

# Knowledge Tracingに Deep Learningと忘却モデルを 組み合わせた学生モデリング手法の提案

長谷川 駿一

2023年2月21日

木更津工業高等専門学校  
情報工学科 大枝研究室

# 目次

<b>第1章</b>	<b>はじめに</b>	<b>1</b>
1.1	研究背景 . . . . .	1
1.2	研究目的 . . . . .	1
<b>第2章</b>	<b>項目反応理論 (Item Response Theory:IRT)</b>	<b>2</b>
2.1	概要 . . . . .	2
2.2	正規累積モデル . . . . .	2
2.2.1	1 母数正規累積モデル . . . . .	2
2.3	ロジスティックモデル . . . . .	3
2.4	Rasch モデル . . . . .	3
<b>第3章</b>	<b>RNN(Recurrent Neural Network)</b>	<b>4</b>
3.1	概要 . . . . .	4
3.2	LSTM と GRU . . . . .	5
3.2.1	LSTM . . . . .	5
3.2.2	GRU . . . . .	6
3.3	Bi-RNN . . . . .	7
<b>第4章</b>	<b>エビングハウスの忘却曲線</b>	<b>8</b>
4.1	概要 . . . . .	8
4.2	節約率 . . . . .	8
<b>第5章</b>	<b>先行研究手法</b>	<b>10</b>
5.1	Bi-GRU モデル . . . . .	10
5.2	先行実験結果 . . . . .	11
<b>第6章</b>	<b>提案手法</b>	<b>12</b>
6.1	概要 . . . . .	12
6.2	実装方法 . . . . .	13
<b>第7章</b>	<b>実験</b>	<b>14</b>
7.1	実験方法 . . . . .	14
7.2	人工データ . . . . .	14
7.3	実験結果 . . . . .	15
7.4	考察 . . . . .	16
<b>第8章</b>	<b>おわりに</b>	<b>17</b>
8.1	まとめ . . . . .	17
8.2	今後の展望 . . . . .	17

# 第1章 はじめに

## 1.1 研究背景

今日の教育現場において、大量に蓄積された教育データに対して機械学習の手法を活用して学習者のスキル状態や設問の難易度などの情報の抽出を行い、効果的な教育を促進させる試みである、EDM(Educational Data Mining) が注目されている [1].

Educational Data Mining の手法の一つである KT(Knowledge Tracing) とは、学習者のスキル状態を隠れ変数、設問に関する解答結果 (正答・誤答) を観測値とする隠れマルコフモデル (Hidden Markov Model:HMM) で表現される手法であり、学習履歴から学習者の潜在能力を推定して学習者の未知の設問への正答予測を行う。近年、この KT にディープラーニングの手法を組み込むモデルが普及しており、比較的高い精度で正答予測を行うことが可能となった [2]。その中でも、KT の手法として双方向 RNN(Bi-directional Recurrent Neural Network) を用いたモデルでは、他のディープラーニングを組み込んだ手法と比較してより精度の高い予測結果を実現していることが先行研究によって明らかとなった [3].

その反面、ディープラーニングを取り入れたモデルではブラックボックス性が高いことが問題となっており、教育心理面での説明可能性に乏しい。そのために十分な性能が発揮できていないことがある。それに対して、説明可能性の高いモデルとして項目反応理論 (Item Response Theory:IRT) という手法が存在する。このモデルは実用上においても良く用いられており、教育系の統計モデルとして確立したモデルとなっている。

## 1.2 研究目的

本研究では、先行研究のモデルである双方向 RNN の手法を用いた Knowledge Tracing に対して、説明可能性の高いモデルの一つである IRT に忘却の概念を取り入れたモデルを組み込み、精度の向上と説明可能性の関連を目標としたモデルを提案する。忘却の概念にはエビングハウスの忘却曲線の理論を採用した。

実験には人工データを用いて、先行研究のモデルと提案手法のモデル、そして、忘却の概念を取り入れなかった、先行研究のモデルと IRT を組み合わせたのみのモデルの 3 種類を用意し、提案手法の有用性を示す。

## 第2章 項目反応理論 (Item Response Theory:IRT)

### 2.1 概要

項目反応理論 (Item Response Theory:IRT) は、心理・教育系の統計モデルとして、研究上においても実用上においても非常によく用いられている現代テスト理論である [4]。従来の古典的テスト理論では、項目分析および標準化からの知見が被験者の特性の分布に完全に依存していたが、IRT はこれを克服している。例を挙げると、語学試験において、「どこで受験したか」「いつ受験したか」「どの問題を解いたか」に関係なく試験の結果を留学の可否に利用することが可能になる。つまり、異なる受験者が、異なる項目を、異なる日時に、異なる場所で受験したにもかかわらず、被験者は統一された処遇を受けることができる。実際、IRT は TOEIC や TOEFL などの試験作成に使用されている。

### 2.2 正規累積モデル

IRT モデルは、古典的テスト理論の「学力別集団が受験者集団の特性値の分布に依存している」という欠点を、受験者集団に依存しない特性値  $\theta$  を導入することで解決している。この特性値は被験者母数とも呼ばれ、被験者の能力を表しており、直接観測できないため潜在特性 (latent state) である。モデリングの際、横軸に潜在特性  $\theta$  を、縦軸に正答確率を配する。このままでは、横軸が直接観測することができないため具体的な関数形は一通りに決定できない。そこで、統計学で最も頻繁に利用される標準正規分布の密度関数 (2.1) の累積分布関数 (2.2) を利用する。

