

**テスト情報量と移動距離を最適化する  
モバイル・テストング・システム**

**宮澤 芳光**

電気通信大学大学院 情報システム学研究科

学位申請論文 博士(工学)

2014年3月



テスト情報量と移動距離を最適化する  
モバイル・テストング・システム

博士論文審査委員会

主査: 植野 真臣 教授  
委員: 大須賀 昭彦 教授  
委員: 栗原 聡 教授  
委員: 田原 康之 准教授  
委員: 笠井 裕之 准教授

著作権所有者

宮澤 芳光

2014 年

# Mobile Testing System Optimizing Test Information and Movement Distance

Yoshimitsu Miyasawa

## **Abstract**

Knowledge is embedded in various situations rather than existing on its own. Moreover, knowledge is known to be acquired in conjunction with past experiences. These observations suggest that mobile technologies enhance learning in situations. Some effective learning support systems using mobile technologies have been developed. However, assessment methods for these learning styles have not been discussed sufficiently. Effective assessment is expected to be embedded in the same situations. For this purpose, mobile testing, which provides test items on a mobile device in the field, is known to be effective. However, accurate assessment of student's ability requires a sufficient number of responses to items. The mobile testing might not be sufficient to provide enough test items because it sometimes consumes time for movement in the field. Therefore, mobile testing must use a more efficient test item presentation strategy. A more effective presentation method is known as Computerized Adaptive Testing (CAT). In fact, CAT progressively estimates the examinee's ability from the answer history and uses an item bank to present test items that maximize the amount of item information with regard to an ability estimate of the user. However, tradi-

tional adaptive testing does not guarantee optimization of the examinee's movement distance during testing in the field. The purpose of this study is to improve the accuracy of estimating an examinee's ability for tests in the field. Therefore, we propose a mobile CAT system that optimizes both test information and movement distance in the field. For this purpose, we use the traveling purchaser problem (TPP), an optimization problem using graph theory. Furthermore, we provided some simulation and actual experiments to demonstrate more accurate measuring performance of the proposed method.

---

# テスト情報量と移動距離を最適化するモバイル・テストイング・システム

宮澤 芳光

## 和文概要

本研究では、状況に埋め込まれた学習の効率的な評価を目的とする。具体的には、現実の状況における観察や探索を通じた学習の後、同じ状況で学習を効率的に評価する。ここでは、効率的な評価を実現するため、テスト理論に基づく適応型テストを用いて受検者の能力値を推定する。適応型テストとは、受検者の能力値を逐次的に推定し、その能力値に対して情報量が最大の項目を出題する手法である。しかし、状況に埋め込まれた学習の評価のために受検者は学習と同じ場所まで移動する必要があるため、移動時間が増加し、受検者が十分な項目数を解答する前にテストが終了するかもしれない。そこで、本研究では、最適化問題の一つである Traveling Purchaser Problem (TPP) が組み込まれた適応型テストを提案し、この適応型テストを用いたモバイル・テストイング・システムを開発する。具体的には、制限時間内で移動距離最小化とテスト情報量最大化を同時に満たす項目を逐次的に選択する手法を提案する。本システムの利点として以下が挙げられる。(1) 移動経路の最適化により、テストの解答所要時間に対する移動時間の割合が減少する。(2) 移動時間の減少により受検者が解答できる項目数が増加する。(3) 情報量が高い項目を出題できるため、高精度な能力測定が期待できる。本研究では、シミュレーション実験および被験者実験から上記の利点を検証し、本システムの有効性を示した。

# 目次

<b>第 1 章</b>	<b>緒言</b>	1
<b>第 2 章</b>	<b>適応型テストを用いたモバイル・テストング・システム</b>	4
2.1	はじめに . . . . .	4
2.2	先行研究 . . . . .	6
2.3	適応型テスト . . . . .	7
2.4	適応型テストを用いたモバイル・テストング・システム . . .	15
2.5	アイテムバンク . . . . .	16
2.5.1	アイテムバンクのスキーマ . . . . .	16
2.5.2	対象の寺院 . . . . .	21
2.5.3	アイテムバンクの開発 . . . . .	22
2.6	評価実験 . . . . .	25
2.6.1	テストング機能の有効性評価 . . . . .	27
2.6.2	ナビゲーション機能の有効性評価 . . . . .	28
2.6.3	観察・探索活動の分析 . . . . .	28
2.7	おわりに . . . . .	36
<b>第 3 章</b>	<b>テスト情報量と移動距離を最適化する</b>	
	<b>モバイル・アダプティブ・テストング</b>	39
3.1	はじめに . . . . .	39
3.2	制限時間を制約として組み込んだ適応型テスト . . . . .	41



---

3.3	Traveling Purchaser Problem を用いた適応型テスト . . . . .	43
3.3.1	Traveling Purchaser Problem(TPP) . . . . .	43
3.3.2	テスト情報量と移動距離を最適化する TPP の定式化 . . . . .	44
3.3.3	動的計画法を用いた最適解探索アルゴリズム . . . . .	45
3.3.4	Traveling Purchaser Problem を用いた適応型テストの アルゴリズム . . . . .	47
3.4	シミュレーション実験 . . . . .	48
3.5	評価実験 . . . . .	55
3.5.1	実験の目的 . . . . .	55
3.6	おわりに . . . . .	60
<b>第 4 章</b>	<b>結言</b> . . . . .	<b>72</b>
	<b>参考文献</b> . . . . .	<b>75</b>

