

# AIアプローチにおけるIRT

植野 真臣

電気通信大学 情報理工学研究科

本発表は一貫してグラフィカルモデルとしてのIRTの提案を行っているので 以下のタイトルと受け取ってもらってもよいです。

# グラフィカルモデルアプローチに おけるIRT

植野 真臣

電気通信大学 情報理工学研究科

# 1. 内容 確率的アプローチ v.s. 深層学習アプローチ

## 深層学習アプローチ

1. 反応予測精度と解釈性を満たすDeep-IRT: Deep Learning + IRT
2. 能力変化を反映したKnowledge TracingのためのDeep-IRT

## 確率的アプローチ

3. 正答確率予測を保証するBayesian network Classifier (BNC)と Sliding window Bayesian network Knowledge Tracing
4. Bayesian network IRT

## 2. 能力推定・反応予測の従来手法

### 確率的アプローチ

- 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, pp. 253-278(1995)
- W. J. van der Linden: *Handbook of Item Response Theory, Volume One: Models*, 1st ed., Boca Raton, FL, USA: Chapman & Hall (2016).
- R. Pelanek : Conceptual Issues in Mastery Criteria: Differentiating Uncertainty and Degrees of Knowledge. *Artificial Intelligence in Education* , pp. 450-461 (2018).

### ディープラーニングアプローチ

- C.Piech,J.Bassen,, S.Ganguli,"Deep Knowledge Tracing", *Advances in Neural Information Processing Systems 28 (NIPS)*,2015.
- 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, 2017.
- Song Cheng and Qi Liu "Enhancing Item Response Theory for Cognitive Diagnosis," *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, ACM* , 2019
- Converse G., Pu S., Oliveira S. "Incorporating Item Response Theory into Knowledge Tracing," *Artificial Intelligence in Education. AIED 2021*.
- 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*,2019.
- A. Ghosh, N. Heernan, and A. S. Lan. Context-aware attentive knowledge tracing. In *Proceedings of the 26<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020.

# 2.1. 項目反応理論 (Item Response Theory : IRT)

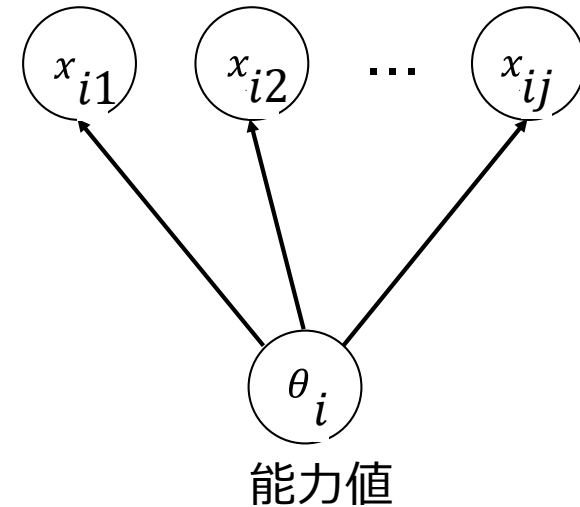
学習者*i*の項目*j*への予測正答確率

$$P(u_j = 1 | \theta_i) = \frac{1}{1 + \exp(-a_j \theta_i + b_j)}$$

項目識別力 (red box) points to  $a_j$   
項目の難易度 (orange box) points to  $b_j$   
学習者の能力値 (green box) points to  $\theta_i$

$a_j$ は能力と項目の関係性 (項目識別力)  
 $b_j$ は項目*j*の難しさ (項目困難度)

学習者の課題への反応



長所：解釈性の高さ

短所：

- Deep Learningに比べて 反応予測精度が低い
- 局所独立性が仮定されているが、現実には項目連鎖、文脈依存を回避できない
- 受検者の同一母集団からの独立ランダムサンプリングが仮定されているが、現実には偏ったテストデータがサンプルされることが多い

## 2.2. Deep Learning による反応予測

- C. Piech et al., “Deep knowledge tracing,” in Proc. Adv. Neural Inf. Process. Syst., NeurIPS2015, pp. 505–513.
- J. Zhang, X. Shi, I. King, and D.-Y. Yeung, “Dynamic key-value memory network for knowledge tracing,” in Proc. 26th Int. Conf. World Wide Web, WWW 2017, pp. 765–774。
- A. Ghosh, N. Heffernan, and A. S. Lan, “Context-aware attentive knowledge tracing,” in Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, KDD 2020, pp. 2330–2339 : Transformerモデル

長所：反応予測精度が高い

短所：解釈性がない

# 3.1. Deep-IRT( Tsutsumi 2021)

学習者の能力値パラメータと項目難易度パラメータを  
独立のニューラルネットワークで推定することで予測精度の高いままパラ  
メータの解釈性を持つ

学習者*i*の能力値 $\theta^{(i)}$

$$\theta_1^{(i)} = \tanh(W^{(\theta_1)} s_i + \tau^{(\theta_1)}).$$

$$\theta_2^{(i)} = \tanh(W^{(\theta_2)} \theta_1^{(i)} + \tau^{(\theta_2)}).$$

$$\theta_3^{(i)} = W^{(\theta_3)} \theta_2^{(i)} + \tau^{(\theta_3)}.$$

課題*j*の難易度 $\beta^{(j)}$

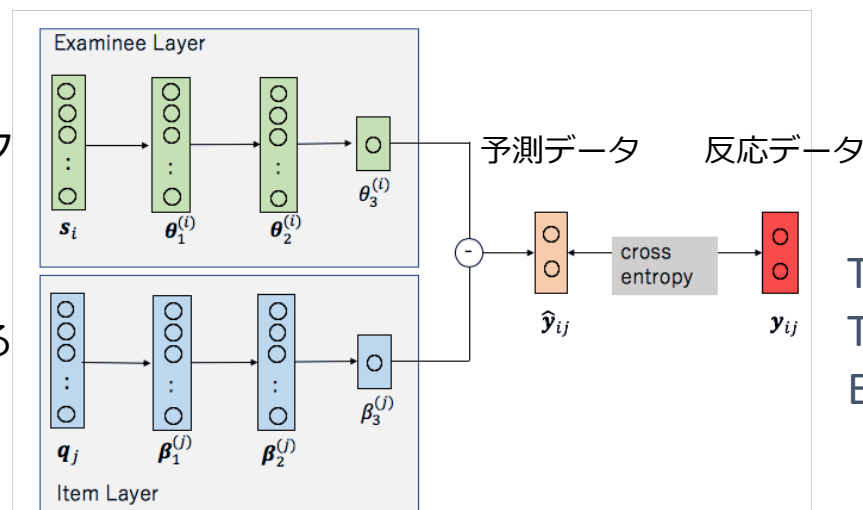
$$\beta_1^{(j)} = \tanh(W^{(\beta_1)} q_j + \tau^{(\beta_1)}).$$

