した。

表 4 に各条件のデータセットについて算出した相関係数を示す。提案手法は全てのデータセットにおいて TCN の層数 8, $k_{last} = 2$ を用いた。表 4 から、能力変化の分散が小さい $\sigma = \{0.1, 0.3\}$ では提案手法は既存手法と比較して高い推定精度を示した。 σ が小さい場合には現在の能力値が過去の能力値に大きく依存するために、長期の過去の能力値を考慮する Skill Convolutional Network が効果的に機能し、推定精度が向上したと考えられる。一方で、能力の分散 σ が大きい ($\sigma = \{0.5, 1.0\}$) ということは、現在の能力値が過去の能力値に強く依存していないことを示しており、急激な能力値変動が起きやすくなっている。した

がって、この場合、長期ではなく直近の能力値と反応データのみから能力推定を行う既存手法 (Tsutsumi-DI, Tsutsumi-HN) が高い相関を示している。このように、現在の能力値が過去の能力値に依存する場合は提案手法が優位になり、依存が弱い場合は既存手法が優位になる傾向がある。

5.2 能力パラメータ解釈性評価

本章では、提案手法を用いて推定した学習者の能力値推移を時点ごとに可視化し、能力パラメータが解釈性をもつことを示す。本実験では、Tsutsumi ら [18], [20] と同様に ASSISTments2009 データセットを用いて、提案手法を用いて学習者の各スキルの能力推移を推定した。実データでの評価実験では学習者の真の能力値が未知であるため、推定された能力値が実際の反応データに対して妥当であるかを評価する。

図 5 は、ある学習者の 23 項目 (単一スキル) に対する反応データから提案手法と Tsutsumi らの手法 (DeepIRT-HN) [19] を用いて推定した能力変化である。縦軸右側は学習者の能力値、横軸は学習者の解答項目を表し、左から右に学習が進む様子を表す。また、学習者が項目に正答した場合は “●” に、項目に誤答した場合は “○” で表す。図 5 は、学習者が数問正答後に誤答する反応パターンを繰り返し、最後に 5 問連続で解答する反応データである。DeepIRT-HN は学習の初期段階で大幅に推定能力値が向上した後、反応データを反映した能力変動はあるものの、高い値からほとんど変化しない。DeepIRT-HN は現時点の能力状態が一つ前の能力状態のみに依存するため、特に正答 (誤答) が連続する場合には極端に高い (低い) 能力値に収束してしまい、過学習する可能性が考えられる。一方、提案手法では、過去の能力変化を考慮しながら能力推