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \quad (2.1)$$

$$\Phi(f(\theta)) = \int_{-\infty}^{f(\theta)} \phi(z) dz \quad (2.2)$$

このように横軸に潜在特性を配し、縦軸に項目への正答確率を配した関数を**項目特性曲線 (Item Characteristic Curve:ICC)**、あるいは**項目反応関数 (Item Response Function:IRF)** という。また、式 (2.2) を、特に正規累積モデルという。  $f(\theta)$  としては、  $\theta$  に関する単調増加関数を選ぶ。このとき、  $\Phi(f(\theta))$  も同様に、  $\theta$  に関する単調増加関数となる。

#### 2.2.1 1 母数正規累積モデル

項目  $j$  の性質として最も基本となるのは、その項目が易しいのか、難しいのかという性質である。この性質を表現するために、式 (2.1) の  $\theta$  の関数を式 (2.3) で表し、項目  $j$  の ICC を式 (2.4) で表現する。

$$f(\theta) = a(\theta - b_j) \quad (2.3)$$

$$p_j(\theta) = \Phi(a(\theta - b_j)) \quad (2.4)$$

$b_j$  は項目  $j$  の難しさを決める母数であり、難易度母数、または項目難易度 (item difficulty) と呼ばれる。ここで  $a$  はすべての項目に共通の値である。母数とは統計モデルの性質を決める数指標であり、ここでは  $b_j$  という単一の母数によって ICC の統計的性質が決まる。このため、このモデルを **1 母数正規累積モデル** という。

## 2.3 ロジスティックモデル

正規累積モデルは、式 (2.1) で示したように積分を含んでいる。基本となるモデル式が積分を含んでいるとその後の取り扱いが複雑になってしまう。そこで、ロジスティック分布の分布関数を利用した ICC が提案された。式 (2.1) に対して、式 (2.5) のような近似公式を適応する。

$$\int_{-\infty}^{f(\theta)} \phi(z) dz \simeq \frac{1}{1 + \exp(-Df(\theta))} \quad (2.5)$$

$D$  は尺度因子であり、 $D = 1.7$  のときに、 $\theta$  の全域にわたって食い違いが 0.01 以下になることが知られている。この近似公式は精度が高いため、式 (2.5) の右辺と左辺のどちらを使用してもテストの運用上は ICC に区別はない。

## 2.4 Rasch モデル

式 (2.3) と式 (2.5) を利用した以下の式を **Rasch モデル** という。

$$p_j(\theta) = \frac{1}{1 + \exp(-1.7(\theta - b_j))} \quad (2.6)$$

IRT では項目の形式に合わせてさまざまなモデルが考案されているが、Rasch モデルは、項目への反応を 2 種類に分類可能なテスト (正答・誤答) を対象にしている。

## 第3章 RNN(Recurrent Neural Network)

### 3.1 概要

**RNN(Recurrent Neural Network)** とは、時系列データの予測を得意とする DNN(Deep Neural Network) の構造の一つである。時系列データの予測においては、入力それぞれ独立しておらず一連の入力データとして扱われる。この特性を持つ時系列データを扱うことのできる RNN は、文章生成などの自然言語処理や音声認識、画像解析など、幅広い場面で活用されている。

RNN の概要図を図 3.1 に示す。層自体は「入力層-隠れ層 (中間層)-出力層」と一般的な DNN と相違ないが、時刻  $t$  における入力  $x(t)$  に加え、時刻  $t-1$  における隠れ層の値  $h(t-1)$  も時刻  $t$  における隠れ層に伝搬させる点が従来の DNN と大きく異なる。時刻  $t$  の状態を  $t-1$  の状態として保持しフィードバックさせるので、過去の隠れ層の値  $h(t-1)$  には再帰的 (recurrent) に過去の状態がすべて反映されることになる。

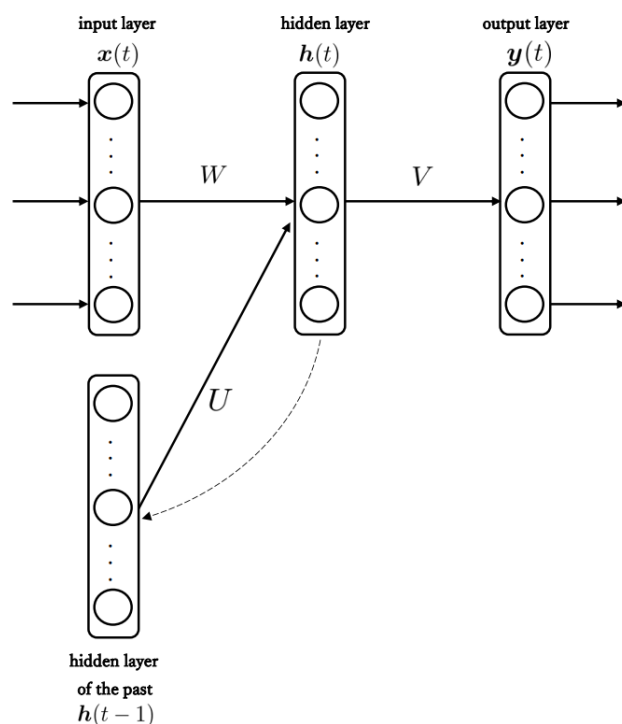


図 3.1 RNN の概要図

よって、モデルを表す式は、隠れ層を  $h(t)$ 、出力層を  $y(t)$  とすると、以下の通りである。

$$h(t) = f(Wx(t) + Uh(t-1) + b) \quad (3.1)$$

$$y(t) = g(Vh(t) + c) \quad (3.2)$$

ここで、 $f(\cdot)$ 、 $g(\cdot)$  は活性化関数、 $b$ 、 $c$  はバイアスベクトルである。

## 3.2 LSTM と GRU

RNN の 1 種として、LSTM(Long Short-Term Memory)[6] と GRU(Gated Recurrent Unit)[7] が存在する。