## 目次

2.1	識別力のパラメータ $a_i$ を変化させた項目反応関数 . . . . .	9
2.2	難易度のパラメータ $b_i$ を変化させた項目反応関数 . . . . .	10
2.3	当て推量のパラメータ $c_i$ を変化させた項目反応関数 . . . . .	11
2.4	フィッシャー情報量 . . . . .	12
2.5	適応型テストのアルゴリズム . . . . .	14
2.6	モバイル・テストのアルゴリズム . . . . .	17
2.7	システム構成図 . . . . .	18
2.8	ナビゲーション機能の画面例 . . . . .	19
2.9	項目表示機能の画面例 . . . . .	19
2.10	アイテムバンクの属性 . . . . .	20
2.11	アイテムバンクのスキーマ . . . . .	21
2.12	展示物の知識を評価するアイテム・バンクの記述例 . . . . .	22
2.13	特定の状況における知識を評価するアイテム・バンクの記述例	23
2.14	難易度のパラメータ $b_i$ の推定に用いた反応データ . . . . .	26
2.15	深大寺の写真 . . . . .	27
2.16	真の能力値が 1.0 のときの推定の過程 . . . . .	29
2.17	真の能力値が 0.8 のときの推定の過程 . . . . .	29
2.18	真の能力値が 0.6 のときの推定の過程 . . . . .	30
2.19	真の能力値が 0.4 のときの推定の過程 . . . . .	30
2.20	真の能力値が 0.2 のときの推定の過程 . . . . .	31

---

2.21	真の能力値が-0.2 のときの推定の過程 . . . . .	31
2.22	真の能力値が-0.4 のときの推定の過程 . . . . .	32
2.23	真の能力値が-0.6 のときの推定の過程 . . . . .	32
2.24	真の能力値が-0.8 のときの推定の過程 . . . . .	33
2.25	真の能力値が-1.0 のときの推定の過程 . . . . .	33
2.26	深大寺の地図 . . . . .	34
2.27	仮想環境の配置図 . . . . .	36
3.1	ダイナミックプログラミングによる最適解の探索 . . . . .	47
3.2	最適化問題を組み込んだ適応型テストのアルゴリズム . . . . .	49
3.3	1 項目目のパス例 . . . . .	56
3.4	2 項目目のパス例 . . . . .	56
3.5	3 項目目のパス例 . . . . .	56
3.6	4 項目目のパス例 . . . . .	56
3.7	5 項目目のパス例 . . . . .	56
3.8	6 項目目のパス例 . . . . .	56
3.9	提案手法のパス . . . . .	57
3.10	時間のみ制約とした適応型テストのパス . . . . .	58
3.11	項目が特定の場所に偏ったアイテムバンクの例 . . . . .	59

## 表目次

2.1	眼球運動計測装置を用いた注視の計測結果 . . . . .	35
2.2	ビデオデータからの計測結果 (**有意水準 1% で有意差あり).	37
3.1	各適応型テストの平均テスト情報量 (標準偏差) の例 . . . . .	63
3.2	各適応型テストの平均解答項目数 (標準偏差) の例 . . . . .	64
3.3	各適応型テストの平均移動時間 (標準偏差) の例 . . . . .	65
3.4	拡張した提案手法の平均テスト情報量 (標準偏差) 分析 . . . . .	66
3.5	ランダム生成されたアイテムバンクを用いた各適応型テスト の平均テスト情報量 (標準偏差) 分析 . . . . .	67
3.6	容易な項目が多いアイテムバンクを用いた各手法の平均テス ト情報量 (標準偏差) 分析 . . . . .	68
3.7	難しい項目が多いアイテムバンクを用いた各手法の平均テス ト情報量 (標準偏差) 分析 . . . . .	69
3.8	特定の一か所に項目が集まったアイテムバンクを用いた各手 法の平均テスト情報量 (標準偏差) 分析 . . . . .	70
3.9	実験の結果 (** 有意水準 1% の有意差, * 有意水準 5% の有 意差) . . . . .	71
1	被験者実験に用いた項目 . . . . .	83
2	被験者実験に用いた項目 . . . . .	84
3	被験者実験に用いた項目 . . . . .	85

# 第1章

## 緒言

知識はそれ単体で存在するのではなく，状況に埋め込まれて存在する [1]. また，新しい知識は，現実の状況における学習者自身の経験に融和して獲得される [2]. このような知識観に基づき，現実での経験を重視する学習は「状況に埋め込まれた学習」と呼ばれ，近年注目されている.

状況に埋め込まれた学習は，これまで，数多くの携帯端末を用いたシステムにより支援されている. 例えば，屋外や博物館で展示物の学習を支援するため，学習者が持つ携帯端末に学習教材を配信し，学習教材に関連する展示物まで学習者を誘導するシステム [3][4] [5] や，RFID(Radio Frequency IDentification) タグや QR コードを用いて生徒が学習している対象物を同定し，その対象物に関する学習教材を携帯端末に配信するシステム [6][7][8][9] が開発されている.

これらの先行研究では，状況に埋め込まれた学習の支援にのみ着目しており，学習に対する評価手法については十分に議論されていない. 状況に埋め込まれた学習の評価では，現実の状況における，観察や探索を通じた学習そのものを評価することが重要であるため，学習と同様に現実の状況で行うべきである.

状況に埋め込まれた学習の評価を実現するシステムとして，Santos ら [10] は，GPS を用いて受検者の現在地を同定し，現在地に対応する項目を観察や探索を通して解答させるシステムを開発している. Huang ら [11] は，植物園にお

いて現在地周辺に咲いている草花の種類に関する知識を評価するため、RFIDを用いて草花を同定し、解答履歴に応じて適応的に項目を出題するシステムを開発している。

一方、信頼性の高い評価には、十分な項目数を受検者に解答させる必要がある。しかし、受検者が対象物間を移動する時間は増加するため、これらのシステムでは多くの項目を解答させることが困難である。

少数の項目で信頼性の高い評価を実現する手法として、項目反応理論を用いた適応型テスト (Computerized adaptive testing:CAT) が知られている [12, 13, 14, 15, 16]。適応型テストとは、受検者の解答履歴から知識状態を逐次的に測定し、その知識状態に対して情報量が最大の項目を出題する手法である。適応型テストを用いることにより、知識状態の測定精度を減少させずに、出題項目数を減らすことができる。