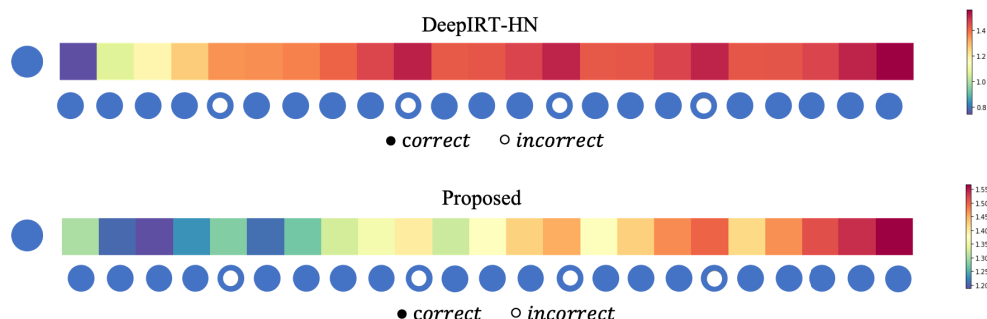


図5 単一のスキル項目における能力パラメータの比較

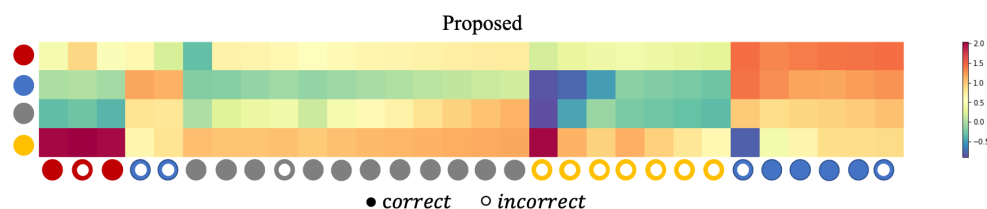


図6 提案手法での多次元の能力パラメータの変化

定を行うため、正答する項目が増えるにつれて徐々に能力値が上昇し、最終時点で最も能力値が高くなるように推定されている。

次に、ASSISTments2009 データセットの反応データを用いて、学習者が多次元のスキルを学習した際の推定能力値の変化を分析する。多次元のスキルに対する能力値変化を可視化することは、指導者が学習者の習熟度や苦手分野を把握し、適切な支援を行うために重要である。図6は、提案手法を用いて、ある学習者の30項目に対する反応データから推定した能力値推移を示す。縦軸左側がスキル、縦軸右側は学習者の能力値を表し、横軸は学習者の解答項目を表す。また、学習者が項目に正答した場合は“●”に、項目に誤答した場合は“○”で表す。学習者は四つのスキルに回答しており、それぞれのスキルは“ordering fractions”(オレンジ色)、“equation solving more than two steps”(青色)、“equation solving two or fewer steps”(灰色)、“finding percents”(黄色)で示されている。提案手法は学習者の能力の傾向とともにスキル間の関連性を考慮して能力値推定を行うため、反応データの正誤に伴って該当するスキルだけでなく他のスキルも変化する傾向があることがわかる。例えば、“equation solving more than two steps”(青色)と“equation solving two or fewer steps”(灰色)は類似のスキルであるため、灰色のスキルに正答し続けると、青色のスキルの能力も共に上昇する。提案手法は

過去の能力変化を考慮しているため、推定能力値は各スキルにおいて反応データを徐々に反映するように能力値が遷移していることがわかる。一方で、解答する項目のスキルが変わると推定能力値が大幅に変動する部分が見られたが、実際には、一つのスキルの能力値が急激に変化することは起こりにくいと考えられるため、多次元の能力値の解釈性向上については今後の課題としたい。

6. む す び

本研究では、Temporal Convolutional Network (TCN) を用いて学習者の過去の潜在能力値の畳み込みを行う Skill Convolutional Network を組み込んだ新たな Deep-IRT を提案した。提案手法は過去の多次元の能力状態を保存し、TCN で過去の潜在能力値の畳み込みを行うことにより、学習者の時系列変化する能力を考慮した反応予測を行う。また、Causal Dilated Convolution と Residual Connection と呼ばれる手法を用いることで、より長期的な学習データに対してパラメータ数の増加を抑えながら反応予測を行うことが可能である。評価実験では、実データを用いて反応予測精度の比較を行い、既存手法と比較して提案手法の反応予測精度が向上したことを示した。特に学習者数の多い大規模なデータに対して高い反応予測精度を示した。また、時系列 IRT で生成したシミュレーションデータを用いた

能力推定精度比較では、真の能力が一時点前の能力に依存するような学習過程の場合において、提案手法が既存手法より高精度に能力推定を行うことを示した。

今後の研究では、本研究においては探索によって決定した TCN の層数や最終層のカーネルサイズを自動で推定する手法を開発し、より高い精度での反応予測と能力値推定を目指す。また、近年研究されている学習者のレベルにあったヒントを提示する足場がけ学習システム [60], [61] への応用を行っていきたい。

謝辞 本研究は JSPS 科研費 JP19H05663, P22K19825, JP22J15279, 22KJ1368 の助成と 2022 年度岡太彬訓研究助成を受けたものです。