### 3.2.1 LSTM

LSTM は、従来の RNN の問題であった長期記憶における勾配消失問題を解決したものである。この問題を解決したことにより、LSTM では長期の時間依存性と短期の時間依存性の用法を学習することができる。

LSTM の概要図を図 3.2 に示す。LSTM は、RNN の勾配消失問題を解決するためのニューロンである CEC(Constant Error Carousal) と、3 つのゲート機構 (gate mechanism)(入力ゲート, 出力ゲート, 忘却ゲート) を取り入れたモデルである。CEC は隠れ層の誤差がどれだけ時間を遡っても 1 となり続けるように、隠れ層における各ニューロンと並列して用意されている。これにより、長期依存性をネットワーク内に織り込むことができるようになった。また、入力ゲートと出力ゲートは、時間依存性がある信号を受け取ったときのみ活性化するゲートであり、CEC の入力部分に入力ゲート、出力部分に出力ゲートを導入することで、入出力ともに過去の情報が必要になったときのみゲートを開けて信号を伝搬し、それ以外はゲートを閉じておくことで過去の情報を保持することが可能となる。そして、忘却ゲートでは、CEC に保存された過去の情報を忘れ去る機能を有しており、入力となる時系列データ内でパターンが大きく異なる場合に CEC の誤差を受け取ることによって、CEC の値を直接書き換えている。

LSTM のモデルを表す式を以下に示す。ここで、 $c(t)$  は CEC、 $i(t)$  は入力ゲート、 $o(t)$  は出力ゲート、 $f(t)$  は忘却ゲート、 $h(t)$  は最終的な出力を示している。

$$c(t) = i(t) \odot a(t) + f(t) \odot c(t-1) \quad (3.3)$$

$$i(t) = \sigma(W_i x(t) + U_i h(t-1) + b_i) \quad (3.4)$$

$$o(t) = \sigma(W_o x(t) + U_o h(t-1) + b_o) \quad (3.5)$$

$$f(t) = \sigma(W_f x(t) + U_f h(t-1) + b_f) \quad (3.6)$$

$$h(t) = o(t) \odot g(c(t)) \quad (3.7)$$

ただし、

$$a(t) = f(W_a x(t) + U_a h(t-1) + b_a) \quad (3.8)$$

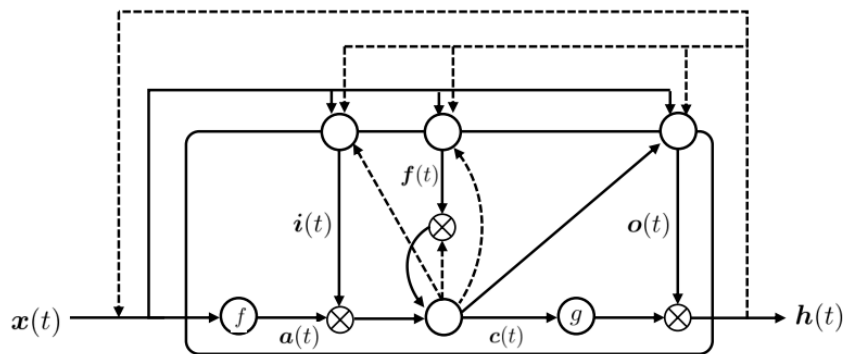


図 3.2 LSTM の概要図

### 3.2.2 GRU

LSTM では、長期の時間依存性を扱えるようになった分、パラメータの数が多く計算に時間がかかるという問題が発生した。GRU はこの問題を解決し、かつ LSTM と同等あるいはそれ以上の性能を持つ。

GRU の概要図を図 3.3 に示す。GRU ではリセットゲートと更新ゲートの 2 つのゲートのみで構成されている。ここで、 $r(t)$  はリセットゲート、 $z(t)$  は更新ゲートである。以下の式からもわかるように、GRU のパラメータの数は 9 個であるため、パラメータの数が 12 個である LSTM よりも計算量が少なく済むことがわかる。

$$r(t) = \sigma(W_r \mathbf{x}(t) + U_r \mathbf{h}(t-1) + \mathbf{b}_r) \quad (3.9)$$

$$z(t) = \sigma(W_z \mathbf{x}(t) + U_z \mathbf{h}(t-1) + \mathbf{b}_z) \quad (3.10)$$

$$\mathbf{h}(t) = \mathbf{z}(t) \odot \mathbf{h}(t-1) + (t) \odot \tilde{\mathbf{h}}(t) \quad (3.11)$$

ただし、

$$\tilde{\mathbf{h}}(t) = f(W_h \mathbf{x}(t) + U_h (r(t) \odot \mathbf{h}(t-1)) + \mathbf{b}_h) \quad (3.12)$$