そこで本研究では、状況に埋め込まれた学習を効率的に評価するため、適応型テストを用いたモバイル・テスト・システムを開発する。

第2章では、適応型テストを用いたモバイル・テスト・システムについて詳述する。本システムでは、受検者の知識状態を効率的に測定するテスト機能と、展示物が設置された場所へ受検者を短時間で誘導するナビゲーション機能を持つ。本システムにより期待される利点は以下の通りである。1. 受検者の知識状態を効率的に測定できる。2. 展示物が設置された場所へ受検者を短時間で誘導できる。3. 状況に埋め込まれた学習そのものを評価することができる。具体的には、現実の状況で観察や探索を通して受検者が項目に解答することができる。ここでは、眼球運動計測装置を用いて受検者が観察や探索をしているかを検証する。評価実験では、シミュレーションおよび被験者実験から上記の利点を検証し、本システムの有効性を示した。

---

第3章では、状況に埋め込まれた学習の評価に費やされる移動時間を短縮させるため、移動距離とテスト情報量の最適化を組み込んだ適応型テストを提案する。具体的には、Traveling Salesman Problem の一般化の一つである Traveling Purchaser Problem (TPP) を組み込んだ適応型テストを提案する。TPP は、複数の商品が売られている店舗が点在するとき、商品の価格と移動距離が最小なパスを探索する最適化問題である。本研究では、TPP において定義されている商品と店を項目と項目が出題される場所として置き換え、テスト情報量最大化と移動距離最小化を同時に満たす最適なルートの探索を目的とする。しかし、TPP では、購入すべき商品数を制約としているが、本研究では、テストの制限時間を制約とする必要がある。本研究では、制限時間を制約とした TPP を提案し、この最適化問題を用いた適応型テストを提案する。本手法により、移動プロセスを効率化し、受検者の能力測定精度の向上が期待される。シミュレーション実験及び被験者実験により、先行研究と比較して能力測定精度が高いことを示す。

## 第2章

# 適応型テストを用いたモバイル・テスト ング・システム

### 2.1 はじめに

知識は状況に埋め込まれて存在し [1], 新しい知識は現実の状況における学習者の経験に融合して獲得されるという知識観に基づき [2], 状況に埋め込まれた学習が近年注目されている.

状況に埋め込まれた学習は, これまで, 数多くの携帯端末を用いたシステムにより支援されてきた [6][3][7][8][9][4][5]. 同様に, 状況に埋め込まれた学習の評価では, 現実の状況における観察や探索を通じた学習そのものを評価することが重要であるため, 学習と同様に現実の状況で行うべきである. 状況に埋め込まれた学習の評価のため, 携帯端末を用いたシステムが効果的であることが報告されている [10, 11]

一方, 信頼性の高い評価には, 十分な項目数を受検者に解答させる必要がある. しかし, 受検者は対象物間を移動する必要があるため, これらのシステムでは多くの項目を解答させることが困難である.

少数項目で信頼性の高い評価を実現する手法として, 項目反応理論を用いた適応型テスト (Computerized adaptive testing:CAT) が知られている [12, 13, 14, 15, 16]. 適応型テストは, 受検者の解答履歴から知識状態を逐次的



に測定し、その知識状態に対して情報量が最大の項目を出題する手法である。

しかし、状況に埋め込まれた学習を効率的に評価するため、適応型テストが用いられた研究は見当たらない。また、先行研究では、受検者が持っている知識のみで解答し、現実の状況で探索や観察を通して解答していないかもしれない。

そこで、本研究では、効率的に知識状態を測定するため、テスト理論に基づく適応型テストを用いたモバイル・テストング・システムを開発する。加えて、眼球運動計測装置を用いて受検者が探索や観察を通して項目に解答しているかを検証する。本システムは、適応型テストを用いたテストング機能と出題された項目に対応する場所へ受検者を誘導するナビゲーション機能を持つ。

本システムにより期待される利点は、以下の通りである。

1. テスティング機能は、項目反応理論に基づき知識状態を逐次的に測定し、その知識状態に対して情報量が最大の項目を出題することにより、受検者の知識状態を効率的に測定できる。
2. ナビゲーション機能は、出題された項目に対応する場所まで受検者を短時間で誘導できる。
3. 受検者が事前に持っている知識のみで項目に解答させるのではなく、現実の状況で観察や探索を通して項目に解答させることができる。

シミュレーション実験および被験者実験により、本システムの有効性を示す。特に、利点3は、受検者が観察や探索しているか眼球運動計測装置を用いて検証する。

## 2.2 先行研究

状況に埋め込まれた学習は、これまで、数多くの携帯端末を用いたシステムにより支援されてきた。Hwang らは、博物館において学習に躓いている生徒を支援するため、QR コードを用いて学習中の展示物を同定し、既に学習し終えた生徒を紹介するシステムを開発している [6]。Chiou らは、博物館での学習において展示物の学習人数が制限されているため、各生徒が展示物を学習する経路を最適化し、展示物まで生徒を誘導するモバイルシステムを開発している [3]。Ogata らは、QR コードを用いて学習プロセスを同定し、学習プロセスに一致するビデオ教材を提示するモバイルシステムを開発している [7]。Hwang らは、RFID(Radio Frequency Identification) を用いて生徒の学習プロセスを同定し、そのプロセスに応じて学習を支援するモバイルシステムを開発している [8]。Chu らは、RFID を用いて生徒が学習している植物を同定し、生徒の質問への反応に応じて適応的に学習を支援するモバイルシステムを開発している [9]。Chu らは、モバイルマインドツールを開発し、屋外での小学校の科学の授業に用いている [4]。Chen らは、バードウォッチで生徒の学習を足場掛けするため、支援のレベルを調整するモバイルシステムを開発している [5]。このように、状況に埋め込まれた学習は、多様なモバイル技術を用いて促進されている。

同様に、状況に埋め込まれた学習の評価では、現実の状況における観察や探索を通じた学習そのものを評価することが重要であるため、学習と同様に現実の状況で行うべきである。

Santos らは、特定の場所において観察や探索を通して受検者に解答させるため、GPS(Global Positioning System) を用いて受検者の現在地を同定し、現在地に対応した項目が出題されるシステムを開発している [10]。また、Huang

らは、屋外において草花の種類に関する知識状態を測定するため、RFIDを用いて草花を同定し、適応的に項目を出題するシステムを開発している [11].

