

認知診断モデルの変分ベイズ推定とその応用について

○岡田 謙介¹ 岡 元紀² 土方 啓一郎¹ 福島 健太郎¹ 東口 琳妃¹ 山口 一大³

¹東京大学大学院教育学研究科

²London School of Economics and Political Science

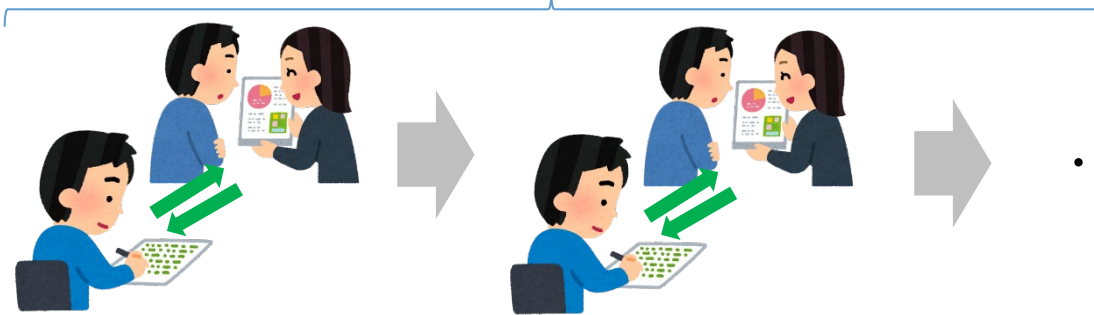
³筑波大学人間系

形成的評価と総括的評価

■ 認知診断モデル(CDM)

$$P(y_{ij} = 1 | \alpha_i) = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}}, \eta_{ij} = \prod_{k=1}^K \alpha_{c_{ik}}^{q_{jk}}$$

形成的評価



■ 項目反応モデル(IRT)

$$P(y_{ij} = 1 | \theta_i) = \text{logit}^{-1}(a_j(\theta_i - b_j))$$

総括的評価



	認知診断モデル CDM (cognitive diagnostic model, diagnostic classification model)	項目反応モデル IRT (item response theory)
利用目的	形成的評価	総括的評価
潜在変数	2値 (習得 / 未習得)	連続
-の粒度	小さめ (例: 分母の有理化)	大きめ (例: 数学力)
-の次元数	多次元	基本は1次元

本日の流れとTakeaways

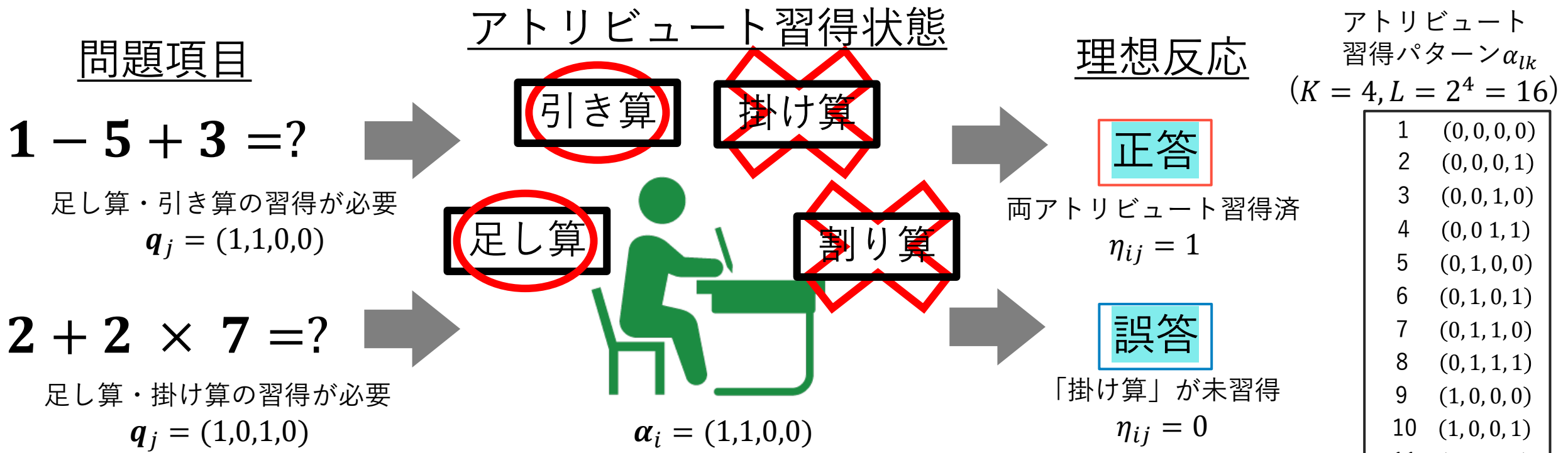
- 認知診断モデルの概要（DINAモデル）
- 提案する変分ベイズ推定法
- 拡張: 1. 一般化CDMでの推定
2. テスト構造（Q行列）推定
3. 多肢選択情報活用モデル
- 応用 1. Rパッケージ
2. 漢字の学習診断