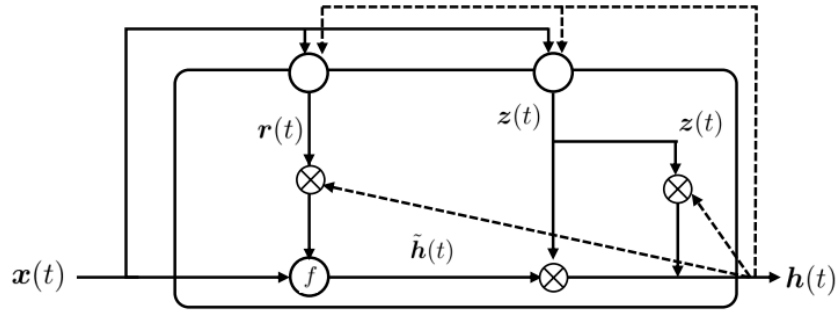


図 3.3 GRU の概要図



### 3.3 Bi-RNN

一般的な RNN が時刻  $t-1$  から  $t$  への隠れ層の状態を伝搬する、つまり過去から未来への 1 方向の流れを前提としていた。これに対して、**Bi-GRU**(**Bidirectional Recurrent Neural Network**) では、過去から未来の 1 方向だけでなく、未来から過去の方向も考慮した 2 方向の時系列の依存関係を考慮している。

Bi-RNN の概要図を図 3.4 に示す。過去から未来と未来から過去の 2 種類の隠れ層がそれぞれ独立して繋がっていることがわかる。過去から未来の隠れ層を「前向き層」、未来から過去の隠れ層を「後ろ向き層」といい、それぞれの層は以下のように矢印の向きを対応させて表現する。

$$\vec{h}(t) = f(\vec{W}x_t + \vec{U}h(t-1) + \vec{b}) \quad (3.13)$$

$$\overleftarrow{h}(t) = f(\overleftarrow{W}x_t + \overleftarrow{U}h(t-1) + \overleftarrow{b}) \quad (3.14)$$

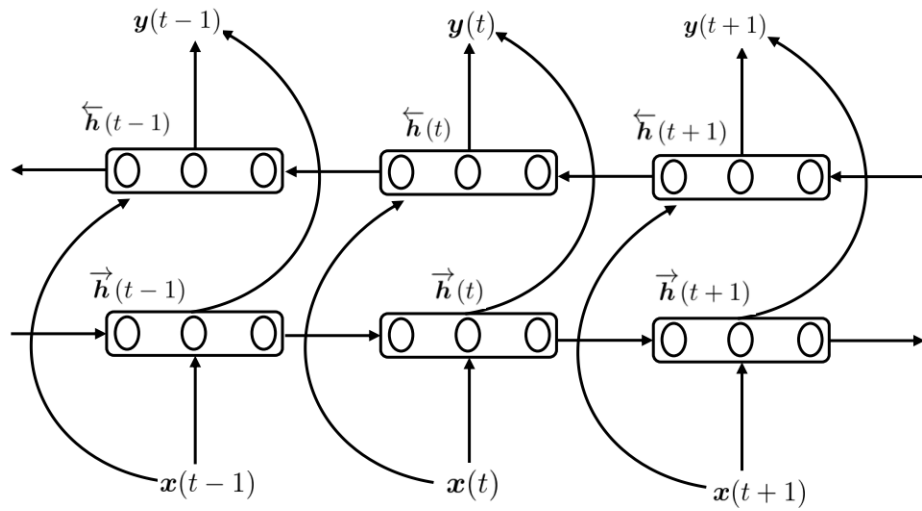


図 3.4 Bi-RNN の概要図

## 第4章 エビングハウスの忘却曲線

### 4.1 概要

**エビングハウスの忘却曲線** (Ebbinghaus' Forgetting Curve) とは、ヘルマン・エビングハウス (Hermann Ebbinghaus, 1850~1909) が、1885 年に行った時間経過による忘却度合いの変化の研究 [8] の実験結果より考察された、人間の記憶、特に中長期的記憶の忘却についてモデル化した曲線である。研究ではエビングハウス自身が被験者となって学習を行い、最初の学習と間隔を空けた再学習で覚え直しに必要な所要時間の比率 (節約率) を調査した。エビングハウスの忘却曲線は、調査結果において横軸を経過時間、縦軸を節約率としてグラフにプロットしたものが元となっている。

### 4.2 節約率

**節約率**とは、一度覚えた内容を再び完全に記憶し直すまでに必要な時間を、どれだけ節約できたかを表す割合である。この値は最初の学習に要した時間から再学習に要した時間の差をとり、その結果を最初の学習に要した時間で割ることにより算出される。節約率が高いほど、学習内容の覚え直しが簡単であるといえる。

エビングハウスは実験内で「子音・母音・子音」から成り立つ意味の無い音節を用意し、その音節を最初に記憶してから再学習するまでの時間間隔を変化させ、再び記憶するまでの比率から節約率を計算した。その結果、エビングハウスは自身の節約率に関して以下の式を提唱した。

$$f = 100 \times \frac{1.84}{(\log_{10} t)^{1.25} + 1.84}. \quad (4.1)$$

変数  $t$  は経過時間 (分) を、 $f$  は節約率を表している。この式をプロットした結果を図 4.1 に示す。エビングハウスの実験では 20 分後の節約率を約 58% としており、20 分以降も引き続き急激に忘却が進んで、約 9 時間 (540 分) 後あたりから緩やかな減少傾向になる。

エビングハウスが用意したデータに対する式 (4.1) の正確性については、後の研究 [9] において計算機を用いて検証した結果、二つのパラメータである 1.25 と 1.84 は計算機でそれぞれ 1.8 と 1.21 というほぼ近い値を算出しており、正確性が高いことが示されている。また、データに対して、決定係数  $R^2$  が 0.998、AIC (Akaike Information Criterion: 赤池情報量規準) が -30.6 で、他のデータに式 (4.1) の適応させたときの AIC と比較して最も値が小さいため (AIC は値が小さいほど良いモデルであることを示している)、式 (4.1) のモデルはエビングハウスのデータをよく示していると言える。

