

深層学習によるテスト理論：Deep Response Model

A test theory based on deep learning: Deep Response Model

木下涼 植野真臣
Ryo Kinoshita Maomi Ueno

電気通信大学情報理工学研究所

Graduate school of Informatics and Engineering, The University of Electro-Communications.

In this article, Deep Response Model(DRM), a novel deep neural networks model for estimating student characteristic parameters and predicting student's response patterns is proposed. It consists of two independent neural networks, Student Layer and Item Layer, combines their outputs and predict student's response. Then, values of the last student layer are considered as student parameters. It can estimate student parameters and predict their responses with high accuracy even when it is difficult to calibrate item parameters.

1. まえがき

近年、Web上でテストを実施するeテストが注目されている。eテストには、適応型テストのような動的テストを実現できるなど、様々な利点があるため、様々なテスト場面において、その導入が進んでいる。

一般的に、eテストの運用では、項目反応理論(IRT)[Lord 68]をもとに、難易度や識別力などの項目パラメータが既知のテスト項目を、アイテムバンクと呼ばれるデータベースに蓄積する必要がある。eテストでは、このアイテムバンクが極めて重要な役割を果たす。

しかし、項目パラメータは、その受験者集団によって尺度が異なるため、アイテムバンクを作成する際には、全ての項目パラメータを同一尺度上に揃える等化が必要となる[Lord 80]。等化を行う際には、各テストの実施前に、受験者あるいはテスト項目の一部が共通するようにテストの設計を行うなどの等化計画を行う必要があり、eテスト導入の妨げとなっている。

また、近年、タブレット教育が学校現場などに取り入れられており、大量の学習データを獲得することが可能となっている。しかし、これらの学習データは非常にスパースであることが多く、既存の手法では、学習支援に活用することが難しい。

一方で、近年の深層学習技術の発展に伴い、教育データ解析において、深層学習の応用が盛んに行われており、高い精度が報告されている[Piech 15]。

これらを踏まえて、本研究では、綿密な等化計画を行っていない場合でも、高精度に能力パラメータを推定し、受験者の反応予測が可能な深層学習モデルであるDeep Response Model(DRM)を提案する。特に適応型テストや学習支援では、学習者の能力推定・解答予測を動的に行う必要があり、高精度に解答を予測する技術が重要であるため、提案手法が活用できる可能性が高い。

提案手法では、受験者とテスト項目を独立したニューラルネットワークの入力とし、それぞれの出力を組み合わせることで受験者の反応を予測する。本研究では、受験者を入力とするニューラルネットワークの最終層の値を能力パラメータとみなす。

提案モデルでは、反応パターンの予測誤差が小さくなるよう

に、一度に共通する重みを更新するため、IRTと異なり、受験者間・テスト項目間に依存関係があることが大きな特徴である。

シミュレーション実験により、提案手法は、等化計画が十分でない場合に、高精度に能力パラメータを推定し、受験者の反応を予測可能であることを示す。

2. 項目反応理論

本章では、まずはじめに、一般的にアイテムバンクの作成に用いられる項目反応理論とその等化手法について述べる。項目反応理論は、様々な評価場面で実用化が進められている数理モデルを用いたテスト理論の一つである[Lord 68]。

一般に項目反応理論は、正誤判定問題や多肢選択式問題など、データが正誤の2値となる反応データに適用されることが一般的である。このような2値データに適用できる項目反応モデルとして、2パラメータロジスティックモデル(2PLM)が古くから広く利用されてきた。

2.1 等化

IRTによって推定されるパラメータは、学習者集団ごとに尺度が異なるため、異なる集団のパラメータを比較する際には、パラメータを同一尺度上に変換する等化処理が必要となる。

IRTに基づいて等化を行う際には、テストを実施する前に、次の等化計画のいずれかを行う必要がある。1) 複数のテストを受験する受験者をテストに含める共通受験者計画。2) 複数のテストに共通するテスト項目を含める共通項目計画。3) 係留テストと呼ばれる共通項目群を用意し、係留テストと各尺度に共通受験者を用意する係留テスト計画。