■ Takeaways:

- 認知診断モデルは、多次元の学習要素習得状態を推定し、形成的評価に活用するための有用な枠組み
- 変分ベイズ推定により高速で実用的な推定が実現できる
- 一般化モデル/Q行列推定/多肢選択型モデルなど、多様な状況への適用が可能で様々な拡張が提案できる
- パッケージ開発や共同研究・社会実装により応用の進展が期待される

DINAモデル (最も基本的な認知診断モデル)

(deterministic inputs, noisy "and" gate model; Junker & Sijtsma, 2001)



■ 理想反応に確率的ノイズが加わり項目反応 y_{ij} が観測される

- 項目特徴 {
- Guessing (当て推量) の確率 g_j
 - Slip (うっかりミス) の確率 s_j

DINAモデル（最も基本的な認知診断モデル）

DINAモデルの項目反応関数は

$$P(y_{ij} = 1 | s_j, g_j, \mathbf{z}_i; \mathbf{q}_j, \boldsymbol{\alpha}_l) = \left((1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}} \right)^{y_{ij}} \left(s_j^{\eta_{ij}} (1 - g_j)^{1 - \eta_{ij}} \right)^{1 - y_{ij}}$$

- y_{ij} : 項目反応（2値, $N \times J$; 観測データ）
- q_{jk} : Q行列（2値, $J \times K$; 事前指定済みのテスト構造）
- $L = 2^K$: 潜在クラス数（可能な全アトリビュート習得状態パターン数）
- α_{lk} : アトリビュート習得パターン（2値, $L \times K$ ）
- z_{il} : クラス所属インディケータ（2値, $N \times L$, $\sum_{l=1}^L z_{il} = 1$; 潜在変数）
- $\eta_{lj} = \sum_{k=1}^K \alpha_{lk}^{q_{jk}}$: 潜在クラス l の理想反応（項目 j へ本来正答できるか）
- $\eta_{ij} = \boldsymbol{\eta}_j^T \mathbf{z}_i = \sum_{l=1}^L \eta_{lj} z_{il}$: 解答者 i の理想反応
- $g_j = P(y_{ij} = 1 | \eta_{ij} = 0)$: 当て推量での正答確率（連続, $J \times 1$, パラメータ）
- $s_j = P(y_{ij} = 0 | \eta_{ij} = 1)$: うっかりの誤答確率（連続, $J \times 1$, パラメータ）

DINAモデルの事前分布設定と変分ベイズ推論

- $P(s_j | \alpha_s, \beta_s) \propto s_j^{\alpha_s - 1} (1 - s_j)^{\beta_s - 1}$ (ベータ分布)

- $P(g_j | \alpha_g, \beta_g) \propto g_j^{\alpha_g - 1} (1 - g_j)^{\beta_g - 1}$ (ベータ分布)

