

平成 30 年度 情報数理工学コース卒業論文概要

平成 27 年度入学	学籍番号 1511107
指導教員 植野真臣	氏名 塩野谷周平
題目 学習者・課題パラメータを持つ LogisticHMM の提案	

概要

近年、教育の現場において、学習者に適切な支援を行うために、学習者の知識の習得度を推定することが重要な課題となっている。長年、知的学習支援システム (ITS) の分野では、学習者の知識の理解度を推定するモデルが提案されてきた。Bayesian Knowledge Tracing (BKT) は、学習過程が隠れマルコフに従うと仮定したモデルで、時系列データから学習者の理解度を推定することができる。しかし、BKT では学習者の理解度を測ることができるが、学習者全体で共通のパラメータを用いるため、学習者の理解度が正しく推定されず、学習者の過大評価もしくは過小評価が起りやすい。また、LogisticHMM では、知識理解に個人の特性を考慮せず、課題の特徴を反映する課題パラメータがすべての課題で共通にされており、学習者に十分な学習支援が行えない可能性が高い。本研究では、遷移確率を個人化させ、課題パラメータをそれぞれの課題に依存したパラメータとした LogisticHMM を用いて、学習者の知識状態の予測精度を高めるモデルを提案する。知識状態の予測精度の評価のために、学習者の課題に対する正誤予測の精度を向上させることを目的とする。

学習者・課題パラメータを持つ LogisticHMM の提案

平成 31 年 2 月 28 日

情報数理工学コース

学籍番号 1511107

塩野谷周平

指導教員 植野真臣

目次

1	まえがき	2
2	Bayesian Knowledge Tracing	4
2.1	従来の Bayesian Knowledge Tracing	4
2.2	個人パラメータをもつ Bayesian Knowledge Tracing	6
3	LogisticHMM	6
3.1	従来の LogisticHMM	6
3.2	提案モデル	9
4	パラメータの推定	10
4.1	状態 Z_{ij} のサンプリング	11
4.2	パラメータ a, b のサンプリング	12
4.3	アルゴリズム	13
5	評価実験	14
5.1	データ	14
5.2	正誤予測の精度	15
5.3	考察	17
6	むすび	17

1 まえがき

近年，学習者への適切な支援を行うために，学習者の知識の習得度を，正確に推定することが重要な課題となっている．知的学習支援システム (Intelligent Tutoring System, ITS と呼ぶ) の分野では，学習者の未習熟な課題を同定し，個々の学習者に未習熟課題を提示するシステムが数多く提案された [1]～[6]．ITS の中でも，学習者の課題に対する知識の理解度を推定することで，学習者に適切な課題を選択させるアルゴリズムや [1]～[3]，ベイジアンネットワークを用いた手法 [4]，さらに教育データマイニングの分野では新しい学習者モデルの開発が進められてきた [5], [6]．また ITS では，学習者の発達に対する学習援助を調節するために，問題解決の各段階で学習者の適応的な支援を提供するシステムの開発も行われた [7]～[11]．これらのモデルは，学習者が間違ったステップから次のステップに向けて，何をすべきかを記されたヒントが提示される．

ITS の分野において，学習者の知識の理解度を推定するために，Corbett and Anderson (1995) [1] が開発した，Bayesian Knowledge Tracing (BKT) と呼ばれる数理モデルが広く用いられてきた [1], [12]～[14]．BKT では学習者の知識状態は「理解している」か「理解していない」の2値を取るものとし，知識の習得を，その知識を理解している確率で表すことができる．BKT には，知識を事前に理解している確率，「理解していない状態」から「理解している状態」へ遷移する確率，学習者が知識を理解していなくても課題に正答する確率，理解していても課題に誤答する確率の合計4つのパラメータを持ち，これらを用いて，知識を理解している確率を求めることができる．BKT では学習者の知識の理解度を推定することができ，特定の知識に関して，適切な学習支援を行うことが可能となる．また BKT では学習者の課題に対する正誤予測も行うことができる．

しかし，標準的な BKT では，パラメータはスキルにのみ依存するパラメータしか用いられず，学習者に依存するパラメータがないため，個々の学習者に対して，適切な学習支援が行われにくい．そこで，BKT の4つのパラメータを，学習者に依存するパラメータに変更させた BKT が提案された [15]～[17]．Yudelson, Koedinger and Gordon (2013) [17] は，BKT パラメータの個人化を

検討し、初期確率（最初の課題で、すでに理解している確率）と遷移確率（「理解していない状態」から「理解している状態」へ遷移する確率）の個人化を提案した。彼らが行った、実データを用いた実験により、遷移確率を個人化させた BKT が、従来の BKT に比べ、学習者の課題に対する正誤予測の精度が向上したことが報告された。正誤予測の向上により、個人パラメータを導入した BKT の方が、従来の BKT に比べ、知識状態の予測精度が向上したといえる。

しかしながら、BKT は学習者の知識の理解度を求めることができるモデルだが、学習者が取り組んだ課題の特徴を推定することはできない。また BKT では知識状態は 2 値とされており、学習者の知識が正しく推定されず、学習者を過大もしくは過小評価してしまい、適切な学習支援が行われない可能性が高い。そこで Pelánek (2018) [18] は、BKT では 2 値としていた学習者の理解度を多値に拡張し、観測データの出力確率にロジスティック関数を用いることで、課題の特徴を反映するパラメータを推定することができる LogisticHMM モデルを提案した。Pelánek は LogisticHMM を用いることで、BKT よりも学習者の知識の習得度をより高精度に推定できることを示した。

しかし、

- (1) Pelánek が提案した LogisticHMM は、遷移確率がスキルのみ依存している。
- (2) 課題の特徴を反映する課題パラメータが、すべての課題で共通のものである。

遷移確率は個人特性により変化し、課題パラメータはそれぞれの課題に依存すると考えられる。従来の LogisticHMM では、学習者・課題パラメータを持っていないので、それぞれの特徴を使った推定ができない。

