

## 修士論文の和文要旨

研究科・専攻	大学院 情報理工学研究科 情報・ネットワーク工学専攻 博士前期課程		
氏名	佐々木極昇	学籍番号	2431076
論文題目	Adaptive Learning におけるヒント付き問題項目の Deep-Neural-Network 反応予測		
要旨	<p>古くから学習工学分野では、効果的な学習を提供するために、学習者の能力に適した支援を行う Adaptive Learning が研究されてきた。様々な Adaptive Learning システムが提案される一方で、これらのシステム過剰な支援は課題の困難度を低下させ、自律的な問題解決を阻害しうる問題点が存在した。最適な支援を提示するために、項目反応理論 (Item Response Theory; IRT) に基づき、支援後の学習者反応を予測して学習効率が最大となる支援を決定するシステムの有効性が報告されている。また、予測正答確率が 0.5 に最も近くなる支援の提示が学習効果に資することが示されている。しかし既存システムでは、能力の時系列変化や複数スキル間の関係性を十分に表現できず、反応予測精度が低下する可能性がある。さらに、課題反応の局所独立性を仮定するため、課題間に局所依存性が存在する場合にパラメータ推定精度が低下する。そこで本研究では、局所独立性を仮定せず、多次元スキルと能力の時系列変化を考慮した能力値推定が可能な Student network に、課題の難易度を推定する Item network と各支援提示後の難易度を推定する Hint network を統合し、ヒント提示後の正答確率を予測する Deep-Hint-IRT を提案する。提案手法は、課題と潜在スキルの関連を深層学習により表現し、課題解決までに使用したヒント情報を能力更新に反映させることで、多次元能力値とヒントによる困難度変化を同時に推定する。実データおよびシミュレーションデータを用いた比較実験により、正答に必要な最小限のヒント段階の予測において、提案手法が既存手法に比べて反応予測精度を向上させることを示した。</p>		

2025 年度 修士論文

Adaptive Learning におけるヒント付き問題項目の  
Deep-Neural-Network 反応予測

電気通信大学大学院情報理工学研究科  
情報・ネットワーク工学専攻 情報数理工学プログラム

学籍番号 2431076

佐々木 極昇

主任指導教員：植野 真臣

指導教員：宇都 雅輝

提出年月日：2026 年 1 月 26 日

# 目次

<b>1</b>	<b>まえがき</b>	<b>2</b>
<b>2</b>	<b>確率的足場かけシステム</b>	<b>3</b>
2.1	ヒント提示システム . . . . .	3
2.2	ヒント段階反応モデル . . . . .	4
2.3	最適ヒント提示システム . . . . .	5
2.4	隠れマルコフIRTモデル . . . . .	5
<b>3</b>	<b>Deep-Hint-IRT</b>	<b>7</b>
3.1	パラメータ推定 . . . . .	8
3.2	能力値パラメータ更新 . . . . .	9
<b>4</b>	<b>評価実験</b>	<b>10</b>
4.1	データセット . . . . .	10
4.2	正答に必要な最小限のヒント数予測精度 . . . . .	11
4.3	局所独立性検定 . . . . .	12
4.4	学習効果を最大化するヒント予測実験 . . . . .	15
<b>5</b>	<b>まとめ</b>	<b>17</b>

# 1 まえがき

教育における難しさの一つに、教師が学習者に教えすぎても教えなさすぎても、学習者の発達を十分に促すことができないという問題がある [1, 2]. Vygotsky[3, 4] の提唱した「最近接発達領域」(Zone of Proximal Development: ZPD) では、学習者が自力で解決できない課題でも、教師の最適な支援を受けて取り組むことで学習者の能力が成長すると考えられている. Brown and Ferrara[5], Campione[6] は、あらゆる学習課題において、学習者が課題を解決するために必要な最小限の支援を提示するときに、最大の学習効果が得られると報告している.

近年、教育現場では、学習者の特性に合わせて適応的な学習支援を提供する ITS(Intelligent Tutoring System) が多く利用されている [7–17]. しかし、従来の ITS における学習支援システムは、学習者に提示するヒントの量が最適化されておらず、過剰な支援を提供するために課題の困難度が低下し、効果的な学習を妨げる可能性があることが指摘されている [18]. この原因として Ueno and Miyazawa[1, 2] は、従来の学習支援システムが (1) 学習者の課題達成の程度を予測しておらず、(2) 能力から最適な支援を選択する戦略を持っていない点を挙げている.

この問題を解決するために、植野・松尾 [19], Ueno and Miyazawa[1, 2] は、項目反応理論 (IRT; Item Response Theory) を用いて学習者の学習履歴データから課題への正答確率を予測し、適応的にヒントを提示する適応的ヒントシステムを開発した. Ueno and Miyazawa[1, 2] は、ヒント提示時に学習者の学習効率が最適になるような正答確率  $p$  が存在し、学習者が独力では解けない課題を提示し、正答確率が  $p$  になるようなヒントを提示することで学習効率が向上すると仮定した. 彼らは膨大な実験により、予測正答確率が  $p = 0.5$  に最も近づく量のヒントを提示することで最大の学習効果を得ることを示した. Ueno and Miyazawa[1, 2] が提案したシステムについて、Thomann and Deutscher[20] が 68 個の適応的ヒントシステムと比較し、トップクラスの学習成果を達成できることが報告されている.

しかし、Ueno and Miyazawa[1, 2] が用いた IRT モデルは、能力値の時系列変化を考慮しておらず、長期の学習過程に対して予測正答確率の誤差が増加する問題があった. そこで堤ら [21, 22] は、学習過程における能力値の時系列変化を考慮するために、学習者の能力が時点ごとに変動するプロセスを IRT に組み込んだ隠れマルコフ IRT モデル (HMIRT; Hidden Markov IRT) を提案した. HMIRT はある時点での能力値が影響する時間 (課題数) を表すウィンドウサイズと学習者の能力の変動の程度を反映する変動パラメータをもつ. これらの最適値をデータから推定することで、ヒント提示後の学習者の正答確率の予測精度が向上することを示し、大学における離散数学の授業で HMIRT を用いた適応的ヒントシステムを実際に利用し、その有効性を示している [22]. その他にも学習者の能力値時系列変化を追跡するために、潜在能力値が隠れマルコフ過程に従う確率モデルが複数提案されている [23–25]. しかし、これらの確率モデルには以下の問題がある.

1. 複数のスキル間の関係性を考慮した能力値推定や困難度推定ができない。
2. 課題の局所独立性を仮定しているため、課題間に局所依存性が存在する場合にパラメータ推定精度が下がる [26, 27].

近年では、課題の局所独立性を仮定せず、複数の潜在スキルを考慮して学習者の課題への反応予測を行う深層学習手法が複数提案されている [28–31]. これらの深層学習手法は確率モデルと比較して高精度に学習者の反応予測を行うことが報告されているが、IRT のような解釈可能な能力値パラメータ・困難度パラメータを持たないため、教育への応用には限界があった。この問題を解決するために、堤ら [32–36] は学習者の能力値パラメータと困難度パラメータを独立のニューラルネットワークを用いて推定することにより、IRT モデルと同等の解釈性をもつ新たな深層学習手法、Deep-IRT を提案した。しかし、堤らの Deep-IRT はヒントを含まない課題に対する学習者の反応予測（正答・誤答予測）を目的としているため、段階的なヒントを与えた時の課題に対する正答確率を予測できない。

本研究では、ヒント提示後の学習者の正答確率を予測するために、堤らの Deep-IRT に Hint network を追加した Deep-Hint-IRT を提案する。提案手法はヒント提示後の課題の困難度と学習者の時系列変化する多次元の能力値を推定し、ヒント提示後の課題への正答確率を予測することができる。実データとシミュレーションデータを用いた数値実験で、提案手法と既存手法を用いて学習者が使用したヒントに対する予測精度の比較を行い、提案手法の有効性を示す。