テスト実施後、得られた反応パターンから、各パラメータを同一尺度上に変換していく。その際に用いられる手法として、等化係数推定法[Loyd 80]、同時尺度推定法[Bock 97]、固定項目パラメータ法などが知られている。

IRTを用いた等化では、パラメータを特定の集団やテスト項目に依存することなく推定できるが、実施には綿密な等化計画と莫大なコストが必要になることが多い。

3. 提案手法

本章では、提案手法について詳しく述べる。提案手法の概要図を図1に示す。提案手法では、受験者 $i \in \{1, \dots, I\}$ の項目 $j \in \{1, \dots, J\}$ に対する反応パターン $u_{ij} \in \{0, \dots, C\}$ を、以下のように予測する。

なお、提案手法は多値反応のテストデータに用いることもできるが、本研究では2値のテストデータを想定する。従って、

連絡先: 木下涼, 電気通信大学情報理工学研究所, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, 182-8585, Japan, kinoshita@ai.is.uec.ac.jp

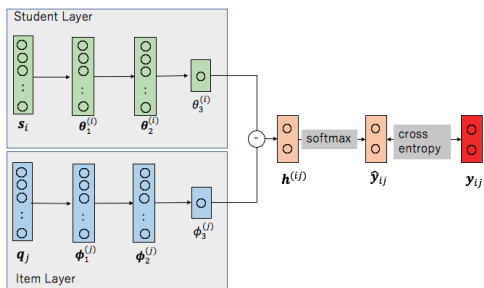


図 1: 提案手法

u_{ij} は、受験者 i が項目 j に正答の場合 1, 誤答の場合 0 をとる。

3.1 モデル定義

提案手法では、受験者の能力パラメータと項目パラメータを分離するため、受験者層 (Student Layer) と項目層 (Item Layer) の二つの独立したニューラルネットワークを組み合わせる。

受験者層では、 i 番目の要素のみが 1, 他の要素が 0 の one-hot vector である受験者ベクトル $\mathbf{s}_i \in \mathbb{R}^I$ を入力として、次のように 3 層のニューラルネットワークを計算する。

$$\theta_1^{(i)} = \tanh(\mathbf{W}^{(\theta_1)} \mathbf{s}_i + \mathbf{b}^{(\theta_1)}) \quad (1)$$

$$\theta_2^{(i)} = \tanh(\mathbf{W}^{(\theta_2)} \theta_1^{(i)} + \mathbf{b}^{(\theta_2)}) \quad (2)$$

$$\theta_3^{(i)} = \mathbf{W}^{(\theta_3)} \theta_2^{(i)} + b^{(\theta_3)} \quad (3)$$

ここで、受験者 i の隠れ層ベクトル $\theta_1^{(i)}, \theta_2^{(i)}$ の次元数は任意の値に設定することができる。

さらに、活性化関数として、以下のハイパボリックタンジエント関数を用いている。

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \quad (4)$$

$\mathbf{W}^{(\theta_1)}, \mathbf{W}^{(\theta_2)}$ は以下のような重みパラメータ行列である。

$$\mathbf{W}^{(\theta_1)} = \begin{pmatrix} w_{11}^{(\theta_1)} & w_{12}^{(\theta_1)} & \cdots & w_{1I}^{(\theta_1)} \\ w_{21}^{(\theta_1)} & w_{22}^{(\theta_1)} & \cdots & w_{2I}^{(\theta_1)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{|\theta_1|1}^{(\theta_1)} & w_{|\theta_1|2}^{(\theta_1)} & \cdots & w_{|\theta_1|I}^{(\theta_1)} \end{pmatrix}$$

$$\mathbf{W}^{(\theta_2)} = \begin{pmatrix} w_{11}^{(\theta_2)} & w_{12}^{(\theta_2)} & \cdots & w_{1|\theta_1|}^{(\theta_2)} \\ w_{21}^{(\theta_2)} & w_{22}^{(\theta_2)} & \cdots & w_{2|\theta_1|}^{(\theta_2)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{|\theta_2|1}^{(\theta_2)} & w_{|\theta_2|2}^{(\theta_2)} & \cdots & w_{|\theta_2||\theta_1|}^{(\theta_2)} \end{pmatrix}$$