一方、信頼性の高い評価には、十分な項目数を受検者に解答させる必要がある。しかし、受検者が対象物間を移動する必要があるため、これらのシステムでは多くの項目を解答させることが困難である。

## 2.3 適応型テスト

本研究では、効率的に知識状態を測定するために適応型テストを用いる。適応型テストとは、受検者の解答履歴から知識状態を逐次的に測定し、その知識状態に対して最も情報量が高い項目を出題する CBT(Computer Based Testing) である [12, 13, 14, 15, 16]。適応型テストの利点は以下の通りである [17][18].

1. 知識状態の測定に最適な項目を出題することにより測定精度を向上させる。
2. 難しすぎる項目や易しすぎる項目を出題しないため、受検者のフラストレーションやテスト不安を減少させる。
3. 測定精度が向上するため、ペーパーテストと同じ測定精度を保ったまま出題項目数を減少させることができ、受検者の負担を減少させることができる。

適応型テストは、項目反応理論に基づいて受検者の知識状態を測定する。項目反応理論とは、受検者の項目への反応データから、受検者の知識状態と項目の特性を測定するテスト理論である [19, 20, 21]。項目反応理論の利点として、それぞれの受検者が異なる項目で作成されたテストを受けた場合において

も同一の尺度上に配置し、比較できることが挙げられる。また、反応データが欠測値であるときも知識状態を測定できる。

ここでは、項目を  $i = 1, \dots, I$ 、項目の出題順序を  $k = 1, \dots, K$  と表し、 $k$  番目に出題した項目を  $i_k$  とする。  $u_{i_k}$  を、項目  $i_k$  に正答したとき 1、それ以外のとき 0 をとる確率変数とすると、反応データは  $U_i = \{u_{i_1}, \dots, u_{i_K}\}$  と定義できる。

受検者の知識状態は  $\theta \in (-\infty, \infty)$  と表し能力値と呼ばれる。受検者が項目  $i$  に正答する確率には、以下の 3 パラメータロジスティックモデルを仮定する。

$$p(U_i = 1|\theta) = c_i + (1 - c_i) \frac{1}{1 + \exp[-1.7a_i(\theta - b_i)]} \quad (2.1)$$

ここで、  $a_i \in [0, \infty)$  が項目  $i$  の識別力のパラメータ、  $b_i \in (-\infty, \infty)$  が項目  $i$  の難易度のパラメータ、  $c_i \in [0, 1]$  が項目  $i$  の当て推量のパラメータと呼ばれる。式 (2.1) において、  $c_i = 0$  としたものが 2 パラメータロジスティックモデルと呼ばれ、さらに、  $a_i = 1$  としたものが 1 パラメータロジスティックモデルと呼ばれる。

図 (2.1) と図 (2.2)、図 (2.3) は、項目の特性を表す項目反応関数 (item response function : IRF) の例である。横軸は受検者の能力値、縦軸は正答確率である。図 (2.1) に、識別力のパラメータ  $a_i$  が異なる 3 つの項目反応関数を示す。識別力のパラメータ  $a_i$  が低い項目 1 は、傾きが小さく、能力値上で正答確率の変化が少ないため、受検者の能力値を十分に反映しない。例えば、正答か誤答かが運で決まるような項目である。反対に、識別力のパラメータ  $a_i$  が高い項目反応関数は、急勾配であり、受検者の能力値を分けることができる。図 (2.2) には、難易度のパラメータ  $b_i$  が異なる 3 つの項目反応関数が示されて