$$\beta_2^{(j)} = \tanh(W^{(\beta_2)} \beta_1^{(j)} + \tau^{(\beta_2)}).$$

$$\beta_3^{(j)} = W^{(\beta_3)} \beta_2^{(j)} + \tau^{(\beta_3)}.$$

能力値を推定する  
学習者ネットワーク

項目難度を推定する  
項目ネットワーク



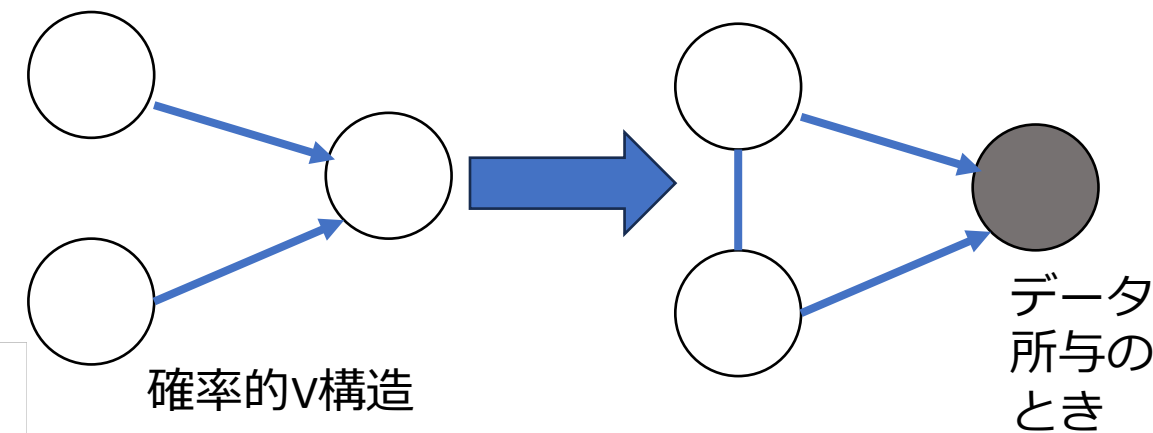
Tsutsumi, E., Kinoshita, R., and Ueno, M., Deep Item Response Theory as a Novel Test Theory Based on Deep Learning. Electronics, no.10, pp.1020, (2021)

# 3.2. Deep-IRTの数理的特徴

学習者の能力値パラメータと項目難易度パラメータを

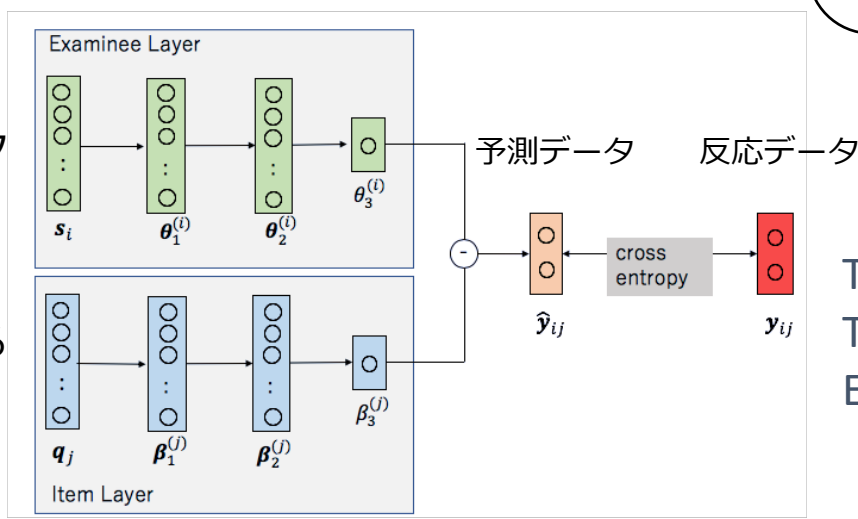
独立のニューラルネットワークで推定することで予測精度の高いままパラメータの解釈性を持つ

1. 従来のIRTは確率生成モデルであるのに対し、Deep-IRTは識別モデルである。
2. Deep-IRTは グラフィカルモデルの数学的性質の応用である。グラフィカルモデルのV構造を持ち、パラメータ推定時には学習者ネットワークと項目ネットワークのすべてのパラメータは相互依存するが、予測時には学習者ネットワークと項目ネットワークは独立に働く。



能力値を推定する  
学習者ネットワーク

項目難度を推定する  
項目ネットワーク

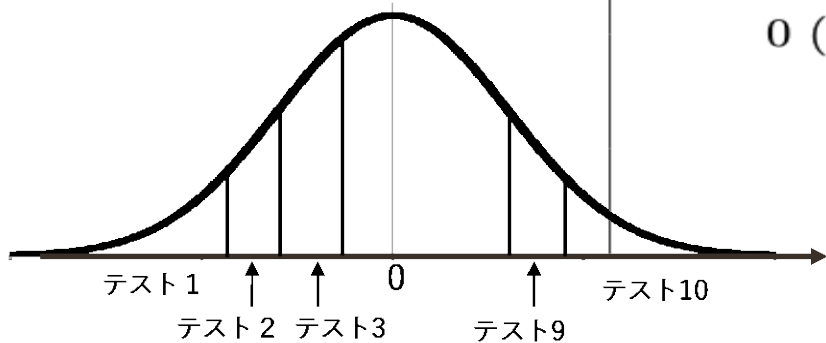


Tsutsumi,E., Kinoshita,R., and Ueno,M., Deep Item Response Theory as a Novel Test Theory Based on Deep Learning. Electronics,no.10, pp.1020,(2021)



# 3.3. シミュレーション(ランダムサンプリングなし)での予測精度、信頼性の比較

項目数	共通項目数	受検者数	手法	RMSE	Pearson	Kendall
50	5 (455)	50 (500)	DeepIRT	<b>0.376</b>	<b>0.929</b>	<b>0.802</b>
			IRT	0.426	0.909	0.760
		100 (1000)	DeepIRT	<b>0.393</b>	<b>0.923</b>	<b>0.811</b>
			IRT	0.805	0.750	0.543
	500 (5000)	DeepIRT	<b>0.372</b>	<b>0.930</b>	<b>0.810</b>	
		IRT	1.044	0.454	0.282	
	1000 (10000)	DeepIRT	<b>0.392</b>	<b>0.914</b>	<b>0.798</b>	
		IRT	0.923	0.512	0.342	
0 (500)	50 (500)	DeepIRT	<b>0.635</b>	<b>0.798</b>	<b>0.599</b>	
		IRT	0.782	0.694	0.489	
	100 (1000)	DeepIRT	<b>0.408</b>	<b>0.916</b>	<b>0.785</b>	
		IRT	0.612	0.812	0.532	
	500 (5000)	DeepIRT	<b>0.421</b>	<b>0.891</b>	<b>0.765</b>	
		IRT	0.598	0.822	0.495	
	1000 (10000)	DeepIRT	<b>0.411</b>	<b>0.901</b>	<b>0.785</b>	
		IRT	0.602	0.829	0.498	