また、 $\mathbf{W}^{(\theta_3)} = (w_1^{(\theta_3)}, w_2^{(\theta_3)}, \dots, w_{|\theta_2|}^{(\theta_3)})$ は重みパラメータベクトルである。

$\mathbf{b}^{(\theta_1)} = (b_1^{(\theta_1)}, b_2^{(\theta_1)} \dots b_{|\theta_1|}^{(\theta_1)})$, $\mathbf{b}^{(\theta_2)} = (b_1^{(\theta_2)}, b_2^{(\theta_2)} \dots b_{|\theta_2|}^{(\theta_2)})$ はバイアスパラメータベクトル, $b^{(\theta_3)}$ はバイアスパラメータスカラーである。これらのパラメータは全て、反応パターンから学習される。

本研究では、受験者層の最後の隠れ層である $\theta_3^{(i)}$ を受験者 i の能力パラメータとみなす。

同様に、項目層では、 j 番目の要素のみが 1, 他の要素は 0 の one-hot vector である項目ベクトル $\mathbf{q}_j \in \mathbb{R}^J$ を入力として、次のように 3 層のニューラルネットワークを計算する。

$$\phi_1^{(j)} = \tanh(\mathbf{W}^{(\phi_1)} \mathbf{q}_j + \mathbf{b}^{(\phi_1)}) \quad (5)$$

$$\phi_2^{(j)} = \tanh(\mathbf{W}^{(\phi_2)} \phi_1^{(j)} + \mathbf{b}^{(\phi_2)}) \quad (6)$$

$$\phi_3^{(j)} = \mathbf{W}^{(\phi_3)} \phi_2^{(j)} + b^{(\phi_3)} \quad (7)$$

ここで、項目 j の隠れ層ベクトル $\phi_1^{(j)}, \phi_2^{(j)}$ の次元数は任意の値に設定することができる。

$\mathbf{W}^{(\phi_1)}, \mathbf{W}^{(\phi_2)}$ は、以下のような重みパラメータ行列である。

$$\mathbf{W}^{(\phi_1)} = \begin{pmatrix} w_{11}^{(\phi_1)} & w_{12}^{(\phi_1)} & \cdots & w_{1J}^{(\phi_1)} \\ w_{21}^{(\phi_1)} & w_{22}^{(\phi_1)} & \cdots & w_{2J}^{(\phi_1)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{|\phi_1|1}^{(\phi_1)} & w_{|\phi_1|2}^{(\phi_1)} & \cdots & w_{|\phi_1|J}^{(\phi_1)} \end{pmatrix}$$

$$\mathbf{W}^{(\phi_2)} = \begin{pmatrix} w_{11}^{(\phi_2)} & w_{12}^{(\phi_2)} & \cdots & w_{1|\phi_1|}^{(\phi_2)} \\ w_{21}^{(\phi_2)} & w_{22}^{(\phi_2)} & \cdots & w_{2|\phi_1|}^{(\phi_2)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{|\phi_2|1}^{(\phi_2)} & w_{|\phi_2|2}^{(\phi_2)} & \cdots & w_{|\phi_2||\phi_1|}^{(\phi_2)} \end{pmatrix}$$

また、 $\mathbf{W}^{(\phi_3)} = (w_1^{(\phi_3)}, w_2^{(\phi_3)}, \dots, w_{|\phi_2|}^{(\phi_3)})$ は、重みパラメータベクトルである。