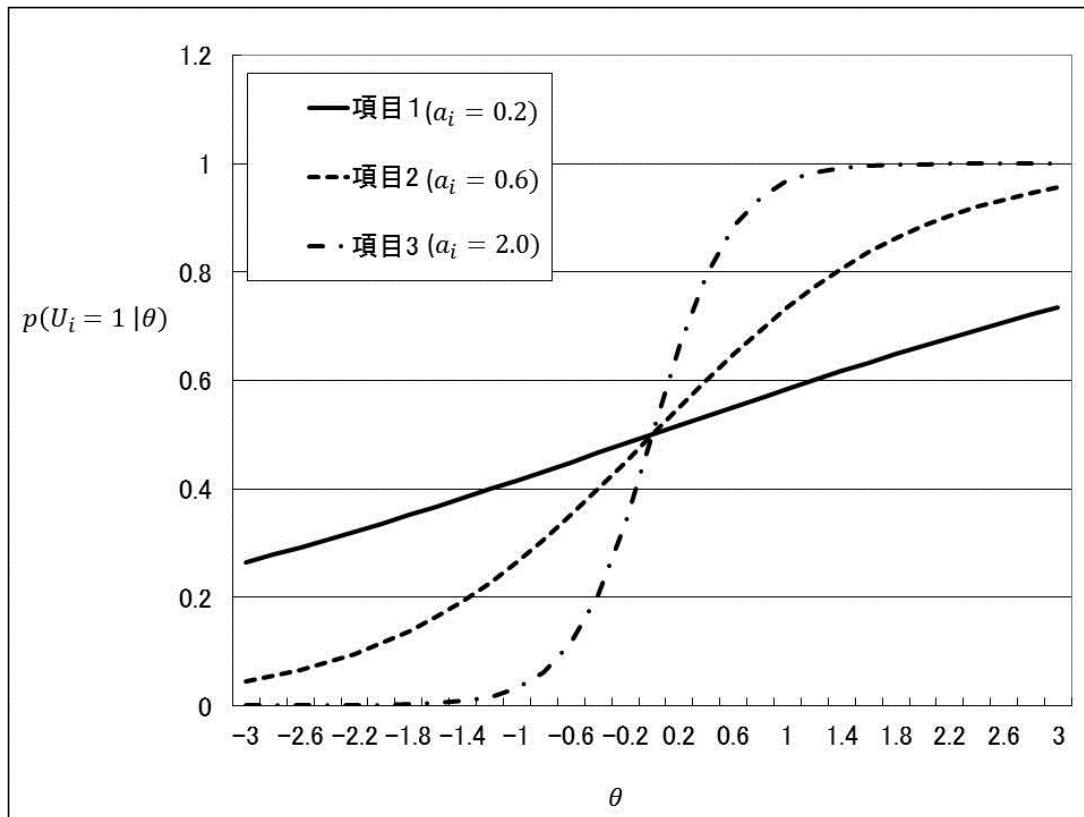


図 2.1 識別力のパラメータ  $a_i$  を変化させた項目反応関数

いる。難易度のパラメータ  $b_i$  が高い項目 3 は、項目 1, 2 より右にシフトし、能力値上において正答確率が低く、難しい項目である。難易度のパラメータ  $b_i$  と能力値が等しい値であるとき、正答確率が 0.5 であり、項目反応関数が最も急勾配になる。図 (2.3) は、当て推量のパラメータ  $c_i$  が異なる 3 つの項目反応関数を示している。当て推量のパラメータ  $c_i$  が高い項目 3 は、能力値が低い受検者の正答確率が高く、偶然に正答する確率が高い。

$k - 1$  番目までの項目に対する反応データ  $u_{i_1}, \dots, u_{i_{k-1}}$  の尤度関数は以下の通りである。

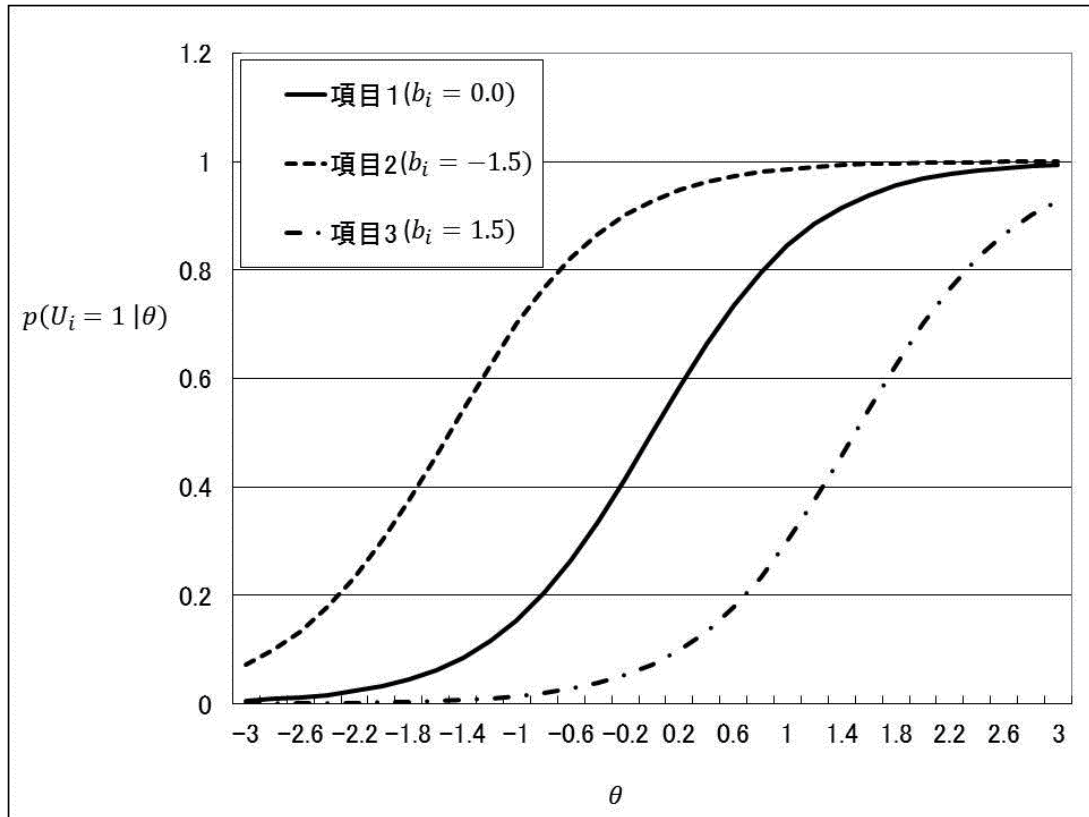


図 2.2 難易度のパラメータ  $b_i$  を変化した項目反応関数

$$L(\theta|u_{i_1} \dots u_{i_{k-1}}) \equiv \prod_{j=1}^{k-1} \frac{\{\exp[a_{i_j}(\theta - b_{i_j})]\}^{u_{i_j}}}{1 + \exp[a_{i_j}(\theta - b_{i_j})]} \quad (2.2)$$

尤度の二次導関数は、尤度関数の曲率を表し、この導関数の負数は、情報量として知られている。

$$J_{u_{i_1} \dots u_{i_{k-1}}}(\theta) \equiv -\frac{\partial}{\partial \theta} L(\theta|u_{i_1} \dots u_{i_{k-1}}) \quad (2.3)$$