- $P(\mathbf{z}_i | \boldsymbol{\pi}) \propto \prod_l \pi_l^{z_{il}}$ (カテゴリカル分布)

- $P(\boldsymbol{\pi} | \boldsymbol{\delta}) \propto \prod_l \pi_l^{\delta_l - 1}$ (ディリクレ分布)

- 事後分布に対して次の仮定 (平均場近似) を置くことを提案

- $q(\mathbf{Z}, \boldsymbol{\pi}, \mathbf{s}, \mathbf{g}) = q(\mathbf{Z})q(\boldsymbol{\pi}, \mathbf{s}, \mathbf{g}) = (\prod_i q(\mathbf{z}_i))q(\boldsymbol{\pi})(\prod_j q(s_j)q(g_j))$

⇒ 最適変分事後分布の各要素が事前分布と同じ分布族になる

- たとえば $q^*(s_j) \propto s_j^{\alpha_{s_j}^* - 1} (1 - s_j)^{\beta_{s_j}^* - 1}$ (ベータ分布)

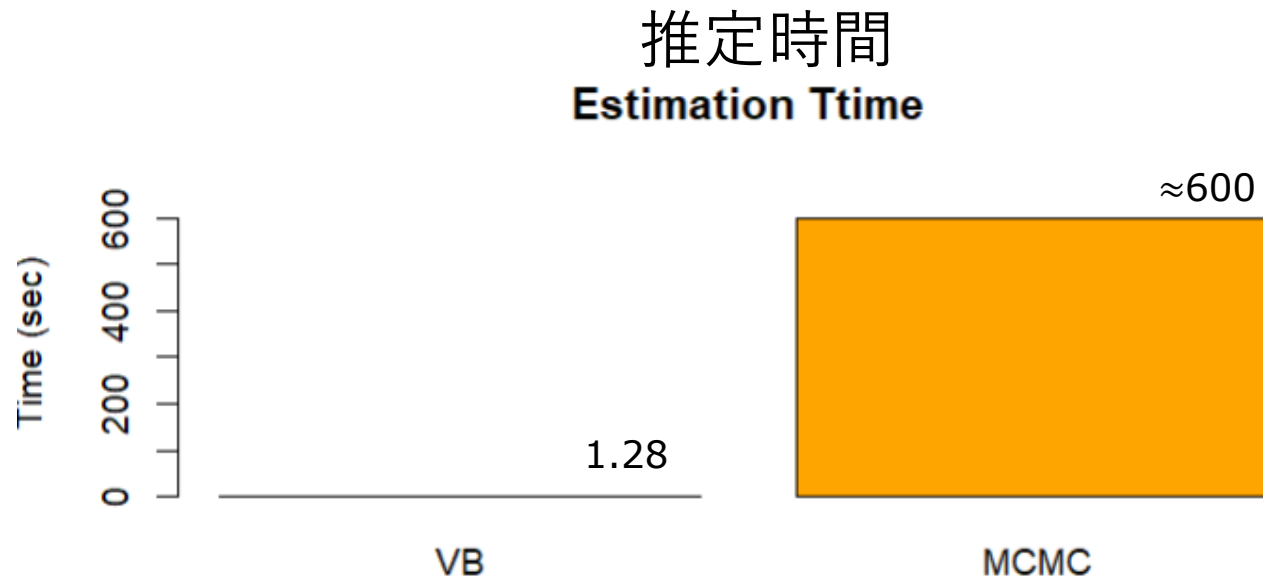
$$\alpha_{s_j}^* = \sum_i \boldsymbol{\eta}_j^\top \mathbb{E}_{q(\mathbf{z}_i)}[\mathbf{z}_i] (1 - y_{ij}) + \alpha_s, \beta_{s_j}^* = \sum_i \boldsymbol{\eta}_j^\top \mathbb{E}_{q(\mathbf{z}_i)}[\mathbf{z}_i] y_{ij} + \beta_s$$