## 2 確率的足場かけシステム

### 2.1 ヒント提示システム

Ueno and Miyazawa [1, 2] は、個々の学習者に与える最適な支援の量を測定するために、学習者が課題に誤答した際に段階的なヒントを提示するヒント提示システムを開発した。彼らのヒント提示システムでは、はじめに学習者が課題科目の基礎知識を学習した後、いくつかの課題に解答する。学習者が課題に誤答した場合、システムは図 1 のようにプログラミングの文法の説明やコードの意味など、課題を解くためのヒントを段階的に提示する。

次に、Ueno and Miyazawa [1, 2] のヒント提示システムで得られるデータについて説明する。システムには、課題  $j = \{1, 2, \dots, J\}$  に対し  $K^j$  個の段階ヒント  $k = \{1, 2, \dots, K^j\}$  が用意されている。初めはヒントを表示しない状態で学習者  $i = \{1, 2, \dots, I\}$  に課題  $j$  を提示する。学習者が課題  $j$  に誤答した場合はヒント  $k = 1$  を提示し、更に誤答するごとにヒント  $k = K^j$  までを順次提示する。学習者が課題に正答するか最後のヒント  $k = K^j$  提示後に誤答した場合は、課題の解答と解説を提示し、その後課題  $j + 1$  を提示する。したがって、学

Programming Test

**Programming Code**

Answer the final values of **a, b, and c** after the program runs.

```

1 public class Question1_2 {
2     public static void main(String args[]){
3         int a = 2;
4         int b = 4;
5         int c = 6;
6
7         c = a + a * b;
8         b = c / a;
9         a += b;
10        a++;
11    }
12 }

```

**Hint2:** c = 2 + 2 \* 4; c = 2 + 8; c = 10;

**Hint3:** b = 6 / 2; b = 3;

**Hint4:** a = a + b; a = 2 + 4; a = 6;

**Answer**

a	1 ▼	b	1 ▼	c	1 ▼
---	-----	---	-----	---	-----

**Hint1**

Variable
= Assignment Statement
Operator

++ Increment
+= Addition assignment operator

++ Increment

**++ : Increment**  
a ++; ⇒ a = a + 1;

Code

```

a = a + 1;
a = 5 + 1;
a = 6;

```

[- top](#)

図1 ヒント提示システム

習者  $i$  が課題  $j$  に対する反応データ  $u^{(i,j)}$  は以下のように表される。

$$u^{(i,j)} = \begin{cases} k & \text{ヒント } k \text{ を提示して正答} \\ K^j + 1 & \text{最後のヒント } K^j \text{ を提示後に誤答} \\ 0 & \text{ヒントなしで正答} \end{cases} \quad (1)$$

## 2.2 ヒント段階反応モデル

学習者に課題提示を行い、ヒントを提示した後の正答確率を評価するために、Ueno and Miyazawa[1, 2] は、IRT の段階反応モデル [37] を用いて学習者の能力値評価と各ヒント提示後の正答確率予測を行った。ヒント段階反応モデルは学習者  $i$  が課題  $j$  にヒント  $k$  で正答する確率  $P(u^{(i,j)} = k)$  を次式で求める。

$$P(u^{(i,j)} = k) = \begin{cases} p^{(i,j,0)} & k = 0 \\ 1 - p^{(i,j,K^j)} & k = K^j + 1 \\ p^{(i,j,k)} - p^{(i,j,k-1)} & \text{otherwise} \end{cases} \quad (2)$$

$$p^{(i,j,k)} = \frac{1}{1 + \exp(-\alpha^j (\theta^i - \beta^{(j,k)}))} \quad (3)$$

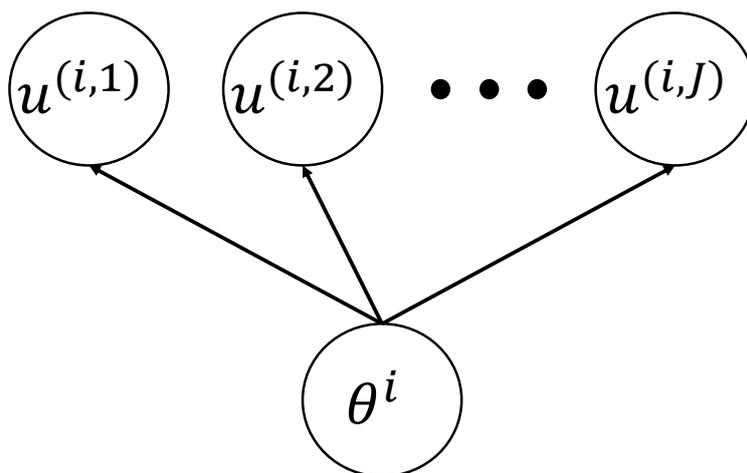


図2 ヒント段階反応モデル

ここで、 $\alpha^j$  は課題  $j$  の識別力パラメータ、 $\beta^{(j,k)}$  は  $k=0$  のとき課題  $j$  でヒントを提示しない場合の困難度、 $1 \leq k \leq K^j$  のとき課題  $j$  でヒント 1 からヒント  $k$  までを提示した場合の困難度、 $\theta^i$  は学習者  $i$  の能力パラメータをそれぞれ表す。ただし、 $\beta^{(j,0)} > \beta^{(j,k)} > \beta^{(j,K^j)}$  である。

## 2.3 最適ヒント提示システム

推定した能力から最適な支援を選択する手法として、Ueno and Miyazawa[1,2] はヒント提示後に学習者の学習効率が最大化する正答確率  $p$  が存在すると仮定し、学習者が課題に誤答した際に、正答確率が  $p$  に最も近いヒントを提示する方法を提案した。そして、最適な正答確率  $p$  を明らかにするために、ヒント提示後の正答確率が異なる複数の学習者グループで実験を行った結果、予測正答確率が 0.5 に近いヒントを提示する支援が最も学習効率が高いことを示した。

以上より、最適なヒント提示するためには正答確率を正確に予測することが重要である。しかし、Ueno and Miyazawa[1,2] で用いられたヒント段階反応モデル [37] は、学習者の真の能力値が学習過程で固定値であることが仮定されている。したがって、能力値の時系列変化を考慮できず、長期の学習過程に対して予測正答確率の誤差が増加する問題があった。

## 2.4 隠れマルコフ IRT モデル

Ueno and Miyazawa[1,2] のヒント段階反応モデルの問題を解決するために、堤ら [21,22] は学習者の能力が時点ごとに変化するプロセスを IRT に組み込んだ隠れマルコフ IRT モデル (HMIRT) を提案した。HMIRT のグラフィカルモデルを図 3 に示す。HMIRT では学習者  $i$  の時点  $t = \{1, 2, \dots, T\}$  における能力値  $\theta^{(i,t)}$  が時点  $t-1$  での能力値  $\theta^{(i,t-1)}$  に依存