そこで本研究では、遷移確率を個人化させ、課題パラメータをそれぞれの課題に依存させたパラメータとした LogisticHMM を提案する。

提案モデルと、従来の BKT、遷移確率を個人化させた BKT、従来の LogisticHMM で、課題の正誤予測を比較し、BKT や従来の LogisticHMM よりも学習者の課題に対する正誤予測の精度が高いことを示し、学習者の知識状態を正

確に推定できるモデルであることを示す。また本研究では提案モデルのパラメータ推定をマルコフ連鎖モンテカルロ法的一种である周辺化ギブスサンプリングとメトロポリスヘイスティング法を合わせた手法で行う。

2 Bayesian Knowledge Tracing

2.1 従来の Bayesian Knowledge Tracing

Corbett and Anderson (1995) [1] は, Bayesian Knowledge Tracing (BKT) という, 学習過程が隠れマルコフに従うと仮定するモデルを提案した。BKT モデルでは, 隠れ変数は 2 値で表し, 学習者 i が課題 j 問目で「理解している」か「理解していない」の 2 通りを表す。学習者数を I , 課題数を J とすると, 隠れ変数は次のように変数 Z_{ij} で表せる。

$$Z_{ij} = \begin{cases} 1 & (\text{学習者 } i \text{ が課題 } j \text{ を理解している}) \\ 0 & (\text{学習者 } i \text{ が課題 } j \text{ を理解していない}) \end{cases}$$

$$\mathbf{Z} = \{Z_{ij}\}, (i = 1, \dots, I, j = 1, \dots, J)$$

また, 学習者 i の課題 j に対する反応データは次の観測変数 X_{ij} で表される。

$$X_{ij} = \begin{cases} 1 & (\text{学習者 } i \text{ が課題 } j \text{ に正答}) \\ 0 & (\text{学習者 } i \text{ が課題 } j \text{ に誤答}) \end{cases}$$

$$\mathbf{X} = \{X_{ij}\}, (i = 1, \dots, I, j = 1, \dots, J)$$

BKT には次の 4 つのパラメータを含む。

- $p(L_0)$ … 最初の課題で知識を理解している確率
- $p(T)$ … 知識を「理解していない状態」から, 「理解している状態」へ遷移する確率
- $p(G)$ … 知識を理解していないにも関わらず, 課題を正答する確率
- $p(S)$ … 知識を理解しているにも関わらず, 課題を誤答する確率

表 1: BKT における初期確率

$Z_{i0} = 0$	$Z_{i0} = 1$
$1 - p(L_0)$	$p(L_0)$

表 2: BKT における遷移確率

	$Z_{ij} = 0$	$Z_{ij} = 1$
$Z_{i,j-1} = 0$	$1 - p(T)$	$p(T)$
$Z_{i,j-1} = 1$	0	1

表 3: BKT における観測データ出力確率

	$X_{in} = 0$	$X_{in} = 1$
$Z_{in} = 0$	$1 - p(G)$	$p(G)$
$Z_{in} = 1$	$p(S)$	$1 - p(S)$

これら BKT パラメータの関係は表 1, 2, 3 のようになる。

BKT では、学習者の「理解している状態」から「理解していない状態」への遷移（いわゆる知識の忘却）は起こらないものとする。BKT モデルのグラフィカル表現を図 1 に示す。グラフは時系列になっており、学習者 i の学習過程を表す。

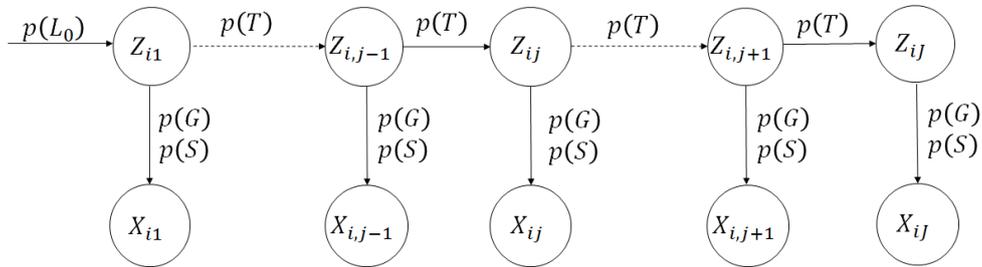


図 1: BKT モデルのグラフィカル表現

BKT モデルは、学習者の知識の習熟度を推定することができるモデルである。学習者 i が、課題 j を理解している確率を $p(Z_{ij} = 1)$ とすると、 $p(Z_{ij} = 1)$

は式 (2), または式 (3) を用いて, 式 (1), (4) のように表せる.

$$p(Z_{i1} = 1) = p(L_0) \quad (1)$$

$$p(Z_{ij} = 1|X_{ij} = 1) = \frac{p(Z_{ij} = 1)(1 - p(S))}{(1 - p(Z_{ij} = 1))p(G) + p(Z_{ij} = 1)(1 - p(S))} \quad (2)$$

$$p(Z_{ij} = 1|X_{ij} = 0) = \frac{p(Z_{ij} = 1)p(S)}{(1 - p(Z_{ij} = 1))(1 - p(G)) + p(Z_{ij} = 1)p(S)} \quad (3)$$

$$p(Z_{ij} = 1) = p(Z_{i,j-1} = 1|X_{i,j-1}) + (1 - p(Z_{i,j-1} = 1|X_{i,j-1}))p(T) \quad (4)$$

また, 学習者 i が, 課題 j を正答する確率を $p(X_{ij} = 1)$ とすると, $p(X_{ij} = 1)$ は次のように表せる.

$$p(X_{ij} = 1) = (1 - p(Z_{ij} = 1))p(G) + p(Z_{ij} = 1)(1 - p(S)) \quad (5)$$

2.2 個人パラメータをもつ Bayesian Knowledge Tracing

従来の BKT では, パラメータは 4 つしかないため, 学習者の理解度を過大もしくは過小評価されてしまい, 学習者の課題の正誤を正確に予測することができない. そこで, Yudelson ら (2013)[17] は BKT パラメータのうち, 遷移確率 $p(T)$ を次のように変更させた.