文 献

- [1] M. Ueno, "Data mining and text mining technologies for collaborative learning in an ILMS "samurai"," Proc. IEEE Int. Conf. Advanced Learning Technologies, ICALT 2004, Joensuu, Finland, Aug. 30 – Sept. 1, 2004.
- [2] A.T. Corbett and J.R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," User Model. User-Adapt. Interact., vol.4, no.4, pp.253–278, Dec. 1995.
- [3] M.M. Khajah, R. Wing, R.V. Lindsey, and M.C. Mozer, "Integrating latent-factor and knowledge-tracing models to predict individual differences in learning," Educational Data Mining, 2014, <https://api.semanticscholar.org/CorpusID:225023>
- [4] J. Lee and E. Brunskill, "The impact on individualizing student models on necessary practice opportunities," Proc. 5th Int. Conf. Educational Data Mining, pp.118–125, Jan. 2012.
- [5] M.A.S. Pedro, R.S.J. deBaker, and J.D. Gobert, "Incorporating scaffolding and tutor context into bayesian knowledge tracing to predict inquiry skill acquisition," EDM, 2013.
- [6] C. Yeung, "Deep-IRT: Make deep learning based knowledge tracing explainable using item response theory," Proc. 12th Int. Conf. Educational Data Mining, EDM, 2019.
- [7] R. Pelánek, "Conceptual issues in mastery criteria: Differentiating uncertainty and degrees of knowledge," Proc. 19th Int. Conf. Artificial Intelligence in Education, vol.1, pp.450–461, June 2018.
- [8] Z. Pardos and N. Heffernan, "Modeling individualization in a bayesian networks implementation of knowledge tracing," Proc. 18th Int. Conf. User Modeling, Adaption, and Personalization, pp.255–266, June 2010.
- [9] Z.A. Pardos and N.T. Heffernan, "KT-IDEM: Introducing item difficulty to the knowledge tracing model," Proc. 19th Int. Conf. User Modeling, Adaptation and Personalization (UMAP 2011), pp.243–254, Jan. 2011.
- [10] R. Weng and D. Coad, "Real-time bayesian parameter estimation for item response models," Bayesian Analysis, vol.13, Dec. 2016.
- [11] X. Wang, J. Berger, and D. Burdick, "Bayesian analysis of dynamic item response models in educational testing," Annals of Applied Statistics, vol.7, no.1, pp.126–153, 2013.
- [12] 堤瑛美子, 宇都雅輝, 植野真臣, "ダイナミックアセスメントのための隠れマルコフ IRT モデル," 信学論 (D), vol.J102-D, no.2, pp.79–92, Feb. 2019.
- [13] 堤瑛美子, 植野真臣, "Knowledge Tracing のための Sliding Window 隠れマルコフ IRT," 信学論 (D), vol.J103-D, no.12, pp.894–905, Dec. 2020.
- [14] M.V. Yudelson, K.R. Koedinger, and G.J. Gordon, "Individualized bayesian knowledge tracing models," Artificial Intelligence in Education, pp.171–180, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [15] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L.J. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracing," Advances in Neural Information Processing Systems 28, eds. by C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, and R. Garnett, pp.505–513, Curran Associates, 2015.
- [16] J. Zhang, X. Shi, I. King, and D.-Y. Yeung, "Dynamic key-value memory network for knowledge tracing," Proc. 26th Int. Conf. World Wide Web, pp.765–774, WWW '17, International World Wide Web Conferences Steering Committee, 2017.
- [17] A. Ghosh, N. Heffernan, and A.S. Lan, "Context-aware attentive knowledge tracing," Proc. 26th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining, 2020.
- [18] E. Tsutsumi, R. Kinoshita, and M. Ueno, "Deep-IRT with independent student and item networks," Proc. 14th Int. Conf. Educational Data Mining (EDM), 2021.
- [19] E. Tsutsumi, Y. Guo, and M. Ueno, "DeepIRT with a hypernetwork to optimize the degree of forgetting of past data," Proc. 15th Int. Conf. Educational Data Mining (EDM), 2022.
- [20] 堤瑛美子, 郭 亦鳴, 植野真臣, "学習データの忘却を最適化する Hypernetwork を組み込んだ DeepIRT," 信学論 (D), vol.J106-D, no.2, pp.72–83, Feb. 2023.
- [21] F.B. Baker and S.H. Kim, Item Response Theory: Parameter Estimation Techniques, Second Edition, Statistics: A Series of Textbooks and Monographs, Taylor & Francis, 2004.
- [22] K.H. Wilson, Y. Karklin, B. Han, and C. Ekanadham, "Back to the basics: Bayesian extensions of irt outperform neural networks for proficiency estimation," 9th Int. Conf. Educational Data Mining, vol.1, pp.539–544, June 2016.
- [23] C. Ekanadham and Y. Karklin, "T-SKIRT: Online estimation of student proficiency in an adaptive learning system," CoRR, vol.abs/1702.04282, 2017.
- [24] S. Pandey and G. Karypis, "A self-attentive model for knowledge tracing," Proc. Int. Conf. Education Data Mining, 2019.
- [25] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "Meta-learning with memory-augmented neural networks," Proc. 33rd Int. Conf. Machine Learning - Vol.48, pp.1842–1850, ICML'16, JMLR.org, 2016.
- [26] 木下 涼, 植野真臣, "深層学習によるテスト理論: Item Deep Response Theory," 信学論 (D), vol.J103-D, no.4, pp.314–329, April 2020.
- [27] 堤瑛美子, 木下 涼, 植野真臣, "独立な学習者・項目ネットワークをもつ Deep-IRT," 信学論 (D), vol.J104-D, pp.596–608, July 2021.
- [28] E. Tsutsumi, R. Kinoshita, and M. Ueno, "Deep item response theory as a novel test theory based on deep learning," Electronics, vol.10, no.9, 2021.
- [29] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, vol.86,