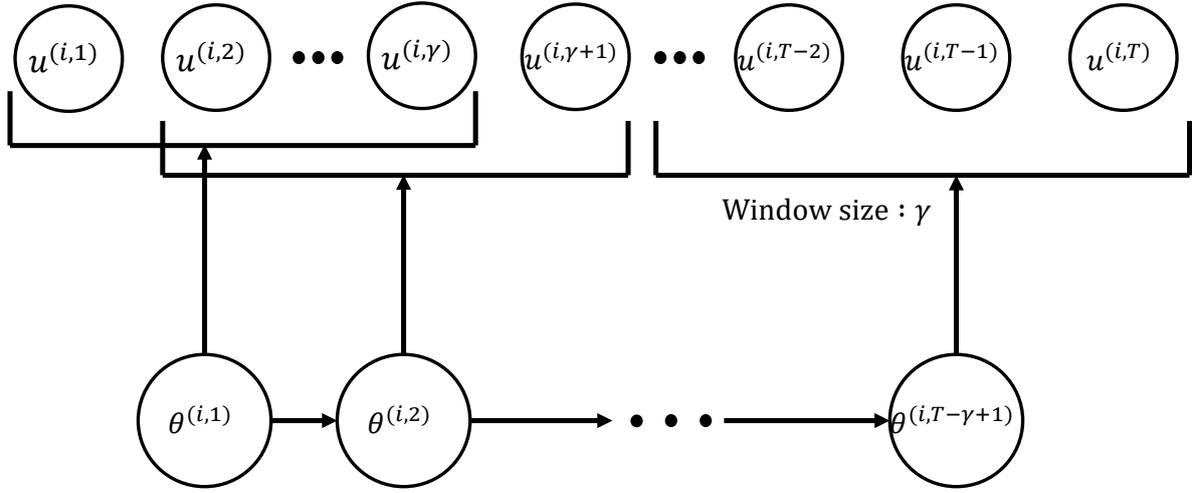


図3 隠れマルコフ IRT モデル

して変動する [21, 22]. また, 能力値  $\theta^{(i,t)}$  の変動モデルにスライディングウィンドウ方式 [38, 39] を用いている. Sliding Window 方式は画像処理や音声処理, 通信工学などの分野で用いられる手法であり, 指定した長さの Window が学習者の学習履歴データ上を移動し, Window 内の学習履歴データのみを用いて能力値推定を行う [21, 40]. これにより, 学習者が解答した直近の課題数 (ウィンドウサイズ)  $\gamma$  個以前の反応データを忘却した能力値推定が可能となる.

HMIRT は時点  $t$  において学習者  $i$  が課題  $j$  にヒント  $k$  で正答する確率  $P(u^{(i,t)} = k)$  を次式で求める.

$$P(u^{(i,t)} = k) = \begin{cases} p^{(i,0,t)} & k = 0 \\ 1 - p_t^{(i,K^j,t)} & k = K^j + 1 \\ p^{(i,k,t)} - p^{(i,k-1,t)} & otherwise \end{cases} \quad (4)$$

$$p^{(i,k,t)} = \frac{1}{1 + \exp(-\alpha^j (\theta^{(i,t)} - \beta^{(j,k)}))} \quad (5)$$

ここで,  $\theta^{(i,t)}$  は時点  $t$  での学習者  $i$  の能力パラメータをそれぞれ表す. ただし,

$$\theta^{(i,t)} \sim N(\theta^{(i,t-1)}, \delta) \quad (6)$$

$$\theta^{(i,t)} \sim N(0, 1) \quad (7)$$

ここで,  $N(\mu, \sigma)$  は平均  $\mu$ , 標準偏差  $\sigma$  の正規分布を表し,  $\delta$  は能力値の変動パラメータである.

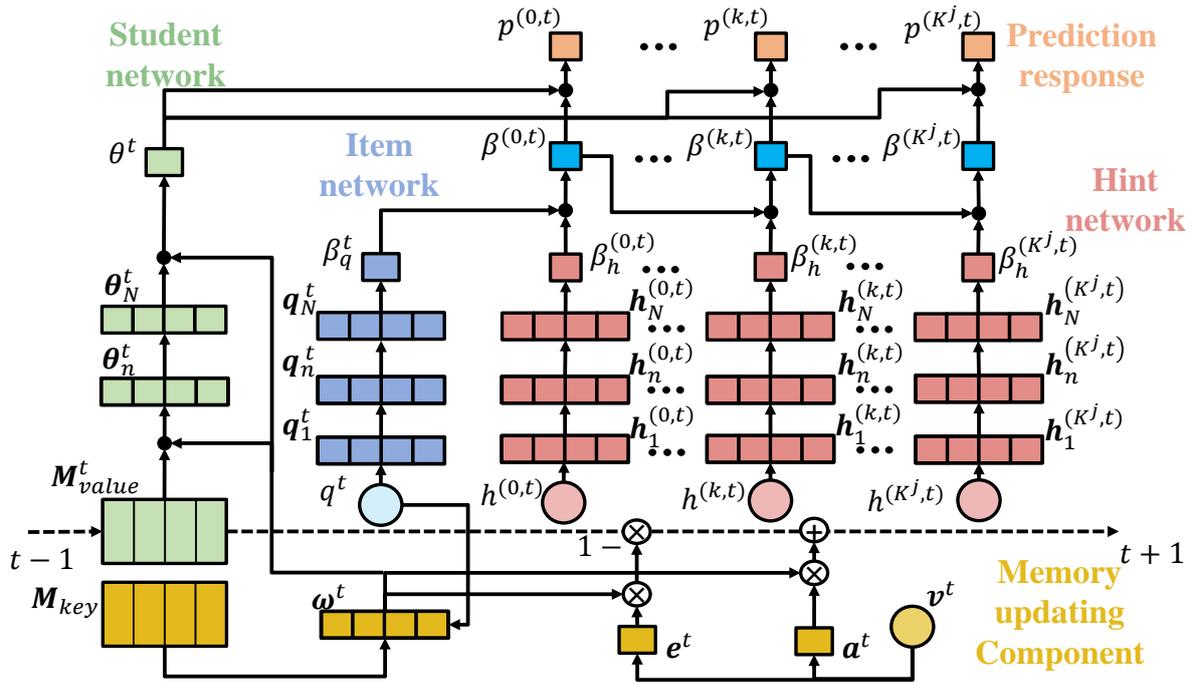


図4 提案手法の構造

堤ら [21,22] は HMIRT を用いて学習者のヒント提示後の反応予測を行い、従来のヒント段階反応モデルの予測精度を大きく改善することを示した。Ueno, Tsutsumi, Miyazawa [22] は、大学における離散数学の授業で HMIRT を用いた適応的ヒントシステムを実際に利用し、その有効性を示している。しかし、堤らの HMIRT や従来の項目反応モデルは複数のスキル間の関係性を考慮して能力値推定や困難度推定ができないために、正答率の予測精度が低下している可能性がある。また、IRT の確率モデルは課題の局所独立性を仮定しているが、Yen [26,27] では、課題間に局所依存性が存在する場合、パラメータの推定精度が低下することが理論、実験の双方で示されている。

### 3 Deep-Hint-IRT

近年、学習者の解答履歴データから、学習者の習熟度と未知の課題に対する学習者の反応を予測する knowledge Tracing が盛んに研究されている [23, 28–33, 41–47]。その中でも深層学習を用いた手法は、学習者の時系列変化する多次元の潜在能力推定と複数のスキル・課題間の関係性を考慮した反応予測を行うことで、確率モデルを用いた手法より高精度な予測精度を示すことが知られている。最新の研究において堤らは、高い反応予測精度とパラメータの教育的解釈性を実現するために、深層学習手法と IRT を組み合わせた Deep-IRT を提案した [32–36]。堤らの Deep-IRT は時点ごとの学習者の能力値を推定する Student network と課題の困難度を推定する Item network の2つの独立したニューラルネットワークで構成することで、IRT と同等のパラメータ解釈性を実現した。しかし、堤らの Deep-IRT はヒ

ントを含まない課題に対する学習者の反応予測（正答・誤答予測）のみを目的としているため、ヒントを提示した際の正答確率予測には適応できない。

本研究では、学習効果を最大化する足場掛けを実現するために、堤らの Deep-IRT[32–36] に Hint network を追加し、ヒント提示後の課題への正答確率を予測する新たな Deep-IRT を提案する。さらに、提案手法は課題間の局所独立性を仮定せず、学習者が課題解決までに用いたヒントを考慮した多次元の能力値更新を行う。

