

日本テスト学会賞記念講演

AI時代のCBTと その未来

植野真臣

電気通信大学大学院

情報理工学研究科 教授

講演内容

- ① 植野真臣の略歴
- ② これまでのCBT研究
- ③ 電通大でのCBT入試プロジェクト
- ④ AI時代のCBTとその未来

2. 2000年前後 e-learning system "Samurai"

様々な人工知能技術を組み込んだ高機能 adaptive learning eLearning system "Samurai"を開発

日本で最初のeラーニングの論文

世界最大のeラーニングの国際会議"eLearn"で

2004, 2005, 2007年 Best Paper賞。



3. eテストニング研究

日本テスト学会論文賞：

ソムアン・ポクポン, 植野真臣：統合型eテストニング・システムの開発と実践, テスト学会誌, 4巻, 54-64 (2008)

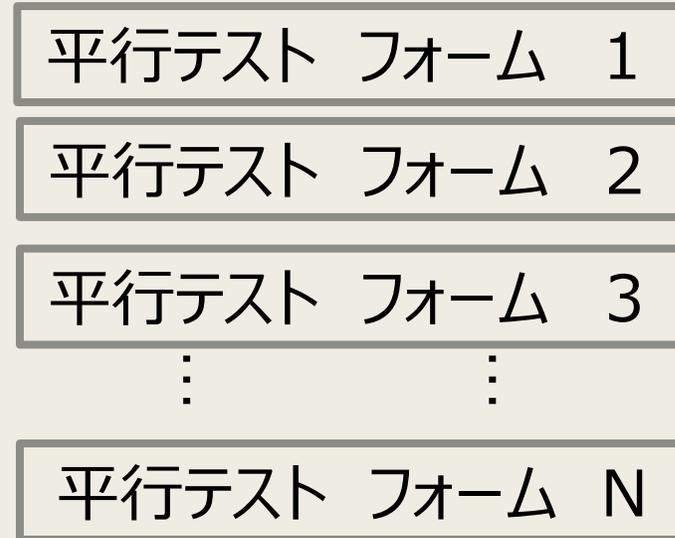
経済産業省 情報処理技術者試験 のeテストニング ITパスポート
の開発

eテストティングとは？

- 同一能力の受検者が異なるテストを何度受検しても同スコアを返すCBT
- アイテムバンク（データよりパラメータ推定された問題項目データベース）から数理モデルを用いてスコアの予測誤差を小さくするように等質のテスト（**平行テスト**と呼ぶ）を人工知能技術を用いてできるだけ多く生成する技術

平行テスト（等質の複数テスト）生成

できるだけ多くの平行テストを生成

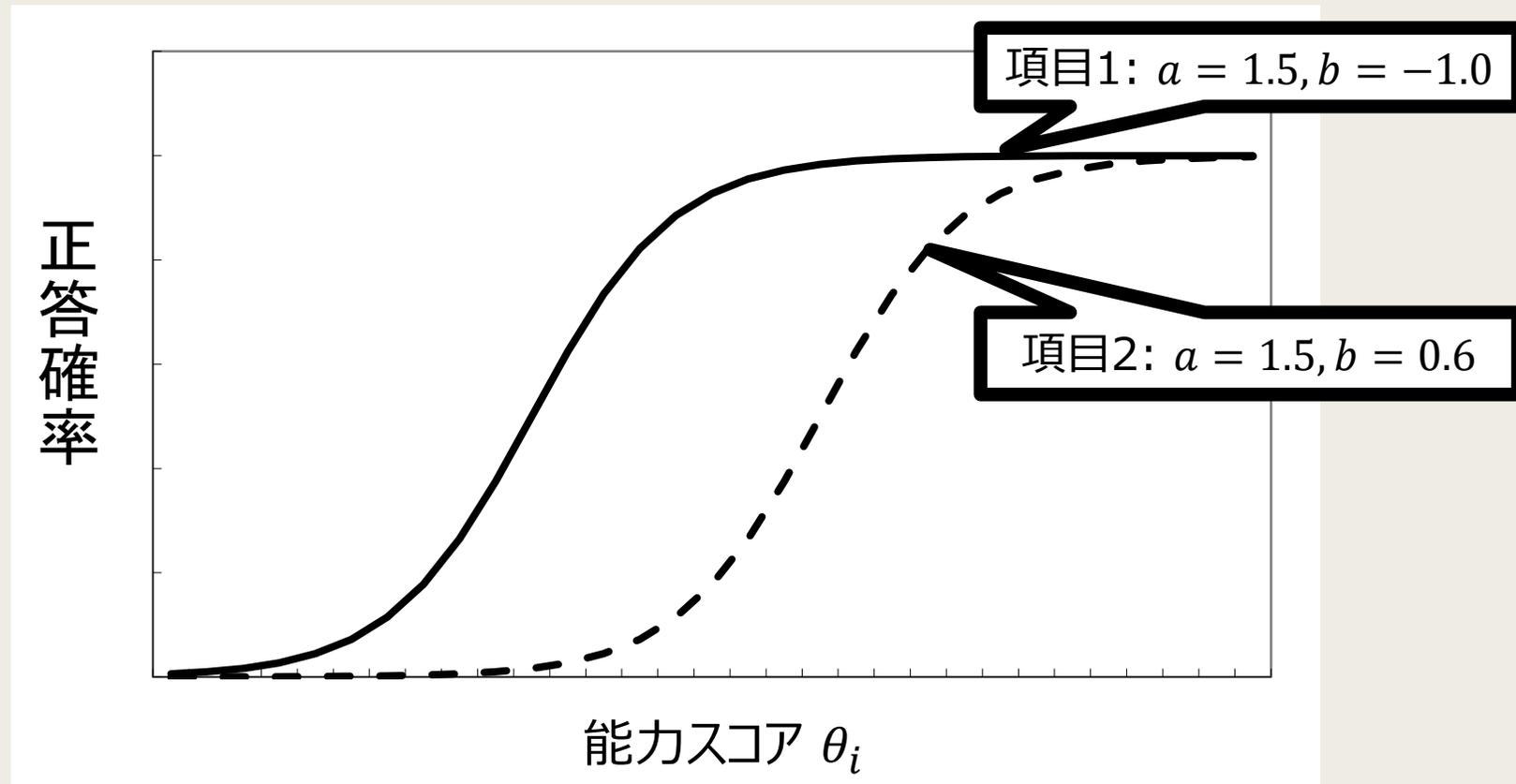


我が国のテストでは20000問程度の
問題項目が蓄積されていることが多い。
毎年項目は追加されるが、作問した問
題項目がすべて登録されるわけではなく非
常に貴重！！