- これを利用して、潜在変数とパラメータの更新、期待値の更新を繰り返す変分EMアルゴリズムを構築できる

DINAモデルの変分ベイズ推論

推定値

- 分数の引き算の解答データ (Tatsuoka, 2002) を分析. $N = 536, K = 8$



Item	VB		MCMC	
	Guessing (SD)	Slip (SD)	Guessing (SD)	Slip (SD)
1	.040 (.013)	.079 (.016)	.045 (.016)	.078 (.018)
2	.023 (.010)	.041 (.011)	.024 (.015)	.040 (.012)
3	.007 (.005)	.133 (.019)	.007 (.007)	.132 (.020)
4	.234 (.024)	.111 (.020)	.237 (.026)	.110 (.021)
5	.306 (.029)	.159 (.022)	.309 (.030)	.151 (.024)
6	.154 (.034)	.040 (.009)	.198 (.054)	.038 (.010)
7	.033 (.011)	.200 (.025)	.036 (.014)	.200 (.028)
8	.435 (.047)	.172 (.018)	.432 (.048)	.164 (.019)
9	.221 (.040)	.249 (.021)	.236 (.059)	.248 (.023)
10	.035 (.011)	.199 (.025)	.036 (.012)	.191 (.028)
		⋮		⋮

- 提案手法はMCMC法と同等の推定値をはるかに**高速に推定**
- 平均場近似による事後SDの過小評価は見られるが、その程度は小さくなく、提案した変分ベイズ推定は**実用的方法**と言える

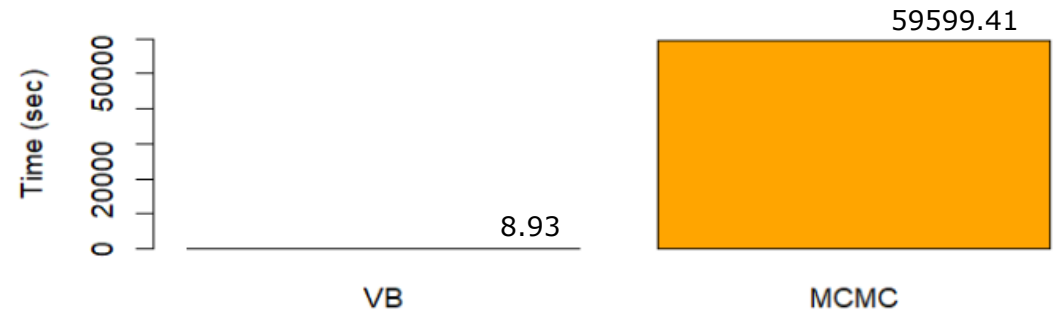
一般化モデルへの拡張

- 一般化CDM (G-DINA model, de la Torre, 2011; LCDM, Henson et al., 2009) へ提案を拡張

$$\text{logit}[P(y_{ij} = 1 | \delta_j, \alpha_l)] = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'} \alpha_{lk} \alpha_{lk'} + \dots + \delta_{j12\dots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}$$

切片 効果 1次の相互作用 最高次の相互作用

- DINAモデルをはじめとする各種の2値CDMは、この一般化CDMの下位モデルとして表現できる



- 英語運用能力を診断するテストデータ (ECPE) の分析例 (フルモデル, $N = 2922, K = 3$)

Item	Variational expectation-maximization algorithm				Markov chain Monte Carlo algorithm			
	P(0)	P(1)			P(0)	P(1)		
	P(00)	P(10)	P(01)	P(11)	P(00)	P(10)	P(01)	P(11)
1	.697 (.013)	.801 (.016)	.515 (.065)	.937 (.007)	.692 (.016)	.814 (.028)	.810 (.075)	.930 (.012)
2	.732 (.012)	.905 (.007)			.738 (.015)	.907 (.010)		
3	.415 (.016)	.500 (.017)	.604 (.072)	.782 (.013)	.416 (.019)	.503 (.022)	.624 (.123)	.786 (.016)
4	.461 (.016)	.822 (.009)			.466 (.019)	.825 (.010)		