- $p(T)_i$ …… 学習者 i が, 「理解していない状態」から「理解している状態」へ遷移する確率

この個人化された遷移確率を用いることにより, 学習者 i が, 課題 j を理解している確率を $p(Z_{ij} = 1)$ は次のようになる.

$$p(Z_{ij} = 1) = p(Z_{i,j-1} = 1|X_{ij}) + (1 - p(Z_{i,j-1} = 1|X_{ij}))p(T)_i \quad (6)$$

この個人化された遷移確率を用いることにより, Yudelson らは, 従来の BKT よりも学習者の課題に対する正誤予測の精度が向上したことを示した.

3 LogisticHMM

3.1 従来の LogisticHMM

本章では LogisticHMM モデルについて説明する. LogisticHMM は Pelánek (2018) [18] によって開発された新しいモデルである. BKT では学習者の知識の

理解度を推定することができたが、学習者が取り組んだ課題の特徴を推定することができなかった。しかし、LogisticHMM を用いれば、課題の特徴を反映する課題パラメータを推定することができる。さらに BKT では学習者の知識状態は「理解していない」と「理解している」の2種類しか定義していなかったが、LogisticHMM では知識状態は多値に変更している。多値に変更することで BKT よりも学習者の知識の習得度を高精度に評価することが可能になった。

LogisticHMM では、学習者 i が、隠れ変数である知識状態が s であるとき、課題 j に正答する確率は

$$p(X_{ij} = 1 | Z_{ij} = s) = \frac{1}{1 + \exp(-a(s/(S-1) - b))} \quad (7)$$

$$(s = 0, \dots, S-1)$$

とモデル化する。ここで、 S は学習者の知識状態数、 a は識別力パラメータ、 b は難易度パラメータを表す。識別力パラメータは、学習者の能力を識別する力を表し、難易度パラメータは、課題の難しさを表す。これらのパラメータはすべての課題において共通の値を用いる。図2に、知識状態数 $S = 5$ 、識別力パラメータ $a = 5.0$ 、難易度パラメータ $b = 0.5$ としたときの反応関数の例を示す。横軸は学習者の知識状態を表し、縦軸は、知識状態が s だったときの、学習者 i が課題 j に正答する確率 $p(X_{ij} = 1 | Z_{ij} = s)$ を表す。図2より、学習者の知識状態が低位のときは正答確率は低く、高位のときは正答確率が高い。

次に LogisticHMM の遷移確率に関して、知識の遷移を示すパラメータ l を用いて、次のように定める。

- $p(Z_{ij} = s | Z_{i,j-1} = s) = 1 - l \dots$ 知識状態の遷移がないとき
- $p(Z_{ij} = s | Z_{i,j-1} = s - 1) = l \dots$ 知識状態が $s - 1$ から s へ遷移するとき
- 上記以外 $\dots 0$ (知識の忘却や、2段階以上の知識の遷移はないものとする)

遷移確率に関しては、表4のように表現できる。

LogisticHMM モデルでは、学習者の知識の習熟度を推定することができる。学習者 i が、課題 j で、状態 $Z_{ij} = s$ となる確率は、式(8)、あるいは式(9)を

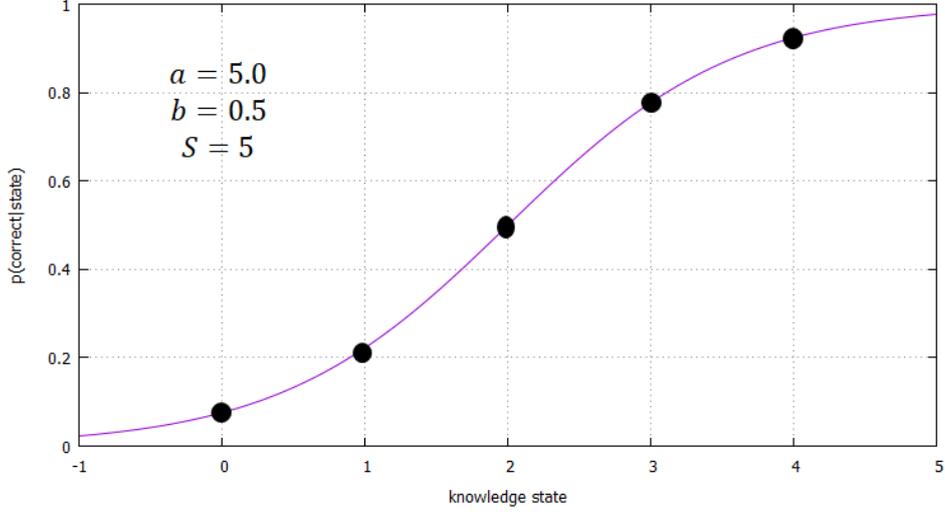


図 2: LogisticHMM の反応関数の例

表 4: LogisticHMM における遷移確率

	$Z_{ij} = 0$	$Z_{ij} = 1$	$Z_{ij} = 2$	\dots	$Z_{ij} = S$
$Z_{i,j-1} = 0$	$1 - l$	l	0	\dots	0
$Z_{i,j-1} = 1$	0	$1 - l$	l	\dots	0
$Z_{i,j-1} = 2$	0	0	$1 - l$	\dots	0
\dots	\dots	\dots	\dots	\dots	\dots
$Z_{i,j-1} = S$	0	0	0	\dots	1

用い、式 (10) のように求められる。

$$p(Z_{ij} = s | X_{ij} = 1) = \frac{p(X_{ij} = 1 | Z_{ij} = s)p(Z_{ij} = s)}{\sum_{s'=0}^{S-1} p(X_{ij} = 1 | Z_{ij} = s')p(Z_{ij} = s')} \quad (8)$$

$$p(Z_{ij} = s | X_{ij} = 0) = \frac{p(X_{ij} = 0 | Z_{ij} = s)p(Z_{ij} = s)}{\sum_{s'=0}^{S-1} p(X_{ij} = 0 | Z_{ij} = s')p(Z_{ij} = s')} \quad (9)$$

$$p(Z_{ij} = s) = \sum_{s'=0}^{S-1} p(Z_{ij} = s | Z_{i,j-1} = s')p(Z_{i,j-1} = s' | X_{i,j-1}) \quad (10)$$

また、学習者 i が、課題 j を正答する確率は、次のようになる。

$$p(X_{ij} = 1) = \sum_{s'=0}^{S-1} p(X_{ij} = 1 | Z_{ij} = s')p(Z_{ij} = s') \quad (11)$$