項目反応理論

2パラメータロジスティックモデル

$$P(x_j = 1 | a_j, b_j, \theta_i) = \frac{1}{1 + \exp(-1.7a_j\theta_i + b_j)}$$

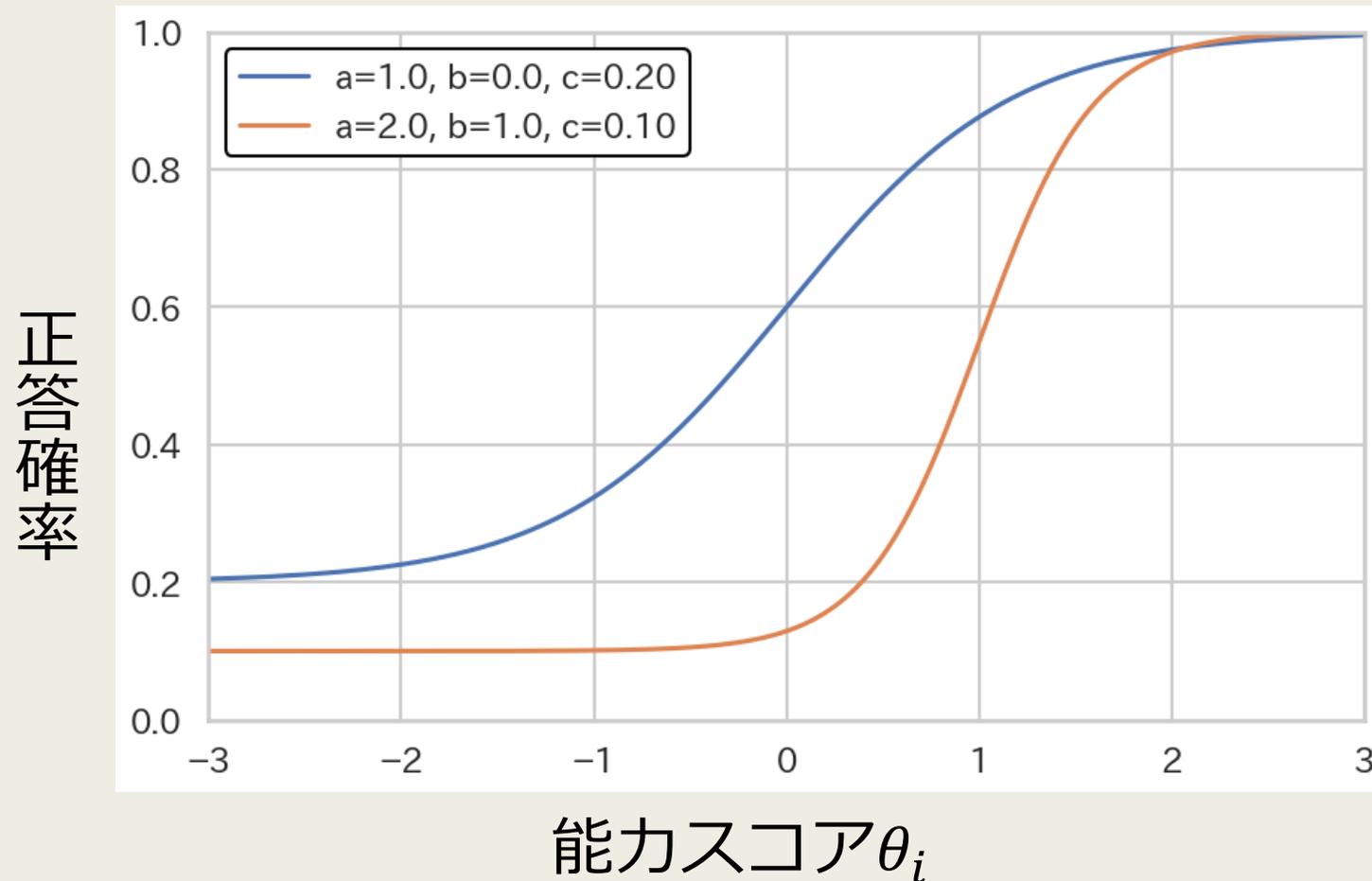


国際標準：今はこれを使います！！

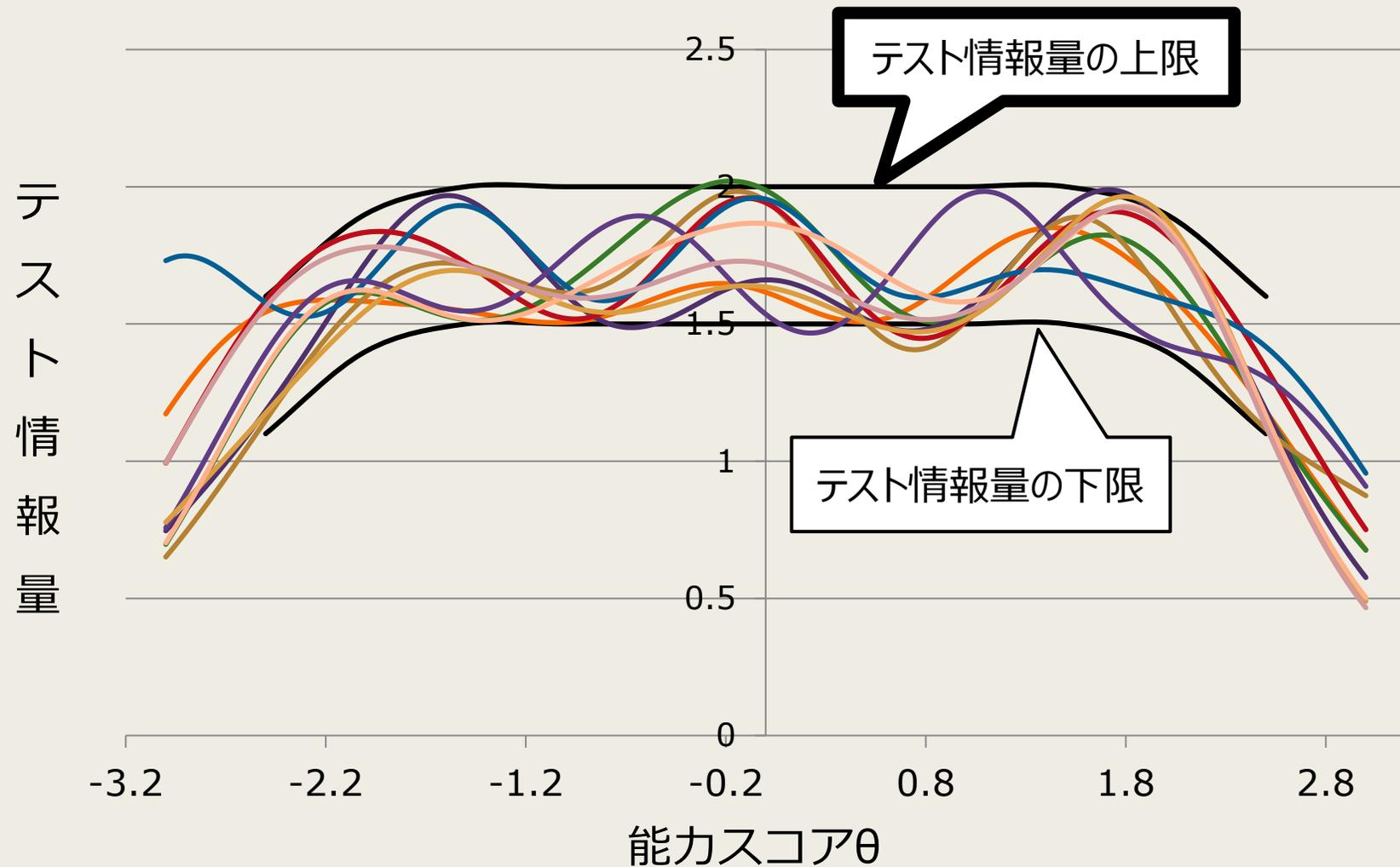
3パラメータロジスティックモデル

$$P_{ij}(u_{ij} = 1 | a_i, b_i, c_j, \theta_j) = c_j + \frac{1 - c_j}{1 + \exp(-1.7a_i\theta_j + b_i)}$$

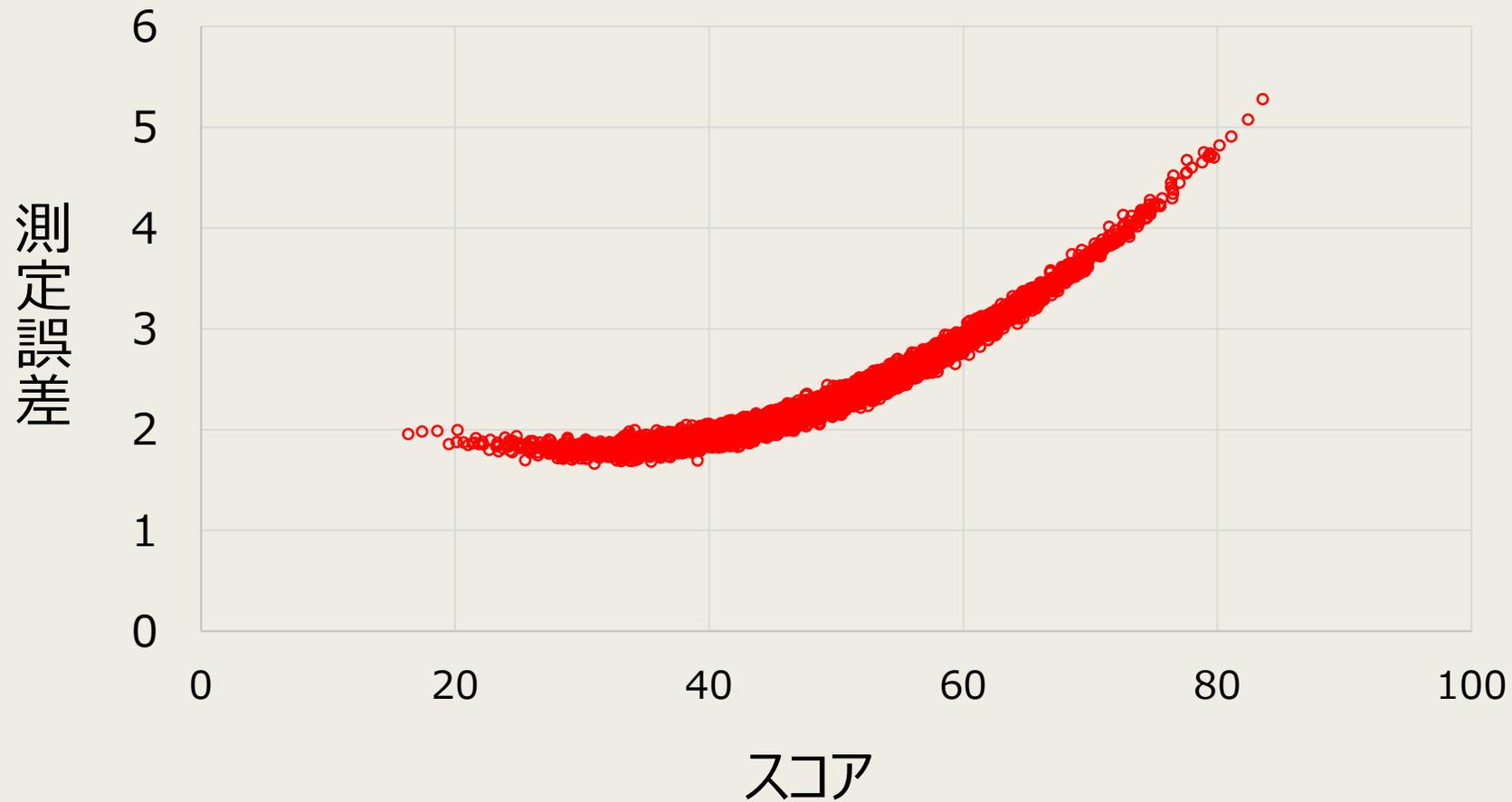
当て推量正答確率



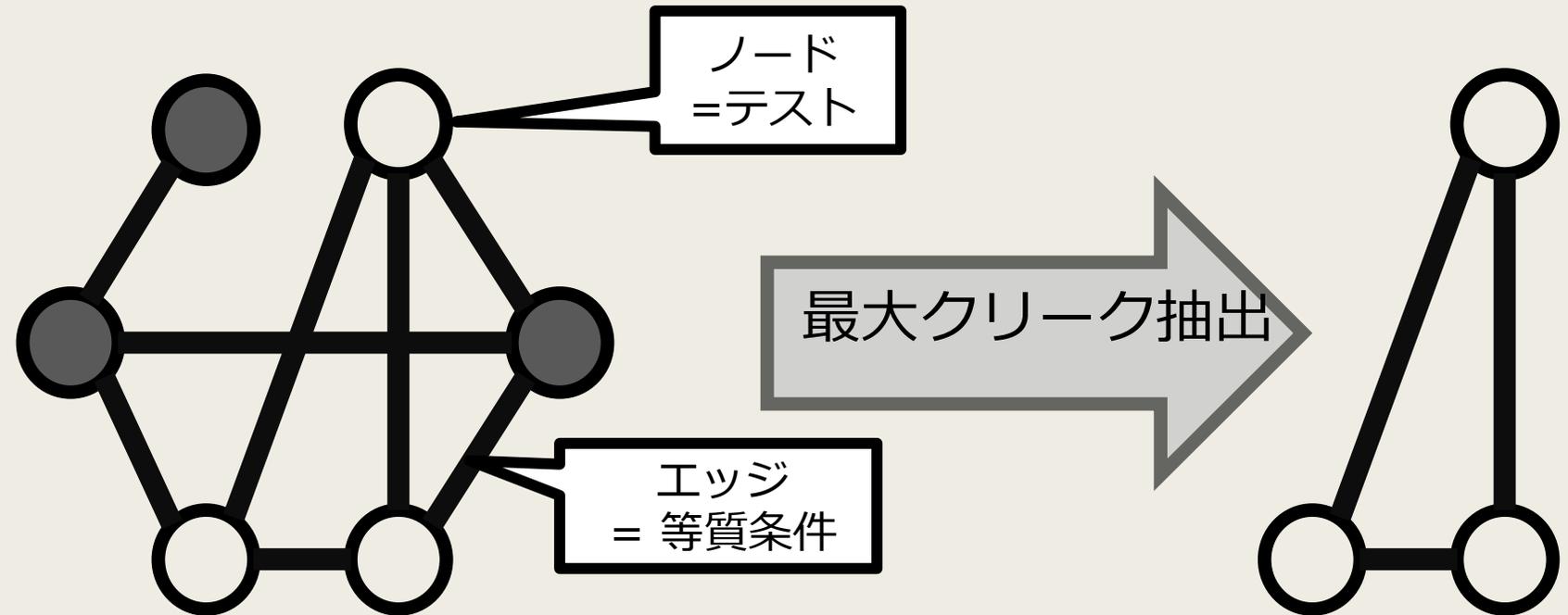
テスト情報量 = スコアの予測誤差の 逆数



あるハイスタークスな公的試験における等質テスト のスコアの測定誤差



最大クリーク問題に帰着（10万以上のテスト生成） 2018年度電子情報通信学会論文賞受賞



1. 石井隆稔・赤倉貴子・植野真臣“複数等質テスト構成における整数計画問題を用いた最大クリーク探索の近似法”, 電子情報通信学会論文誌D Vol.J100-D(1), pp.47-59, 2017.
2. Takatoshi Ishii, Maomi Ueno, "Algorithm for Uniform Test Assembly Using a Maximum Clique Problem and Integer Programming", International Conference on Artificial Intelligence in Education (AIED), LNAI 10331, pp. 102-112. 2017
3. Takatoshi Ishii, Pokpong Songmuang, Maomi Ueno, "Maximum Clique Algorithm and its approximation for Uniform Test Form Assembly", IEEE Transactions on Learning Technologies, Vol.7(1), pp.83-95, 2014.

その後 150万以上の平行テスト生成に成功

1. Fuchimoto, Kazuma, Shin-ichi Minato, and Maomi Ueno. "Automated Parallel Test Forms Assembly using Zero-suppressed Binary Decision Diagrams." IEEE Access (2023).
2. Kazuma Fuchimoto, Takatoshi Ishii, and Maomi Ueno: Hybrid Maximum Clique Algorithm Using Parallel Integer Programming for Uniform Test Assembly, IEEE Transactions on Learning Technologies, vol. 15, no. 2, pp. 252-264, 1 (2022)
3. 湊 真一, 植野 真臣. Zero-suppressed Binary Decision Diagrams を用いた自動テスト構成, 人工知能学会論文誌, 37.5 A-M23_1, (2022).
4. 湊 真一, 植野 真臣. 等質テスト構成における整数計画法を用いた最大クリーク探索の並列化. 電子情報通信学会論文誌D, Vol.J103-D, No.12, pp. 881-893, (2020).