- 提案手法の高速性、実用性がDINAの場合と同様示された

各種の2値CDM

モデル	反応データ	アトリビュートのカテゴリ水準	アトリビュート以外の潜在変数	補償・非補償	応用研究例	1項目あたりの項目パラメータ数
DINA	2値	2値	無	非補償	Lee, Park, & Taylan (2011)	2
DINO	2値	2値	無	補償	Templin & Henson (2006)	2
NIDA	2値	2値	無	非補償	なし	$2K_j^*$
NIDO	2値	2値	無	補償	なし	$2K_j^*$
LLM	2値	2値	無	補償	Chen & de la Torre (2014)	$1 + K_j^*$
A-CDM	2値	2値	無	補償	Chen & de la Torre (2014)	$1 + K_j^*$
R-RUM	2値	2値	無	非補償	Jang (2009)	$1 + K_j^*$
C-RUM	2値	2値	無	補償	山口 (2016)	$1 + K_j^*$
G-DINA	2値	2値	無	飽和	鈴木他 (2015)	2^{K_j}
LCDM	2値	2値	無	飽和	Templin & Hoffman (2013)	2^{K_j}

多値への拡張・アトリビュート以外の潜在変数を含む拡張も各種提案されている

Q行列推定への拡張

- Q行列で指定されるテスト構造（下表）は、従来のCDMの枠組みでは、専門家の知見や予備調査に基づき事前指定済とされていた

項目数 $J=4$, アトリビュート数 $K=4$ の場合の
テスト構造を表す2値行列(Q行列)の例

4次元の学習要素（アトリビュート）

項目	足し算	引き算	かけ算	割り算
3 + 2 - 1 = ?	1	1	0	0
$\frac{6}{3} = ?$	0	0	0	1
5 × 3 - 3 = ?	0	1	1	0
4 + 16 = ?	1	0	0	0

4項目

- しかしこれは強い仮定の一つと言え、専門家でも確信がないことも多い
⇒ テスト構造のデータドリブンでスケーラブルな推定法が求められる
従来の課題：解空間の離散性と広さに由来する計算コスト

推定対象は0と1の値をとる2値(離散)行列であり、その可能なパターン数は $(2^K - 1)^J$ 通り

たとえば5次元20項目のテストの場合、解は $6.7179053e+29$ 通り存在

Q行列推定への拡張

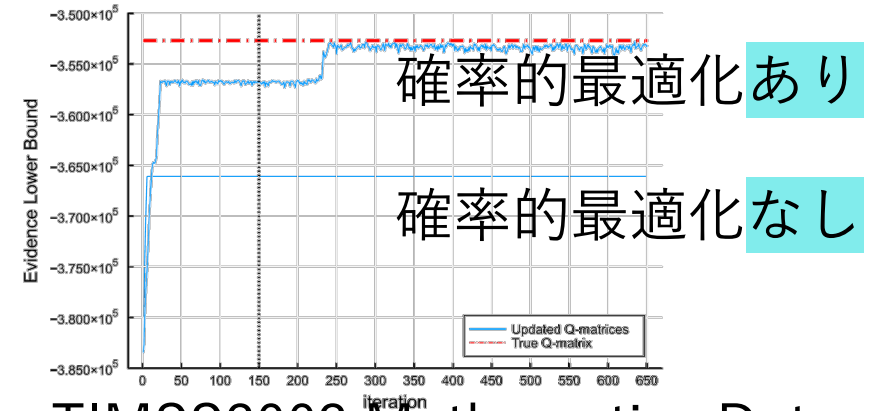
- 問題を対数周辺尤度の周辺化問題として定式化し、確率的近似と変分近似を利用して対数周辺尤度を最大化する、スケーラブルな反復計算による推定アルゴリズムを開発