$\mathbf{b}^{(\phi_1)} = (b_1^{(\phi_1)}, b_2^{(\phi_1)} \dots b_{|\phi_1|}^{(\phi_1)})$, $\mathbf{b}^{(\phi_2)} = (b_1^{(\phi_2)}, b_2^{(\phi_2)} \dots b_{|\phi_2|}^{(\phi_2)})$ はバイアスパラメータベクトル, $b^{(\phi_3)}$ はバイアスパラメータスカラーである。これらのパラメータは全て、反応パターンから学習される。

本研究では、項目層の最後の隠れ層である $\phi_3^{(j)}$ を項目 j の項目パラメータとみなす。

その後、受験者パラメータと項目パラメータの差を用いて、反応パターンの予測を行う。具体的には、以下のように、反応パターン u_{ij} に対する反応確率 $\hat{y}_{i,j} = (\hat{y}_{i,j}^{(0)}, \hat{y}_{i,j}^{(1)})$ を算出し、モデルの出力とする。

$$\mathbf{h}^{(i,j)} = (\mathbf{W}^{(y)})^T (\theta_3^{(i)} - \phi_3^{(j)}) + \mathbf{b}^{(y)} = (h_0^{(i,j)}, h_1^{(i,j)}) \quad (8)$$

$$\hat{y}_{i,j}^{(c)} = \text{softmax}(\mathbf{h}^{(i,j)}) = \frac{\exp(h_c^{(i,j)})}{\sum_c \exp(h_c^{(i,j)})} \quad (9)$$

ここで、 $\mathbf{W}^{(y)} = (w_1^{(y)}, w_2^{(y)})$, $\mathbf{b}^{(y)} = (b_1^{(y)}, b_2^{(y)})$ は、それぞれ重みパラメータベクトル, バイアスパラメータベクトルであり、反応パターンから学習される。

3.2 パラメータ学習

本節では、各パラメータの学習方法について述べる。提案手法では、反応確率 $\hat{\mathbf{y}}$ と真の反応である \mathbf{y} から、以下の分類誤差を表すクロスエントロピー Loss を算出し、損失関数とする。

$$\text{Loss} = - \sum_i \sum_j \sum_{c \in (0,1)} \mathbf{y}_{i,j}^{(c)} \log \hat{\mathbf{y}}_{i,j}^{(c)} \quad (10)$$

ここで、 $\mathbf{y}_{i,j}^{(c)}$ は、 $u_{ij} = c$ のとき 1, そうでないとき 0 をとる。

提案手法は、全ての反応パターンをもとに、Adam と呼ばれる最適化アルゴリズムに従って、損失関数が小さくなるよう、全てのパラメータを更新する。

4. 評価実験

本章では、シミュレーション実験と実データ実験で提案手法の有効性を示す。

本実験では、提案手法の実装にニューラルネットワークのフレームワークの一つである Chainer^{*1} を用いて、バッチ学習でパラメータを更新した。各隠れ層のノード数は 5, エポック数は 300 に固定した。また、 $\epsilon, \beta_1, \beta_2$ には、先行研究 [Kingma 14] で提案されている値を用いた。

一方、IRT のモデルには 2PLM を用いた。IRT のパラメータ推定は、マルコフ連鎖モンテカルロ法を用いた期待事後確率

*1 <https://chainer.org/>

推定法で行ない、等化には同時尺度推定法を用い、すべてのパラメータを同時に推定した。また、それぞれのパラメータの事前分布には次の分布を用いた。

$$\theta \sim N(0, 1), \quad \log a \sim N(0, 1), \quad b \sim N(1, 0.4) \quad (11)$$

ここで、 $N(\mu, \sigma)$ は平均 μ 、標準偏差 σ の正規分布を表している。

4.1 シミュレーションデータによる精度評価

等化に関する研究では、実データを収集するのに多大なコストを要することから、現実に近い条件のもと、シミュレーションにより評価を行うことが一般的である。したがって、本節では、シミュレーションデータをもとに、提案手法と IRT の等化精度を比較する。