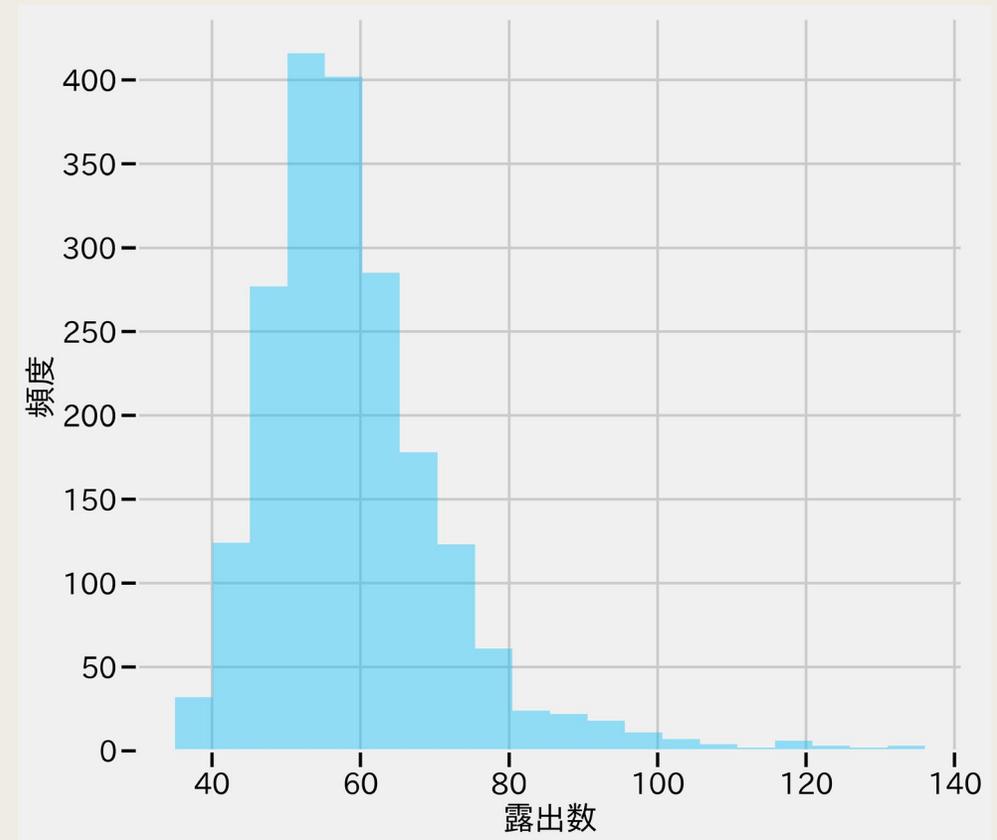
2万項目のアイテムバンクから**24時間で約150万のテスト生成**

項目露出数（出題数）のヒストグラム

- 特定の良問を集中的に選択してしまふことで良問の露出数（出題数）が増加
- 項目の特性が良問から大きく変化



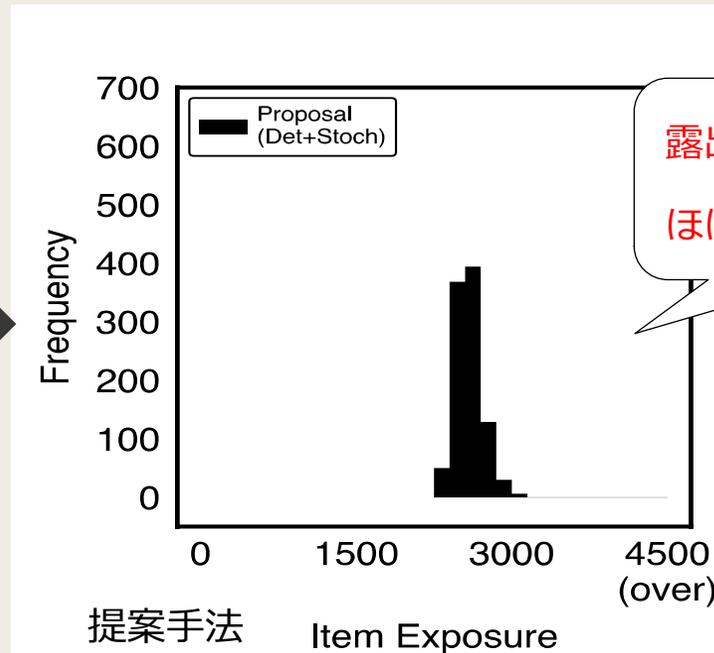
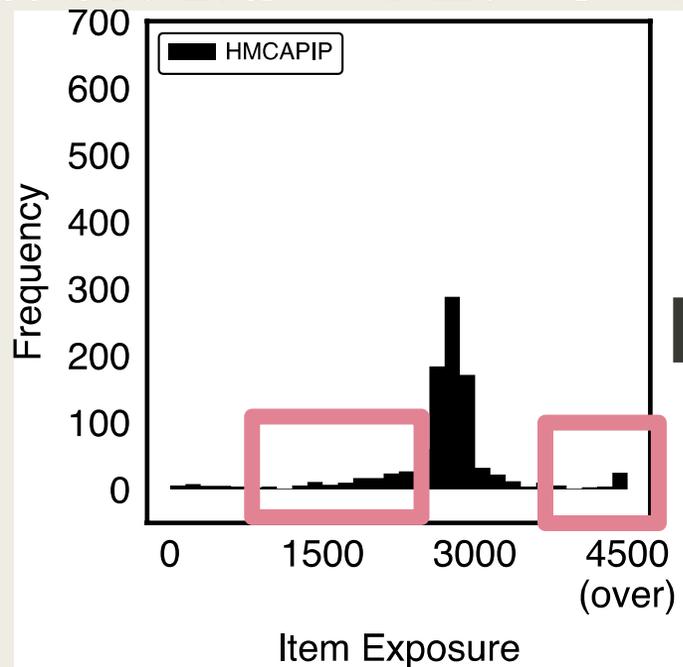
測定誤差の劣化が急速に進行



露出数の偏り軽減のための整数計画法

$$\text{Maximize } \sum_{i=1}^n \left(\lambda_i x_i - \underbrace{\frac{1}{1+e^{-z_i}}}_{\textcircled{1}} - \underbrace{M_{\{i,p\}}}_{\textcircled{2}} \right) x_i$$

- 露出数の大きさに応じたロジスティック関数による
①決定論的ペナルティおよび②確率的ペナルティを与えた
整数計画法を用いて逐次的にテスト生成



準備完了！！

eテスト実施のための準備
はそろった！！

4. 文部科学省委託事業 電通大でのCBT入試

世界最高峰技術を集積したCBTシステム

受験勉強や暗記では対処できないAI時代の本物の能力の評価を目指す。個人的興味で深く育ててきた高度で光り輝く実践能力（例えばAIアルゴリズムなどを理解した上でのプロ級のプログラミング能力など）までを幅広く精度高く評価し、将来のAI人材やデータサイエンス人材を発掘する。

日本ではCBT専門家の不足が予測できる。電気通信大学で高度な技術を持つCBT人材を多く育てる。

日本初の文部科学省公認のCBT入試（文部科学省委託事業）。電気通信大学は日本の大学入試改革に最先端技術で貢献したい。

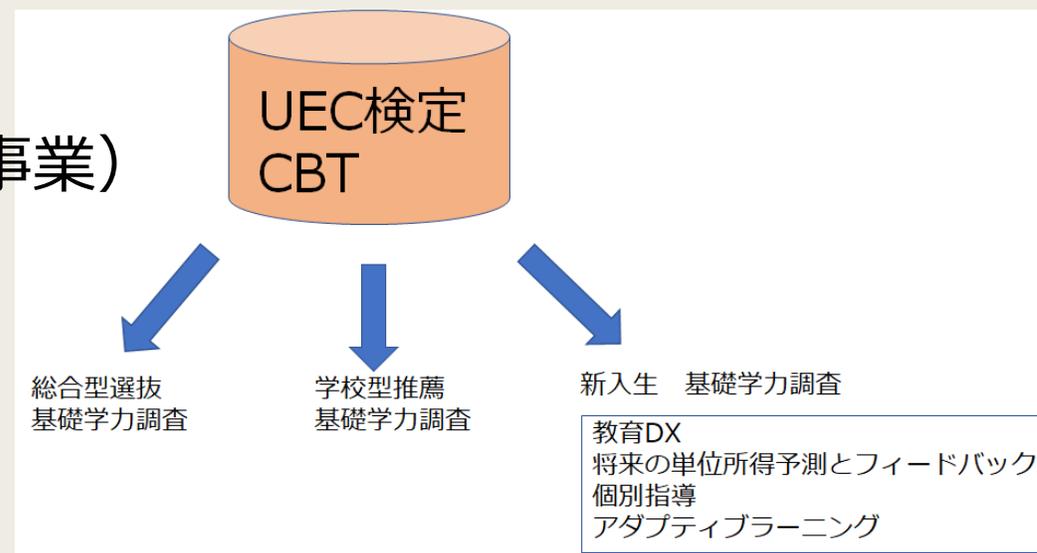
AI時代に打ち勝つ、受験勉強では育たない本物の情報能力を生徒が普段から育てていく文化と環境の重要性をメッセージとして社会に送り、ICT版科学技術立国の文化構築に寄与したい。

4.1. 電通大のアイテムバンク式CBT運用モデル

1. 新入生基礎学力調査（UEC検定）

CBTによる新入生基礎学力調査（UEC検定）（2023年対象者703名）2023年度より新入生の「数学」「物理」「情報」（順次「化学」「英語」も検討）の基礎学力を評価し，成績データから個別の単位所得を予測し，個別指導，教学指導を行う。

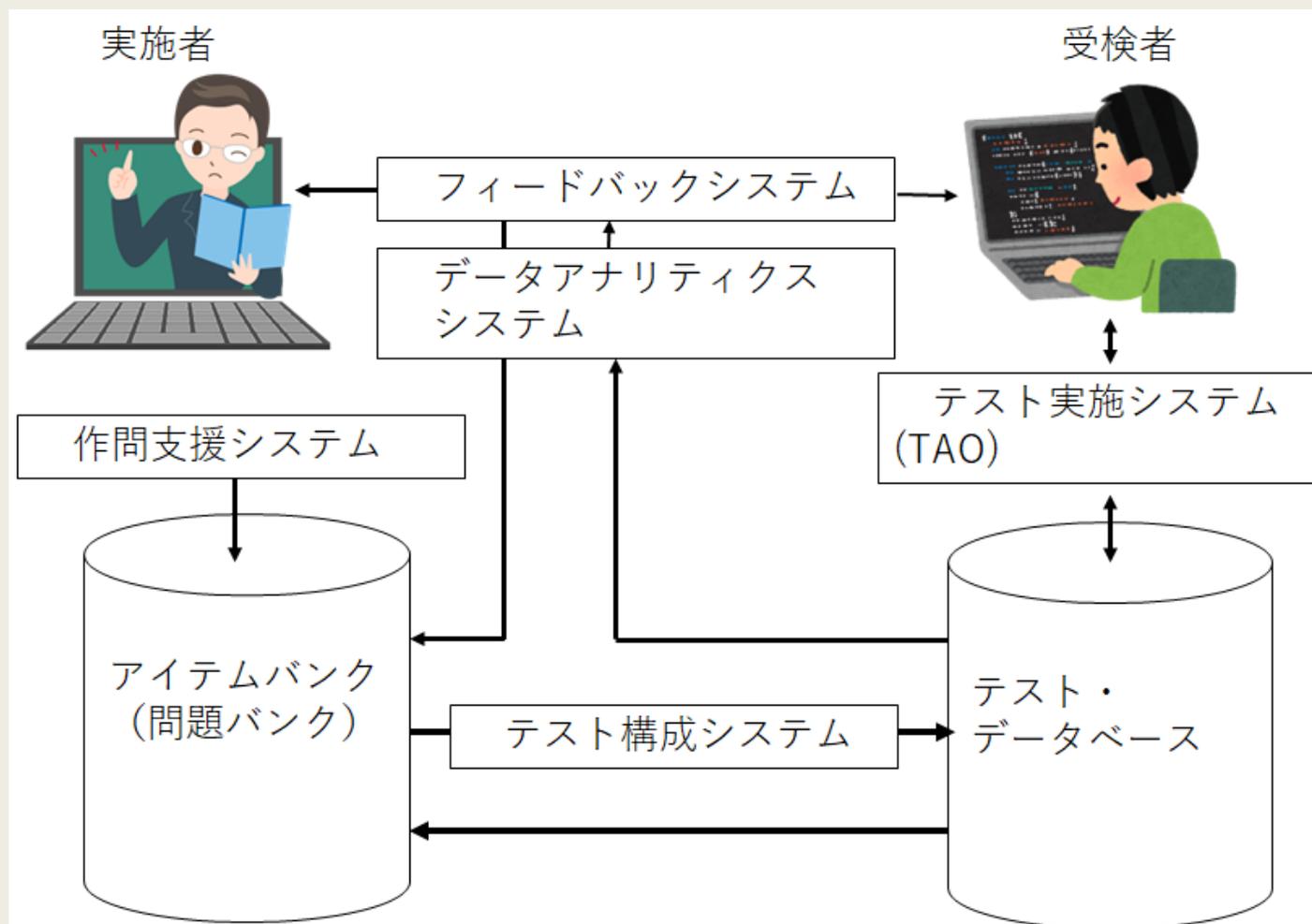
2. CBT入試（文部科学省委託事業）



4.2. 電通大でのCBTの特徴

- ① CBTの実施や運営に関する**世界標準**（ISO/IEC 2007, JIS X 7221）に従い、項目反応理論を用いた問題バンク方式をとる。**世界最高精度の測定精度**を持つテスト構成技術を持つ最先端人工知能搭載システムを開発する。
- ② 日本で初めて「**情報I**」を含むCBT入試を 総合型選抜、学校推薦型選抜で、実施し、**1. 実際のプログラミング環境**でプログラムを編集・実行しながら解答を求める問題、**2. データ解析ツールを用いて実際のデータを分析しながら解答を求める**問題など、従来の紙による試験や口頭試験だけでは測定ができなかった実践的な力を評価できるシステムを用いる。
- ③ プログラミング問題やデータ解析問題の 受検者の実行しエラーを表示するなどの試行錯誤の**解答プロセスも保存され、自動採点の評価**に加味できる。

4.3. CBTシステム



ソナムアン・ポクポン, 植野真臣: 統合型eテストティング・システムの開発と実践, テスト学会誌, 4巻, 54-64 (2008)

4.4. アイテムバンクシステム

TAOのみでは管理できない
IRTなど項目の属性情報を管理するシステム

The screenshot shows the '項目属性変更入力' (Item Attribute Change) form. The left sidebar contains navigation items: ユーザー管理, 細目表管理, メディア管理, IRTモデル管理, 評価観点管理, 問題タイプ管理, 項目属性管理 (highlighted), and TAO登録作業管理. The form fields include: 項目ID (01010100000003), 登録作業 (てすと太郎), 作問者 (必須) (てすと太郎), 教科 (必須) (情報01), 分野 (必須) ((1) 情報社会の問題解決), 領域 (必須) (1. 情報の特性), 領域2 (選択して下さい), 評価観点 (0. 知識及び技能) with three radio button options, 評価観点 (1. 思考力, 判断力, 表現力) with three radio button options, and 項目名 (empty).