対数周辺尤度

変分近似

$$\log p(\mathbf{X}|\mathbf{Q}) = E \left[\frac{N}{S} \log p(\mathbf{X}_S|\mathbf{Q}) \right] \geq E \left[\frac{N}{S} L_{\max}(\mathbf{X}_S|\mathbf{Q}) \right]$$

確率的近似 これを最大化



Fraction Subtraction Dataset (Tatsuoka, 1983)

TIMSS2003 Mathematics Dataset

	Proposed method	Gibbs sampler (Chung, 2019)	EM-based algorithm (Chen et al., 2015)
AIC	7460.52	7520.70	7489.86
BIC	7670.44	7730.62	7699.78
WAIC	6268.58	6390.34	6320.25
WBIC	310.10	317.60	313.36
Negative ELBO	3834.33	3855.40	3840.05
Strictly Identifiable?	Yes	No	No
		(p-partially identifiable)	(p-partially identifiable)
Time (seconds)	24	1065	23

	Proposed method	Expert knowledge (Su et al., 2003)
AIC	35420.71	35743.76
BIC	73552.78	73875.83
WAIC	18581.60	19291.64
WBIC	803.34	855.29
Negative ELBO	10752.03	10929.17
Strictly Identifiable?	No	No
Time	4h 30min	-

Note. To compute WAIC and WBIC, we run MCMC estimation using four chains with 5,000 iterations and 2,000 burn-in.

Note. To compute WAIC and WBIC, we run MCMC estimation using four chains with 5,000 iterations and 2,000 burn-in.

➡ 提案手法は高速に高適合な解を探索できていた

多枝選択解答情報を活用できるCDMへの拡張

- 従来型のCDMは観測データを「正答／誤答」の2値でコーディング
⇒ どの選択肢を選んで誤答したのかという情報を活用できない

		観測データ
Nicholas has a very tiring job and always looks very ().	1 tiring 2 tired 3 excited 4 exciting	$y_1 = 0$ (誤答)
Janis () New Orleans when I realized that.	1 has already left 2 had already left 3 had already left from 4 has already left from	$y_2 = 1$ (正答)
She is the lady () the day before yesterday.	1 whom I meet 2 that I met 3 which I met 4 which I meet	$y_3 = 0$ (誤答)
		⋮

各問題が持つ項目パラメータ値 (例: slip, guessing) を考慮したCDMで習得状態を推定

時制○ 語彙× 関係詞×

- ⇒ 構造化された多枝選択型DINAモデル(Ozaki, 2015)に基づき、部分的な知識習得状態を表現できるよう拡張を行う

多肢選択解答情報を活用できるCDMへの拡張

- 習得が求められる **学習要素** (欠損でない列)
 - 当該要素を習得済のとき **選べる選択肢** (1の要素)
- を各選択肢に付与

		時制	語彙	分詞	関係詞
Nicholas has a very tiring job and always looks very ().	1 tiring		1	0	
	2 tired		1	1	
	3 excited		0	1	
	4 exciting		0	0	
Janis () New Orleans when I realized that.	1 has already left	0	1		
	2 had already left	1	1		
	3 had already left from	1	0		
	4 has already left from	0	0		
She is the lady () the day before yesterday.	1 whom I meet	0			1
	2 that I met	1			1
	3 which I met	1			0
	4 which I meet	0			0

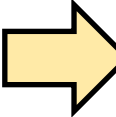
MC-PK-DINA
モデル

{ PK1-S
PK1-M*

$$\begin{aligned}
 & P(y_{ijc} = 1 | \alpha_i) \\
 &= \frac{1 - \sigma_j}{R_{ij}} \eta_{ijc} + \frac{\sigma_j}{C_j}
 \end{aligned}$$