しかし、この忘却曲線の式は、被験者がエビングハウス自身のみのデータから算出されたため信頼性に欠けると指摘され、今日に至るまでさまざまな忘却曲線が提案されてきた。しかし、後の研究 [9] によると、エビングハウスのデータを含めた 4 つのデータに対して、式 (4.1) のモデルは  $R^2$  が 0.96 以上であるため、どのデータに対しても当てはまりが高いといえる。また、式 (4.1) のモデルは、提案された他の忘却曲線よりもパラメータが比較的少ないという利点を持つ。

よって、本研究では式 (4.1) のモデルを忘却の概念として取り入れることにする。

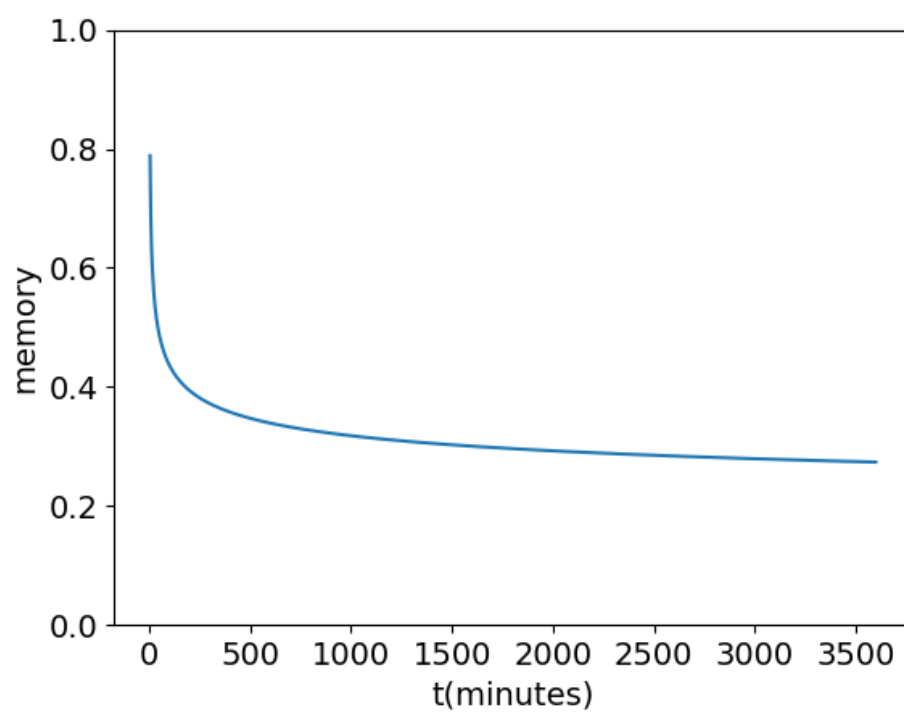


図 4.1 エビングハウスの忘却曲線

## 第5章 先行研究手法

### 5.1 Bi-GRU モデル

2021 年に開催された The 14th International Conference on Educational Data Mining において、Knowledge Tracing に GRU の双方向モデルを使用した **Bi-GRU** モデルが発表された。

このモデルの概要図を図 5.1 に示す。このモデルの入力値は、時刻  $t$  において被験者がある設問で試されたスキル  $q_t$  と、その設問を解いたときの結果  $r_t$  の 2 種類である。 $r_t$  は、正解を 1、不正解を 0 の二値で表している。出力は、学習者の新たなテスト結果  $\hat{r}_t$  である。

2 種類の入力値にそれぞれ記載されている添え字  $L$  は、このモデルの一回の学習で  $L$  時点前の結果までを入力とすることを示している。 $L$  の値が小さいほど対象のモデルに対する学習回数が増えるため、精度は高くなるが計算時間が長くなる傾向にある。

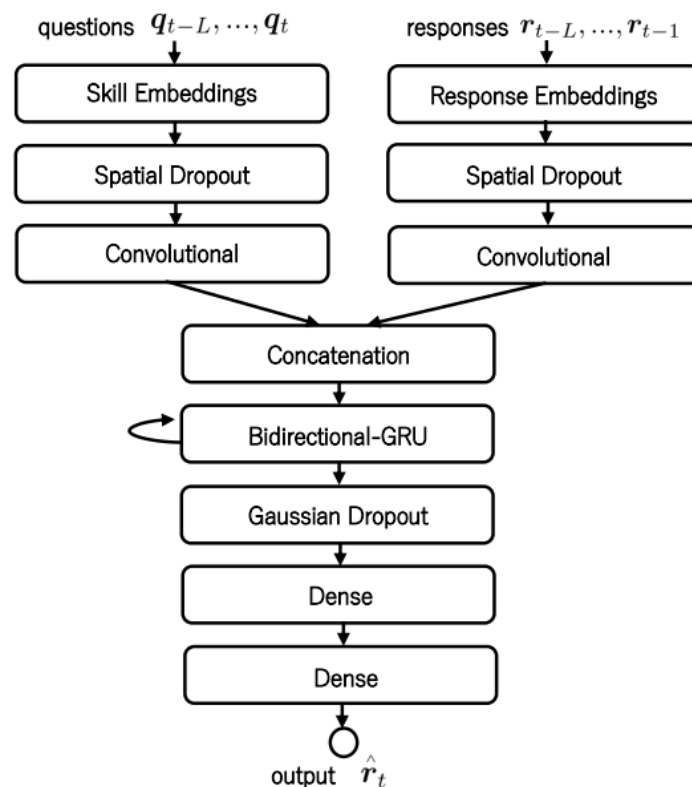


図 5.1 Bi-GRU モデルの概要図

このモデルでは、入力に対してまず埋め込み (Embedding) 層に伝搬し、次に SpatialDropout で埋め込み層に対してドロップアウトを行う。M.Delianidi らはこのドロップアウト率はほとんどのデータセットに対して 0.2 を設定している。続いて、カーネルサイズ 3、ストライド 1、活性化関数に ReLU を用いた、100 個のフィルターからなる畳み込み (Convolutional) 層に送り込み、前の埋め込み層からの入力次元を縮小している。彼女らは、これら一連の動作は過学習を抑制することに役立っていると報告している。