TAOと内部的にデータ連携して動作する
(問題内容の登録・編集はTAOで行う)

The screenshot shows the 'アイテム編集' (Item Edit) interface. The top navigation bar includes: taos, アイテム, テスト, 受験者, グループ, デリバリ, 結果, Assets, and イベントログ. The main content area shows a tree view with 'Item' and 'QTI Interactions' folders, and 'Item 1' selected. Below the tree is a toolbar with icons for: 新規クラス, 削除, インポート, エクスポート, コピー, 指定コピー, 移動先の選択, and アクセス制限. The right panel shows the 'アイテム編集' form with fields for: Label* (itemID), リソース識別子 (http://tao.127.0.0.1.xip.io/default.rdf#i62f35), and Item Model (QTI selected). A '保存' (Save) button is at the bottom right.

4.5.情報Iプログラミング問題例

実際のプログラミング環境でプログラムを編集・実行しながら解答を求める。データ解析バージョンもある。プログラミング問題やデータ解析問題の受検者の実行エラーを見て修正するなどの試行錯誤の解答書き直しプロセスも保存され、自動採点の評価（IRTのスコア）に加味される。PCIは大学入試センターとの共同開発。

以下の文章を読み、短冊型コード選択肢を解答欄に並び替えてプログラムを完成せよ。

ただし、全ての短冊型コード選択肢を使うとは限らない。

配列Aに格納されている5個の整数について、その値が偶数のとき「偶数」と出力し、素数以外の奇数であるとき「奇数」と出力し、素数であるとき「素数」と出力するプログラムを作成せよ。

The screenshot shows a programming problem interface with two main panels: '短冊型コード選択肢' (Code Block Selection) and '解答欄' (Answer Field). The '短冊型コード選択肢' panel contains several code blocks: a for loop for i from 2 to x-1, an if statement for x % i == 0, two printf statements for even and odd numbers, another if statement for x % i == 0 with a 'なければ' (if not) label, and three blocks for flag = 0, int flag = 1, and break, followed by int x = A[j]. The '解答欄' panel shows the same code blocks arranged in a sequence: int A[5] = {101,345,332,233,221};, for(int j = 0; j < 5; ++j), int x = A[j];, if(x % 2 == 0), printf("A[%d]は偶数\n", j);, a 'なければ' label, int flag = 1;, for(int i = 2; i < x; ++i), if(x % i == 0), flag = 0;, break;, if(flag == 0), and printf("A[%d]は奇数\n", j);. Below the code is the '実行結果' (Execution Result) section, which displays the output: A[0]は素数, A[1]は奇数, A[2]は偶数, A[3]は素数, A[4]は奇数, and プログラムの実行が完了しました。 (Program execution completed.). At the bottom are 'リセット' (Reset) and '実行' (Execute) buttons.

4.5. データ解析ツールを用いて実際のデータを分析しながら解答を求める問題の例

問 次の文章を読み、空欄に入れる最も適切な語句を答えなさい。

太郎さんは、ある年のサッカーワールドカップ大会のデータ（Wカップデータ）を分析した。分析対象となるデータは、各チームについて、以下のとおりである。

- 試合数...大会期間中に行った試合数
- 獲得点...大会で行った試合すべてで獲得した得点の合計
- ショートパス本数...全試合で行った短い距離のパスのうち成功した本数の合計
- ロングパス本数...全試合で行った長い距離のパスのうち成功した本数の合計
- 反則回数...全試合において審判から取られた反則回数の合計

チームごとに試合数が異なるので、各項目を1試合当たりの値に変換して、項目間の相関を調べたところ、予選結果（決勝進出または予選敗退）によって次のような興味深い結果が得られた。

「試合あたりの反則回数が増えると1試合あたりの獲得点が増える傾向が強い。」

Wカップデータ

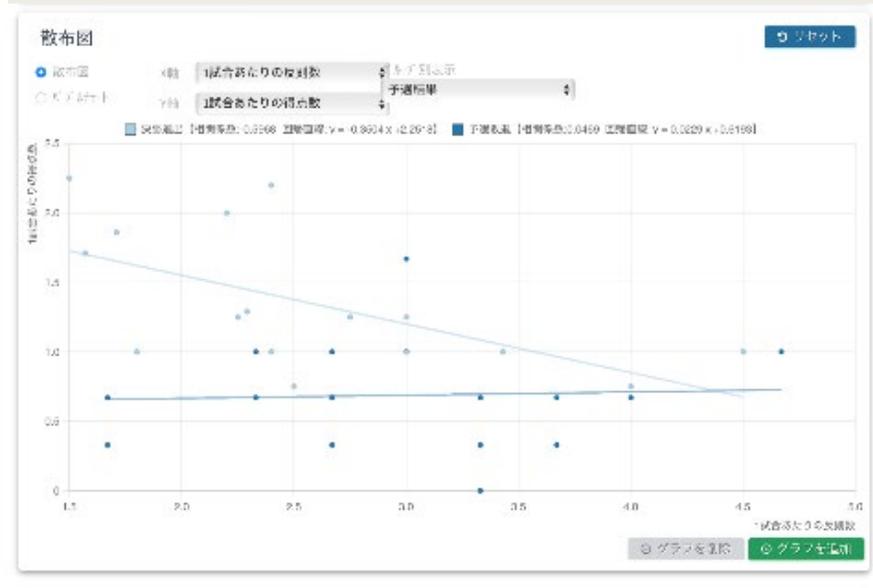
チーム	試合数	獲得点	ショートパス本数	ロングパス本数	反則回数	予選結果	1試合あたりの獲得点	1試合あたりのショートパス本数	1試合あたりのロングパス本数	1試合あたりの反則回数
イタリア	7	12	2,257	711	11	決勝進出	1.71	322.43	101.57	1.57
フランス	7	9	2,273	553	16	決勝進出	1.29	324.71	79	2.29

散布図

散布図

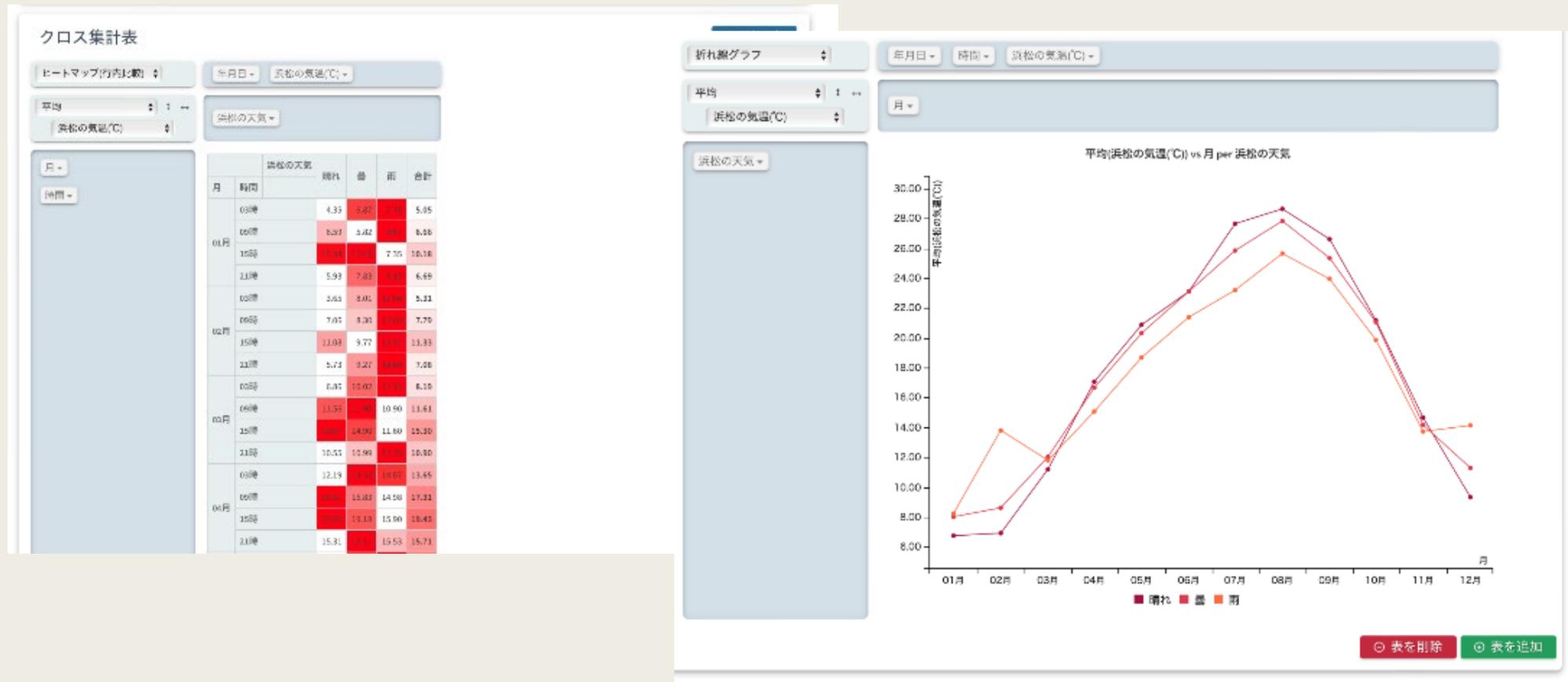
X軸: グループ別表示:

Y軸:



PCIは大学入試センターとの共同開発。

データ解析ツールを用いて実際のデータを分析しながら解答を求める問題の例



PCIは大学入試センターとの共同開発.

4.5. CBT実施動画

tao TAO: CBT説明会

172.21.210.230:8080/taoProctoring/DeliveryServer/runDeliveryExecution?deliveryExecution=http%3A%2F%2Ftao.127.0.0.1.xip.io%2Fdefault.rdf%23i655d9d974c643136576bfbde92acd79f3f

Test 27 - Section 1

0% 無線

Test-taker 25 ログアウト

次に示す論理回路の真理値表として適当なものを、下の選択肢の中から1つ選べ。

図. 論理回路

(ア)

入力		出力
A	B	X
0	0	1
1	0	1
0	1	1
1	1	1

(イ)

入力		出力
A	B	X
0	0	1
1	0	1
0	1	0
1	1	0

(ウ)

入力		出力
A	B	X
0	0	0
1	0	1
0	1	1
1	1	0

(エ)

入力		出力
A	B	X
0	0	1
1	0	0
0	1	0
1	1	1

(オ)

入力		出力
A	B	X
0	0	1
1	0	0
0	1	1
1	1	1

(カ)

入力		出力
A	B	X
0	0	1
1	0	1
0	1	0
1	1	1

(キ)

入力		出力
A	B	X
0	0	0
1	0	0
0	1	1
1	1	0

(ク)

入力		出力
A	B	X
0	0	0
1	0	0
0	1	0
1	1	1

(ア) (イ) (ウ) (エ) (オ) (カ) (キ) (ク)

▶ 次へ スキップ

© 2013 - 2023 - 2021.11 - Open Assessment Technologies S.A. - All rights reserved.