# 3.4. 実データでの予測精度、信頼性の比較

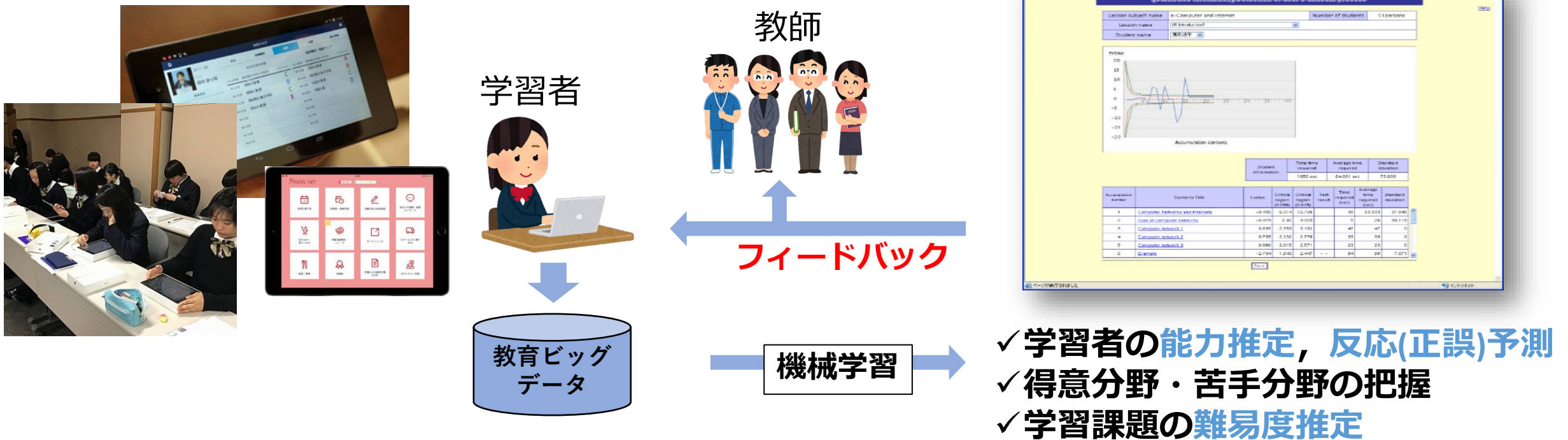
Dataset	Method	RMSE	Pearson	Kendall
Information 1	IRT	<b>0.466</b>	<b>0.891</b>	<b>0.685</b>
	DeepIRT	0.514	0.867	0.687
Information 2	IRT	0.562	0.841	0.668
	DeepIRT	<b>0.555</b>	<b>0.845</b>	<b>0.662</b>
Critical Thinking	IRT	1.064	0.464	0.318
	DeepIRT	<b>1.025</b>	<b>0.474</b>	<b>0.327</b>
Program 1	IRT	0.890	0.599	0.403
	DeepIRT	<b>0.864</b>	<b>0.622</b>	<b>0.417</b>
Program 2	IRT	0.752	0.713	0.468
	DeepIRT	<b>0.720</b>	<b>0.737</b>	<b>0.475</b>
Practice_Math	IRT	<b>0.589</b>	<b>0.748</b>	<b>0.533</b>
	DeepIRT	0.744	0.723	0.514
Practice_Physics	IRT	<b>0.884</b>	<b>0.609</b>	<b>0.424</b>
	DeepIRT	0.911	0.585	0.411
ECPE	IRT	0.875	0.615	0.435
	DeepIRT	<b>0.874</b>	<b>0.618</b>	<b>0.440</b>

Dataset	Method	RMSE	Pearson	Kendall
TIMSS	IRT	0.753	0.716	0.525
	DeepIRT	0.753	0.716	<b>0.523</b>
Information Ethics	IRT	0.394	0.920	0.643
	DeepIRT	<b>0.382</b>	<b>0.925</b>	<b>0.712</b>
Engineer Ethics	IRT	0.544	0.850	0.403
	DeepIRT	<b>0.517</b>	<b>0.865</b>	<b>0.313</b>
Classi_Physics	IRT	1.053	0.444	0.299
	DeepIRT	<b>0.943</b>	<b>0.554</b>	<b>0.403</b>
Classi_Chemistry	IRT	1.077	0.420	0.297
	DeepIRT	<b>0.923</b>	<b>0.574</b>	<b>0.439</b>
Classi_Biology	IRT	1.020	0.475	0.326
	DeepIRT	<b>0.748</b>	<b>0.717</b>	<b>0.531</b>
Average	IRT	0.764	0.680	0.451
	DeepIRT	<b>0.742</b>	<b>0.707</b>	<b>0.495*</b>

\* p<0.05

# 4. Knowledge Tracing

教育ビッグデータ(学習履歴データ)を分析することにより、学習過程における学習者の習熟度や理解度を把握し、未知の問題への反応予測を行い、学習者個人に最適な課題やヒントを提供する



# 4.1. Knowledge TracingのためのDeep-IRT

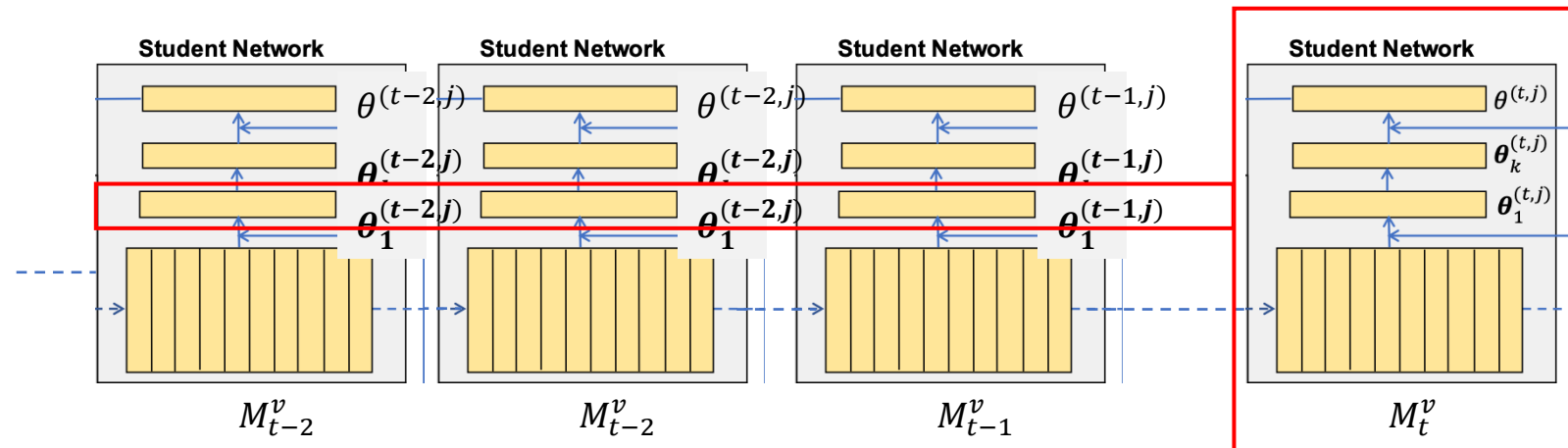
- 学習による能力値の時系列変化
- スキル間の関係性を考慮した多次元スキルの推定
- アテンションを用いて潜在的な多次元スキルに対する能力を自動的に推定（認知診断を自動的に行う）

1. Emiko Tsutsumi, Tetsurou Nishio and Maomi Ueno: Deep-IRT with temporal convolutional network for comprehensive reflection of student ability history data. Proceedings of 17th International Conference on Artificial Intelligence in Education, AIED, 250-264, (2024)
2. Tsutsumi,E., Guo,Y., Kinoshita,R., and Ueno,M., Deep knowledge tracing incorporating a hypernetwork with independent student and item networks. IEEE Transactions on Learning Technologies, 10.1109/TLT.2023.3346671 (2023)
3. 堤瑛美子, 西尾徹朗, 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.J107-D,No.03, (2024). **電子情報通信学会論文賞**
4. Tsutsumi,E., Guo,Y., and Ueno,M., Deep knowledge tracing incorporating a hypernetwork with independent student and item networks, In the proceeding of the International Conference on Educational Data Mining, EDM, (2022)
5. Tsutsumi,E., Kinoshita,R., and Ueno,M., Deep-IRT with independent student and item networks, In the proceeding of the International Conference on Educational Data Mining, EDM, (2021)

# 4.2. Deep-IRT with Temporal Convolutional Network (Tsutsumi, Nishio, Ueno 2024)

## 畳み込みニューラルネットワークを用いたDeep-IRT

- 各時点の潜在能力値 $M_t^v$ から出力された多次元の能力状態 $\theta_1^{(t,j)}$ を保存し、Temporal Convolutional Network(TCN)で過去の能力状態を畳み込む
- 長期の過去の能力値に依存した能力推定を行う