3.2 提案モデル

Pelánek が提案した LogisticHMM モデルを用いることで、BKT では推定することができなかった課題の特徴を反映するパラメータを推定できるようになり、知識状態を多値にすることで、学習者の理解度を BKT よりも高精度に評価できるようになった。しかしながら、

- (1) LogisticHMM は、遷移確率がスキルのみ依存しているものである。つまり、学習者全体で共通の遷移確率を用いるため、学習者ごとの特性が反映されていない。
- (2) LogisticHMM は、識別力、難易度パラメータが、すべての課題で共通のものであり、課題ごとの特性が、反映されていない。

といった問題点が挙げられる。そこで、本研究では、遷移確率を個人化させ、さらに課題に依存した識別力、難易度パラメータを含む LogisticHMM を提案する。

Yudelson ら (2013)[17] の、BKT の遷移確率を個人化することで、従来の BKT に比べ、課題の正誤予測の精度が向上した研究結果が報告されている。また、学習者の正誤予測を評価するモデルに項目反応理論 (Item Response Theory, IRT と呼ぶ) [19], [20] があり、IRT では、LogisticHMM のように、予測正答確率はロジスティック関数で表される。IRT では識別力、難易度は課題ごとに依存したパラメータである。識別力パラメータは、課題 j が学習者の能力を識別する力を表し、難易度パラメータは、課題 j の難しさを表すものである。従来の LogisticHMM をこれらのパラメータに拡張することで、課題の正誤予測の精度が向上し、学習者に適切な学習支援が行えるモデルになると期待される。

a_j を、課題 j における識別力パラメータ、 b_j を課題 j における難易度パラメータとすると、知識状態が s であるとき、学習者 i が、課題 j に正答する確率は

$$p(X_{ij} = 1 | Z_{ij} = s) = \frac{1}{1 + \exp(-a_j(s/(S-1) - b_j))} \quad (12)$$
$$(s = 0, \dots, S-1)$$

となる。

4 パラメータの推定

本章では，提案モデルである，遷移確率を個人化させ，さらに課題に依存した識別力，難易度パラメータを含む LogisticHMM の推定方法について説明する．BKT や LogisticHMM のような隠れマルコフモデルでは，パラメータの推定方法として，Baum-welch アルゴリズムによる最尤推定法や，マルコフ連鎖モンテカルロ法 (MCMC) による，ベイズ推定法が一般に利用されている [21], [22]．一般に MCMC によるベイズ推定法のほうが頑健であるとされており，本研究では，MCMC の中でも，高効率なアルゴリズムとして知られている周辺化ギブスサンプリング (Collapsed Gibbs Sampling) を用いた方法で，パラメータの推定を行った．CGS はパラメータ集合の一部を周辺消去することによって，効率的にサンプリングができ，隠れマルコフモデルの学習に使われてきた [23], [24]．

CGS は，各パラメータの条件付き事後分布からパラメータ値のサンプリングを行い，得られたサンプルを用いて，パラメータの事後分布を近似する．ここで，CGS における条件付き事後分布とは，結合分布から特定のパラメータ集合を周辺化し，着目しているパラメータ以外をすべて所与とした分布である．

LogisticHMM において，学習者 i が，状態 s から状態 s' へ遷移する確率を $A_{iss'}$ (ただし， $0 \leq A_{iss'} \leq 1$, $\sum_{s'=0}^{S-1} A_{iss'} = 1$) とし， $\mathbf{A}_i = \{A_{i00}, \dots, A_{i,S-1,S-1}\}$ とする．また， $j = 1$ のとき状態 $Z_{i1} = s$ となる確率を π_s (ただし， $0 \leq \pi_s \leq 1$, $\sum_{s=0}^{S-1} \pi_s = 1$) とし， $\boldsymbol{\pi} = \{\pi_1, \dots, \pi_s\}$ とする．また，データ X_{ij} の生起確率は次のように書ける．

$$\begin{aligned} p(X_{ij}|Z_{ij} = s, a_j, b_j) \\ = p(X_{ij} = 0|Z_{ij} = s, a_j, b_j)^{1-X_{ij}} \cdot p(X_{ij} = 1|Z_{ij} = s, a_j, b_j)^{X_{ij}} \end{aligned} \quad (13)$$

LogisticHMM において，周辺化ギブスサンプリングによって，初期確率 $\boldsymbol{\pi}$ と遷移確率 \mathbf{A}_i は周辺化され，隠れ変数 $\mathbf{Z} = \{Z_{11}, \dots, Z_{IJ}\}$ を条件付き事後分布からサンプリングすることで，パラメータの推定値を求める．また，識別力パラメータ $\mathbf{a} = \{a_1, \dots, a_J\}$ と，難易度パラメータ $\mathbf{b} = \{b_1, \dots, b_J\}$ はメトロポリスヘイスティング法を用いてサンプリングを行った．なお，パラメータ $\mathbf{A}_i, \boldsymbol{\pi}$

の事前分布のパラメータをそれぞれ α, β とした。

4.1 状態 Z_{ij} のサンプリング

状態 Z_{ij} の条件付き事後分布は $j = 1$ と $j < 1$ のときで導出される。

1. $n < 1$ のとき

条件付き事後分布は着目しているパラメータ以外をすべて所与としている。 $\mathbf{X}^{\setminus ij} = \mathbf{X} \setminus X_{in}, \mathbf{Z}^{\setminus ij} = \mathbf{Z} \setminus Z_{ij}$ とすると、 $Z_{ij} = s$ が得られる分布は次のようになる。

$$\begin{aligned} p(Z_{ij} = s | X_{ij}, \mathbf{X}^{\setminus ij}, \mathbf{Z}^{\setminus ij}) \\ \propto p(X_{ij} | Z_{ij} = s, \mathbf{X}^{\setminus ij}, \mathbf{Z}^{\setminus ij}) \cdot p(Z_{ij} = s | \mathbf{X}^{\setminus ij}, \mathbf{Z}^{\setminus ij}) \end{aligned} \quad (14)$$