- no.11, pp.2278–2324, 1998.
- [30] F. Yuan, Z. Zhang, and Z. Fang, “An effective cnn and transformer complementary network for medical image segmentation,” *Pattern Recognition*, vol.136, p.109228, 2023. <https://www.sciencedirect.com/science/article/pii/S0031320322007075>
- [31] E. Arkin, N. Yadikar, X. Xu, A. Aysa, and K. Ubul, “A survey: object detection methods from cnn to transformer,” *Multimedia Tools and Applications*, vol.82, Oct. 2022.
- [32] Y. Bai, J. Mei, A. Yuille, and C. Xie, “Are transformers more robust than cnns?,” 35th Conf. Neural Information Processing Systems, 2021.
- [33] Z. Wang, Y. Ma, Z. Liu, and J. Tang, “R-transformer: Recurrent neural network enhanced transformer,” *ArXiv*, vol.abs/1907.05572, 2019.
- [34] D. Walther, J. Viehweg, J. Haeisen, and P. Mäder, “A systematic comparison of deep learning methods for eeg time series analysis,” *Frontiers in Neuroinformatics*, vol.17, 2023. <https://www.frontiersin.org/articles/10.3389/fninf.2023.1067095>
- [35] S. Wang, “A stock price prediction method based on BiLSTM and improved transformer,” *IEEE Access*, vol.11, pp.104211–104223, 2023. doi: 10.1109/ACCESS.2023.3296308.
- [36] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, “Wavenet: A generative model for raw audio,” *Arxiv*, 2016. <https://arxiv.org/abs/1609.03499>
- [37] S. Bai, J.Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” 2018. <https://arxiv.org/abs/1803.01271>
- [38] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol.9, no.8, pp.1735–1780, 1997.
- [39] J. Chung, Ç. Gülçehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” *CoRR*, vol.abs/1412.3555, 2014. <http://arxiv.org/abs/1412.3555>
- [40] A. Kanazawa, J.Y. Zhang, P. Felsen, and J. Malik, “Learning 3d human dynamics from video,” *Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [41] T. Afouras, J.S. Chung, A. Senior, O. Vinyals, and A. Zisserman, “Deep audio-visual speech recognition,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.44, no.12, pp.8717–8727, 2022.
- [42] C. Lea, M.D. Flynn, R. Vidal, A. Reiter, and G.D. Hager, “Temporal convolutional networks for action segmentation and detection,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [43] W. Zhao, Y. Gao, T. Ji, X. Wan, F. Ye, and G. Bai, “Deep temporal convolutional networks for short-term traffic flow forecasting,” *IEEE Access*, vol.7, pp.114496–114507, 2019.
- [44] Z. Chen, M. Ma, T. Li, H. Wang, and C. Li, “Long sequence time-series forecasting with deep learning: A survey,” *Information Fusion*, vol.97, p.101819, 2023. <https://www.sciencedirect.com/science/article/pii/S1566253523001355>
- [45] C. Sheng, G. Kuang, L. Bai, C. Hou, Y.K. Guo, X. Xu, M. Pietikäinen, and L. Liu, “Deep learning for visual speech analysis: A survey,” *ArXiv*, vol.abs/2205.10839, 2022.
- [46] Q. Ma, Z. Liu, Z. Zheng, Z. Huang, S. Zhu, Z. Yu, and J.T. Kwok, “A survey on time-series pre-trained models,” *ArXiv*, vol.abs/2305.10716, 2023.