この操作をそれぞれの入力で行ったあと、バッチ正規化 (batch normalization) と ReLU による活性化を行い、それぞれの入力に連結された後に Bi-GRU 層に通される。その後、ガウシアンドロップアウトを行い、最終的な出力値  $\hat{r}_t$  を出力する。彼女らはさまざまなドロップアウト手法及び活性化関数、ドロップアウト率を変えて実験した結果、図 5.1 に示すモデルが最も良い精度で予測を行うと結論づけた。

## 5.2 先行実験結果

先行研究 [3] では、このモデルに対して、“ASSISTment09”<sup>1</sup>，“ASSISTment09 corrected”<sup>1</sup>，“ASSISTment12”<sup>2</sup>，“ASSISTment17”<sup>3</sup>，“FSAI-F1toF3”<sup>4</sup>の5つのデータセットを用いて、他の Knowledge Tracing のモデルとの予測精度の比較を行った。表 5.1 に先行実験における AUC (Area Under the Curve) を示す。

表 5.1: 先行実験の AUC(%)

Dataset	DKT	DKVMN	Deep-IRT	Bi-GRU
ASSISTment09	81.56	81.61	81.65	<b>82.55</b>
ASSISTment09 corrected	74.27	74.06	73.41	<b>75.27</b>
ASSISTment12	69.40	69.26	<b>69.73</b>	68.40
ASSISTment17	66.85	70.25	70.54	<b>73.76</b>
FSAI-F1toF3	69.42	68.40	68.69	<b>70.47</b>

5つのうち4つのデータセットで他の3つのモデルよりも高い AUC を出力しており、Bi-GRU モデルは予測精度が比較的高いモデルであることが示されている。

<sup>1</sup><https://sites.google.com/site/assistmentsdata/home/assistment-2009-2010-data/skill-builder-data-2009-2010>

<sup>2</sup><https://sites.google.com/site/assistmentsdata/home/2012-13-school-data-with-affect>

<sup>3</sup><https://sites.google.com/view/assistmentsdatamining/data-mining-competition-2017>

<sup>4</sup><https://www.4littletrees.com/>

## 第6章 提案手法

### 6.1 概要

本研究では、5章で紹介した先行モデルを改良し、忘却の概念を取り入れたモデルを考案した。本研究で使用するモデルの概要図を図6.1に示す。

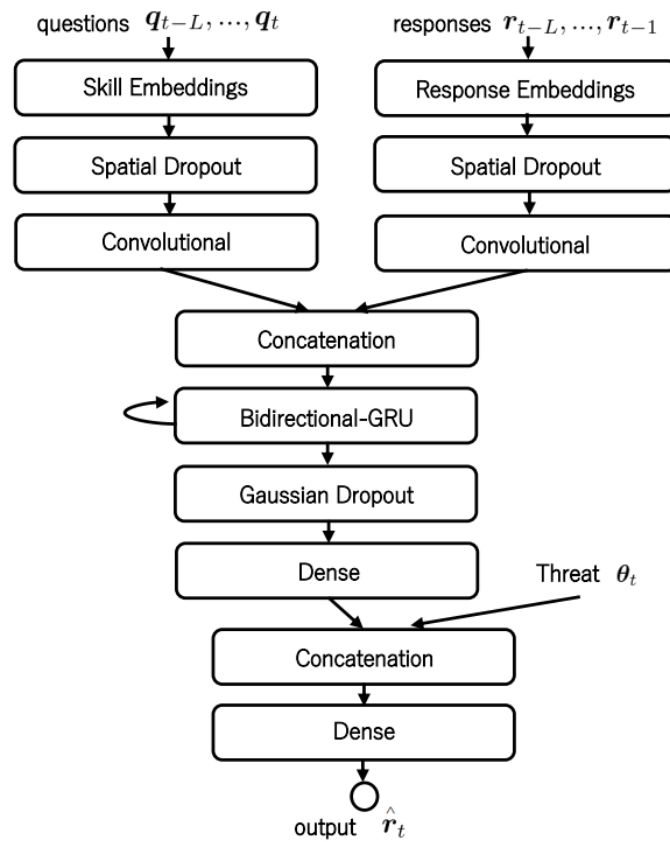


図 6.1 本研究のモデルの概要図

先行研究のモデルと大きく違うところは、モデルの最後のほうで3つ目の入力  $\theta_t$  を加えている点である。この  $\theta_t$  は、予測したい解答結果  $\hat{r}_t$  において、その問題を解くために必要なスキル  $q_t$  に対する特性値を示している。特性値  $\theta_t$  は、2章で説明したIRTのモデルであるRaschモデルを元に予測を行う。本研究では、 $\theta_t$  の入力をBi-GRU層よりあとの層で結合させることによって、出力値  $\hat{r}_t$  を出力している。