ここで、式 (14) の第 1 項は式 (13) で得られる。式 (14) の第 2 項は、

$$\begin{aligned} p(Z_{in} = s | \mathbf{X}^{\setminus ij}, \mathbf{Z}^{\setminus ij}) &= p(Z_{ij} = s | \mathbf{Z}^{\setminus ij}) \propto p(Z_{ij} = s, Z_{i,j+1} | \mathbf{Z}^{\setminus i,j,j+1}) \\ &\propto p(Z_{i,j+1} | Z_{ij} = s, \mathbf{Z}^{\setminus i,j,j+1}) \cdot p(Z_{ij} = s | Z_{i,j-1}, \mathbf{Z}^{\setminus i,j-1,j,j+1}) \end{aligned} \quad (15)$$

となる。ここで、 $\mathbf{Z}^{\setminus i,j,j+1} = \mathbf{Z} \setminus \{Z_{ij}, Z_{i,j+1}\}$, $\mathbf{Z}^{\setminus i,j-1,j,j+1} = \mathbf{Z} \setminus \{Z_{i,j-1}, Z_{ij}, Z_{i,j+1}\}$ と定義する。式 (15) の右辺第 1 項は次のようになる。

$$\begin{aligned} p(Z_{i,j+1} | Z_{ij} = s, \mathbf{Z}^{\setminus i,j,j+1}) \\ \propto \int p(Z_{i,j+1} | \mathbf{A}_{is}) \cdot p(\mathbf{A}_{is} | \mathbf{Z}^{\setminus i,j,j+1}) d\mathbf{A}_{is} = \frac{n_{i,s,Z_{i,j+1}}^{\setminus j,j+1} + \alpha}{\sum_{s'=0}^{S-1} (n_{i,s,s'}^{\setminus j,j+1} + \alpha)} \end{aligned} \quad (16)$$

ここで、 $n_{i,s,s'}^{\setminus j,j+1}$ は $\mathbf{Z}^{\setminus i,j,j+1}$ のうち、学習者 i が、状態 s から状態 s' へ遷移した頻度を表す。

また、式 (15) の右辺第 2 項は次のようになる。

$$\begin{aligned} p(Z_{ij} = s | Z_{i,j-1}, \mathbf{Z}^{\setminus i,j-1,j,j+1}) \\ \propto \int p(Z_{ij} = s | \mathbf{A}_{i,Z_{i,j-1}}) \cdot p(\mathbf{A}_{i,Z_{i,j-1}} | \mathbf{Z}^{\setminus i,j-1,j,j+1}) d\mathbf{A}_{i,Z_{i,j-1}} \\ = \frac{n_{i,Z_{i,j-1},s}^{\setminus j-1,j,j+1} + \alpha}{\sum_{s'=0}^{S-1} (n_{i,Z_{i,j-1},s'}^{\setminus j-1,j,j+1} + \alpha)} \propto (n_{i,Z_{i,j-1},s}^{\setminus j-1,j,j+1} + \alpha) \end{aligned} \quad (17)$$

ここで, $n_{i,s,s'}^{j-1,j,j+1}$ は, $\mathbf{Z}^{i,j-1,j,j+1}$ のうち, 学習者 i が状態 s から, 状態 s' へ遷移した頻度を表す.

以上より, $j < 1$ における Z_{ij} の条件付き事後分布は次のように表せる.

$$p(Z_{ij} = s | X_{ij}, \mathbf{X}^{\setminus ij}, \mathbf{Z}^{\setminus ij}) \propto p(X_{ij} | Z_{ij} = s, a_j, b_j) \cdot \frac{n_{i,s,Z_{i,j+1}}^{j,j+1} + \alpha}{\sum_{s'=0}^{S-1} (n_{i,s,s'}^{j,j+1} + \alpha)} \cdot (n_{i,Z_{i,j-1},s}^{j-1,j,j+1} + \alpha) \quad (18)$$

2. $n = 1$ のとき

$n = 1$ のとき, (14) の右辺第 2 項は次のようになる.

$$p(Z_{i1} = s | \mathbf{X}^{\setminus i1}, \mathbf{Z}^{\setminus i1}) = p(Z_{i1} = s | \mathbf{Z}^{\setminus i1}) \propto p(Z_{i2} | Z_{i1} = s, \mathbf{Z}^{\setminus i,1,2}) \cdot p(Z_{i1} = s | \mathbf{Z}^{\setminus i,1,2}) \quad (19)$$

式 (19) の右辺第 2 項に関しては, 次のように表せる.

$$p(Z_{i1} = z | \mathbf{Z}^{\setminus i,1,2}) \propto \int p(Z_{i1} = s | \boldsymbol{\pi}) \cdot p(\boldsymbol{\pi} | \mathbf{Z}^{\setminus i,1,2}) d\boldsymbol{\pi} = \frac{n_s^{\setminus i,1,2} + \beta}{\sum_{s'=0}^{S-1} (n_{s'}^{\setminus i,1,2} + \gamma)} \propto (n_s^{\setminus i,1,2} + \beta) \quad (20)$$

$n_s^{\setminus i,1,2}$ は, $\mathbf{Z}^{\setminus i,1,2}$ のうち, 状態 $Z_{i,1}$ が状態 s となる頻度を表す. ($\setminus i$ は i 以外の学習者)

以上より, 状態 Z_{i1} の条件付き事後分布は次式で表せる.

$$p(Z_{i1} = s | X_{i1}, \mathbf{X}^{\setminus i1}, \mathbf{Z}^{\setminus i1}) \propto p(X_{i1} | Z_{i1} = s, a_1, b_1) \cdot \frac{n_{i,z,Z_{i,2}}^{\setminus i,1,2} + \alpha}{\sum_{s'=0}^{S-1} (n_{s,s'}^{\setminus i,1,2} + \alpha)} \cdot (n_s^{\setminus i,1,2} + \beta) \quad (21)$$

4.2 パラメータ \mathbf{a}, \mathbf{b} のサンプリング

パラメータ \mathbf{a}, \mathbf{b} は, メトロポリスヘイスティング法を用いてサンプリングを行う.