4.6. 新入生の学力調査

数学90分 (30問ほど) 、物理80分 (30問ほど) 、非認知能力テスト10分
クラスごとに異なるテストを実施
情報は 後期に開始

4.7. フィードバック例

フィードバック画面

試験名 UEC検定 第1回 科目 数学 表示

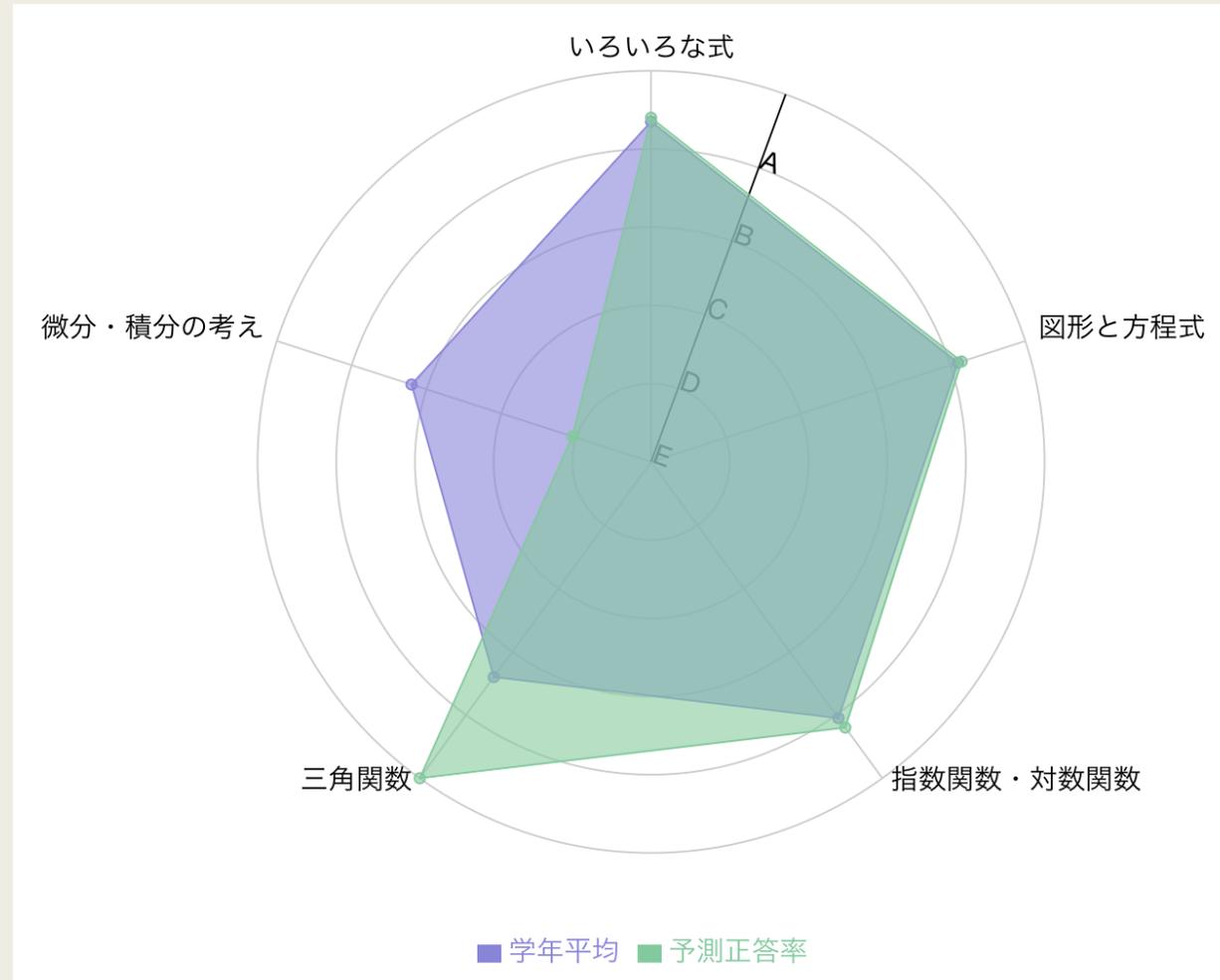
平均予測正答率	習熟度	解説
81~100%	A	内容を十分習熟できています。
61~80%	B	内容を習熟できています。
41~60%	C	習熟度がやや不足しています。自習を行いましょう。必要に応じて補習に参加しましょう。
21~40%	D	習熟度が不足しています。補習に参加しましょう。
0~20%	E	知識が不足しています。基礎から自習を行った後、補習に参加しましょう。

① あなたのスコア偏差値は50です。判定はC、平均予測正答率は59%です。

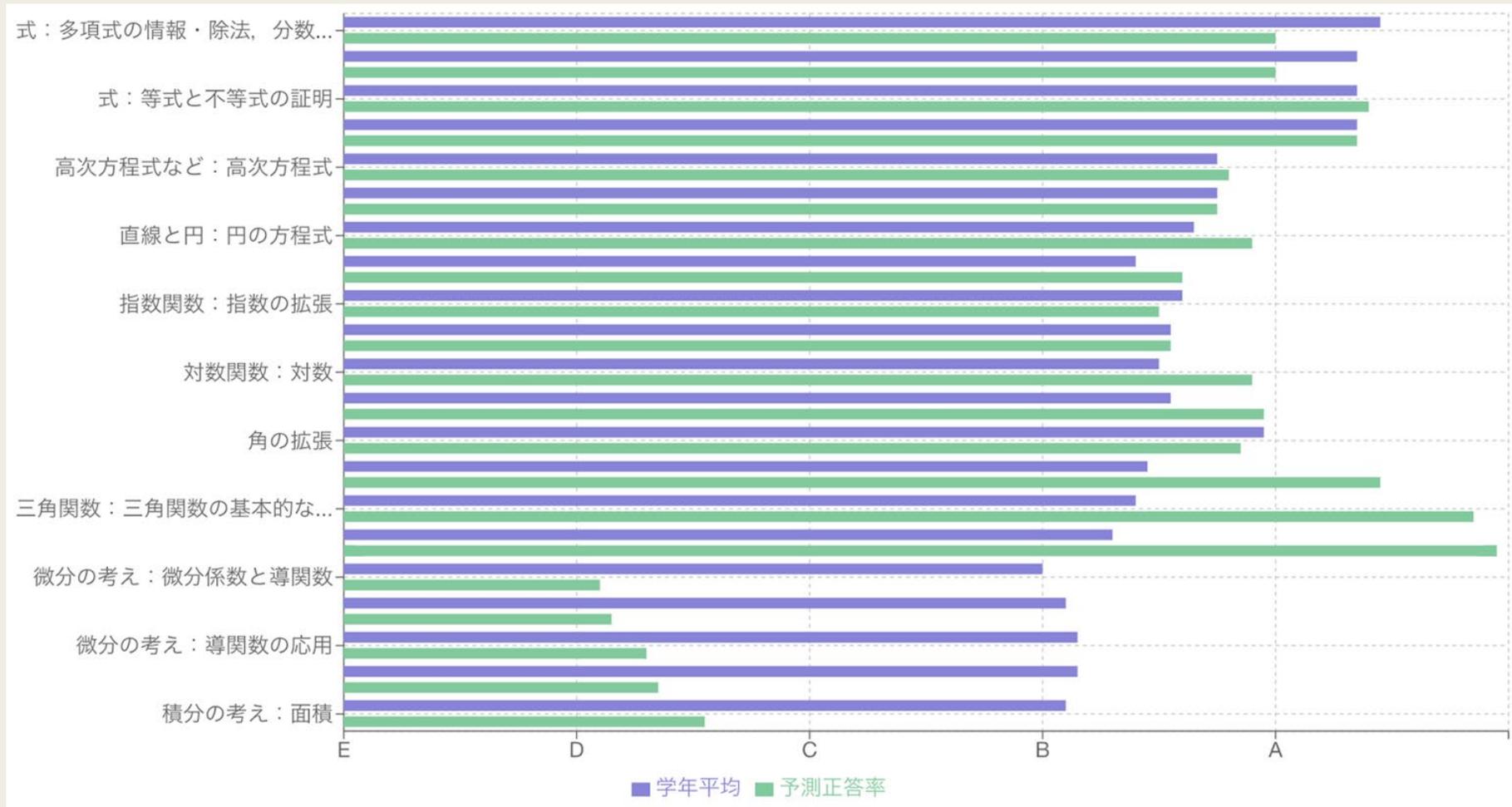
- ②
- 数学I
 - 数学II
 - 数学III
 - 数学A
 - 数学B
 - 数学C

- ① : 該当科目のスコア偏差値、判定及びあなたの回答から算出した平均予測正答率
- ② : クリックすると、次ページ以降の分野別・領域別のフィードバック

フィードバックの実施例（分野別）



フィードバックの実施例（領域別）



5. AI時代のテストとその未来

5.1. AI時代

「米国企業が人工知能（AI）の活用を理由にした従業員の解雇に動いている。大手通信会社などはAIで代替できる事務部門の雇用を削減。IBMはリスキリング（学び直し）の機会を提供し、配置転換を加速する。AIが理由の人員削減は全米で今年8月までに約4000人にのぼり、「AI時代」到来に向けた人材の流動化が始まった。」

2023/9/24

日本経済新聞 電子版

5.2. 日本の少子高齢化と労働人口の減少

日本は深刻な少子高齢化により労働人口の減少が危惧される。AIは日本の労働人口減少の危機の救世主となることが期待される。

特に今後CBTが普及した場合CBT運営に関する人材は日本は不足する。AIによるCBT運用は重要課題となる。

5.3. AI時代のテスト

労働人口減少時代のAI時代に適応できる社会づくりが重要.

テスト分野では,

- AI時代には, 人間が必要とする知識が変容し, 教育目標, テストの評価目標が変容
- AIによるテスト業務支援により少人数の専門家で効率的にテスト運用

5.3. AI時代に必要なスキルを測定するテスト

- 工業社会（従来の日本型） How to do

価値を語ることがテストでは評価されない時代

正誤が明確な問題のみしか扱わない

- AI時代 What to do とWhy

価値創造と説明力こそ重要（世の中を変えられる力）

そのようなスキルを評価するためには 論文形式やプレゼンテーション形式、質疑応答の評価は必須となるであろう。

価値を伴う論文形式の業績が今後、学者でなくても個人の大きな成果となるであろう。

5.4. LLM・生成AIによる 思考力を問う問題項目の自動採点

- 論述式問題項目
 - 数学証明問題
 - 図形の記述問題
- 等

5.6. 深い思考重視の日本型CBTは誤差大

- 日本型テストでは深い思考を問う問題が多く、所要時間が長く、十分な項目数が出題できない → 欧米型CBTに比べて テストの誤差が増大
- 難易度が能力値の平均付近のみに対応している問題項目が多いため、能力値に対して測定誤差の偏りが大きい。

やはり 適応型テストしかない！！

CAT (Computerized Adaptive Testing; 適応型テスト)

受検者の問題項目への反応履歴（正誤履歴）より逐次能力推定し、情報量を最大にする問題項目を提示するCBT.

利点

1. 測定精度を落とさずに出題項目数を減少できる
2. 各受検者への提示問題数を減らすことができるので暴露総数を減じることができる
3. 難しすぎる問題項目や易しすぎる問題項目が出題されなくなり受検時のストレスを減じる
4. 能力値の違いによる測定誤差の偏りを緩和できる

欠点

同一能力の受検者には同じ問題項目を出題するので、結果として問題項目の露出偏りが大きくなる。

5.7.項目露出を制御する等質適応型テスト

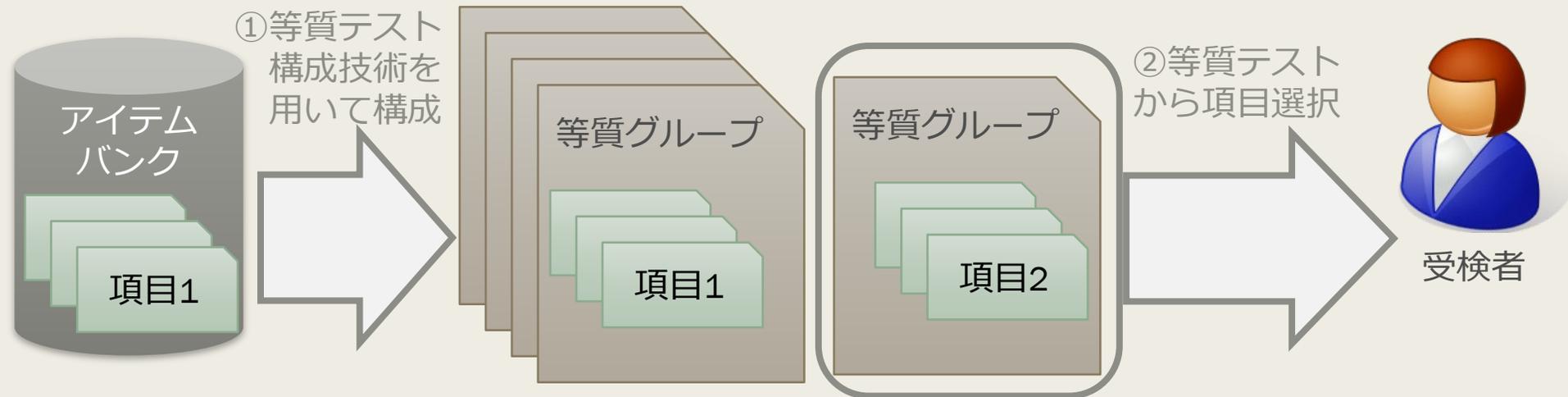
1. Wakaba Kishida, Kazuma Fuchimoto, Yoshimitsu Miyazawa, Maomi Ueno (2023) Item difficulty constrained uniform adaptive testing. International Conference on Artificial Intelligence in Education (AIED), Communications in Computer and Information Science, vol. 1831, pp. 568–573.
2. 宮澤 芳光, 植野 真臣 (2023) 高精度能力推定を保証する2段階等質適応型テスト, 電子情報通信学会論文誌D, Vol. J106-D, No.1, pp. 34–46.
3. Maomi Ueno, Yoshimitsu Miyazawa (2022) Two-Stage Uniform Adaptive Testing to Balance Measurement Accuracy and Item Exposure. International Conference on Artificial Intelligence in Education (AIED), Lecture Notes in Computer Science, vol. 13355, pp. 626–632.
4. Yoshimitsu Miyazawa, Maomi Ueno (2020) Computerized Adaptive Testing Method Using Integer Programming to Minimize Item Exposure. Advances in Artificial Intelligence: Selected Papers from the Annual Conference of Japanese Society of Artificial Intelligence (JSAI), pp.105–113.
5. Maomi Ueno, Yoshimitsu Miyazawa (2019) Uniform adaptive testing using maximum clique algorithm. International Conference on Artificial Intelligence in Education (AIED), Lecture Notes in Computer Science, vol. 11625, pp. 482–493.