等化精度の評価方法は様々であるが、ここでは、推定した能力パラメータと真の能力パラメータとの平均平方二乗誤差 (RMSE)、Pearson の積率相関係数と Kendall の順位相関係数について示す。ただし、RMSE を求める際には、推定したパラメータと真のパラメータを平均 0、分散 1 の分布に標準化して算出する。

4.1.1 シミュレーションデータの生成

本実験では、 J 個の項目で構成された K 個のテストを、それぞれ I 人からなる受験者集団が解答した状況を想定する。ここでは、 k 番目のテストが、 $k-1$ 番目、 $k+1$ 番目のテストとのみ共通項目を持つようにシミュレーションデータを生成する。

シミュレーションデータの生成は以下の分布に従ってパラメータを発生させ、2 PLM に従って行う。

$$\theta \sim N(0, 1), \quad \log a \sim N(0, 1), \quad b \sim N(1, 0.4) \quad (12)$$

具体的には、能力パラメータと項目パラメータを次の手法のいずれかを用いて受験者・項目に割り振る。

ランダム割り当て

式 (12) に従って、受験者数または項目数だけパラメータを発生させ、受験者または項目に無作為に割り振る。

システム割り当て

1) 式 (12) に従って、受験者数または項目数だけパラメータを発生させる。

2) 発生させたパラメータを昇順に並び替え、受験者集団数あるいはテスト数にパラメータを等分割する。

3) k 番目の受験者集団あるいはテストに、 k 番目のパラメータ分割内からパラメータを割り振る。

4.1.2 受験者の偏りとテスト受験者数による等化精度

本節では、受験者の偏りと受験者数による等化精度の変化を報告する。ここでは、受験者パラメータをシステム割り振りした場合を、受験者に偏りがある条件、ランダム割り振りした場合を受験者に偏りが無い条件とした。

本実験では、50 項目からなる 10 個のテストをそれぞれ異なる受験者が受験する状況を想定して、シミュレーションデータを発生させる。なお、識別力パラメータと困難度パラメータはランダム割り当てし、テスト間の共通項目数は 5 項目に固定した。

ここで、提案手法と IRT の等化精度を比較した結果を表 1 に示す。

表 1 より、受験者に偏りが無い場合、受験者数に関わらず、提案手法よりも IRT の精度が高いことがわかる。これは、全てのパラメータをランダム割り当てした場合、常にパラメータは等化されているとみなせるため、特別な等化処理が不要であることが原因であると考えられる。

一方で、受験者に偏りが存在する場合、いずれの受験者数で

も、提案手法の精度が IRT を大きく上回る結果となった。特に、受験者数が増えると IRT では精度が急激に低下するのに対し、提案手法では精度の差が見られなかった。

したがって、受験者集団間に能力の偏りがある状況下で、IRT は各テストの受験者数に大きく影響を受けるのに対し、提案手法は受験者数の変化に頑健な手法であることが示唆された。

等質な受験者集団を用意するには莫大なコストがかかることが多く、等化計画を行っていない場合、受験者集団間に能力の偏りがあることが一般的なため、提案手法の傾向は望ましいといえる。

4.1.3 テスト項目数による等化精度

本節では、各テストを構成する項目数を変化させた際の等化精度を比較する。ここでは、テスト間の共通項目が 5 項目である 10 個のテストをそれぞれ異なる 100 人が受験する状況を仮定して、シミュレーションデータを発生させる。また、能力パラメータのみをシステム割り当てし、項目パラメータはランダムに割り当てた。

ここで、各テストを構成する項目数を変化させて提案手法と IRT の等化精度を比較した結果を表 2 に示す。

表 2 より、各テストを構成する項目数が 10 から 50 の場合に、提案手法は IRT と比較して精度が高いことがわかる。一方で、項目数が 100 以上になると IRT の精度が上回ることが明らかとなった。従って、提案手法はテスト項目数が比較的小さい場合に有効な手法であることが示唆された。