### 3.1 パラメータ推定

提案手法の構造を図 4 に示す。提案手法は Student network, Item network, Hint network の 3 つの独立なニューラルネットワークから構成される。また、提案手法では既存手法と同様に  $L$  個の潜在スキルを仮定しており、各課題と潜在スキルとの関係を key memory  $\mathbf{M}_{key} \in \mathbb{R}^{L \times d_{key}}$  へ保存し、学習者  $i$  の時点  $t$  の各潜在スキルに対する能力値を value memory  $\mathbf{M}_{value}^t \in \mathbb{R}^{L \times d_{value}}$  に保存する [32–36]。ここで、 $d_{key}, d_{value}$  はチューニングパラメータである。

Student network では、既存手法と同様に  $N$  層のニューラルネットワークを用いて  $\mathbf{M}_{value}^t$  から時点  $t$  で課題  $j$  に解答する際の能力値  $\theta^t$  を推定する [32–36]。

$$\mathbf{w}^t = \text{softmax}(\mathbf{M}_{key} \mathbf{q}^t) \quad (8)$$

$$\boldsymbol{\theta}_1^t = \sum_{l=1}^L \mathbf{M}_{value}^{t,l \top} \quad (9)$$

$$\boldsymbol{\theta}_n^t = \tanh(\mathbf{W}^{\theta_n} \boldsymbol{\theta}_{n-1}^t + \boldsymbol{\tau}^{\theta_n}) \quad (10)$$

$$\theta^t = \mathbf{w}^{t \top} \boldsymbol{\theta}_N^t \quad (11)$$

ここで、 $\mathbf{w}^t$  は課題のスキルと潜在スキルの関係性の強さを表す attention である。Item network では、既存手法と同様に  $N$  層のニューラルネットワークを用いて、課題  $j$  の課題 ID から計算された埋め込みベクトル  $\mathbf{q}^t \in \mathbb{R}^{d_{key}}$  を入力とし、課題  $j$  の困難度  $\beta_{item}^j$  を推定する [32–36]。

$$\mathbf{q}_1^t = \tanh(\mathbf{W}^{q_1} \mathbf{q}^t + \boldsymbol{\tau}^{q_1}) \quad (12)$$

$$\mathbf{q}_n^t = \tanh(\mathbf{W}^{q_n} \mathbf{q}_{n-1}^t + \boldsymbol{\tau}^{q_n}) \quad (13)$$

$$\beta_q^t = \mathbf{W}^{\beta_q} \mathbf{q}_N^t + \boldsymbol{\tau}^{\beta_q} \quad (14)$$

Hint network では、課題  $j$  が持つヒント  $k$  から埋め込み処理を行った  $h^{(k,t)} \in \mathbb{R}^{d_{key}}$  を入力とし、段階ヒントを提示した際の困難度の差分  $\beta_h^{(k,t)}$  を推定する。

$$\mathbf{h}_1^{(k,t)} = \tanh(\mathbf{W}^{h_1} h^{(k,t)} + \boldsymbol{\tau}^{h_1}) \quad (15)$$

$$\mathbf{h}_n^{(k,t)} = \mathbf{W}^{h_n} \mathbf{h}_{n-1}^{(k,t)} + \tau^{h_n} \quad (16)$$

$$\begin{aligned} \beta_h^{(k,t)} &= \text{softplus} \left( \mathbf{W}^{\beta_h} \mathbf{h}_N^{(k,t)} + \tau^{\beta_h} \right) \\ &= \log \left( \exp \left( \mathbf{W}^{\beta_h} \mathbf{h}_N^{(k,t)} + \tau^{\beta_h} \right) + 1 \right) \end{aligned} \quad (17)$$

したがって、課題  $j$  のヒント提示後の困難度  $\beta^{(k,t)}$  は次式で求める。

$$\beta^{(k,t)} = \begin{cases} \beta_q^t - \beta_h^{(k,t)} & k = 0 \\ \beta^{(k-1,t)} - \beta_h^{(k,t)} & 1 \leq k \end{cases} \quad (18)$$

各ネットワークの出力から学習者  $i$  が時点  $t$  で課題  $j$  にヒント  $k$  までを用いて正答する確率は以下で求める。

$$p^{(k,t)} = \sigma \left( \theta^t - \beta^{(k,t)} \right) \quad (19)$$

ここで、 $\sigma$  はシグモイド関数を表す。

## 3.2 能力値パラメータ更新

次に、提案手法における時点ごとの能力値更新の方法について説明する。Deep-IRT[32,33] は、Memory updating component と呼ばれる機構で学習者の最新の反応を用いて、学習者の潜在能力パラメータ  $\mathbf{M}_{value}^t$  を更新する。しかし、能力値更新にヒントの情報を用いないために、学習者が課題解決までに使用したヒントを考慮した能力値を更新ができない。本研究では学習者が使用したヒントの情報を用いて能力値パラメータの更新を行う。始めに、学習者が課題  $j$  に正答するまで使用したヒント数を埋め込み処理したベクトル  $\mathbf{v}^t$  を求める。

$$\eta^{(j,t)} = \begin{cases} 1 & \text{時点 } t \text{ で課題 } j \text{ に解答} \\ 0 & \text{それ以外} \end{cases} \quad (20)$$

$$\phi^{(k,t)} = \begin{cases} 1 & \text{課題に対しヒント } k \text{ で正答} \\ 0 & \text{それ以外} \end{cases} \quad (21)$$

$$\boldsymbol{\eta}^t = [\eta^{(1,t)}, \dots, \eta^{(j,t)}, \dots, \eta^{(J,t)}] \quad (22)$$

$$\boldsymbol{\phi}^t = [\phi^{(0,t)}, \dots, \phi^{(k,t)}, \dots, \phi^{(K,t)}] \quad (23)$$

$$\mathbf{v}^t = \mathbf{W}_\eta \boldsymbol{\eta}^t + \mathbf{W}_\phi \boldsymbol{\phi}^t + \tau_v \quad (24)$$

ここで、 $K = \max\{K^1, \dots, K^j, \dots, K^J\}$  である。続いて、堤ら [32,33] と同様に、 $\mathbf{v}^t$  を用いて value memory  $\mathbf{M}_{value}^t$  を更新する。

$$\mathbf{e}^t = \sigma \left( \mathbf{W}_e \mathbf{v}^t + \tau_e \right) \quad (25)$$

$$\mathbf{a}^t = \tanh \left( \mathbf{W}_a \mathbf{v}^t + \tau_a \right) \quad (26)$$

表1 データセットの詳細

データセット名	学習者数	課題数	欠損率	平均ヒント数	平均各ヒント利用率
プログラミング [21]	75	18	1.19%	9.55	3.42%
物理 [48]	99	64	50.63%	2.18	13.30%
離散数学 [22]	436	123	18.78%	5.07	6.40%

$$\tilde{M}_{value}^{t+1} = M_{value}^t \otimes (1 - \mathbf{w}^t \mathbf{e}^t) \quad (27)$$

$$M_{value}^{t+1} = \tilde{M}_{value}^{t+1} + \mathbf{w}^t \mathbf{a}^{t\top} \quad (28)$$

ここで、 $\mathbf{W}$  は重みパラメータ、 $\boldsymbol{\tau}$  はバイアスパラメータ、 $\otimes$  はアダマール積を表す。また、 $\mathbf{e}^t$  は  $M_{value}^t$  の値をどの程度保存するかを決定し、 $\mathbf{a}^t$  は時点  $t$  の反応データをどの程度更新に反映するかを決定するパラメータである。

## 4 評価実験

### 4.1 データセット

本研究では、プログラミング [21]・物理 [48]・離散数学 [22] の3科目に関する実反応データと、HMIRT[21,22] により生成されたシミュレーションデータを実験に用いた。

実データの詳細を表1に示す。表中の平均各ヒント利用率とは課題が持つ各ヒントを使用した学習者の割合の平均を表し、式(29)で算出した。

$$\begin{aligned} & \text{課題 } j \text{ のヒント利用率} \\ &= \frac{\text{課題 } j \text{ でヒントを使用して正答した解答数}}{\text{課題 } j \text{ の解答数} \times K^j} \end{aligned} \quad (29)$$