■ 受賞 (2件)

1. 日本テスト学会大会発表賞 (2022) 宮澤 芳光, 湊本 志真, 植野 真臣. 「等質テスト構成の並列化技術を用いた2段階等質適応型テスト」
2. 人工知能学会 JSAI Excellence Award: International Session (2019) Yoshimitsu Miyazawa, Maomi Ueno, Computerized Adaptive Testing Method using Integer Programming to Minimize Item Exposure.

(1) 等質適応型テスト (Ueno and Miyazawa, 2019)

各受検者に異なる問題項目を適応的に出題し，項目露出の偏りを増大させず測定精度を落とさずに項目出題数を減じ，等質のテストを保証する。



Maomi Ueno, Yoshimitsu Miyazawa (2019) Uniform adaptive testing using maximum clique algorithm.

International Conference on Artificial Intelligence in Education (AIED),
Lecture Notes in Computer Science, vol. 11625, pp. 482–493.

(2) 2段階等質適応型テスト

(Ueno and Miyazawa, 2022, Kishida 他 2023,
宮澤・植野, 2023)

テストの前半 (第1段階) に等質な項目集合から
項目を選択

テストの後半 (第2段階) にアイテムバンク全体
を第1段階の能力推定値の近傍の難易度お項目集合か
ら出題する

Wakaba Kishida, Kazuma Fuchimoto, Yoshimitsu Miyazawa, Maomi Ueno (2023)
Item difficulty constrained uniform adaptive testing. International Conference on
Artificial Intelligence in Education (AIED), Communications in Computer and
Information Science, vol. 1831, pp. 568–573.

宮澤 芳光, 植野 真臣 (2023) 高精度能力推定を保証する2段階等質適応型テスト,
電子情報通信学会論文誌D, Vol. J106-D, No.1, pp. 34–46.

Maomi Ueno, Yoshimitsu Miyazawa (2022) Two-Stage Uniform Adaptive Testing to
Balance Measurement Accuracy and Item Exposure.

International Conference on Artificial Intelligence in Education (AIED), Lecture Notes
in Computer Science, vol. 13355, pp. 626–632.

5.8. 実験結果

テストの長さ	手法	グループ数	暴露数の平均値		MSE	未出題項目数
30	CAT	-	1986.75	(2145.36)	0.07 (0.10)	827
	KZ(30)	32	1578.95	(1616.64)	0.08 (0.13)	788
	IP	-	753.77	(381.24)	0.12 (0.18)	580
	Prob.	-	867.05	(396.43)	0.12 (0.17)	632
	TI	-	306.75	(1005.18)	0.12 (0.18)	0
	UAT-RIPMCP(30)	743	306.75	(83.40)	0.27 (0.39)	0
	UAT-HMCAPIP(30)	106570	305.75	(68.51)	0.27 (0.39)	0
	TUAT[UAT-AT]-RIPMCP(30,0.25)	743	495.05	(1202.79)	0.07 (0.11)	375
	TUAT[UAT-AT]-HMCAPIP(30,0.25)	106570	484.65	(1197.73)	0.07 (0.10)	359
	TUAT[UAT-TM]-RIPMCP(30,0.25)	743	307.06	(768.36)	0.07 (0.11)	1
TUAT[UAT-TM]-HMCAPIP(30,0.25)	4264	304.54	(731.16)	0.07 (0.10)	0	
50	CAT	-	2192.98	(2270.93)	0.05 (0.07)	750
	KZ(25)	39	1474.93	(1897.74)	0.05 (0.08)	639
	IP	-	803.86	(348.43)	0.09 (0.14)	356
	Prob.	-	929.37	(388.18)	0.09 (0.12)	440
	TI	-	511.25	(1313.30)	0.13 (0.18)	0
	UAT-RIPMCP(50)	726	511.25	(105.83)	0.16 (0.24)	0
	UAT-HMCAPIP(50)	89339	511.15	(88.87)	0.16 (0.25)	0
	TUAT[UAT-AT]-RIPMCP(50,0.075)	726	641.85	(1291.37)	0.05 (0.08)	288
	TUAT[UAT-AT]-HMCAPIP(50,0.075)	89339	641.67	(1303.78)	0.05 (0.07)	200
	TUAT[UAT-TM]-RIPMCP(50,0.075)	726	512.29	(512.29)	0.05 (0.07)	2
TUAT[UAT-TM]-HMCAPIP(50,0.075)	4872	509.10	(1203.94)	0.05 (0.07)	0	

適応型テスト (CAT・AT)、アイテムバンク分割法 (KZ)、制約付き適応型テスト (IP)、適格確率を用いた適応型テスト (Prob)、 v 等質適応型テスト (UAT)、二段階等質適応型テスト (TUAT)、

所望の測定精度の項目集合から露出率が最小の項目を選択 (TM)

5.9. 制限時間を考慮した等質適応型テスト

5.9.1. Deep-IRT

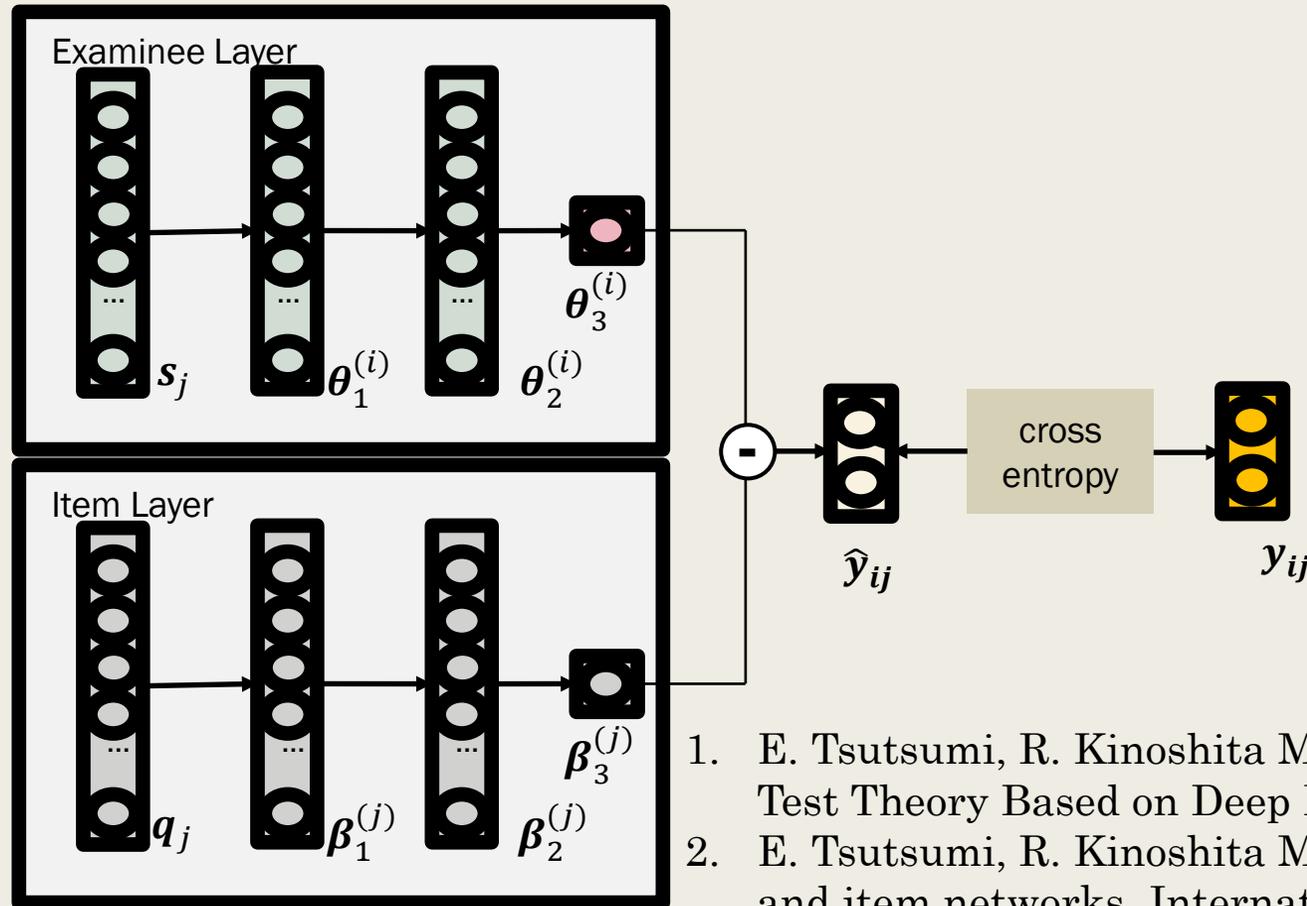
用いるツール： Deep-IRT

1. E. Tsutsumi, R. Kinoshita M. Ueno (2021) Deep Item Response Theory as a Novel Test Theory Based on Deep Learning. *Electronics*, no.10, pp.10-20
2. E. Tsutsumi, R. Kinoshita M. Ueno (2021) Deep-IRT with independent student and item networks, International Conference on Educational Data Mining (EDM)
3. E. Tsutsumi, Y. Guo, M. Ueno (2022) Deep knowledge tracing incorporating a hypernetwork with independent student and item networks, International Conference on Educational Data Mining (EDM)
4. E. Tsutsumi, Y. Guo, M. Ueno (2023) Deep knowledge Tracing incorporating hypernetwork with independent student and item networks, *IEEE Transactions on Learning Technologies*, doi: 10.1109/TLT.2023.3346671

(1) Deep-IRT (Tsutsumi et al 2021)

パラメータの解釈性と高精度な反応予測を両立するDeep-IRTの開発

2つの独立ニューラルネットワークを用いることで局所独立性や受検者のランダムサンプリングの仮定なしでIRTにおける能力値パラメータと難易度パラメータと同様に解釈できるパラメータ推定を実現



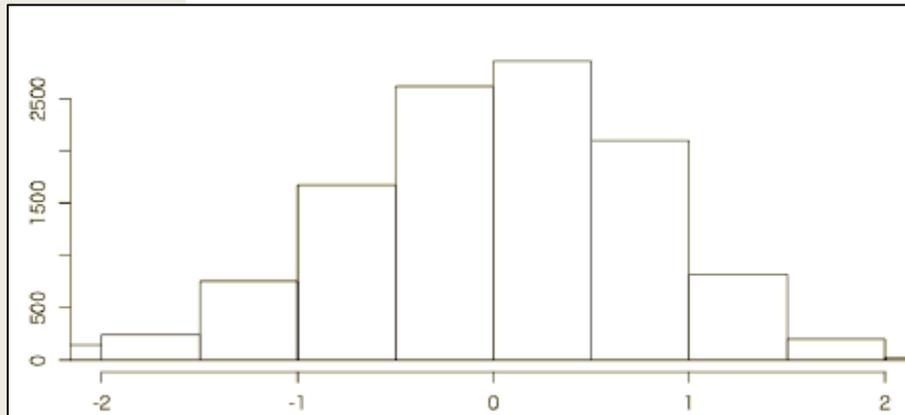
利点

1. 局所独立性の仮定がいらぬ
2. 受検者のランダムサンプリングの仮定がいらぬ
3. 予測精度がIRTに比較して非常に高い
4. 多次元能力とスキルを自動的に割付可能
5. 能力変動の時系列モデルに容易に拡張可能
6. LLMなどの自然言語処理. 所要時間などのデータ処理への拡張が容易