$k-1$  番目の項目のフィッシャー情報量は、以下の通りである。

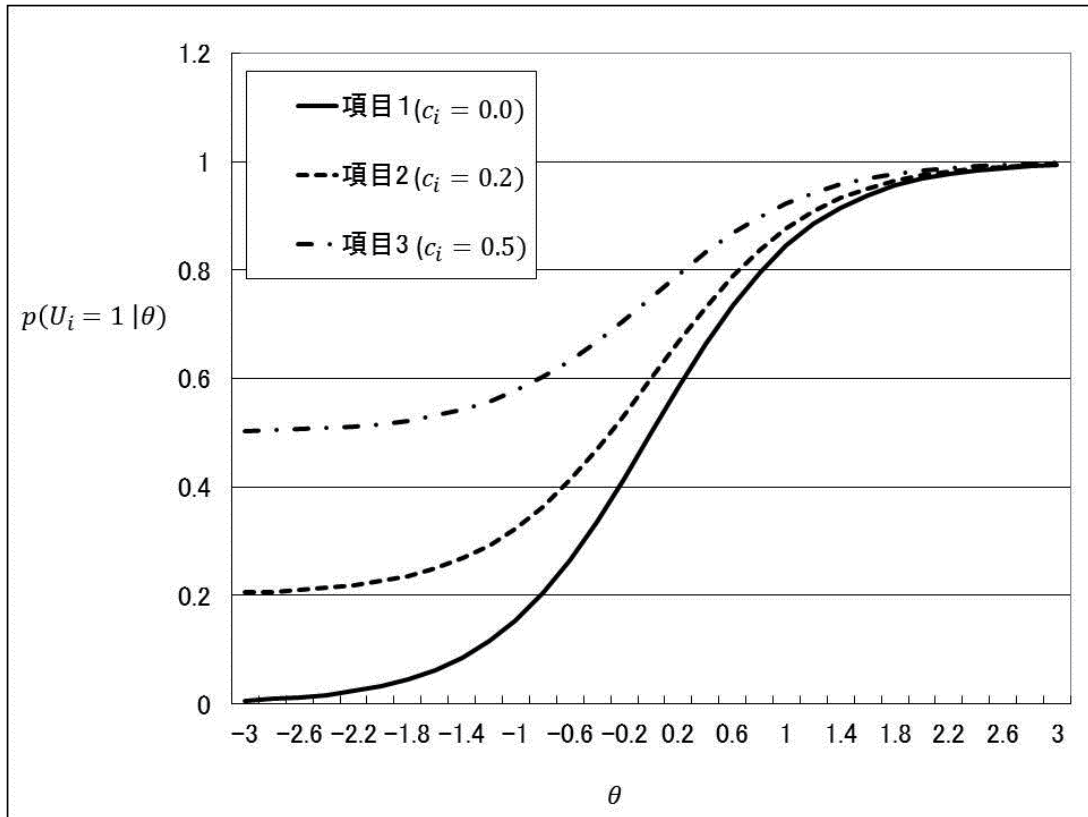


図 2.3 当て推量のパラメータ  $c_i$  を変化させた項目反応関数

$$\begin{aligned}
 I_{U_{i_1} \dots U_{i_{k-1}}}(\theta) &\equiv E[J_{u_{i_1} \dots u_{i_{k-1}}}(\theta)] & (2.4) \\
 &= \sum_{j=1}^{k-1} \frac{[p'(U_{i_j} = 1|\theta)]^2}{p(U_{i_j}|\theta)[1 - p(U_{i_j} = 1|\theta)]}
 \end{aligned}$$

ただし

$$p'(U_{i_j} = 1|\theta) \equiv \frac{\partial}{\partial \theta} p(U_{i_j} = 1|\theta) \quad (2.5)$$

フィッシャー情報量の逆数は、能力推定値の漸近的な標準誤差に一致する。図(2.4)に、2つの項目のフィッシャー情報量を示す。横軸は受検者の能力値、

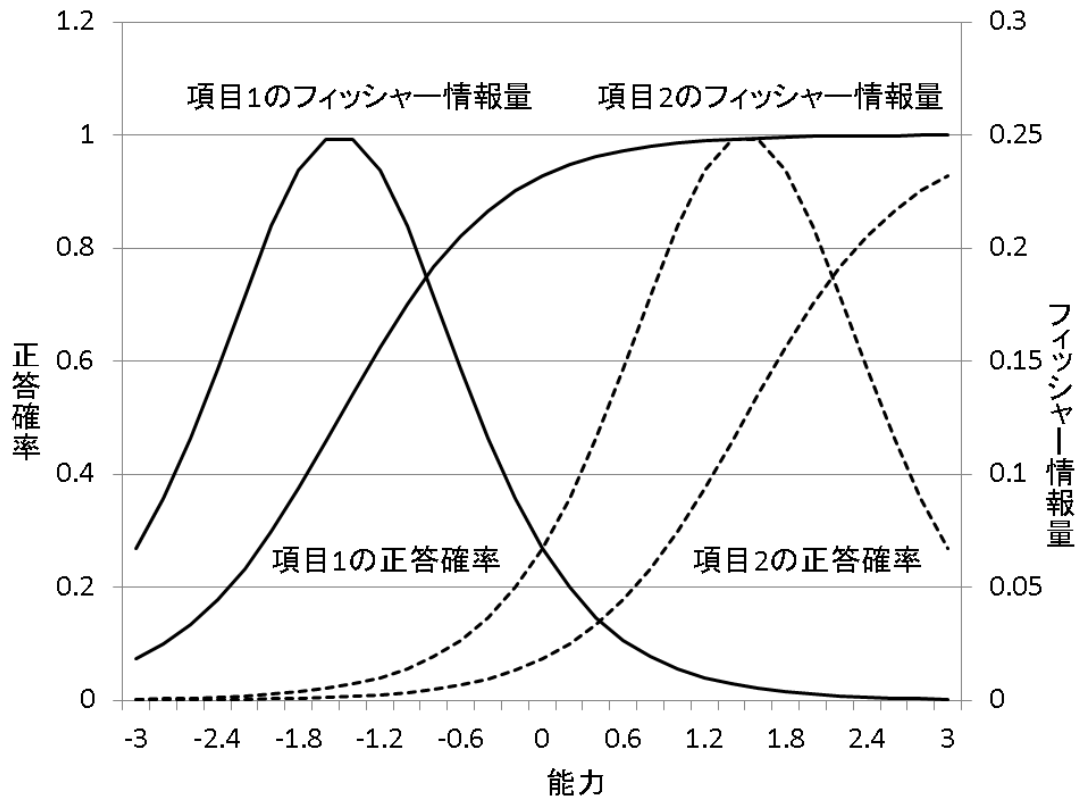


図 2.4 フィッシャー情報量

左の縦軸は正答確率，右の縦軸はフィッシャー情報量である．図 (2.4) から，項目反応関数の傾きが大きく，正答確率が 0.5 になる能力値で情報量が多いことがわかる．