$M_t^v$ : 多次元の潜在能力値

# 4.2. Deep-IRT with Temporal Convolutional Network (Tsutsumi, Nishio, Ueno 2024)

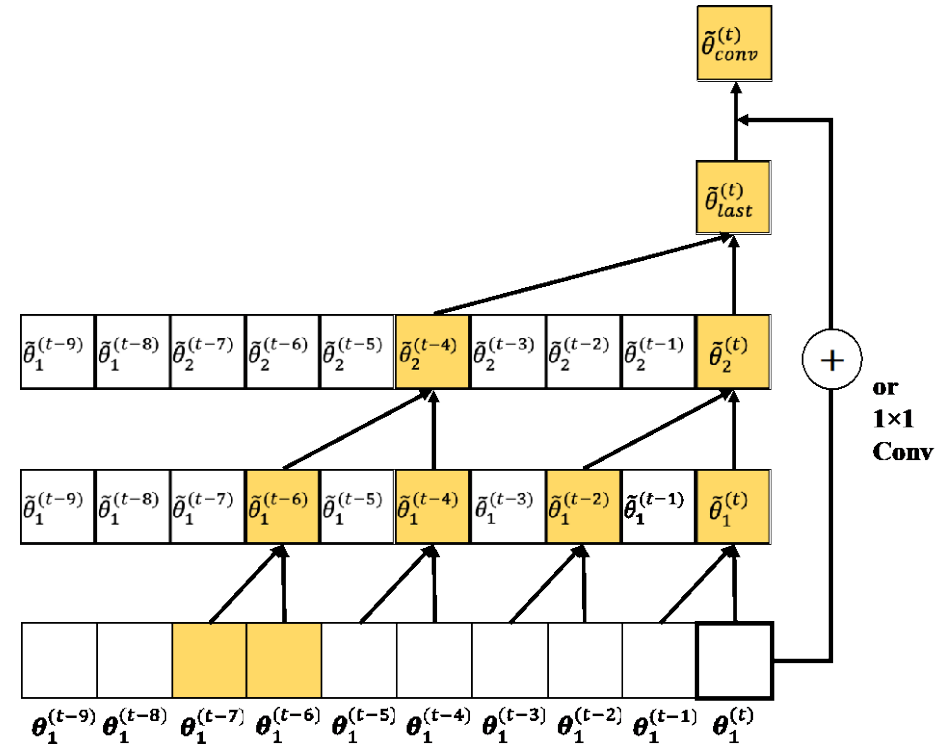
## 多次元能力値の畳み込み

- 長期の過去の能力値に依存した能力推定を行う
- 過去に推定された能力値をSliding Window方式でTCNに入力する

### □ Dilated Causal Convolution

入力値の各要素を間隔を空けて畳み込む

パラメータ数の増加を抑えながらより長い時系列データの畳み込みを行える



多次元能力値の畳み込み

### □ Residual Connection

1層目を最後の出力に加算することで層が深い場合の勾配消失を防ぐ

# 4.3. 比較実験

データセット詳細

Dataset	No. students	No. skills	No. Items	Rate Correct	Learning length
ASSISTments2009	4151	111	26684	63.6%	52.1
ASSISTments2017	1709	102	3162	39.0%	551.0
Statics2011	333	1223	N/A	79.8%	180.9
Junyi	48925	705	N/A	82.78%	345
Eedi	80000	1200	27613	64.25%	177

提案手法の条件

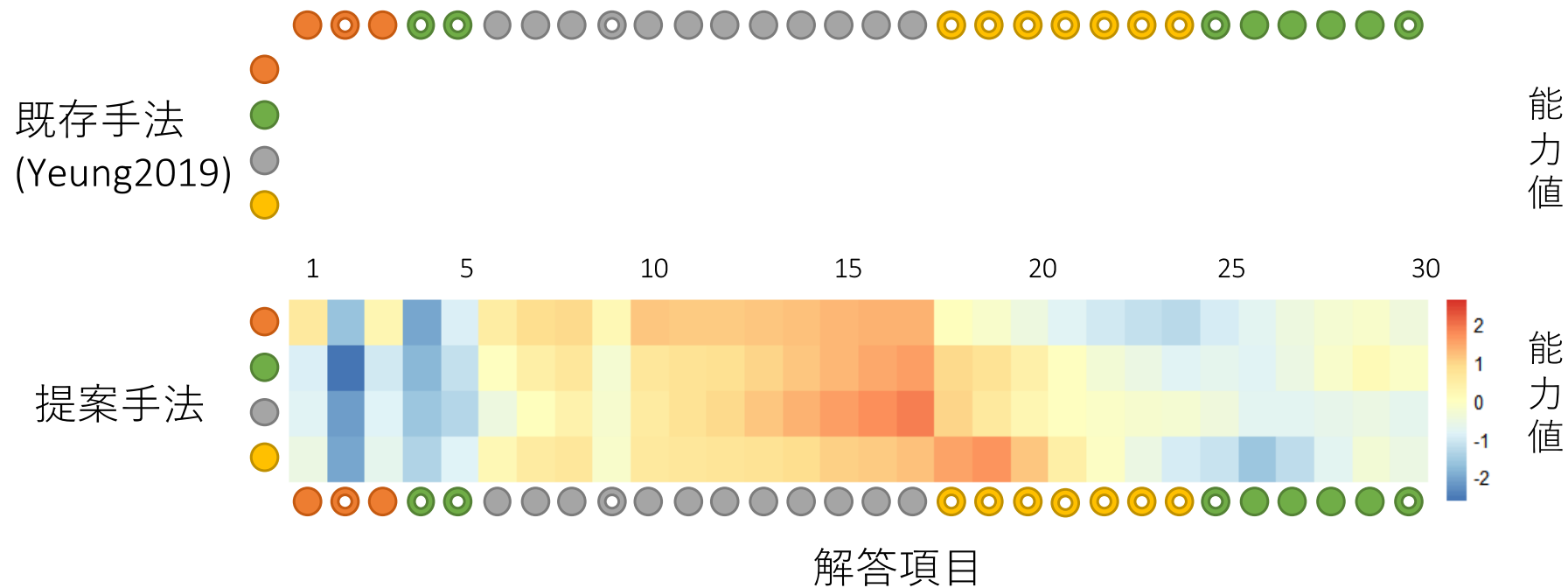
Data set	層数	$k_{last}$
ASSISTments2009	5	3
ASSISTments2017	8	2
Statics2011	3	7
Junyi	8	2
Eedi	8	2

Dataset	metrics	DeepIRT-KT	AKT	DeepIRT-HN	DeepIRT-TCN
ASSISTments2009	AUC	82.09+/-0.28	82.20+/-0.25	81.98+/-0.54	<b>82.95+/-0.30</b>
	Acc	77.41+/-0.53	77.30+/-0.55	77.15+/-0.55	<b>77.60+/-0.56</b>
ASSISTments2017	AUC	73.56+/-0.27	74.54+/-0.21	75.13+/-0.20	<b>75.49+/-0.36</b>
	Acc	69.78+/-0.41	69.83+/-0.06	70.69+/-0.60	<b>70.85+/-0.50</b>
statics2011	AUC	81.15+/-0.37	<b>82.15+/-0.35</b>	81.57+/-0.50	82.02+/-0.39
	Acc	80.01+/-0.92	<b>80.41+/-0.67</b>	80.11+/-0.92	80.40+/-0.80
Eedi	AUC	78.93+/-0.12	77.58+/-0.21	78.97+/-0.10	<b>79.14 +/-0.11</b>
	Acc	73.38+/-0.17	72.35+/-0.21	73.38+/-0.13	<b>73.55+/-0.13</b>
Junyi	AUC	77.92+/-0.41	78.13+/-0.39	77.91+/-0.37	<b>78.14+/-0.43</b>
	Acc	86.79+/-0.15	86.79+/-0.17	86.65+/-0.15	<b>86.85+/-0.14</b>
Average	AUC	78.73	78.92	79.11	<b>79.60</b>
	Acc	77.47	77.54	77.60	<b>77.81</b>