まず, 各パラメータの初期値から事前分布をサンプリングする. 各パラメータの事前分布は, 次のように設定する.

$$\log a_j \sim N(1.0, 0.3)$$

$$b_j \sim N(0.0, 1.0)$$

ここで $\boldsymbol{\xi} = \{\mathbf{a}, \mathbf{b}\}$ とし, ξ_j の事前分布を $g(\xi_j)$ と表現する. ξ_j を現在の値 ξ'_j からサンプリングする. サンプリングの提案分布には, $N(\xi'_j, \sigma)$ を用い, 本研究では $\sigma = 0.01$ とする.

次に, 以下の採択確率に基づいて, 候補値 ξ_j をサンプリングする.

$$\alpha(\xi_j | \xi'_j) = \min \left(1, \frac{g(\xi_j) \prod_{i=1}^I p(X_{ij} | Z_{ij}, \xi_j)}{g(\xi'_j) \prod_{i=1}^I p(X_{ij} | Z_{ij}, \xi'_j)} \right) \quad (22)$$

4.3 アルゴリズム

周辺消去した初期確率 $\boldsymbol{\pi}$, 遷移確率 \mathbf{A}_i は, 得られたサンプルを用いて, 次式のように表される.

$$\pi_s = \frac{n_s + \beta}{\sum_{s'=0}^{S-1} (n_{s'} + \beta)} \quad (23)$$

$$A_{iss'} = \frac{n_{iss'} + \alpha}{\sum_{s'=0}^{S-1} (n_{iss'} + \alpha)} \quad (24)$$

n_s は初期状態 Z_{i1} が状態 s となる頻度, $n_{iss'}$ は, 学習者 i が状態 s から s' へ遷移した頻度を表す. 遷移確率は提案モデルの LogisticHMM に置き換えると, 学習者 i が状態 s から $s+1$ へ遷移する確率 l_i は,

$$l_i = \frac{\sum_{s=0}^{S-2} (n_{i,s,s+1} + \alpha)}{\sum_{s=0}^{S-2} (\sum_{s'=0}^{S-1} (n_{iss'} + \alpha))} \quad (25)$$

となる.

提案モデルの CGS は, 隠れ変数 \mathbf{Z} と, パラメータ $\boldsymbol{\xi} = \{\mathbf{a}, \mathbf{b}\}$ のサンプリングを繰り返し, 初期確率 $\boldsymbol{\pi}$, 遷移確率 \mathbf{A}_i を求め, 得られたパラメータ値を EAP 推定することで, パラメータを推定する. 分布が収束したとみなされるまでのバーンイン期間は, パラメータの初期値の影響が残るため, 推定には使わず, また自己相関を考慮し, 一定区間でサンプルを間引くバーンイン期間を設けた. アルゴリズムの疑似コードを Algorithm 1 に示す.

Algorithm 1 提案モデルの CGS

```
Initialize  $Z, a, b$   
for  $m = 1$  to  $M$  do  
  for  $i = 1$  to  $I$  do  
    Sample  $Z_{i1}$  from eq(21)  
    for  $j = 2$  to  $J$  do  
      Sample  $Z_{ij}$  from eq(18)  
    end for  
  end for  
  for  $j = 1$  to  $J$  do  
    Sample  $a_j^m \sim N(a_j^{m-1}, \sigma)$   
    Accept  $a_j^m$  with the probability  $\alpha(a_j^m | a_j^{m-1})$   
    Sample  $b_j^m \sim N(b_j^{m-1}, \sigma)$   
    Accept  $b_j^m$  with the probability  $\alpha(b_j^m | b_j^{m-1})$   
  end for  
  if  $m >$  burn-in period and  $m \% \text{interval} = 0$  then  
    Calculate  $\pi, l$  from eq(23), (25)  
    Store  $\pi, l, a$  and  $b$   
  end if  
end for  
return Average values of  $\pi, l, a$  and  $b$ 
```

5 評価実験

5.1 データ

評価実験では2種類のデータを用いた。

- データ A

このデータはプログラミングの初学者の大学生を対象にしたデータで、学習者数は148人、課題数は7項目となっている。課題ごとにそれぞれ4つのヒントが設定されているが、ここではヒントなしで課題に正答した場合を正答 ($X_{ij} = 1$) とし、それ以外の場合を誤答 ($X_{ij} = 0$) とみなした。

- データ B

データ B は e ラーニングシステム Samurai の受講データを用いた。大学学部3年生を対象とした講義「統計工学基礎」において、16人の学習者が

7項目の課題に回答した。回答データは正答 ($X_{ij} = 1$), 誤答 ($X_{ij} = 0$) の2種類である。

5.2 正誤予測の精度

この節では, 学習者の知識状態の予測精度の評価のために, 学習者の課題に対する正誤の予測をモデルごとに求め, 比較し, モデルの評価を行う。ここで, 比較するモデルは次のとおりである。

- (1) ... 従来の BKT
- (2) ... 遷移確率を個人化させた BKT
- (3) ... 従来の LogisticHMM
- (4) ... 遷移確率を個人化させた LogisticHMM
- (5) ... 識別力, 難易度パラメータを, 課題に依存させたパラメータとした LogisticHMM
- (6) ... 遷移確率を個人化させ, 識別力, 難易度パラメータを, 課題に依存させたパラメータとした LogisticHMM (提案)

実験では, 提案モデルに対し, どの要因が予測精度に影響を及ぼしているかを明らかにするために, 遷移確率を個人化させた LogisticHMM と, 識別力・難易度パラメータを課題に依存させたパラメータとした LogisticHMM の2つを加えて行った。

正誤の予測に関しては, 次の2つの方法で評価する。

(1) 予測正答確率と回答データとの誤差

学習者 i の, 課題 j における予測正答確率 $p(X_{ij} = 1)$ と, 回答データ X_{ij} との誤差 ϵ を求める。誤差 ϵ は, 次式で求められる。

$$\epsilon = \frac{1}{J} \sum_{j=1}^J \frac{1}{I} \sum_{i=1}^I |(X_{ij} - p(X_{ij} = 1))| \quad (26)$$

(2) 予測回答と回答データとの一致率

次のように、学習者 i の課題 j における予測回答 δ_{ij} を定義する.

$$\delta_{ij} = \begin{cases} 0 \text{ (誤答)} & \text{予測正答確率が } 0.5 \text{ 未満} \\ 1 \text{ (正答)} & \text{予測正答確率が } 0.5 \text{ 以上} \end{cases}$$

予測回答と回答データとの一致率を調べることで、学習者の課題に対する正誤予測の精度を評価する.