$$\eta_{ijc} = \prod_{k=1}^{K_j} q_{jck}^{\alpha_{ik}}$$

$$R_{ij} = \sum_{c=1}^{C_j} \eta_{ijc}$$


 部分的な知識習得状態を考慮し、誤答情報を活用する
 2種の多肢選択型認知診断モデルを開発した

多肢選択解答情報を活用できるCDMへの拡張

- 検証のため、教育心理の専門家監修のもと、時制・語彙・分詞・関係詞の4学習要素を診断する英文法に関する4肢選択型テストを作成・実施

学習要素の下位構成要素

時制	1-1	時制の選択
	1-2	適切な時制の形への変形
	1-3	動詞の基本的な形への変形
語彙	2-1	自動詞と他動詞の区別
	2-2	単語の意味
	2-3	動詞と前置詞の組み合わせ
	2-4	可算・不可算名詞の区別
分詞	3-1	現在分詞・過去分詞の使い分け
	3-2	受動態の知識・使用
関係詞	4-1	関係詞の選択
	4-2	不完全文とすること
	4-3	省略・語順等の規則の把握
	4-4	関係詞の必要性の判断

実際の出題画面例

2. The flag is () every morning by the person in charge.*

- a. rising
- b. risen
- c. raising
- d. raised

次に進む

Model	WAIC	WBIC
MC-S-DINA1	18,477.3	9,244.8
MC-PK-DINA1-S	18,343.1	9,174.4
MC-PK-DINA1-M	18,338.2	9,170.6

提案モデル2つが、既存モデルよりも優れた適合

Mastery Profile	MC-S-DINA1	MC-PK-DINA1-S	MC-PK-DINA1-M
(0, 0, 0, 0)	48	18	19
(1, 1, 1, 1)	438	332	341
Total	500	500	500

$N = 500$

※解答者には一定の英語力水準を要求した
(英検準2級もしくはTOEIC 450点以上)
そのため正答率は約75%と比較的高かった

既存モデルでは習得状態の過大診断が見られたのに対し、提案モデル2種はより妥当と考えられる認知診断を行った

➡ 提案した多肢選択モデルの有効性が実テストデータで確認された

変分ベイズ推定Rパッケージの開発

- 開発した各種変分ベイズ推定法を実装したRパッケージを開発・CRAN公開

既存の実装よりも **高速・高効率** なことを確認

Package 'variationalDCM'

March 25, 2024

Type Package

Title Variational Bayesian Estimation for Diagnostic Classification Models

Version 2.0.1

Description Enables computationally efficient parameters-estimation by variational Bayesian methods for various diagnostic classification models (DCMs). DCMs are a class of discrete latent variable models for classifying respondents into latent classes that typically represent distinct combinations of skills they possess. Recently, to meet the growing need of large-scale diagnostic measurement in the field of educational, psychological, and psychiatric measurements, variational Bayesian inference has been developed as a computationally efficient alternative to the Markov chain Monte Carlo methods, e.g., Yamaguchi and Okada (2020a) <doi:10.1007/s11336-020-09739-w>, Yamaguchi and Okada (2020b) <doi:10.3102/1076998620911934>, Yamaguchi (2020) <doi:10.1007/s41237-020-00104-w>, Oka and Okada (2023) <doi:10.1007/s11336-022-09884-4>, and Yamaguchi and Martinez (2023) <doi:10.1111/bmsp.12308>. To facilitate their applications, 'variationalDCM' is developed to provide a collection of recently-proposed variational Bayesian estimation methods for various DCMs.

```
# fit the DINA model
Q = sim_Q_J80K5
sim_data = dina_data_gen(Q=Q,I=200)
res = variationalDCM(X=sim_data$X, Q=Q, model="dina")
summary(res)
```