シミュレーションデータはHMIRTを用いて学習者数  $I = \{50, 100, 300, 1000, 2000\}$ 、課題数  $J = \{50, 100, 200, 300\}$ 、最大ヒント数  $K^1 = \dots = K^J = 4$  で生成した。また、ウィンドウサイズ  $\gamma = 10$ 、変動パラメータ  $\delta = 0.05$  と設定した。なお、データ生成に使用したHMIRTの事前分布は以下の通りである。

$$\log \alpha^j \sim N(0.0, 0.2) \quad (30)$$

$$\theta^i \sim N(0.0, 1.0) \quad (31)$$

$$\beta^{(j,k)} \sim N(\mu_{(j,k)}, 0.16) \quad (32)$$

$$\begin{aligned} & \{\mu^{(j,0)}, \dots, \mu^{(j,k)}, \dots, \mu^{(j,K^j)}\} \\ &= \left\{ -2, \dots, -2 + \frac{4k}{K^j}, \dots, 2 \right\} \end{aligned} \quad (33)$$

上記のデータセットを10分割交差検証で訓練データ、検証データ、評価データに分割し、訓練データと検証データでモデルのパラメータを調整した後に評価データで予測精度を算出した。

表2 シミュレーションデータを用いた実験における  $N$  の最適値

	学習者:100	学習者:300	学習者:1000	学習者:2000
課題数:50	3	7	7	6
課題数:100	5	3	4	4
課題数:200	6	3	7	4
課題数:300	4	3	7	7

## 4.2 正答に必要な最小限のヒント数予測精度

本節では、提案手法と比較手法 (IRT, HMIRT[21, 22], 1次元能力モデル) の正答確率  $p^{(k,t)}$  の予測精度を評価するために、実データ, シミュレーションデータを用いて実際に学習者が正答までに必要な最小限のヒント数  $\hat{k}$  を予測する. ここで, 1次元能力モデルは提案手法の潜在スキル数  $L = 1$  にしたモデルで, 多次元能力値を用いた反応予測が有効か調べるために実データを用いた実験で提案手法と予測精度を比較する. また,  $\hat{k}$  は以下の式で算出した.

$$\hat{k} = \underset{k}{\operatorname{argmax}} \begin{cases} p^{(0,t)} & k = 0 \\ 1 - p^{(K^j,t)} & k = K^j + 1 \\ p^{(k,t)} - p^{(k-1,t)} & \text{otherwise.} \end{cases} \quad (34)$$

提案手法のネットワークの層数  $N$  は  $\{2, 3, 4, 5, 6, 7\}$  から交差検証で予測精度を最大化する値を選択した. 実データに対する  $N$  の最適値は, プログラミングおよび物理で  $N = 7$ , 離散数学が  $N = 6$  であった. また, シミュレーションデータにおける  $N$  の最適値は表2の通りであった. また, HMIRT モデルの能力値のウィンドウサイズ  $\gamma$  と変動パラメータ  $\delta$  の値は, 実データを用いた実験では  $\{1, 2, \dots, 10\} \times \{0.1, 0.2, 0.3, 0.5\}$  から最適な組み合わせを選択し, シミュレーションデータを用いた実験では真のモデルと同じパラメータを設定した. 各実データにおける HMIRT の  $\gamma, \delta$  の最適値は, プログラミングで  $\gamma = 2, \delta = 0.1$ , プログラミングで  $\gamma = 5, \delta = 0.1$ , プログラミングで  $\gamma = 9, \delta = 0.1$  であった. なお, IRT の段階反応モデルと HMIRT の事前分布は式 (30) ~ (33) の通りであり, それぞれのパラメータは堤ら [21, 22] と同様にマルコフ連鎖モンテカルロ (MCMC) 法を用いた期待事後確率推定で推定した. 予測精度評価には, Multi-AUC スコア (macro-AUC, micro-AUC), 2次重み付きカッパ係数 (quartic weighted kappa: QWK) を用いた. Multi-AUC は ROC 曲線下の面積を表す指標であり, 多次元クラスの予測精度を評価するために用いられる. micro-AUC は全クラスの真陽性率と偽陽性率を用いて算出し, macro-AUC は各クラスの AUC スコアの平均を算出する [49–52]. 堤ら [21, 22] では予測精度評価に予測ヒント数と実際に正答に必要な最小限のヒント数の一致率 (Accuracy) を用いていたが, Ueno and Miyazawa[1, 2, 22] では, 効果的な支援を行うためには未知の課題に対する正答確率を高精度に予測することが重要であると主張されていることから, 本実験では確率値の精度評価と

して AUC スコアを用いた。また、QWK は自動論文採点といった離散の順序データに対する精度評価に使用される指標であり、予測と正解の差が大きいほどペナルティを課す特徴がある [53–55]。

実データを用いた正答に必要な最小限のヒント数予測実験の結果を表 3 に示す。表 3 から提案手法が最も高い精度を示した。既存手法の IRT は時系列と潜在多次元スキルを考慮した能力値推定ができず、HMIRT と 1 次元能力値モデルは共に潜在多次元スキルを考慮した能力値推定ができないため、予測精度が低下していると考えられる。一方で、提案手法は課題とスキルの関係性を推定し、多次元スキルに対する能力値を推定しているため、既存手法よりも柔軟な能力値表現と正答確率予測が可能である。

シミュレーションデータを用いた正答に必要な最小限のヒント数予測実験の実験結果を表 4 に示す。表 4 から提案手法は学習者数や課題数が多くなるほど既存手法よりも高い精度を示すことがわかった。IRT は時系列変化を考慮した能力値推定が出来ないため、予測精度が低下したと考えられる。また、HMIRT は学習者数、課題数が多くなるほど推定するパラメータ数が増加するため、大規模なデータセットではパラメータ推定が十分に収束せず、予測精度が低下したと考えられる。一方、提案手法は学習者数が増加してもモデル内のパラメータが増加しないため、既存手法よりもパラメータ数増加の影響を受けにくいと考えられる。加えて、深層学習手法はデータ量が増加するほど予測精度が向上することが知られているため、大規模な学習支援システムでの運用においては提案手法の方が既存手法よりも適性があるといえる。

### 4.3 局所独立性検定

本節では、IRT および HMIRT が仮定する課題への反応の局所独立性が反応予測精度に影響を与えているかを検証するために、実データセットを用いて課題間に局所依存性の有無を判定する。具体的には、各実データセットの全ての課題ペアに対して、Yen の  $Q_3$  統計量 [26, 27] を用いた局所独立性検定を実施する。

表 3 正答に必要な最小限のヒント数予測実験の結果

データセット名	評価手法	IRT	HMIRT	1 次元能力モデル	提案手法
プログラミン	macro-AUC	0.497 ± 0.006	0.465 ± 0.016	0.668 ± 0.007	<b>0.690 ± 0.005</b>
	micro-AUC	0.838 ± 0.001	0.852 ± 0.002	0.880 ± 0.002	<b>0.885 ± 0.001</b>
	QWK	<b>0.269 ± 0.014</b>	0.163 ± 0.026	0.190 ± 0.026	0.186 ± 0.013
物理	macro-AUC	0.518 ± 0.006	0.507 ± 0.003	0.510 ± 0.002	<b>0.523 ± 0.008</b>
	micro-AUC	0.643 ± 0.001	<b>0.755 ± 0.003</b>	0.745 ± 0.003	0.748 ± 0.005
	QWK	0.174 ± 0.002	0.399 ± 0.013	0.411 ± 0.008	<b>0.433 ± 0.009</b>
離散数学	macro-AUC	0.585 ± 0.002	0.685 ± 0.002	<b>0.725 ± 0.001</b>	<b>0.725 ± 0.001</b>
	micro-AUC	0.862 ± 0.000	0.903 ± 0.000	<b>0.904 ± 0.000</b>	0.903 ± 0.000
	QWK	0.458 ± 0.001	0.565 ± 0.004	0.600 ± 0.002	<b>0.601 ± 0.003</b>
Average	macro-AUC	0.533	0.552	0.634	<b>0.646</b>
	micro-AUC	0.781	0.837	0.843	<b>0.845</b>
	QWK	0.300	0.376	0.400	<b>0.407</b>