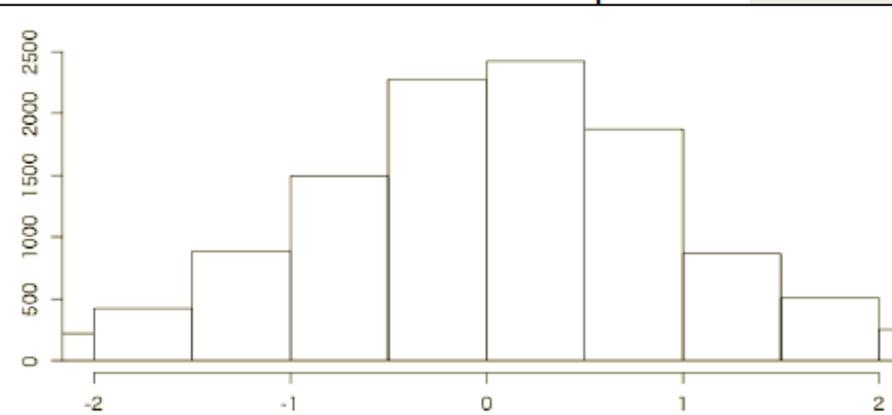
1. E. Tsutsumi, R. Kinoshita M. Ueno (2021) Deep Item Response Theory as a Novel Test Theory Based on Deep Learning. Electronics, no.10, pp.10-20
2. E. Tsutsumi, R. Kinoshita M. Ueno (2021) Deep-IRT with independent student and item networks, International Conference on Educational Data Mining (EDM)

Data	No. examinees	No. items	Rate.Sparse	IRT	Deep-IRT
Information 1	169	50	0%	0.734	0.737
Information 2	266	50	0%	0.699	0.700
Critical Thinking	1221	179	87.8%	0.695	0.689
Program 1	94	13	0%	0.719	0.729
Program 2	74	19	6.8%	0.676	0.685
Practice_Math	12348	48	16.4%	0.783	0.780
Practice_Physics	9172	24	12.0%	0.721	0.710
ASSISTMENTS	3941	2921	84.4%	0.685	0.679
ECPE	2922	28	0%	0.719	0.729
TIMSS	757	24	0%	0.711	0.712
Statistics	26	25	33.8%	0.852	0.893
Information Ethics	31	90	46.3%	0.746	0.803
Engineer Ethics	85	69	26.4%	0.634	0.685
Classi_Physics	239	119	92.4%	0.720	0.721
Classi_Chemistry	1139	364	96.4%	0.710	0.711
Classi_Biology	192	114	93.5%	0.722	0.725
Average				0.719	0.728*

*p<.05



(a) Abilities estimated using IRT

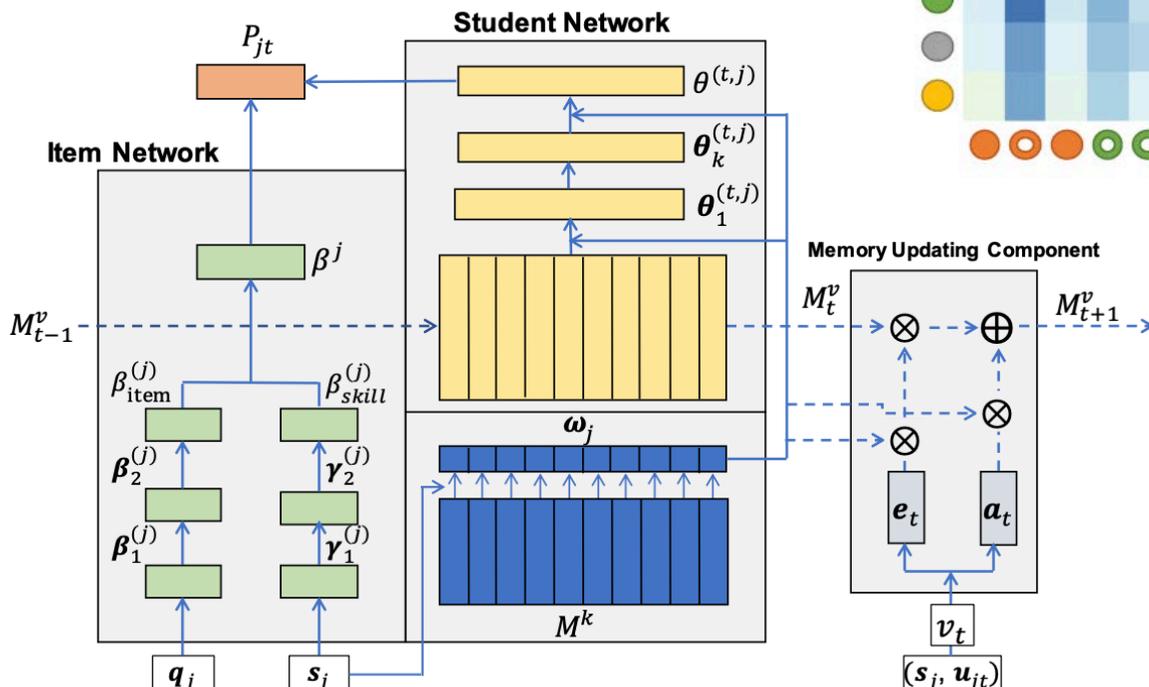


(b) Abilities estimated using Deep-IRT

(2) 多次元スキル, 能力時系列変化への拡張

- Tsutsumi, E., Guo, Y., and Ueno, M., Deep knowledge tracing incorporating a hypernetwork with independent student and item networks, EDM, (2022)
- 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)

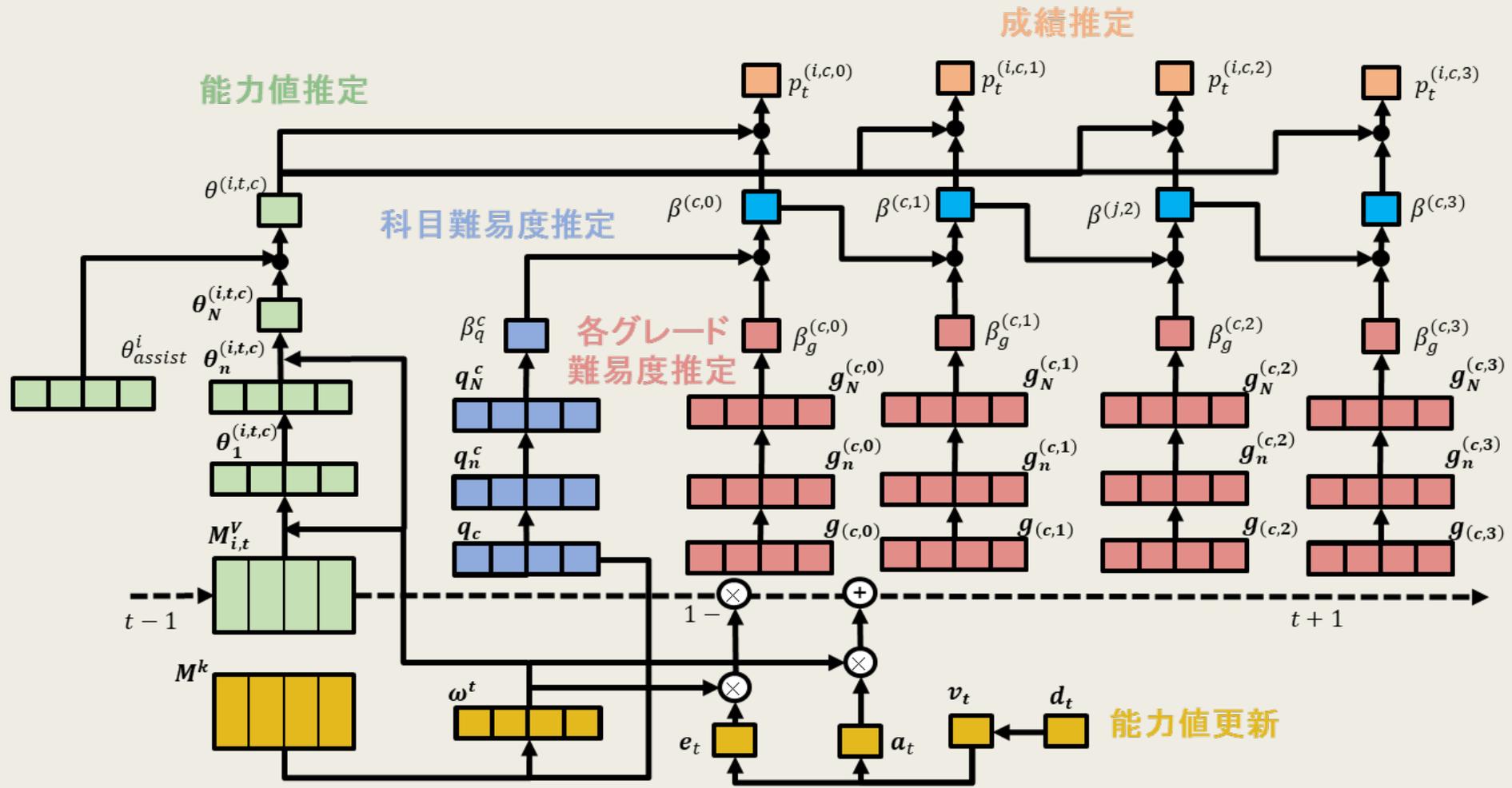
→ 現在世界最高の
アダプティブラーニング予測精度を達成



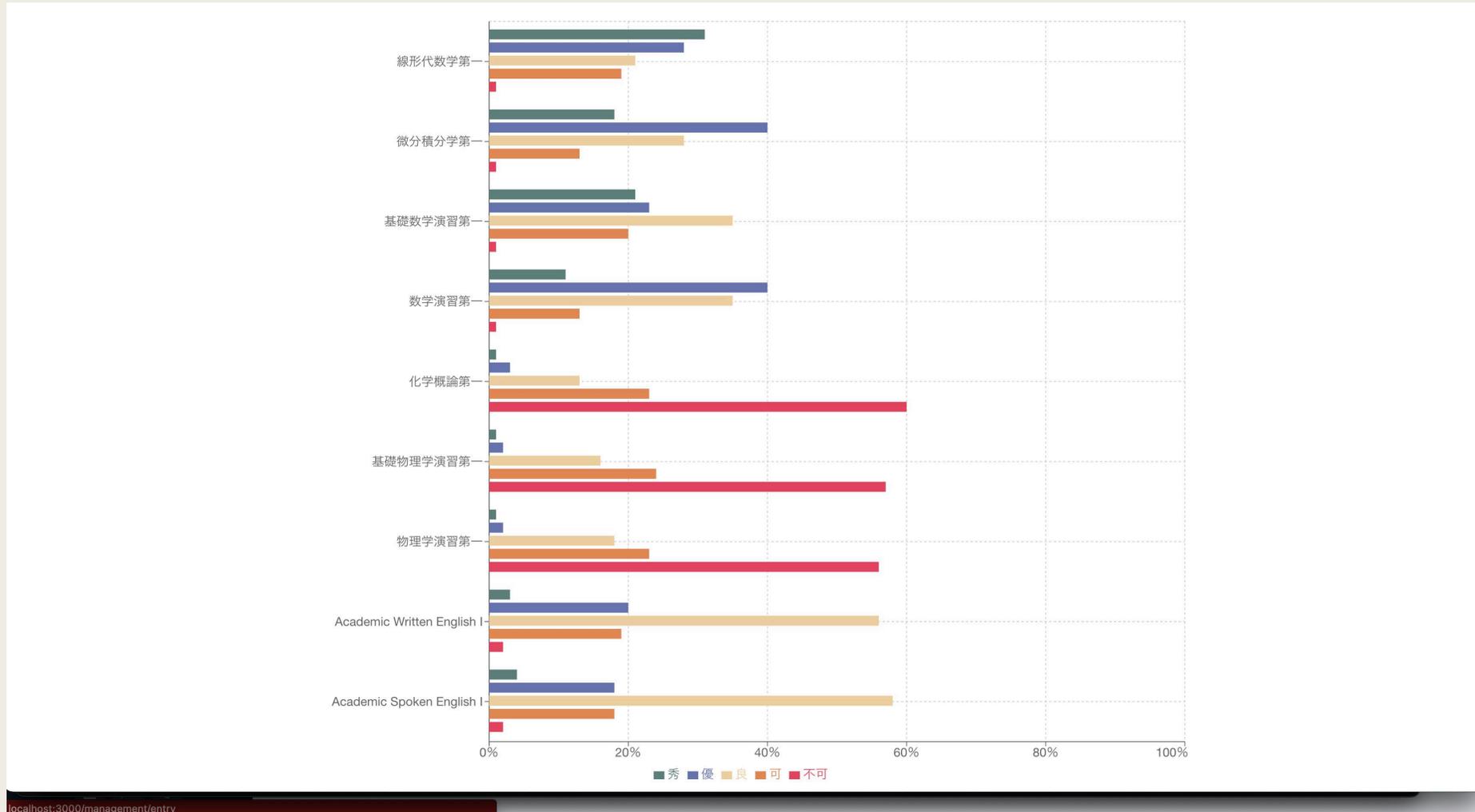
4次元スキルに対する能力変化

- 2021 年度人工知能学会全国大会 大会優秀賞(口頭発表部門)
- 2021 年度教育システム情報学会 大会奨励賞(口頭発表部門)
- 2019 年度人工知能学会全国大会 大会優秀賞(口頭発表部門)