DeepIRT-HNを上回る予測精度を達成

# 4.4. Deep-IRTの多次元能力の推定値分析

- “●”は正答, “○”は誤答を表す
- 学習者はそれぞれオレンジ色, 緑色, 灰色, 黄色の4つのスキルに解答





## 4.5. Deep-IRTのまとめ

- アダプティブラーニングシステムのために、パラメータの解釈性と高精度な反応予測を両立し、**時系列多次元能力変化**を表す新たなDeep-IRTを提案
- 世界トップレベルの反応予測精度とスキル能力値パラメータの解釈性を示した

## 5. 正答確率予測を保証するBayesian network Classifier (BNC)

- Deep-IRTの長所

正誤反応の予測精度及び解釈性が高い

- Deep-IRTの短所

反応（正誤）確率分布の予測精度が低く保証されていない

→ 適応型課題提示、適応的ヒント提示などでは正答確率を用いることが多い。

近年、確率的分類機として最高精度を持ち、確率分布予測の漸近一致性があり、解釈性も持つBayesian network Classifier (BNC:AAAI2024)が提案されている。ここではBNCを用いたIRTを提案する。

# 5. 1. ベイジアンネットワーク分類器 [AAAI2024]

$$p(x_1, x_2, \dots, x_N | G) = \prod_{i=1}^N p(x_i | \text{Pa}(x_i, G))$$

$$p(D|G) = \prod_i^N \prod_j^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + n_{ij})} \prod_k^{r_i} \frac{\Gamma(\alpha_{ijk} + n_{ijk})}{\Gamma(\alpha_{ijk})}$$

分類に影響する目的変数パラメータ数(NCP)の最小化

$$NCP(G) = \sum_{i=0}^n NCP_i(\text{Pa}(X_i, G))$$

$$NCP_i(\text{Pa}(X_i, G)) = \begin{cases} (r_i - 1)q_i & i = 0 \vee X_0 \in \text{Pa}(X_i, G) \\ 0 & otherwise \end{cases}$$

$G$  : 離散確率変数  $V = \{X_0, X_1, \dots, X_n\}$  をノードとし,  
ノード間の依存関係をエッジで表す非循環有向グラフ,

$r_i$  : 各変数  $X_i$  が取りうる状態数,

$\text{Pa}(X_i, G)$  :  $G$  における  $X_i$  の親変数集合,  $q_i$  :  $\text{Pa}(X_i, G)$  が取りうるパターン数

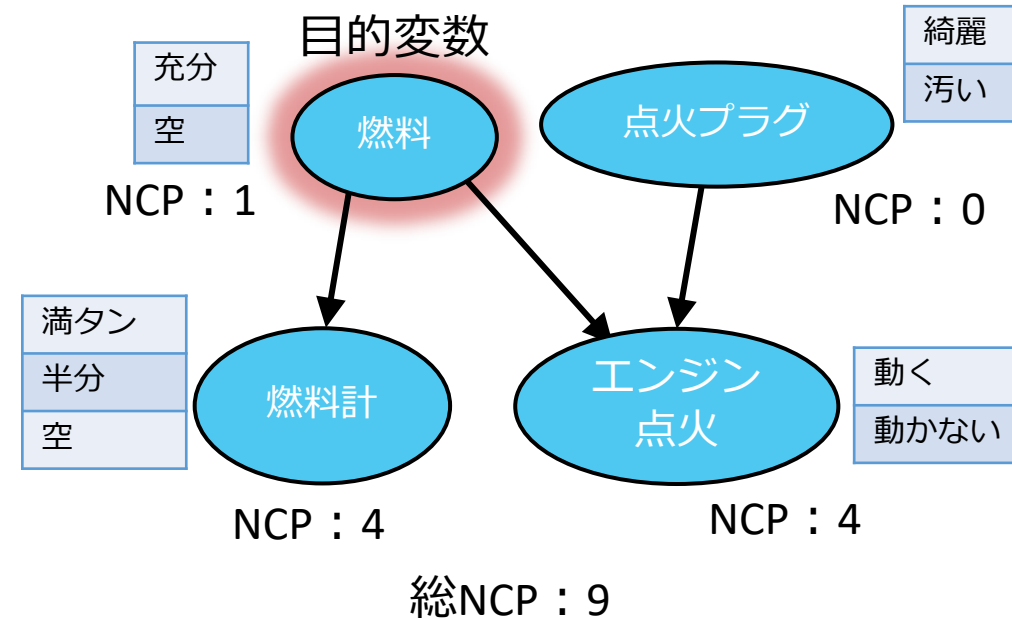


図3: NCPの計算例

## 5.2.BNCの重要な定理

真のモデルがBayesian Networkに従っていない場合でも、目的変数の確率分布の推定値が真の値に漸近収束する保証がある。

## 5.3. 分類問題での深層学習・ランダムフォレスト・ ベイジアンネットワークとの違い

- 深層学習・ランダムフォレストは説明変数を所与として目的変数を予測する識別モデル
- ベイジアンネットワークは同時確率分布を予測する生成モデル  
データ数が十分に大きい場合、ベイジアンネットワークは同時確率分布に対して漸近一致制を持つが、変数数に対して予測しなければならないパターン数が指数的に増加して 識別モデルとしての予測精度は深層学習やランダムフォレストに勝てない。
- BNCは目的変数の確率分布予測のみに集中した新しいモデル

## 5.4. アルゴリズム

1. 目的変数 $X_0$ が親変数を持たず、各説明変数が親として持てる変数を表したすべての制約について、**周辺尤度を最大化する構造**を列挙
  - i. 各変数 $X_i \in V$ と親変数集合のみから成るネットワークの周辺尤度を計算.
  - ii. すべての制約の中で、 $X_i$ のすべての親変数集合から、手順(i)の周辺尤度が最大になるものを決定.

2. 第一ステップで得られた構造の中で**NCP最小の構造**を探索

- NCPをコストとした探索グラフの最短パス探索問題として定式化
- 深さ優先探索で最短パスを求めている.
- 逐次的に最適な構造を更新する深さ優先探索に枝刈りを適用

→NCPを最小にし、真の分類確率に漸近的に一致

- メリット：

1. 枝刈りにより効率的な探索が可能
2. 実行途中にメモリ等のリソースが不足しても、それまでの最適な構造を得られる.