表4 シミュレーションデータを用いたヒント数予測実験の結果

	学習者数:50			学習者数:100			学習者数:300			
	評価手法	IRT	HMIRT	提案手法	IRT	HMIRT	提案手法	IRT	HMIRT	提案手法
課題数:50	macro-AUC	0.526	0.555	<b>0.605</b>	0.530	0.581	<b>0.625</b>	0.557	0.628	<b>0.630</b>
	micro-AUC	0.615	<b>0.682</b>	0.644	0.636	<b>0.705</b>	0.657	0.642	<b>0.715</b>	0.663
	QWK	0.196	<b>0.330</b>	0.298	0.193	<b>0.329</b>	0.325	0.219	<b>0.345</b>	0.334
課題数:100	macro-AUC	0.520	0.503	<b>0.575</b>	0.542	0.557	<b>0.633</b>	0.549	0.583	<b>0.639</b>
	micro-AUC	0.610	<b>0.659</b>	0.607	0.629	<b>0.679</b>	0.662	0.637	<b>0.688</b>	<b>0.668</b>
	QWK	0.174	<b>0.242</b>	0.178	0.231	0.320	<b>0.384</b>	0.199	0.216	<b>0.343</b>
課題数:200	macro-AUC	0.526	0.375	<b>0.606</b>	0.400	0.429	<b>0.637</b>	0.568	0.490	<b>0.660</b>
	micro-AUC	0.617	0.607	<b>0.629</b>	0.603	0.629	<b>0.663</b>	0.652	0.646	<b>0.693</b>
	QWK	0.194	0.116	<b>0.274</b>	0.046	0.102	<b>0.342</b>	0.251	0.091	<b>0.420</b>
課題数:300	macro-AUC	0.548	0.328	<b>0.652</b>	0.555	0.372	<b>0.659</b>	0.569	0.431	<b>0.667</b>
	micro-AUC	0.634	0.585	<b>0.673</b>	0.643	0.609	<b>0.683</b>	0.650	0.620	<b>0.695</b>
	QWK	0.249	0.030	<b>0.371</b>	0.253	0.026	<b>0.426</b>	0.253	0.036	<b>0.447</b>
学習者数:1000										
課題数:50	評価手法	IRT	HMIRT	提案手法	IRT	HMIRT	提案手法	IRT	HMIRT	提案手法
	macro-AUC	0.566	<b>0.641</b>	0.635	0.564	<b>0.640</b>	0.636	0.564	<b>0.640</b>	0.636
	micro-AUC	0.646	<b>0.717</b>	0.699	0.643	<b>0.712</b>	0.711	0.643	<b>0.712</b>	0.711
課題数:100	macro-AUC	0.567	0.614	<b>0.645</b>	0.565	0.610	<b>0.643</b>	0.565	0.610	<b>0.643</b>
	micro-AUC	0.646	0.683	<b>0.713</b>	0.645	0.678	<b>0.719</b>	0.645	0.678	<b>0.719</b>
	QWK	0.232	0.176	<b>0.398</b>	0.215	0.152	<b>0.373</b>	0.215	0.152	<b>0.373</b>
課題数:200	macro-AUC	0.570	0.525	<b>0.654</b>	0.575	0.531	<b>0.660</b>	0.575	0.531	<b>0.660</b>
	micro-AUC	0.648	0.649	<b>0.725</b>	0.654	0.654	<b>0.733</b>	0.654	0.654	<b>0.733</b>
	QWK	0.242	0.075	<b>0.420</b>	0.245	0.062	<b>0.421</b>	0.245	0.062	<b>0.421</b>
課題数:300	macro-AUC	0.575	0.464	<b>0.661</b>	0.580	0.475	<b>0.669</b>	0.580	0.475	<b>0.669</b>
	micro-AUC	0.654	0.634	<b>0.729</b>	0.656	0.637	<b>0.737</b>	0.656	0.637	<b>0.737</b>
	QWK	0.255	0.035	<b>0.435</b>	0.265	0.043	<b>0.445</b>	0.265	0.043	<b>0.445</b>

表5 実データにおける全課題ペアのうち課題間の局所依存性が認められたペアの割合

	$p - value < 0.05$	$p - value < 0.01$
プログラミング	15.44%	5.15%
物理	6.72%	0.99%
離散数学	16.92%	7.33%

Yen の  $Q_3$  統計量は、2つの課題  $j_1, j_2$  に対する学習者の反応と IRT の反応予測との残差の相関  $Q_3^{(j_1, j_2)}$  を表す指標である。  $Q_3^{(j_1, j_2)} = 0$  であれば課題  $j_1, j_2$  は局所独立であると仮定できるため、この帰無仮説を検定によって棄却することで局所依存性の有無を判定できる。一般的には  $Q_3 > 0.2$  の場合に局所依存性が存在すると判断されることが多い。しかし、Christensen ら [56] は、適切な閾値はデータの規模や性質に応じて異なるため、パラメトリック・ブートストラップ法を用いて学習者の反応データをシミュレーションし、局所独立性が存在しないと仮定した  $Q_3$  の帰無仮説の下で、分布の95パーセントタイル値を閾値として推奨している。本実験では、Christensen ら [56] が提案した手法に従い、Yen の  $Q_3$  統計量を用いた局所独立性検定を行う。実データ的全課題ペアに対する局所独立性検定の結果を表5に示す。表5より、全ての実データセットにおいて局所依存性が認められたペアが存在することが確認された。Yen[26, 27] は課題間に局所依存性が存在する場合、IRT におけるパラメータ推定精度が低下することを指摘している。したがって、本実験結果でも IRT および HMIRT のパラメータ推定精度が局所依存性により低下したため、反応予測精度も低下したと考えられる。また、表5より、物理よりもプログラミングおよび離散数学の方が局所依存性が認められた課題ペアの割合が高いことを示した。これは、プログラミング、離散数学では、物理よりも提案手法と IRT・HMIRT との反応予測精度の差が大きかったという結果と一致する。

#### 4.4 学習効果を最大化するヒント予測実験

本節では Ueno and Miyazawa[1, 2] が示した最大の学習効果が得られるヒントの予測精度を評価するために、提案手法と IRT, HMIRT を用いて正答確率  $p = 0.5$  に最も近いヒントの的中率を評価する。実データでは真の正答確率が未知のため、データセットは4.1章で使用したシミュレーションデータのみを用いた。生成したデータセットは4.1章と同様に10分割交差検証でモデルの訓練、パラメータを調整し、評価データで的中率を算出した。また、比較手法の HMIRT のウィンドウサイズ  $\gamma$  と変動パラメータ  $\delta$  は4.1章で生成したモデルのパラメータと同様に設定した。