(3) UEC検定のDeep-IRTによる単位予測とスキル習熟度予測

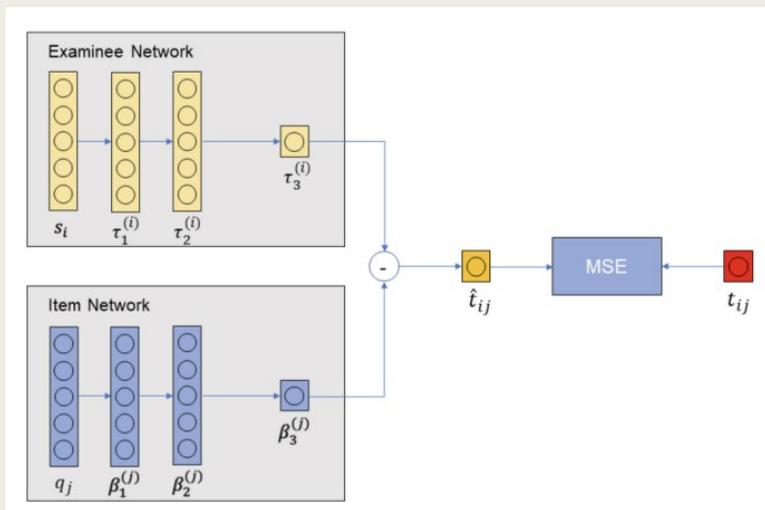


單位成績預測(誤差1.0)



5.9.2 所要時間のためのDeep-IRT

- 制限時間を考慮するためには、各受験者の各項目への所要時間を予測する必要
- 岸田らはDeep-IRT[Tsutsumi 2021]を基に
所要時間を予測可能な手法を提案[岸田 2024]



所要時間のためのDeep-IRTの概要図

s_i : Examinee Networkの入力ベクトル

q_j : Item Networkの入力ベクトル

$$\tau_1^{(i)} = \tanh(\mathbf{W}^{(\tau_1)} s_i + \mathbf{b}^{(\tau_1)}),$$

$$\tau_2^{(i)} = \tanh(\mathbf{W}^{(\tau_2)} \tau_1^{(i)} + \mathbf{b}^{(\tau_2)}),$$

$$\tau_3^{(i)} = \mathbf{W}^{(\tau_3)} \tau_2^{(i)} + \mathbf{b}^{(\tau_3)}.$$

$$\beta_1^{(j)} = \tanh(\mathbf{W}^{(\beta_1)} q_j + \mathbf{b}^{(\beta_1)}),$$

$$\beta_2^{(j)} = \tanh(\mathbf{W}^{(\beta_2)} \beta_1^{(j)} + \mathbf{b}^{(\beta_2)}),$$

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

$$\hat{t}_{ij} = \mathbf{W} (\tau_3^{(i)} - \beta_3^{(j)}) + \mathbf{b}.$$

\mathbf{W} : 回答時間重みパラメータ

\mathbf{b} : バイアスパラメータ

• Tsutsumi, Emiko, Ryo Kinoshita, and Maomi Ueno. 2021. "Deep Item Response Theory as a Novel Test Theory Based on Deep Learning." *Electronics* 10 (9): 1020.

• 岸田若葉, 堤瑛美子, 植野真臣: 所要時間のためのDeep-IRT, 第38回人工知能学会全国大会, 2024 (発表予定)

5.9.3. 所要時間の予測精度改善

- 受検者の各項目への所要時間の予測精度を従来手法よりも改善

従来手法[Linden 2007]との予測時間のRMSEの比較結果

Linden 2007	岸田 2024
220.92	114.36

- 岸田若葉, 堤瑛美子, 植野真臣: 所要時間のためのDeep-IRT, 第38回人工知能学会全国大会, 2024 (発表予定)
- Linden, Wim J. van der. 2007. "A Hierarchical Framework for Modeling Speed and Accuracy on Test Items." *Psychometrika* 72 (3): 287–308.

5.9.4.制限時間を持つCAT

- 提案手法は従来手法よりもテストの所要時間のばらつきを軽減
- 提案手法では約95%の受験者が制限時間内に推定精度を保ち回答を終えた

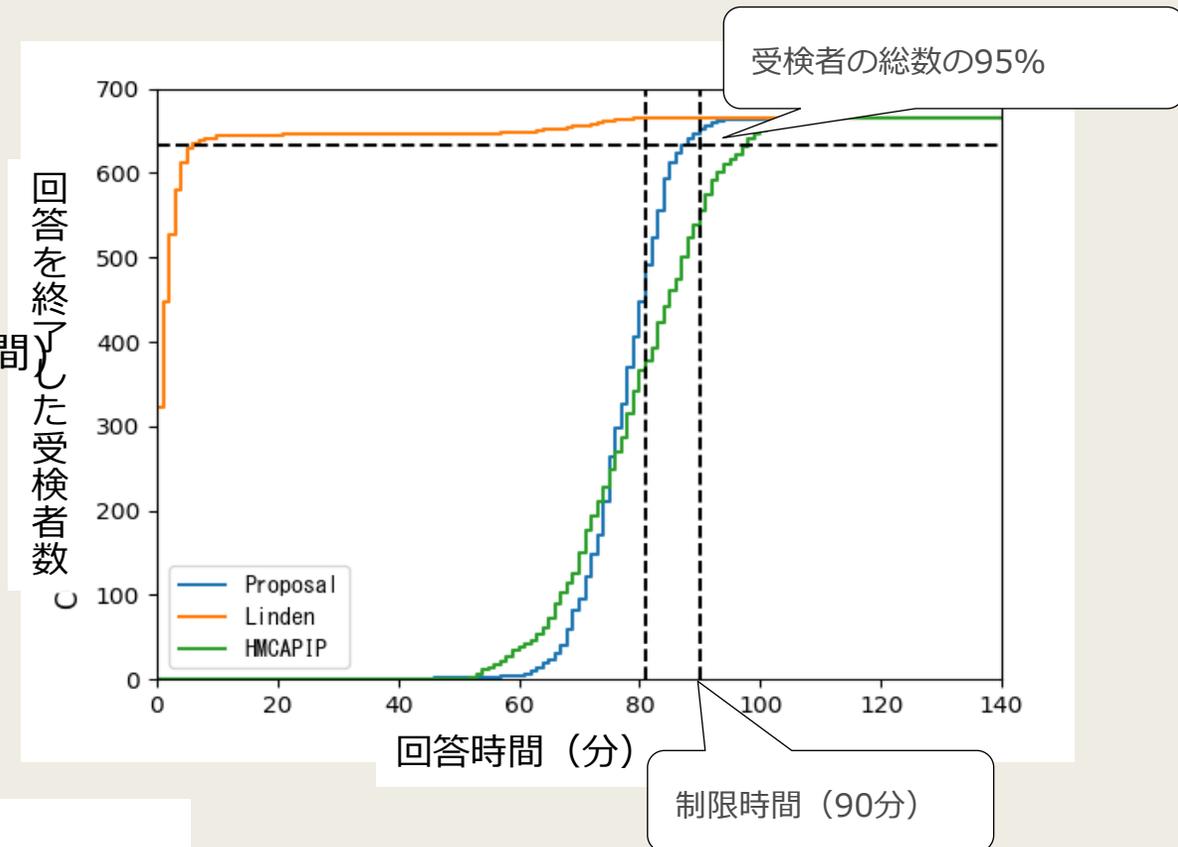
従来手法[Fuchimoto 2022]の整数計画法に、制限時間の制約を追加

subject to $LB_{TL} \leq \sum_{i=1}^n \hat{t}_i x_i \leq UB_{TL}$

LB_{TL} : テストの所要時間の下限

UB_{TL} : テストの所要時間の上限(制限時間)

\hat{t}_i : 各項目の所要時間の予測値

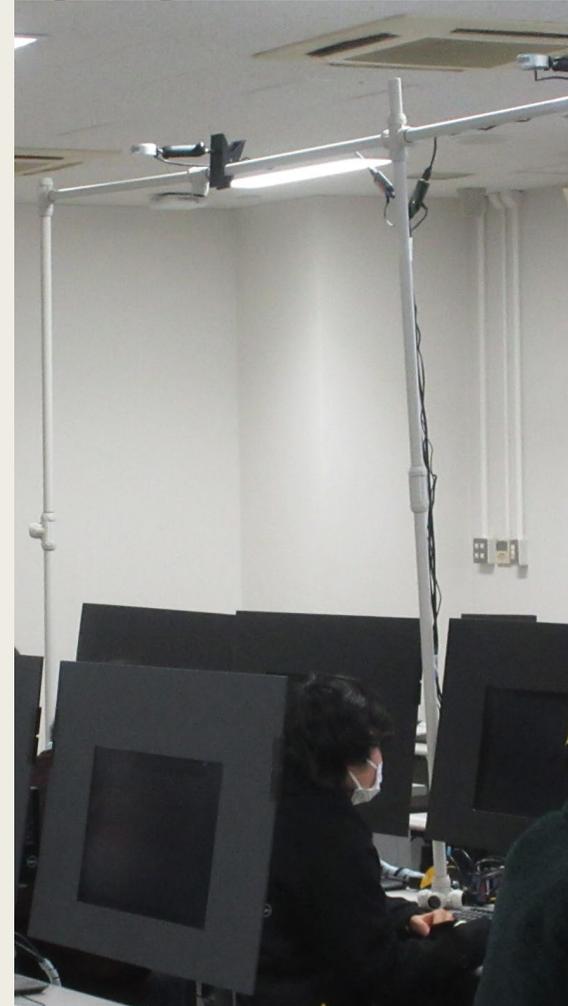


5.10. AIによるCBT運営

AIカメラによる試験監督負担減少！！ 将来的には無人テストセンターも！！

天井方向からカメラ
で撮影

- 不正行為の自動検出



5.11. LLM生成AIによる所望の難易度の問題項目の自動作問

■ 問題項目の特徴量とIRTの難易度によるファインチューニングによる自動作問

問題

ある郵便局には、複数の荷物があります。各荷物には特定の重さがあり、トラックは特定の最大積載量を持っています。以下の短冊から適切に並べ替えて、荷物の重さをすべて載せるために必要なトラックの最小数を計算するプログラムを完成させてください。

短冊

```
1. `def min_trucks(n, packages, max_load):`  
2. `truck_count += 1`  
3. `packages.remove(pkg)`  
4. `current_load = 0`  
5. `return truck_count`  
6. `current_load += pkg`  
7. `if current_load + pkg <= max_load:`  
8. `while len(packages) > 0:`  
9. `for pkg in packages[:]:`  
10. `packages.sort(reverse=True)`  
11. `truck_count = 0`  
12. `packages.append(pkg)`  
13. `if current_load < max_load:`  
14. `for weight in packages:`  
15. `truck_count -= 1`
```

今の時点では人間のチェックが必要であるが、現在でも相当のクオリティの作問を実現できる。

将来、その場で受検者の能力に適応した問題項目（情報量を最大にする問題）を自動生成することも可能である

5.12. AI時代のCBTとその未来のまとめ

- まずは国際標準に従うCBT導入.
- AI時代に生き残れる思考力を問う日本固有のテスト文化を実現する作問とシステム開発
- AIによる作問，テスト運用の負荷軽減により CBT導入コストを削減し，従来のペーパーテストからのコスト減と受検者能力の予測精度向上が期待できる
- 将来はほぼAIで運営できるCBT運営が実現するかもしれない
- 価値を伴う論文形式の業績が今後、学者でなくても個人の大きな成果となるであろう。

■ ご清聴ありがとうございました。