よって、出力値  $\hat{r}_t$  を算出する前に、特性値  $\theta_t$  を予測する必要がある。これには訓練データを特殊な形で分割する。まずはじめに訓練データを2つに分け、被験者の学習した時間の古いほうをIRTにかけ、 $\theta_t$  の初期値を算出する。次に、もう一つの訓練データを時系列順に参照していき、 $\theta_t$  を動的に変動させていく。観測されたレコードに対して、被験者が最後に解いた時間からより多くの時間が経過していたら、 $\theta_t$  を大幅に小さくし、結果を時系列になるように記録する。この  $\theta_t$

の減少の度合いを決めるのにエビングハウスの忘却曲線を使用しており、求められた節約率にパラメータをかけ合わせた値を、 $\theta_t$  から引くことによって忘却を表現している。レコードが観測されて  $\theta_t$  が記録されたあとは、被験者が復習したとみなし  $\theta_t$  を増加してある。

## 6.2 実装方法

本モデルの実装には、Python のディープラーニング向けのライブラリである Keras(2.11.0) を用いた。また、IRT の実装では、統計解析向けのプログラミング言語である R と、R を使用して作成された IRT 専用の OSS である IRT Command Language (ICL)<sup>1</sup>を用いた。ICL の予測では式 2.6 で示した Rasch モデルを想定したパラメータ予測を得意としており、また、IRT のパラメータの予測手法を容易に変更できるという利点がある。本研究ではパラメータの予測手法に、頑健性の高いベイズ推定法を用いた。

---

<sup>1</sup><http://www.openirt.com/b-a-h/software/irt/icl/>

## 第7章 実験

### 7.1 実験方法

本実験では、以下の3つのモデルを用いて、提案手法との予測精度の比較・検証を行った。データには人工データを用い、評価指標には AUC を用いた。

- (I) 先行研究 [3] のモデル
- (II) 提案手法のモデルにおいて、忘却をとり入れなかったモデル
- (III) 提案手法のモデル

2つ目のモデルでは、訓練データ全てを IRT の推定用を使用しているため、6章で説明した特性値  $\theta_t$  の値は固定値となり、各スキルを有する問題に同様の特性値が入力される。つまり、忘却を考慮していないモデルとなっている。

### 7.2 人工データ

本研究では、忘却の概念を取り入れているため、被験者が解答した詳細な時間が必要である。実データでは解答時間を記載してあるデータが少ないことから、人工データを生成して実験を行った。人工データの生成には IRT を用いた。

生成したデータの一部を表 7.1 に示す。一つのレコードにはそれぞれ4つの属性が設定されている。user\_id が被験者の ID、item\_id が問題番号、result が解答結果 (1: 正解, 0: 不正解)、そして date が被験者が解答した時刻を示しており、問題ごとに困難度  $b$  を変えて解答結果に反映させている。問題番号ごとに、その問題を解くために必要なスキルを別のファイルに保存しており、これを **Q-マトリックス**という。実際の学生モデリングでは、Q-マトリックスをその問題分野のエキスパートが設定している。

次節に示す実験結果は、総レコード数を 400,000 件、ユーザ数を 5000 人、問題数を 400 問、そして、スキル数を 10 として人工データを生成した。また、訓練データを 320,000 件、テストデータを 80,000 件としている。320,000 件の訓練データのうち、IRT による特性値  $\theta_t$  の初期値を推定するためのデータを 240,000 件とし、残りの 80,000 件のデータで忘却による  $\theta_t$  の変動を行っている。忘却なしのモデルでは、訓練データすべてを用いて  $\theta_t$  を推定している。

表 7.1: 生成した人工データの一部

user_id	item_id	result	date
237	244	1	2021/8/1 13:01
652	38	0	2021/8/1 13:01
2201	76	1	2021/8/1 13:01
1915	373	0	2021/8/1 13:01
1841	154	0	2021/8/1 13:02



### 7.3 実験結果

訓練データを実行した際の結果を図 7.1，テストデータを実行した際の結果を図 7.2 に示す．検証した 3 つのモデルについて，それぞれエポック数を 5 とし，各エポックごとの AUC 推移を折れ線グラフで示した．先行研究のモデルを破線，忘却なしのモデルを一点鎖線，提案手法のモデルを実線で示している．