## 5.5. 予測精度の比較(RF ランダムフォレスト、DL 深層学習)

No.	Dataset	Variables	Sample		<i>Proposed</i>		
			size	SPP	<i>RF</i>	<i>DL</i>	<i>BFS</i>
13	Heart	14	270	$1.22 \times 10^{-3}$	0.8296	0.8185	0.8037
14	HTRU2	9	17898	$1.56 \times 10^{-3}$	0.9797	0.9793	0.9788
15	Congressional Voting Records	17	232	$1.77 \times 10^{-3}$	0.9567	0.9478	0.9569
16	Solar Flare	11	1389	$3.72 \times 10^{-3}$	0.8294	0.8402	0.8431
17	Glass	10	214	$6.63 \times 10^{-3}$	0.6310	0.6541	0.6946
18	Contraceptive Method Choice	10	1473	$2.66 \times 10^{-2}$	0.4772	0.4970	0.4912
19	Hayes-Roth	5	132	$2.29 \times 10^{-1}$	0.8088	0.7725	0.8258
20	Balance Scale	5	625	$3.33 \times 10^{-1}$	0.8287	0.9840	0.9904
21	Lenses	5	24	$3.33 \times 10^{-1}$	0.8167	0.7667	0.8750
22	Iris	5	150	$6.17 \times 10^{-1}$	0.9333	0.9400	0.9333
23	LED7	8	3200	$2.50 \times 10^0$	0.7269	0.7369	0.7334
24	Banknote authentication	5	1372	$2.72 \times 10^0$	0.9432	0.9403	0.9406
Average accuracy for datasets with large SPP (Nos. 13-24)					0.8134	0.8231	0.8389

# 5.6. 分類確率の推定精度の比較 : 真の分布とのカルバックライブラー

Network	Sample size	<i>RF</i>	<i>DL</i>	<i>Proposed BFS</i>
Cancer (Structure (1))	100	$8.49 \times 10^{-2}$	$5.29 \times 10^{-2}$	$2.70 \times 10^{-2}$
	1,000	$5.33 \times 10^{-2}$	$4.85 \times 10^{-2}$	$2.67 \times 10^{-2}$
	10,000	$1.88 \times 10^{-3}$	$8.38 \times 10^{-3}$	$1.27 \times 10^{-3}$
	100,000	$2.73 \times 10^{-4}$	$5.11 \times 10^{-3}$	$8.46 \times 10^{-5}$
Asia (Structure (2))	100	$1.23 \times 10^{-1}$	$1.51 \times 10^{-1}$	$4.40 \times 10^{-2}$
	1,000	$1.82 \times 10^{-1}$	$2.31 \times 10^{-1}$	$6.38 \times 10^{-2}$
	10,000	$5.62 \times 10^{-2}$	$3.11 \times 10^{-2}$	$2.46 \times 10^{-2}$
	100,000	$3.26 \times 10^{-2}$	$1.92 \times 10^{-2}$	$2.72 \times 10^{-4}$
Markov net (Structure (3))	100	$2.73 \times 10^{-1}$	$1.16 \times 10^{-1}$	$8.18 \times 10^{-2}$
	1,000	$1.38 \times 10^{-1}$	$7.86 \times 10^{-2}$	$6.63 \times 10^{-2}$
	10,000	$1.30 \times 10^{-1}$	$1.36 \times 10^{-1}$	$4.43 \times 10^{-4}$
	100,000	$1.17 \times 10^{-1}$	$1.44 \times 10^{-1}$	$7.94 \times 10^{-5}$

- 推定精度 : RF, DL < 提案手法
- 真のモデルがベイジアンネットワークでない場合(Structure (3))でも提案手法によって推定される確率は真の分布に収束していく



# 6.1 適応的足場がけシステム

Probability-based scaffolding system for discrete mathematics

Incorrect

Prove that the function  $f: \mathbb{N} \rightarrow \mathbb{N}; f(x) = 3x - 1$  (when  $x$  is odd),  $f(x) = \frac{x}{2}$  (when  $x$  is even), is surjective (onto).

The AI system predicts you can solve this problem if you ponder the problem with some perseverance. Believe in yourself and try your best!

Hint 1: Generally, for a mapping  $g: U \rightarrow V; g(x)$ , 'g' is called as a 'surjective' from  $U$  to  $V$  when 'for all  $y$  within  $V$ , there exists  $x$  within  $U$  such that ' $g(x) = y$  ( $\forall y \in V, \exists x \in U. t.g(x)=y$ )' holds.

Hint 2: Express the natural number ' $x$ ' that makes  $f(x) = y$  using ' $y$ '.

Hint 3: When expressing ' $x$ ' that makes  $f(x) = y$  using ' $y$ ', there are two patterns:  $x = 2y$  ( $x$  is an even number) and  $x = \frac{y+1}{3}$  ( $x$  is an odd number).

Sort the options below.

$\forall y \in \mathbb{N},$

$$f(x) = \frac{2y}{2} = y$$

$\exists x \in \mathbb{N},$

Since  $y$  is an element of natural numbers  $\mathbb{N}$ , it follows that  $x$  is also an element of natural numbers  $\mathbb{N}$ .

Regarding  $y$  which is an element of the set of natural numbers  $\mathbb{N}$ .

There exists an  $x$  such that  $x = 2y$ .

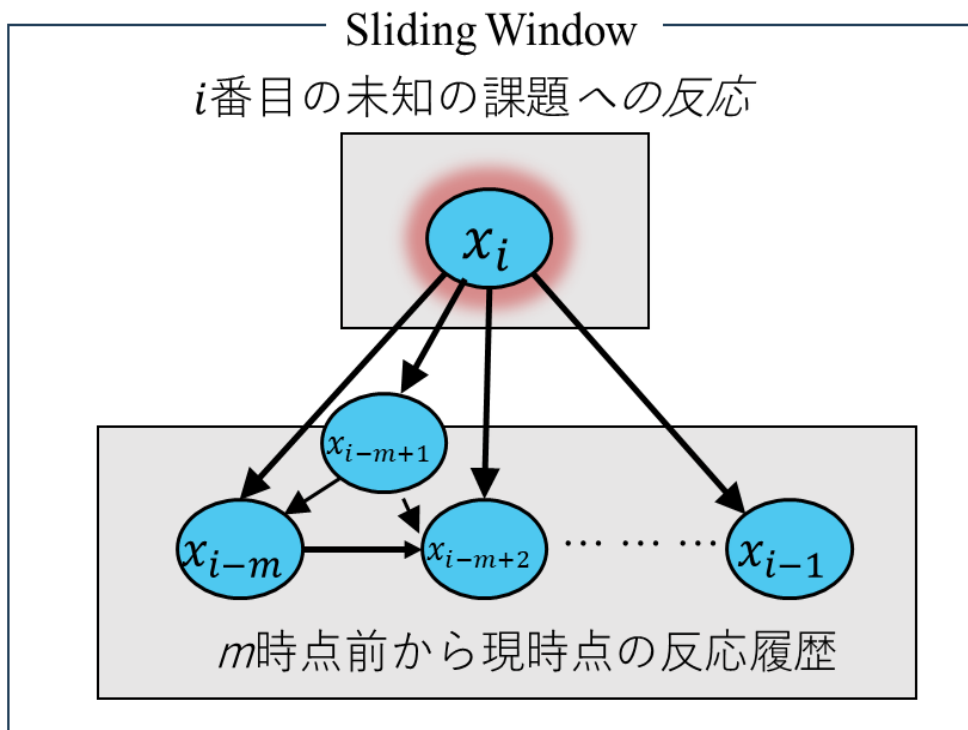
Therefore, the mapping function  $f$  is surjective.