ここでは，一つの項目に対するフィッシャー情報量を項目情報量，テストの全項目に対するフィッシャー情報量をテスト情報量と呼ぶ [22]．

情報量最大化原理に基づく項目選択は， $u_{i_1}, \dots, u_{i_{k-1}}$  までの反応データを用いて推定された能力値のパラメータ  $\theta = \hat{\theta}_{u_{i_1}, \dots, u_{i_{k-1}}}$  においてテスト情報量関数が最大になる  $k$  番目の項目を選択する．



$$i_k = \arg \max_j \{I_{U_1, \dots, U_{k-1}, U_j}(\hat{\theta}_{u_{i_1}, \dots, u_{i_{k-1}}}) : j \in R_k\} \quad (2.6)$$

ここで、 $R_k = \{1, \dots, I\} \setminus \{i_1, \dots, i_{k-1}\}$  を表す。

情報量最大化原理に基づく適応型テストの出題アルゴリズムを図 (2.5) に示す。まず、能力値の初期値を 0 とする。次に、テスト情報量関数が最大になる  $k$  番目の項目をアイテムバンクから選択する。受検者は選択された項目を解答する。解答された項目は、システムにより自動的に正誤判定される。システムは解答履歴から受検者の能力値を推定する。推定された能力値と前に推定された能力値の差が少なければ終了し、そうでなければ、再度受検者に項目を出題する。

能力値の推定には、ベイズ推定法を用いる [19]。ベイズ推定法は、最尤推定法とは異なり、全問正答、または全問誤答のときでも能力値を推定できることが知られている [20]。また、最尤推定法は、一致性、および漸近有効性を持つことが知られており大きなデータには有効であるが、少数データからパラメータを推定することにはあまり適していない。一方、ベイズ推定は、一致性および漸近有効性を持つと同時に少数データからの推定にも適していることが知られている [23]。

能力値のベイズ推定には、 $k$  項目までの反応データを用いて、能力値の事前分布  $g(\theta)$  として標準正規分布が仮定され、以下の事後分布を用いる。

$$g(\theta|u_{1j}, \dots, u_{kj}) = \frac{L(\theta|u_{1j}, \dots, u_{kj})g(\theta)}{\int (L(\theta|u_{1j}, \dots, u_{kj})g(\theta)) d\theta} \quad (2.7)$$

$L(\theta|u_{1j}, \dots, u_{kj})$  は、 $k$  項目の反応データを用いた能力値の尤度である。能力値を推定するため、推定値を  $\hat{\theta}$  とし、EAP(expected a posteriori) 推定を用いた [21]。

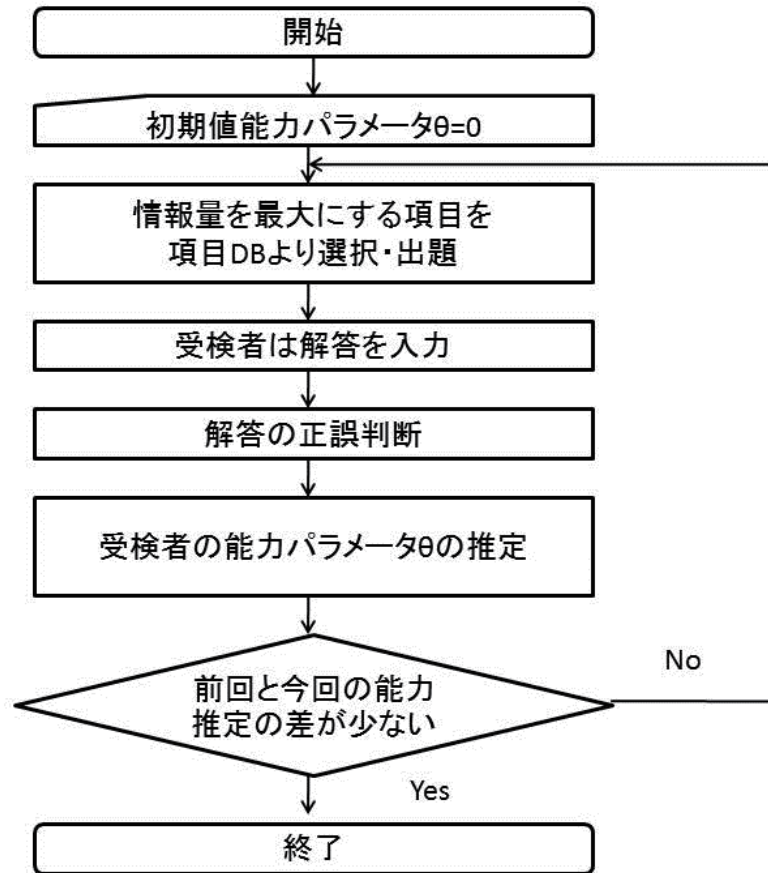


図 2.5 適応型テストのアルゴリズム

$$\hat{\theta} = \int \theta \cdot g(\theta | u_{1j}, \dots, u_{kj}) d\theta : \theta \in (-\infty, \infty) \quad (2.8)$$

本研究では、モバイルテストに適応型テストを適用し、情報量が高い項目を出題することで、高精度な能力推定を目指す。

## 2.4 適応型テストを用いたモバイル・テストング・システム

本節では、モバイル・テストング・システムについて詳述する。

まず、モバイル・テストングのアルゴリズムを図 2.6 に示す。アルゴリズムには、従来の適応型テストのアルゴリズムに加えて、項目が出題される場所まで受検者を誘導することが加えられている。