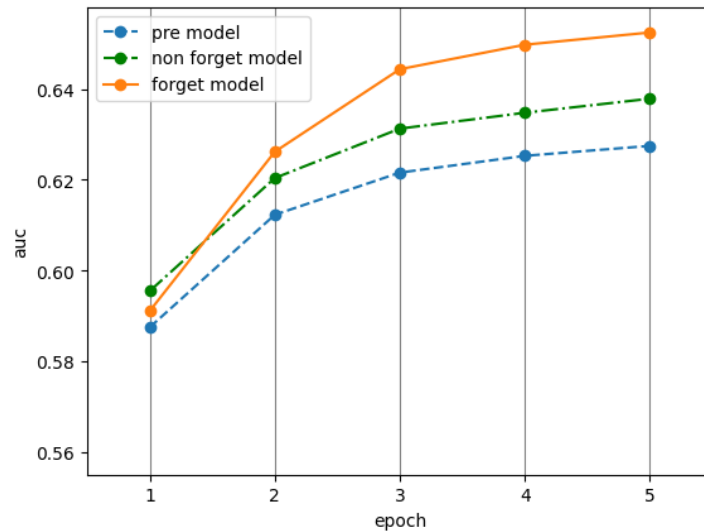


図 7.1 訓練データの AUC

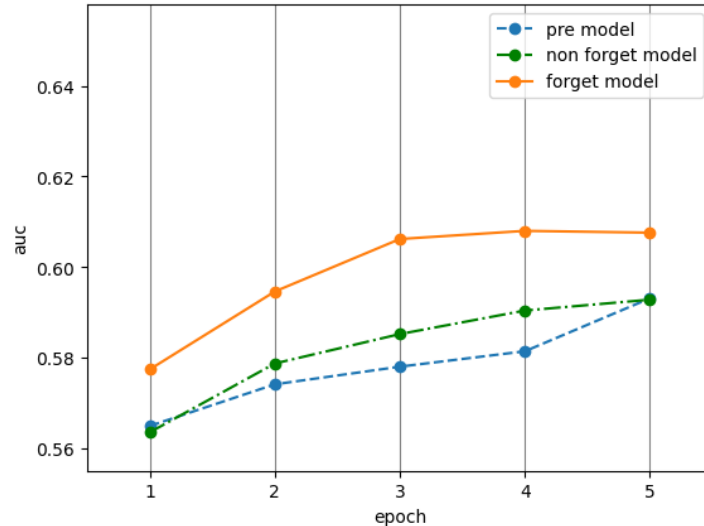


図 7.2 テストデータの AUC

結果として，本研究で提案したモデルが最も高い AUC を出力しており，続いて忘却なしのモデル，先行研究のモデルの順に AUC が低い傾向にあった．テストデータの最終的な AUC は，先行研究のモデルで 0.5931，忘却なしのモデルで 0.5928，本研究で提案したモデルで 0.6076 であった．また，各エポックごとの AUC の推移を見ると本研究で提案したモデルが最も学習速度が速いことがわかる．

## 7.4 考察

前節の結果より，本研究で提案したモデルが最もよい予測精度を出力していたことから，被験者の能力の向上を成長，衰退を忘却としてモデルに取り込むことが，学生モデリングにおいて効果的であることがわかる．従来のモデルではスキル番号とその問題の正誤結果の２種類のみを入力とするモデルが多いため，より人間の教育心理面との関連を追求し，忘却のみならずさまざまな要因を考慮したモデルを考案することが，学生モデリング手法の発展に繋がるであろうと考えられる．

## 第8章 おわりに

### 8.1 まとめ

本研究では、予測精度の向上と教育心理面での説明可能性の両立を目的として、Bi-GRUを用いた Knowledge Tracing と忘却の概念を取り入れた項目反応理論を組み合わせた手法を提案した。実験結果より、忘却の概念が学生モデリング手法における未来の成績予測に有用性があることがわかった。

### 8.2 今後の展望

本研究で提案したモデルには、大きく2つの問題点がある。一つ目は、IRTにおけるパラメータ推定の時間がかかるため、先行研究と比較しても計算時間が大幅に長いことである。そして二つ目は、IRTのパラメータ推定の特性上、スパースなデータには予測精度が悪いことである。特定の被験者が同じスキルを必要とする問題を何回も解いたデータが必要であるため、先行研究で使用していた実データの推定には不向きであることがわかっている。

従って、上記の2つの問題を解決し、かつ、教育心理面での説明可能性の高いモデルを調査し、自身の研究に繋げていきたい。そのためにも、発展的な Knowledge Tracing のモデルで使用されている Attention や Transformer などといったディープラーニングの手法や、記憶の分野に関する認知科学についての勉強に励んでいきたい。

また、私が所属している大枝研究室では、プログラミングに関する学生モデリングの研究も盛んに行われている。今日では ChatGPT<sup>1</sup>などの自然言語生成モデルが人間の要求したソースコードを生成できるまで開発が進んでいる。なので、人間と機械学習モデルが生成したソースコードに対して、学生モデリングの手法を用いた研究も行いたい。

---

<sup>1</sup><https://openai.com/blog/chatgpt/>

# 謝辞

本研究は，JSPS 科学研究 19H01728 の助成を受けたものです．

本研究に際して，指導教官の大枝真一教授から丁寧かつ熱心なご指導を賜りました．ここに感謝の意を表します．また，日常の議論を通じて多くの知識や示唆を頂いた大枝研究室の皆さまに感謝いたします．

## 参考文献

- [1] 柿崎透真, 大枝真一, “Knowledge Tracing と IRT を組み合わせた学生モデリング手法による学習者の潜在状態推定”, 情報処理学会 第 84 回全国大会, 2022.
- [2] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. Guibas, and J. Sohl-Dickstein, “Deep knowledge tracing,” Advances in neural information processing systems, 28:505–513, 2015.
- [3] M. Delianidi, K. Diamantaras, G. Chrysogonidis, and V. Nikiforidis, “Student performance prediction using dynamic neural models,” in Fourteenth International Conference on Educational Data Mining (EDM 2021), pp. 46–54, 2021.
- [4] 豊田秀樹, 項目反応理論 [入門編][第 2 版], 統計ライブラリー, pp.23-31, 2012.
- [5] 巢籠悠輔, 詳解ディープラーニング [第 2 版], マイナビ出版, pp.255-319, 2019.
- [6] S.Hochreiter, J.Schmidhuber, “Long short-term memory,” Neural Computation, 9(8), pp.1735-1780, 1997.
- [7] K.Cho, B.Merruebbouer, C.Gulcehre, F.Bougares, H.Schwenk, Y.Bengio, “Learning phrase representations using rnn encoder-decoder for statistical machine translation,” Proceedings of the Empirical Methods in Natural Language Processing(EMNLP), 2014.
- [8] Ebbinghaus. H, “Über das gedächtnis: untersuchungen zur experimentellen psychologie,” Duncker & Humblot, 1895.
- [9] Murre, Jaap MJ, and Joeri Dros. ”Replication and analysis of Ebbinghaus’ forgetting curve.” PloS one 10.7 (2015): e0120644.