Submit

学習者に一人では解けないレベルの課題を与え、解けるかどうか（正答確率が0.5になるように）ぎりぎりのヒントを与えるシステム

- *Maomi Ueno and Yoshimitsu Miyasawa: Probability Based Scaffolding System with Fading, Artificial Intelligence in Education – 17th International Conference, AIED 2015, 237-246*
- *Maomi Ueno, Yoshimitsu Miyazawa: IRT-Based Adaptive Hints to Scaffold Learning in Programming, IEEE Transactions on Learning Technologies, IEEE computer Society, Vol.11, Issue 4, 415-428 (2018)*

# 6.2. Sliding Bayesian network Knowledge Tracing

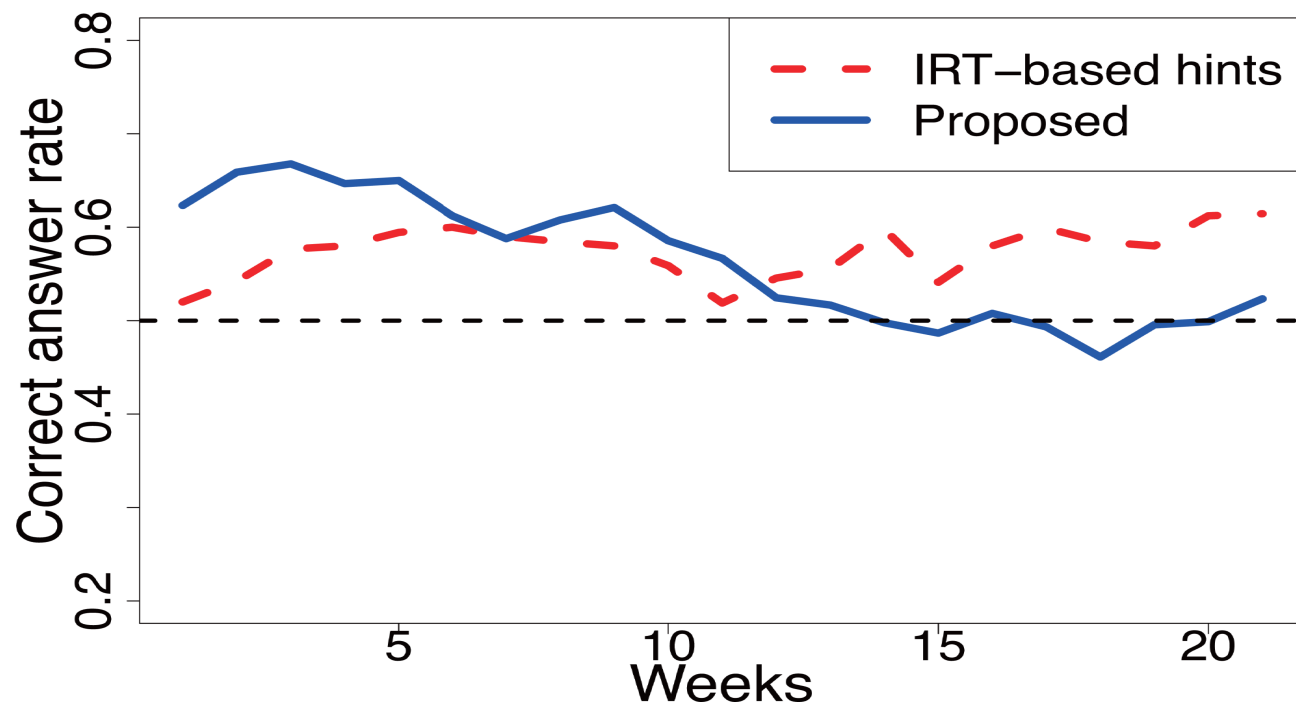


提案手法の正誤予測 AUC 0.8236

Deep-IRT AUC 0.8212

IRTの正誤予測 AUC 0.752

IRTによる従来手法：正答率 0.61  
提案手法：正答率 0.53



## 6.3. Sliding Bayesian network Knowledge Tracing

Group	Graded hints	IRT-based hints	Proposed
No. examinees	31	30	39
Avg. Pre-test	34.00 (12.24)	50.00 (19.07)	45.90 (49.90)
GPA	2.23 (0.48)	2.40 (1.14)	2.45 (1.26)
Avg. Post-test	57.03** (8.92)	61.18** (12.79)	69.33 (11.90)
IPW-Adjusted Avg. Post-test	58.75** (7.21)	60.48** (12.51)	74.99 (13.54)
ATE	-7.46** (1.97)	-7.45** (1.98)	10.43 (1.83)
Avg. Difference abilities	0.116* (0.465)	0.120* (0.423)	0.373 (0.280)
Avg. Attempt	1.814* (0.349)	1.538 (0.221)	1.352 (0.191)
Avg. Response time (s)	243** (219)	230** (215)	305 (377)

Significant difference from the proposed method: \*5%, \*\*1% .

# 7. 目的変数を潜在変数にしたら Bayesian network IRT

項目間の局所依存性を  $\theta$  を潜在変数にもつベイジアンネットワーク構造で表現.

- 親項目への反応によって子項目の難易度パラメータが変化.  
(図1)

$$P(u_{ij} = 1 \mid \theta_i, a_j, b_{jk}, G) = \frac{1}{1 + \exp(-1.7a_j(\theta_i - b_{jk}))}$$

- $b_{jk}$  : 被験者  $i$  が項目  $j$  の親項目群に対して親変数が  $k$  番目の反応パターンを示したときの項目  $j$  の難易度パラメータ

微分最適化モデル → 予測精度の向上の期待

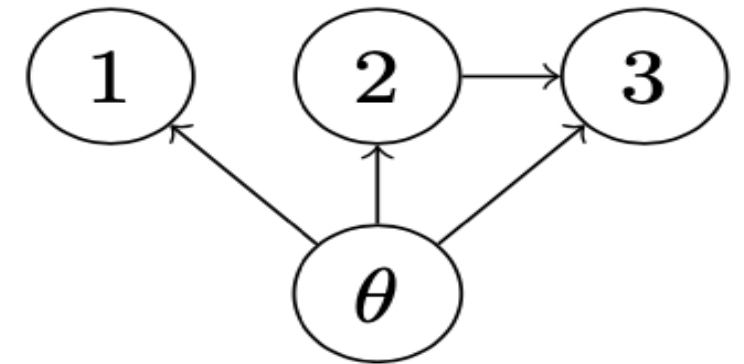


図1: ベイジアンネットワークIRT

[1] M. Ueno, "An extension of the IRT to a network model," Behaviormetrika, vol.29, no.1, pp.59-79, 2002. この時点では構造は所与

## 7.2. 構造推定

### 検定統計量の提案

ベイジアンネットワークIRT学習のためのBICによるBF

$$\log BF(G_1, G_2) \simeq BIC(G_1) - BIC(G_2)$$

$G_1, G_2$  : ベイジアンネットワークIRTの構造

### ベイジアンネットワークIRTの構造 $G$ のBIC

$$BIC(G) = \log \int L(U|G, \hat{\xi}_{MAP}, \boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta} - \frac{k}{2} \log N$$

$U$ : データ,  $\hat{\xi}_{MAP}$ : 項目パラメータのMAP推定値

$k$ : 項目パラメータ数,  $N$ : データ数

## 7.2. アルゴリズム

定理

真のベイジアンネットワークIRTは、  
漸近的にベイジアンネットワークの厳密学習によるグラフの部分グラフになる。

[第1段階]

ベイジアンネットワークの厳密学習により条件付き独立のエッジを削除

[第2段階]

BICによるBFを用いるCIテストを組み込んだベイジアンネットワークIRTの制約ベース構造学習

## 7.3. 評価実験 シミュレーション設定

- シミュレーションのための真のベイジアンネットワークIRTにおいて、各項目は次のいずれか。
  1. 親項目がない (0.3)
  2. 親項目が1つある (0.5)
  3. 親項目が2つある (0.2)※ 括弧内の数字は生成確率
- 親項目は、当該項目より前に出題されたものからランダムに選ぶ.
- $a : \log a \sim N(0,1)$  (ただし,  $0.3 < a < 2.5$ )
- $b$  :
  - (1)親項目を持たない場合  
 $b \sim N(0,1)$
  - (2)親項目を持つ場合  
次の確率変数  $b_{j',c}, d_{j'}$  を用いて難易度パラメータを生成.  
 $b_{j',c} \sim N(0,0.25), d_{j'} \sim N(\mu_d, 0.01)$   
親項目全てに正答... $b_{j',c} - d_{j'}$   
親項目に1つでも誤答... $b_{j',c} + d_{j'}$   
※  $\mu_d$  : 依存性の強さを表すハイパーパラメータ.