実際のテストでは、受験者を考慮して、数百問程度の項目を解かせることは稀なため、提案手法の性質は好ましい。

4.1.4 共通項目数による等化精度

本節では、各テスト間の共通問題数を変化させた際の等化精度を比較する。ここでも、50 項目からなる 10 個のテストをそれぞれ異なる 100 人が受験する状況を仮定して、シミュレーションデータを発生させる。また、能力パラメータのみをシステム割り当てし、項目パラメータはランダムに割り当てた。

ここで、共通項目数を変化させて提案手法と IRT の等化精度を比較した結果を表 3 に示す。

表 3 より、共通項目数が多い場合は、IRT の方が精度が高いことがわかる。一方で、共通項目が 30 以下になると、提案手法がどの等化精度指標でも、IRT を大きく上回っている。従って、提案手法では、共通項目数が少ない状況下で、能力パラメータを高精度で推定できることが明らかになった。また、共通項目数が全くない場合でも、高精度であることが示された。

実際のテストで、数十問を共通項目とすることは、現実的ではないことが多いため、共通項目が少なくても影響の小さい提案手法は有効であるといえる。

4.2 実データ実験

本節では、実データを用いて、提案手法の有効性を示す。具体的には、未知の反応パターンの予測精度を検証する。

実データには、[Ueno 04] が開発した学習管理システム "Samurai" を用いて行ったテストデータ (Samurai) と、批判的思考を主題として行われたテストデータ (CT) を使用する。

Samurai データは、受験者数 169、項目数 50 からなり、欠損値がなく単一の学習者集団から得られたテストデータであるのに対し、CT データは、受験者数 1221、項目数 179 からなり、欠損が多く、等化処理が必要であることが特徴である。

4.2.1 欠損割合による予測精度の変化

次に、実データを用いて、欠損値の反応予測精度を検証する。ここでは、各受験者から得られた反応パターンから、特定の割合の反応パターンをランダムで欠損値とした反応パター

表 1: 能力パラメータの偏りとテスト受験者数による等化精度

θ	受験者数	手法	RMSE	Pearson	Kendall
50		DRM	0.317	0.950	0.882
		IRT	0.251	0.969	0.895
	100	DRM	0.312	0.953	0.882
		IRT	0.243	0.970	0.896
ランダム	150	DRM	0.270	0.964	0.890
		IRT	0.245	0.970	0.895
	200	DRM	0.294	0.957	0.891
		IRT	0.246	0.970	0.898
500	DRM	0.288	0.959	0.894	
	IRT	0.232	0.973	0.901	

表 2: テスト項目数による等化精度

項目数	手法	RMSE	Pearson	Kendall
10	DRM	0.622	0.807	0.629
	IRT	0.779	0.696	0.466
30	DRM	0.501	0.875	0.716
	IRT	0.573	0.836	0.672
50	DRM	0.411	0.916	0.785
	IRT	0.559	0.844	0.661
100	DRM	0.376	0.929	0.810
	IRT	0.325	0.947	0.827
150	DRM	0.305	0.953	0.866
	IRT	0.263	0.965	0.857

表 3: 共通項目数による等化精度

共通項目数	手法	RMSE	Pearson	Kendall
50	DRM	0.255	0.964	0.893
	IRT	0.237	0.972	0.900
40	DRM	0.248	0.969	0.89
	IRT	0.254	0.968	0.886
30	DRM	0.394	0.922	0.787
	IRT	0.360	0.935	0.805
20	DRM	0.389	0.924	0.803
	IRT	0.534	0.856	0.676
10	DRM	0.409	0.917	0.790
	IRT	0.534	0.857	0.677
5	DRM	0.411	0.916	0.785
	IRT	0.559	0.844	0.661
0	DRM	0.408	0.916	0.786
	IRT	0.612	0.812	0.532