表 6 正答確率  $p = 0.5$  に最も近いヒントの中率

		学習者数:50			学習者数:100		
モデル名	IRT	HMIRT	提案手法	IRT	HMIRT	提案手法	
課題数:50	0.195 ± 0.017	<b>0.306</b> ± <b>0.010</b>	0.210 ± 0.034	0.191 ± 0.012	<b>0.300</b> ± <b>0.011</b>	0.248 ± 0.013	
課題数:100	0.170 ± 0.018	<b>0.292</b> ± <b>0.008</b>	0.198 ± 0.036	0.207 ± 0.017	<b>0.320</b> ± <b>0.028</b>	0.306 ± 0.051	
課題数:200	0.188 ± 0.019	0.245 ± 0.021	<b>0.255</b> ± <b>0.055</b>	0.276 ± 0.011	0.250 ± 0.017	<b>0.311</b> ± <b>0.023</b>	
課題数:300	0.207 ± 0.022	0.213 ± 0.007	<b>0.280</b> ± <b>0.062</b>	0.217 ± 0.023	0.238 ± 0.008	<b>0.361</b> ± <b>0.038</b>	
		学習者数:300			学習者数:1000		
モデル名	IRT	HMIRT	提案手法	IRT	HMIRT	提案手法	
課題数:50	0.212 ± 0.013	<b>0.342</b> ± <b>0.012</b>	0.340 ± 0.044	0.206 ± 0.005	<b>0.329</b> ± <b>0.018</b>	0.326 ± 0.004	
課題数:100	0.208 ± 0.007	0.306 ± 0.023	<b>0.327</b> ± <b>0.041</b>	0.187 ± 0.009	0.279 ± 0.013	<b>0.319</b> ± <b>0.021</b>	
課題数:200	0.205 ± 0.009	0.253 ± 0.007	<b>0.336</b> ± <b>0.019</b>	0.200 ± 0.008	0.246 ± 0.006	<b>0.336</b> ± <b>0.015</b>	
課題数:300	0.206 ± 0.009	0.236 ± 0.003	<b>0.329</b> ± <b>0.011</b>	0.197 ± 0.004	0.244 ± 0.006	<b>0.342</b> ± <b>0.003</b>	
		学習者数:2000					
モデル名	IRT	HMIRT	提案手法	IRT	HMIRT	提案手法	
課題数:50	0.209 ± 0.006	<b>0.336</b> ± <b>0.007</b>	0.335 ± 0.014				
課題数:100	0.202 ± 0.006	0.285 ± 0.006	<b>0.352</b> ± <b>0.009</b>				
課題数:200	0.205 ± 0.006	0.257 ± 0.012	<b>0.355</b> ± <b>0.005</b>				
課題数:300	0.205 ± 0.003	0.240 ± 0.009	<b>0.353</b> ± <b>0.002</b>				

正答確率  $p = 0.5$  に最も近いヒントの的中率は以下の計算式を用いて算出した。

$$\hat{k}^{(i,t)} = \underset{k}{\operatorname{argmin}} | p^{(i,k,t)} - 0.5 | \quad (35)$$

$$g(i,t) = \begin{cases} 1 & \hat{k}^{(i,t)} = \bar{k}^{(i,t)} \\ 0 & \text{otherwise} \end{cases} \quad (36)$$

$$Acc_{05} = \frac{\sum_{i=1}^I \sum_{t=1}^{T^i} g(i,t)}{\sum_{i=1}^I T^i} \quad (37)$$

ここで、 $T^i$  は学習者  $i$  の課題への回答数、 $\bar{k}^{(i,t)}$  は学習者  $i$  が時点  $t$  で最も真の正答確率が 0.5 に近いヒントを表す。

実験結果を表 6 に示す。表 6 より、提案手法はデータのサイズが大きくなるほど HMIRT よりも高い予測精度を示した。IRT は能力が変化しない仮定を持つため、能力値が時系列変化する HMIRT で生成したデータセットには適切な予測ができなかったと考えられる。また、HMIRT は MCMC 法を用いてパラメータ推定を行っているため、パラメータ数の増加に伴い事後分布の探索が困難になり、収束性が低下したと考えられる。一方、提案手法は学習者数が増加してもパラメータ数が増加しないため、パラメータ推定の収束性や精度を劣化させないと考えられる。

## 5 まとめ

本研究では、学習効果を最大化する足場掛けを実現するために、深層学習を用いて多次元の能力時系列変化を推定し、最適なヒント予測を行う新たなモデルを提案した。提案手法は Student network で推定した学習者の能力値と、Item network, Hint network で推定した各ヒント提示後の課題の困難度から、各ヒント提示後の課題への正答確率を予測する。さらに、課題解決までに使用したヒント数を Memory updating component に入力することで、ヒント利用情報を反映した能力値更新が可能である。提案手法の利点は下記のとおりである。

1. 複数のスキル間の関係性を考慮した能力値推定や困難度推定ができる。
2. 課題の局所独立性を仮定してないため、課題間に局所依存性が存在する場合でもパラメータ推定精度が下がらない。

実データ、シミュレーションデータを用いた実験では下記の結果が得られた。

1. 正答に必要な最小限のヒント数予測実験では、実データおよびシミュレーションデータの双方において、提案手法は既存手法と比較して高い精度を示した。
2. 実データの局所独立性検定により全データセットで局所依存性を持つ課題ペアが確認され、IRT, HMIRT のパラメータ推定に影響が出ている可能性を示した。

3. 学習効果を最大化するヒント予測実験においても提案手法は既存手法と比べて高い精度を示したため、足場掛け支援において高い学習効果が期待できる。

今後は適応的ヒントシステム [1, 2, 22] に導入して学習効率の評価を実施したい。

## 謝辞

本研究は科学研究費課題番号 22K19825, 24H00739, JST CREST JPMJCR21D1 の助成を受けたものです。また、実験データの一部はエスエイティーティー株式会社から提供していただきました。感謝いたします。

## 参考文献

- [1] M. Ueno and Y. Miyazawa, “Probability based scaffolding system with fading,” International Conference on Artificial Intelligence in Education(AIED), pp.237–246, 2015.
- [2] M. Ueno and Y. Miyazawa, “IRT-based adaptive hints to scaffold learning in programming,” IEEE Transactions on Learning Technologies, vol.11, pp.415–428, Oct. 2018.
- [3] L.S. Vygotsky, Thought and Language, MIT press, 1962.
- [4] L.S. Vygotsky, Mind in Society, MIT press, 1978.
- [5] A. Brown and R. Ferrara, “Diagnosing zones of proximal development,” Culture, communication, and cognition: Vygotskian perspectives, pp.273–305, 1985.
- [6] J.C. Campione, “Assisted assessment: A taxonomy of approaches and an outline of strengths and weaknesses,” Journal of Learning Disabilities, vol.22, no.3, pp.151–165, 1989.
- [7] P. Beed, M. Hawkins, and C. Roller, “Moving learners towards independence: The power of scaffolded instruction,” Leading Teacher, vol.44, no.9, pp.648–655, 1991.
- [8] M.F. Graves, B.B. Graves, and S. Braten, “Scaffolded reading experiences for inclusive classes,” Educational Leadership, vol.53, no.5, pp.14–16, 1996.
- [9] J.W. Bonilla and B. Werchadlo, “Teacher and peer roles in scaffolding first graders’ responses to literature,” The Reading Teacher, vol.52, no.6, pp.598–607, 1999.
- [10] P. Coltman, J.S. Anghileri, and D. Petyaeva, “Scaffolding learning through meaningful tasks and adult interaction,” Early Years: An International Journal of Research and Development, vol.22, no.1, pp.39–49, 2002.
- [11] P. Bell and E.A. Davis, “Designing mildred: Scaffolding students’ reflection and argumentation using a cognitive software guide,” International Conference of the