# 7.4. 結果

Sample size	$\mu_d = 1.95$					$\mu_d = 1.00$				
	提案	制約ベース	ILS (0.01)	ILS (0.05)	ILS (0.10)	提案	制約ベース	ILS (0.01)	ILS (0.05)	ILS (0.10)
$J = 5$										
100	<b>1.6</b> (1.50)	1.9 (1.81)	2.8 (1.25)	2.7 (1.19)	2.5 (1.28)	<b>1.6</b> (1.11)	1.9 (1.04)	4.2 (1.89)	2.0 (1.67)	2.0 (1.48)
500	<b>1.0</b> (1.10)	1.4 (1.36)	2.7 (1.19)	2.3 (1.42)	2.7 (1.19)	<b>0.6</b> (0.66)	1.5 (1.20)	2.8 (0.98)	1.6 (1.02)	1.7 (1.27)
1,000	<b>0.8</b> (0.87)	0.9 (0.94)	2.3 (1.27)	2.4 (1.43)	2.5 (1.28)	<b>0.5</b> (0.81)	0.6 (0.80)	2.5 (1.12)	1.3 (1.27)	1.9 (1.37)
5,000	<b>0.5</b> (0.67)	0.6 (0.80)	1.8 (0.87)	2.2 (1.33)	2.5 (1.28)	0.5 (1.02)	<b>0.2</b> (0.60)	1.8 (1.40)	1.4 (1.28)	1.9 (1.30)
10,000	<b>0.6</b> (0.92)	<b>0.6</b> (0.92)	2.0 (1.00)	2.2 (1.33)	2.5 (1.28)	0.3 (0.64)	<b>0.2</b> (0.60)	1.9 (1.45)	1.3 (1.19)	2.0 (1.26)
20,000	<b>0.3</b> (0.46)	0.4 (0.49)	1.9 (1.04)	2.2 (2.50)	2.5 (1.28)	<b>0.2</b> (0.60)	0.3 (0.64)	2.0 (1.41)	1.4 (1.28)	1.9 (1.37)
$J = 10$										
100	<b>3.0</b> (1.73)	4.6 (2.62)	7.8 (2.32)	6.8 (1.94)	7.2 (1.94)	<b>4.2</b> (1.78)	4.5 (2.77)	10.5 (5.18)	6.1 (1.97)	6.2 (1.99)
500	<b>1.8</b> (2.23)	3.8 (2.68)	6.1 (2.51)	6.7 (1.62)	6.9 (1.58)	<b>1.9</b> (1.76)	3.6 (2.87)	9.1 (6.01)	4.9 (1.87)	5.9 (2.02)
1,000	<b>1.7</b> (1.68)	5.3 (2.33)	5.3 (2.233)	6.4 (1.43)	6.8 (1.54)	<b>0.8</b> (0.63)	3.4 (2.76)	4.6 (6.07)	6.0 (2.33)	6.1 (1.73)
5,000	<b>1.8</b> (1.78)	5.3 (3.58)	4.3 (1.95)	6.3 (2.05)	6.8 (2.27)	<b>2.0</b> (1.79)	3.8 (1.89)	5.1 (3.33)	4.9 (1.76)	6.1 (2.02)
10,000	<b>1.1</b> (1.14)	5.9 (3.11)	3.5 (2.01)	4.2 (1.83)	6.7 (1.90)	<b>0.8</b> (1.17)	6.6 (3.50)	3.7 (2.00)	4.4 (1.80)	6.1 (1.81)
20,000	<b>0.5</b> (0.67)	4.5 (2.42)	4.9 (2.34)	6.3 (1.80)	7.0 (1.61)	<b>0.8</b> (1.08)	8.3 (3.87)	3.8 (1.60)	4.4 (2.11)	6.0 (1.90)
$J = 50$										
100	<b>29.7</b> (7.28)	N/A	65.7 (9.73)	44.6 (1.62)	42.2 (2.19)	<b>32.7</b> (8.17)	N/A	58.7 (10.42)	44.7 (4.62)	44.5 (5.31)
500	<b>16.3</b> (8.25)	N/A	70.4 (8.59)	42.5 (1.47)	41.4 (1.72)	<b>19.1</b> (6.20)	N/A	61.1 (9.54)	43.0 (6.77)	43.1 (5.37)
1,000	<b>11.7</b> (5.62)	N/A	70.3 (10.48)	41.5 (1.83)	41.0 (1.62)	<b>17.8</b> (8.70)	N/A	64.2 (10.07)	41.5 (5.75)	42.8 (5.57)
5,000	<b>7.5</b> (3.56)	N/A	70.5 (8.52)	38.3 (1.90)	39.7 (1.47)	<b>9.3</b> (4.24)	N/A	68.1 (12.64)	40.6 (5.49)	42.1 (6.18)
10,000	<b>6.5</b> (3.14)	N/A	70.9 (8.55)	38.4 (1.90)	39.5 (1.62)	<b>9.6</b> (5.33)	N/A	65.0 (12.86)	40.3 (5.46)	41.7 (5.82)
20,000	<b>7.0</b> (4.31)	N/A	67.9 (10.63)	38.3 (1.60)	39.4 (3.29)	<b>6.3</b> (3.61)	N/A	65.7 (13.68)	38.3 (5.76)	41.6 (6.11)



## 7.5. Bayesian network IRTの課題

1. 数理的性質の解析：ベイジアンネットワークのモデル平均に近いのでは？  
確実に予測精度が向上：

2. 推定された構造の意味？ 通常のベイジアンネットワークではテストデータは完全グラフになりやすい。

共変数である能力の各問題への正誤への影響が大きいためだと考えられる。ベイジアンネットワークIRTは共変数である能力の影響を除外した問題項目間の関係性をスパースな構造としてとらえることができ、教育分野での応用には有利かもしれない。

3. Bayesian networkの漸近的近似として ディレクレ事前分布のハイパーパラメータを変化させると様々なIRTモデルを導出できる。例えば各項目の親変数数をハイパーパラメータに入れると1パラメータロジスティックモデルとなるし、それを一般化すると2パラメータロジスティックモデルに収束する。

同様にBayesian network IRTをなんらかの条件で導出できないか？

## 8. まとめ

1. 反応予測精度と解釈性を満たすDeep-IRT:Deep Learning + IRT
2. 能力変化を反映したKnowledge TracingのためのDeep-IRT
3. 正答確率予測を保証するBayesian network Classifier (BNC)とSliding window Bayesian network Knowledge Tracing
4. Bayesian network IRT

Deep-IRTは実用レベルに至っていると考えられる。最先端研究であるBNCは良い数学的性質を持ち、理想的な状況では最高精度を持つが、実際には計算量の大きさと少数データや欠測データに弱いという欠点がある。

互いの欠点を補うように BNCにDeep Learning手法を組み合わせた手法を現在、研究中である。