表 5, 6 にデータ A, データ B における予測正答確率と回答データとの誤差と予測回答と回答データとの一致率をモデルごとに示す. LogisticHMM では知識の状態数 $S = \{2, \dots, 9\}$ と変化させた.

表 5: データ A における正誤予測

誤差	従来の BKT	状態数 S							
	遷移確率を個人化した BKT	2	3	4	5	6	7	8	9
	従来の BKT	0.436							
	遷移確率を個人化した BKT	0.405							
	従来の LogisticHMM	0.435	0.432	0.422	0.418	0.416	0.414	0.414	0.417
	個人パラメータを持つ LogisticHMM	0.398	0.414	0.426	0.434	0.442	0.449	0.454	0.460
	課題パラメータを持つ LogisticHMM	0.395	0.392	0.401	0.398	0.399	0.402	0.415	0.412
	個人・課題パラメータを持つ LogisticHMM	0.381	0.398	0.396	0.410	0.419	0.413	0.418	0.425
一致率	従来の BKT	状態数 S							
	遷移確率を個人化した BKT	2	3	4	5	6	7	8	9
	従来の BKT	64.6%							
	遷移確率を個人化した BKT	62.7%							
	従来の LogisticHMM	64.6%	58.8%	60.0%	60.0%	60.0%	59.9%	58.2%	57.9%
	個人パラメータを持つ LogisticHMM	63.5%	65.2%	62.8%	59.2%	58.0%	57.6%	57.5%	56.2%
	課題パラメータを持つ LogisticHMM	69.7%	69.0%	69.3%	69.0%	66.4%	68.6%	60.5%	65.2%
	個人・課題パラメータを持つ LogisticHMM	66.5%	68.6%	68.7%	66.6%	64.9%	67.0%	62.8%	63.8%

表 6: データ B における正誤予測

誤差	従来の BKT	状態数 S							
	遷移確率を個人化した BKT	2	3	4	5	6	7	8	9
	従来の BKT	0.481							
	遷移確率を個人化した BKT	0.477							
	従来の LogisticHMM	0.432	0.443	0.452	0.459	0.462	0.466	0.471	0.472
	個人パラメータを持つ LogisticHMM	0.422	0.441	0.453	0.459	0.462	0.466	0.469	0.471
	課題パラメータを持つ LogisticHMM	0.399	0.387	0.398	0.408	0.419	0.409	0.416	0.402
	個人・課題パラメータを持つ LogisticHMM	0.382	0.384	0.403	0.387	0.405	0.402	0.423	0.424
一致率	従来の BKT	状態数 S							
	遷移確率を個人化した BKT	2	3	4	5	6	7	8	9
	従来の BKT	59.8%							
	遷移確率を個人化した BKT	59.8%							
	従来の LogisticHMM	59.8%	63.1%	58.9%	54.5%	55.4%	52.7%	54.5%	55.4%
	個人パラメータを持つ LogisticHMM	59.8%	64.0%	58.9%	54.5%	55.1%	52.4%	54.5%	55.4%
	課題パラメータを持つ LogisticHMM	66.7%	68.2%	66.4%	67.8%	65.8%	67.9%	64.3%	68.2%
	個人・課題パラメータを持つ LogisticHMM	67.9%	70.3%	65.2%	69.3%	66.7%	65.8%	64.9%	63.1%

5.3 考察

遷移確率を個人化させた場合、BKT に関しては、両データとも予測正答確率の向上が見られた。また、LogisticHMM に関しては、状態数の少ない場合は、従来の LogisticHMM より精度が向上しているが、状態数が増えると従来の LogisticHMM より、遷移確率を個人化させた LogisticHMM の方が、誤差、一致率ともに精度が悪化しているのがわかる。BKT において、遷移確率を個人化することで、学習者の課題に対する正誤予測の精度が向上した結果に関しては先行研究 [17] のとおりで、LogisticHMM も、状態数 2 の段階では BKT と同じ結果だったが、状態数が増えると予測精度が悪化している。これは状態数を増やすことによって、正答確率を算出する際に用いるパラメータ数が増加し、精度が悪化してしまったと考える。また、識別力、難易度パラメータを課題に依存させた LogisticHMM は、データ A の状態数 8 以外の場合で、従来の LogisticHMM より精度が向上していたことがわかる。このことから、課題の正誤予測の精度を向上させるために、課題に依存したパラメータを含めることが、とても有効な手段であると考えられる。遷移確率の個人化と識別力、難易度パラメータを課題に依存したパラメータとした LogisticHMM も、状態数の少ないうちは、BKT や従来の LogisticHMM より予測精度は良かったが、状態数が増えるとパラメータ数が増加するため、精度が悪くなってしまったと考える。

6 むすび

Bayesian Knowledge Tracing (BKT) や LogisticHMM は、学習者の知識状態を推定することを目的としたモデルだが、BKT は知識状態数が 2 値になっていることによる予測精度の低さや、課題の特徴を反映する値を求めることができないといった問題点があった。LogisticHMM はこれらの問題点を解消するモデルとして提案されたが、従来の LogisticHMM では

- (1) 遷移確率がスキルのみ依存しており、学習者全体で共通の遷移確率を用いるため、学習者ごとの特性が反映されていない。

(2) LogisticHMM は、識別力、難易度パラメータが、すべての課題で共通のものであり、課題ごとの特性が、反映されていない。

といった問題点があり、学習者の知識状態の推定に対し、過大もしくは過小評価が起き、学習者に適切な支援をすることができないという問題点があった。そこで本論文では BKT の遷移確率を個人化することで、学習者に対する課題の正誤予測の向上ができる研究結果や、項目反応理論の背景から、遷移確率を個人化させ、さらに課題に依存した識別力、難易度パラメータを含む LogisticHMM としたモデルを提案し、学習者に対する課題の正誤予測の精度を向上させることを、実データを用いた実験で示した。その結果、