- Learning Sciences Psychology Press, pp.142–149 2000.
- [12] E.A. Davis, “Scaffolding students’ knowledge integration: Prompts for reflection in kie,” *International Journal of Science Education*, vol.22, no.8, pp.819–837, 2000.
- [13] B.J. Reiser, “Scaffolding complex learning: The mechanisms of structuring and problematizing student work,” *Scaffolding*, pp.273–304, Psychology Press, 2004.
- [14] I. Tabak, “Synergy: A complement to emerging patterns of distributed scaffolding,” *The journal of the Learning Sciences*, vol.13, no.3, pp.305–335, 2004.
- [15] C. Quintana, B.J. Reiser, E.A. Davis, J. Krajcik, E. Fretz, R.G. Duncan, E. Kyza, D. Edelson, and E. Soloway, “A scaffolding design framework for software to support science inquiry,” *Scaffolding*, pp.337–386, Psychology Press, 2004.
- [16] B.R. Johnson and K.R. Koedinger, “Designing knowledge scaffolds to support mathematical problem solving,” *Cognition and Instruction*, vol.23, no.3, pp.313–349, 2005.
- [17] N. Yelland and J. Masters, “Rethinking scaffolding in the information age,” *Computers & Education*, vol.48, no.3, pp.362–382, 2007.
- [18] K. VanLehn, L. Collin, S. Kay, S.J. A, S. Robert, T. Linwood, T. Don, W. Anders, and W. Mary, “The andes physics tutoring system: Lessons learned,” *International Journal of Artificial Intelligence in Education*, vol.15, no.3, pp.147–204, 2005.
- [19] 植野真臣, 松尾淳哉, “項目反応理論を用いて適応的ヒントを提示する足場かけシステム,” *電子情報通信学会論文誌 D*, vol.J98-D, pp.17–29, Jan. 2015.
- [20] H. Thomann and V. Deutscher, “Scaffolding through prompts in digital learning: A systematic review and meta-analysis of effectiveness on learning achievement,” *Educational Research Review*, vol.47, p.100686, 2025.
- [21] 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,” *電子情報通信学会論文誌 D*, vol.J102-D, pp.79–92, 02 2019.
- [22] M.Ueno, E. Tsutsumi, and Y. Miyazawa, “Probability-based scaffolding system using sliding hidden markov irt for longitudinal learning,” *International Conference on Artificial Intelligence in Education(AIED)*Springer, pp.448–462 2025.
- [23] 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.
- [24] 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 International Conference on Educational Data Mining*, vol.1, pp.539–544, 06 2016.
- [25] C. Ekanadham and Y. Karklin, “T-SKIRT: Online estimation of student proficiency in an adaptive learning system,” *CoRR*, vol.abs/1702.04282, 2017.

- [26] W.M. Yen, “Effects of local item dependence on the fit and equating performance of the three-parameter logistic model,” *Applied Psychological Measurement*, vol.8, no.2, pp.125–145, 1984.
- [27] W.M. Yen, “Scaling performance assessments: Strategies for managing local item dependence,” *Journal of educational measurement*, vol.30, no.3, pp.187–213, 1993.
- [28] 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, Inc., 2015.
- [29] J. Zhang, X. Shi, I. King, and D.-Y. Yeung, “Dynamic key-value memory network for knowledge tracing,” *Proceedings of the 26th International Conference on World Wide Web*, pp.765–774, WWW ’17, International World Wide Web Conferences Steering Committee, 2017.
- [30] C. Yeung, “Deep-IRT: Make deep learning based knowledge tracing explainable using item response theory,” *Proceedings of the 12th International Conference on Educational Data Mining, EDM*, pp.683–686, 2019.
- [31] Q. Liu, Z. Huang, Y. Yin, E. Chen, H. Xiong, Y. Su, and G. Hu, “EKT: Exercise-aware knowledge tracing for student performance prediction,” *IEEE Transactions on Knowledge and Data Engineering*, vol.33, pp.100–115, 06 2019.
- [32] 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,” *電子情報通信学会論文誌 D*, vol.J104-D, pp.596–608, July 2021.
- [33] E. Tsutsumi, R. Kinoshita, and M. Ueno, “Deep-IRT with independent student and item networks,” *Proceedings of the 14th International Conference on Educational Data Mining (EDM)*, pp.510–517, 2021.
- [34] E. Tsutsumi, Y. Guo., R. Kinoshita, and M. Ueno, “Deep knowledge tracing incorporating a hypernetwork with independent student and item networks,” *IEEE Transactions on Learning Technologies*, vol.17, pp.951–965, 2024.
- [35] 西尾徹朗, 堤瑛美子, 植野真臣, “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,” *電子情報通信学会論文誌 D*, vol.107, no.3, pp.98–110, 2024.
- [36] E. Tsutsumi, T. Nishio, and M.Ueno, “Deep-IRT with a temporal convolutional network for reflecting students’ long-term history of ability data,” *International Conference on Artificial Intelligence in Education(AIED)*Springer, pp.250–264 2024.
- [37] F. Samejima, “Estimation of latent ability using a response pattern of graded scores,” *Psychometrika monograph supplement*, pp.1–168, 1969.
- [38] J.O. Laguna, A.G. Olaya, and D. Borrajo., “A dynamic sliding window approach for

- activity recognition,” *User Modeling, Adaption and Personalization: 19th International Conference, UMAP 2011, Girona, Spain, July 11-15, Proceedings* 19Springer, pp.219–230 2011.
- [39] I. Sebastiano, A. Ferrante, and R. Modugno, “Hmm based handwritten word recognition system by using singularities,” *2009 10th International Conference on Document Analysis and Recognition* IEEE, pp.783–787 2009.
- [40] 堤瑛美子, 植野真臣, “Knowledge tracing のための sliding window 隠れマルコフ IRT,” *電子情報通信学会論文誌 D*, vol.J103-D, pp.894–905, Dec. 2020.
- [41] 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.
- [42] Z. Pardos and N. Heffernan, “T.: Modeling individualization in a bayesian networks implementation of knowledge tracing,” *Proceedings of the 18th International Conference on User Modeling, Adaption, and Personalization*, pp.255–266, 06 2010.
- [43] Z.A. Pardos and N.T. Heffernan, “KT-IDEM: Introducing item difficulty to the knowledge tracing model,” *Proceedings of 19th International Conference on User Modeling, Adaptation and Personalization (UMAP 2011)*, pp.243–254, 01 2011.
- [44] J. Lee and E. Brunskill, “The impact on individualizing student models on necessary practice opportunities,” *Proceedings of the Fifth International Conference on Educational Data Mining*, pp.118–125, 01 2012.
- [45] H. Nakagawa, Y. Iwasawa, and Y. Matsuo, “Graph-based knowledge tracing: Modeling student proficiency using graph neural network,” *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp.156–163, 2019.
- [46] S. Pandey and J. Srivastava, “RKT: Relation-aware self-attention for knowledge tracing,” *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp.1205–1214, 2020.
- [47] A. Ghosh, N. Heffernan, and A.S. Lan, “Context-aware attentive knowledge tracing,” *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp.2330–2339, 2020.
- [48] 学校法人駿河台学園 駿台予備学校, <https://www2.sundai.ac.jp/> 2025. アクセス日: 2025-11-17.
- [49] D.J. Hand and R.J. Till, “A simple generalisation of the area under the roc curve for multiple class classification problems,” *Machine learning*, vol.45, pp.171–186, 2001.
- [50] T. Fawcett, “An introduction to roc analysis,” *Pattern recognition letters*, vol.27, pp.861–874, 2006.

- [51] F. Li and Y. Hong, “Icd coding from clinical text using multi-filter residual convolutional neural network,” proceedings of the AAAI conference on artificial intelligence, vol.34, no.05, pp.8180–8187, 2020.
- [52] M. Niu, J. Wu, Q. Zou, Z. Liu, and L. Xu, “rbpdl: predicting rna-binding proteins using deep learning,” IEEE Journal of Biomedical and Health Informatics, vol.25, pp.3668–3676, 2021.
- [53] J. Cohen, “Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit,” Psychological bulletin, p.213, 1968.
- [54] V.S. Kumar and D. Boulanger, “Automated essay scoring and the deep learning black box: How are rubric scores determined?,” International Journal of Artificial Intelligence in Education 31, pp.538–584, 2021.
- [55] M. Uto, I. Aomi, E. Tsutsumi, and M. Ueno, “Integration of prediction scores from various automated essay scoring models using item response,” IEEE Transactions on Learning Technologies, pp.1–18, 2023.
- [56] K.B. Christensen, G. Makransky, and M. Horton, “Critical values for  $\chi^2$ : Identification of local dependence in the rasch model using residual correlations,” Applied psychological measurement, vol.41, no.3, pp.178–194, 2017.