ン行列からモデルを学習する。その後、学習したパラメータをもとに、欠損値とした項目の反応予測を行い、その正解率を提案手法と IRT で比較した。この手続きをそれぞれのデータセット、条件に対してずつ 10 回繰り返したときの平均正解率を、表 4 に示す。

ここで、各データセットと欠損割合の条件において、提案手法と IRT の予測精度に差があることを確かめるためにウィルコクソンの符号順位検定を行なった。

表 4 より、Samurai データでは、大部分を欠損値にしたとしても両者の精度に有意な差が見られないことがわかった。これは、単一の受験者集団から得られたデータであり、欠損値を多くしても等化が不要な状況であるため、IRT を用いても十分に精度が高いことが原因と考えられる。

一方で、CT データでは、多くの条件下で提案手法の方が予測精度が有意に高いことが明らかとなった。ただし、欠損をあまりに多くした場合、提案手法でも適切に予測ができないことが示された。

従って、提案手法は、複数の受験者集団が存在する欠損値の多いデータにおいて、反応の予測確率が高いことが示唆された。実際のアイテムバンクにおいて、単一の受験者集団のみが存在する状況は考えにくく、提案手法が有効な場面は多いと考えられる。

5. むすび

本研究では、深層学習を用いて、等化計画が十分でないときでも、高精度に能力パラメータの推定が可能なモデルである

表 4: 欠損値の予測精度

欠損割合	Samurai			
	DRM	IRT	DRM	IRT
1%	77.16%	76.92%	68.39%*	67.66%
10%	77.53%**	78.01%	70.01%**	69.30%
20%	77.48%	77.83%	68.72%**	68.34%
90%	71.71%	71.45%	59.89%	63.28%

* $p < .05$, ** $p < .01$

Deep Response Model を提案した。

具体的には、受験者と項目を独立したニューラルネットワークの入力とし、それぞれの出力を組み合わせることで反応パターンを予測する深層学習モデルを学習し、受験者を入力としたニューラルネットワークの出力を能力パラメータとみなした。

シミュレーション・実データ実験により、提案手法には以下の利点が存在することが明らかとなった。

- 1) 各受験者集団が等質でない場合でも、高精度に能力パラメータの推定が可能である。
- 2) 各テストの受験者数が増減しても、能力パラメータの推定精度への影響が小さい。
- 3) テスト項目数が少ない場合に、高精度に能力パラメータの推定が可能である。
- 4) 共通受験者・共通問題のいずれも存在しない場合でも高精度に能力パラメータの推定が可能である。
- 5) 複数の受験者集団が存在し、欠損値が多いデータにおいて、IRT よりも反応パターンの予測精度が高い。

提案手法は、多段階反応データや時系列解答データ、多次元データへの適応が容易であるため、モデルの拡張を行う。

参考文献

- [Bock 97] Bock, R. D. and Zimowski, M. F.: Multiple Group IRT, pp. 433–448, Springer New York, New York, NY (1997)
- [Kingma 14] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, arXiv:1412.6980 (2014)
- [Lord 68] Lord, F. and Novick, M.: Statistical Theories of Mental Test Scores, Addison-Wesley (1968)
- [Lord 80] Lord, F.: Applications of item response theory to practical testing problems, L. Erlbaum Associates Hillsdale, N.J (1980)
- [Loyd 80] Loyd, B. H. and Hoover, H. D.: Vertical Equating Using the Rasch Model, Journal of Educational Measurement, Vol. 17, No. 3, pp. 179–193 (1980)
- [Piech 15] Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., and Sohl-Dickstein, J.: Deep Knowledge Tracing, in Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R. eds., Advances in Neural Information Processing Systems 28, pp. 505–513, Curran Associates, Inc. (2015)
- [Ueno 04] Ueno, M.: Animated agent to maintain learner's attention in e-learning, in Nall, J. and Robson, R. eds., Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2004, pp. 194–201, Washington, DC, USA (2004), Association for the Advancement of Computing in Education (AACE)