(1) 遷移確率の個人化、識別力、難易度を問題に依存するパラメータとするモデル化は精度予測の向上において、有効である。

(2) ただし、遷移確率の個人化は状態数が増えた場合に、精度が悪くなる。

の2つのことが示された。LogisticHMM では知識の遷移は1段階ずつ行われるが、遷移を2段階以上にしてみたり、長期的な学習では、知識の遷移確率を変化させたり、知識の忘却を考慮してみたりするなど、より学習状況にあったモデルに発展させることを今後の課題としたい。

参考文献

- [1] A.T. Corbett and J.R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User Modeling and User-Adapted Interaction*, vol. 4, no. 4, 1995, pp. 253–278.
- [2] A.T. Corbett, M. McLaughlin, and K.C. Scarpinato, "Modeling student knowledge: Cognitive tutors in high school and college," *User Modeling and User-Adapted Interaction*, vol. 10, no. 2, 2000, pp. 81–108.
- [3] A. Mitrovic and B. Martin, "Evaluating adaptive problem selection," In P.M.E. de Bra and W. Nejdil (Eds.), *Proceeding of the Third International*

- Conference on Adaptive Hypermedia and Adaptive Web-based Systems, AH 2004 (vol. 3137, pp. 185–194), New York: Springer-Verlag (2004).
- [4] C. Conati, A. Gertner, and K. Vanlehn, "Using Bayesian networks to manage uncertainty in student modeling," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, 2002, pp. 371–417.
- [5] Y. Gong, J.E. Beck, and N.T. Hefferman, "How to construct more accurate student models: Comparing and optimizing knowledge tracing and performance factor analysis," *International Journal of Artificial Intelligence in Education*, vol. 21, no. 1-2, 2011, pp. 27–46.
- [6] N. Li, N. Matsuda, W.W. Cohen, and K.R. Koedinger, "A machine learning approach for automatic student model discovery," In M. Pechenizky, T. Calders, C. Conati, S. Ventura, C. Romero, and J. Stamper, (Eds.), *Proceedings of the Fourth International Conference on Educational Data Mining*, 31–40 (2011).
- [7] J.R. Anderson, and B. Reiser, "The LISP tutor," *Byte*, vol. 10, no. 4, 1985, pp. 159–175.
- [8] P. Brusilovsky, "Adaptive and Intelligent technologies for web based education," *Kunstliche Intelligenz*, vol. 4, 1999, pp. 19–25.
- [9] K. Vanlehn, C. Lynch, K. Schultz, J.A. Shapiro, R.H. Shelby, L. Taylor, et al., "The Andes physics tutoring system: Lessons learned," *International Journal of Artificial Intelligence in Education*, vol. 15, no. 3, 2005, pp. 147–204.
- [10] R.S. Crowley, E. Legowski, O. Medvedeva, and E. Tseytlin, "An ITS for medical classification problem-solving: Effects of tutoring and representations," In C.K. Looi, G. McCalla, B. Bredeweg, and J. Breuker, (Eds.), *Proceedings of the 12th International Conference on Artificial Intelligence*, Amsterdam, The Netherlands: AIED (2005).

- [11] R.S. Crowley, and O. Medvedeva, "An intelligent tutoring system for visual classification problem solving," *Artificial Intelligence in Medicine*, vol. 36, no. 1, 2006, pp. 85–117.
- [12] K.R. Koedinger, "Toward evidence for instructional design principles: Examples from Cognitive Tutor Math 6." *Proceedings of PME-NA XXXIII (the North American Chapter of the International Group for the Psychology of Mathematics Education)*, 2002
- [13] J.E. Beck, K.-m. Chang, "Identifiability: A Fundamental Problem of Student Modeling." *Proceedings of the 11th International Conference on User Modeling (UM 2007)*.
- [14] R.S.J.d. Baker, A. T. Corbett, V. Aleven, "More Accurate Student Modeling Through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing." *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 2008, pp. 406-415.
- [15] Z.A. Pardos, and N. T. Heffernan, . "Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing." In: Paul De Bra, Alfred Kobsa, David N. Chin (eds.) *Proceedings of the 18th International Conference on User Modeling, Adaptation, and Personalization (UMAP 2010)*, LNCS vol. 6075 pp. 255- 266. Springer (2010)
- [16] J.I. Lee and E. Brunskill, "The impact on individualizing student models on necessary practice opportunities," In K. Yaccf, O.R. Zaiane, A. Hershkoviz, M. Yudelson, and J.C. Stamer, (Eds.), *Proceedings of the Fifth International Conference on Educational Data Mining*, pp. 118–125 (2012).
- [17] M.V. Yudelson, K.R. Koedinger, G.J. Gordon, "Individualized Bayesian Knowledge Tracing Models," *Artificial Intelligence in Education (2013)*, Springer, pp. 171-180.

- [18] R. Pelánek., “Conceptual Issues in Mastery Criteria: Differentiating Uncertainty and Degrees of Knowledge.” *Artificial Intelligence in Education*, pp. 450-461 (2018).
- [19] F.M. Lord, *Applications of item response theory to practical testing problems*, Mahwah, NJ: Lawrence Erlbaum Associates, Inc,1980.
- [20] F.B. Baker, and S. Kim, *Item Response Theory: Parameter Estimation Techniques*, Second Edition, NY: Marcel Dekker, Inc,2004.
- [21] C.M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, 2006.
- [22] 片平健太郎, 岡ノ谷一夫, 岡田真人, ” 隠れマルコフモデルを用いた行動・神経活動からの状態遷移抽出, ” *日本神経回路学会誌*, vol.18, no.4, pp.204-213, 2011.
- [23] J.Paisley and L.Carin, “Hidden markov models with stick-breaking priors,” *IEEE Transactions on Signal Processing*, vol.57, no.10, pp3905-3917, 2009.
- [24] 宇都雅輝, 宮澤芳光, 加藤嘉浩, 中島功滋, 桑田一, “キーストロークデータに基づくライティングプロセス推定のための執筆者依存型隠れマルコフモデル.” *教育システム情報学会第43回全国大会*. pp59-60, 2018.