- [47] A.D. Martin and K.M. Quinn, “Dynamic ideal point estimation via markov chain monte carlo for the u.s. supreme court, 1953–1999,” *Political Analysis*, vol.10, pp.134–153, 2002.
- [48] P. Chen, Y. Lu, V. Zheng, and Y. Pian, “Prerequisite-driven deep knowledge tracing,” *IEEE Int. Conf. Data Mining, ICDM 2018*, pp.39–48, 2018.
- [49] X. Liangbei and D. Mark, “Dynamic knowledge embedding and tracing,” *Proc. 13th Int. Conf. Educational Data Mining, EDM*, pp.524–530, 2020.
- [50] Y. Lu, D. Wang, Q. Meng, and P. Chen, “Towards interpretable deep learning models for knowledge tracing,” *Proc. 13th Int. Conf. Educational Data Mining, EDM*, pp.185–190, 2020.
- [51] H. Tong, Y. Zhou, and Z. Wang, “Exercise hierarchical feature enhanced knowledge tracing,” *Proc. 13th Int. Conf. Educational Data Mining, EDM*, pp.324–328, 2020.
- [52] Z. Wang, X. Feng, J. Tang, G. Huang, and Z. Liu, “Deep knowledge tracing with side information,” *Proc. 20th Int. Conf. Artificial Intelligence in Education (AIED)*, pp.303–308, 2019.
- [53] X. Xiong, S. Zhao, V. Inwegen, G. Eric, and J.E. Beck, “Going deeper with deep knowledge tracing,” *Proc. Int. Conf. Education Data Mining*, 2016.
- [54] C.K. Yeung and D.-Y. Yeung, “Addressing two problems in deep knowledge tracing via prediction-consistent regularization,” *Proc. 5th ACM Conf. Learning @ Scale*, pp.1–10, 2018.
- [55] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” *Int. Conf. Learning Representations*, 2016.
- [56] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A.L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.40, no.4, pp.834–848, 2018.
- [57] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” *Computer Vision – ECCV 2016*, eds. by B. Leibe, J. Matas, N. Sebe, and M. Welling, pp.630–645, Springer International Publishing, Cham, 2016.
- [58] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp.770–778, IEEE Computer Society, Los Alamitos, CA, USA, June 2016. <https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.90>
- [59] D. Hendrycks and K. Gimpel, “Gaussian error linear units (gelus),” 2016. <https://arxiv.org/abs/1606.08415>
- [60] M. Ueno and Y. Miyazawa, “Probability based scaffolding system with fading,” *Proc. Artificial Intelligence in Education – 17th Int. Conf., AIED*, pp.237–246, 2015.
- [61] M. Ueno and Y. Miyazawa, “IRT-based adaptive hints to scaffold learning in programming,” *IEEE Trans. Learning Technologies*, vol.11, pp.415–428, Oct. 2018.

(2023 年 3 月 22 日受付, 7 月 27 日再受付,
11 月 10 日早期公開)



西尾 徹朗

2023 電気通信大学大学院情報理工学研究
科博士前期課程了。現在、エスエイテー
ティー株式会社 駿台人工知能研究所 研究
員。人工知能の教育応用研究に従事。



堤 瑛美子

2020 電気通信大学大学院情報理工学研究
科博士前期課程了。2023 電気通信大学大学
院情報理工学研究科博士後期課程了。博士
(工学)。2023 より東京大学大学院情報理工
学系研究科特任助教。現在に至る。



植野 真臣 (正員)

1992 神戸大学大学院教育学研究科了。
1994 東京工業大学大学院総合理工学研究科
了。博士(工学)。東京工業大学、千葉大
学、長岡技術科学大学を経て 2006 より電
気通信大学助教授、2013 より教授。現在に
至る。