Item	variationalDCM		dina	
	\hat{s}_j (SD)	\hat{g}_j (SD)	\hat{s}_j (SD)	\hat{g}_j (SD)
1	.079 (.016)	.040 (.013)	.078 (.018)	.045 (.016)
2	.041 (.011)	.023 (.010)	.040 (.012)	.024 (.015)
3	.133 (.019)	.007 (.005)	.132 (.021)	.007 (.007)
4	.111 (.020)	.234 (.024)	.111 (.021)	.237 (.026)
5	.159 (.022)	.306 (.029)	.151 (.024)	.308 (.029)
6	.040 (.009)	.154 (.034)	.038 (.010)	.198 (.052)
7	.200 (.025)	.033 (.011)	.200 (.028)	.036 (.014)
8	.172 (.018)	.435 (.047)	.164 (.019)	.432 (.048)
9	.249 (.021)	.221 (.040)	.248 (.023)	.235 (.059)
10	.199 (.025)	.035 (.011)	.190 (.028)	.036 (.012)
11	.084 (.017)	.069 (.015)	.084 (.019)	.068 (.016)
12	.038 (.010)	.215 (.031)	.038 (.012)	.266 (.048)
13	.335 (.031)	.017 (.007)	.335 (.031)	.017 (.009)
14	.051 (.011)	.125 (.027)	.045 (.013)	.161 (.039)
15	.111 (.019)	.032 (.011)	.112 (.023)	.033 (.016)
16	.099 (.015)	.158 (.030)	.091 (.017)	.184 (.037)
17	.141 (.022)	.042 (.012)	.141 (.023)	.043 (.014)
18	.135 (.022)	.127 (.020)	.135 (.023)	.130 (.021)
19	.234 (.029)	.026 (.009)	.232 (.032)	.027 (.009)
20	.146 (.023)	.018 (.008)	.143 (.026)	.019 (.009)
time (sec.)		2.43		739.93

➡ 開発したスケーラブルな推定法を広く利用可能にした

(Hijikata, Oka, Yamaguchi, & Okada, 2024, PsyArXiv; Okada, Hijikata, Yamaguchi, & Okada, AMPC2024)

社会実装に向けた取り組み

日本漢字能力検定協会との共同研究 (2024/05～)

漢検

日本漢字能力検定

開発した認知診断モデルにより
漢字学習の背後にある学習要素を
特定し、学習支援へ繋げる

(一) 次の一線の漢字の読みをひらがなで記せ。

1 問題の本質を喝破する。
2 自然の恵みを享受する。
3 欠陥のある製品を急ぎ回収する。
4 元大統領が自叙伝を出版した。
5 国の財政が窮迫してきた。

(二) 次の漢字の部首を記せ。

(六) 次の一線のカタカナを漢字に直せ。...

例) 菜 問 門

5 隼 艇 蛭 缶 隸
4
3
2
1

10 乏 豪 閑 義 漠
9
8
7
6

異なる実施回の問題を同一尺度
上で評価する**等化法**の開発

学習対象**アトリビュート**の特定
出題形式に適した**モデル**開発

ことばの学びの**診断的・形成的評価**を
実現するために研究開発中

まとめとTakeaways

- 認知診断モデルの概要（DINAモデル）
- 提案する変分ベイズ推定法
- 拡張: 1. 一般化認知診断モデル
2. テスト構（Q行列）推定
3. 多枝選択型モデル
- 応用 1. Rパッケージ
2. 漢字の学習診断

■ Takeaways:

- 認知診断モデルは、多次元の学習要素習得状態を推定し、形成的評価に活用するための有用な枠組み
- 変分ベイズ推定により高速で実用的な推定が実現できる
- 一般化モデル/Q行列推定/多枝選択型モデルなど、多様な状況への適用が可能で様々な拡張が提案できる
- パッケージ開発や共同研究・社会実装により応用の進展が期待される

謝辞

- 本研究の一部は以下の助成を受けたものです。記して感謝いたします。
 - JSTさきがけ 信頼されるAIの基盤技術領域 「透明性の高い達成度テスト運営基盤の開発」
 - JSPS科研費 基盤(B) 「学習の躓きを診断し克服するための統計学的方法：デジタル問題集の基盤開発と実践」
 - JSPS科研費 若手研究 「テイラーメイド学習環境を実現する認知診断モデルの提案と適用」
 - 漢字能力検定協会共同研究 「『認知診断モデル』を応用した評価ツール開発」