まず、能力値の初期値を 0 とする。次に、システムはテスト情報量関数が最大になる  $k$  番目の項目をアイテムバンクから選択し、項目が出題される場所まで受検者を誘導する。項目が出題される場所に到着したあと、受検者は選択された項目を解答する。システムは受検者が解答した項目を自動的に正誤判定し、解答結果と解答履歴から能力値を推定する。推定された能力値と前回において推定された能力値の差が少なければ終了し、そうでなければ、再度受検者に項目を出題する。

次に、システム構成を図 2.7 に示す。本システムは、受検者の解答履歴を管理するデータベース、アイテムバンク、ナビゲーション機能、項目選択機能、項目表示機能から構成されている。

ナビゲーション機能は、受検者の現在地と項目が出題される場所を Android Maps API の Google マップを用いて表示する。本機能は、GPS 機能を用いて受検者の現在地を取得し、地磁気センサーにより受検者が向いている方向を同定する。画面例を図 2.8 に示す。図の青丸は受検者の現在地を表し、紫のマーカーは項目が出題される場所を指している。受検者は項目が出題される場所に到着したあと、上部の「TEST」ボタンをタップし、項目表示画面へ遷移する。

項目選択機能では、適応型テストに基づいてアイテムバンクから項目を選択する。項目選択の計算が携帯端末では困難であったため、項目選択機能は Web サーバに実装している。

項目表示機能では、項目選択機能が選択した項目を受検者が持つ Android 携帯電話に表示する。この機能の画面例を図 2.9 に示す。

## 2.5 アイテムバンク

### 2.5.1 アイテムバンクのスキーマ

本システムは、アイテムバンクと呼ばれるデータベースを用いて項目を管理している。このように、アイテムバンクに数多くの項目を事前に準備し、テストの目的や受検者に応じてテストを構成する手法をアイテムバンク方式と呼び、近年のテスト手法として注目されている [18]。

一般に、アイテムバンクでは、項目の設問文や選択肢が管理されている。しかし、状況に埋め込まれた学習の評価では、項目が出題される場所や項目に関連する展示物、出題が可能な時期といった多様な情報も管理する必要がある。そこで、本研究では、このような多様な情報を管理するアイテムバンクを構築するため、状況に埋め込まれた学習 [6][3][7][8][9][4][5] やその学習の評価 [10][11] を支援しているシステムで用いられているデータベースの属性を全て抽出・整理し、アイテムバンクのスキーマを構築した。図 2.10 に、先行研究のデータベースから抽出・整理した 6 つの属性を示す。1 つ目は、項目が出題される場所の情報を管理するため、緯度・経度、RFID や QR コードの ID の情報を格納する出題位置の属性である。本研究では、事前に機器を設置する準備が必要ない GPS を用いて、項目が出題される場所を管理する。一方、博物館

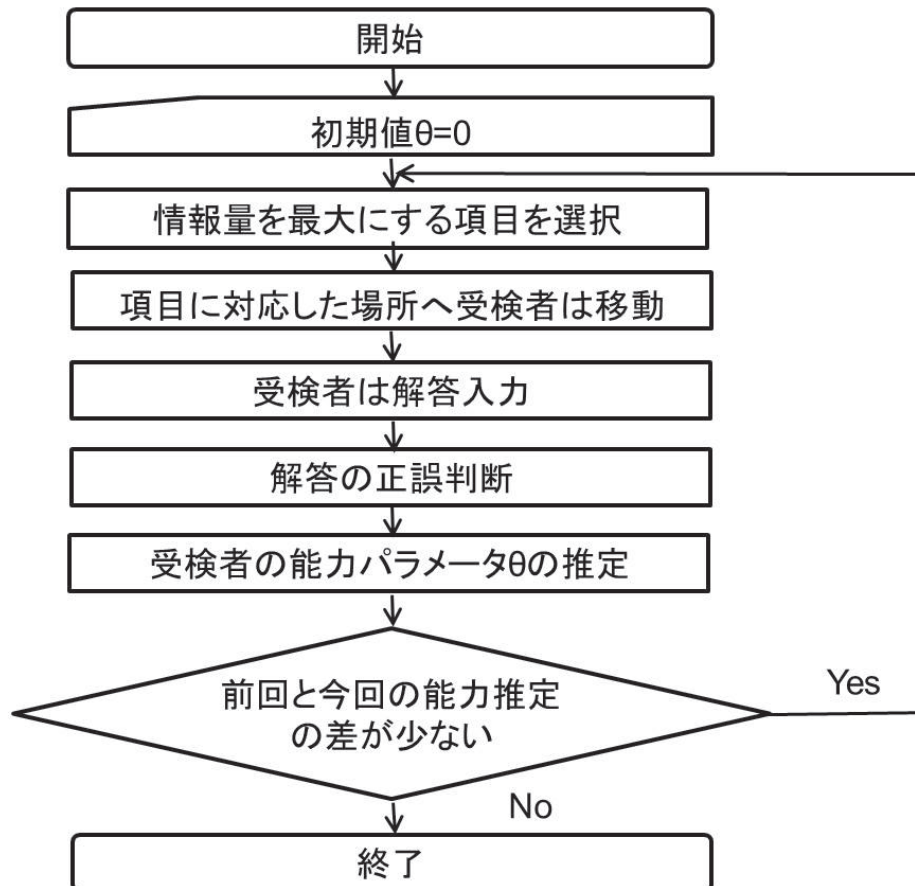


図 2.6 モバイル・テストのアルゴリズム

などの屋内では GPS が使えないため、RFID や QR コードの ID の情報を用いて項目が出題される場所を管理するシステムが開発されている [11]。2つ目、3つ目、4つ目は、項目を出題することができる期間、時間帯、天候の情報を管理する属性である。状況に埋め込まれた学習の評価は、現実の状況や実際の

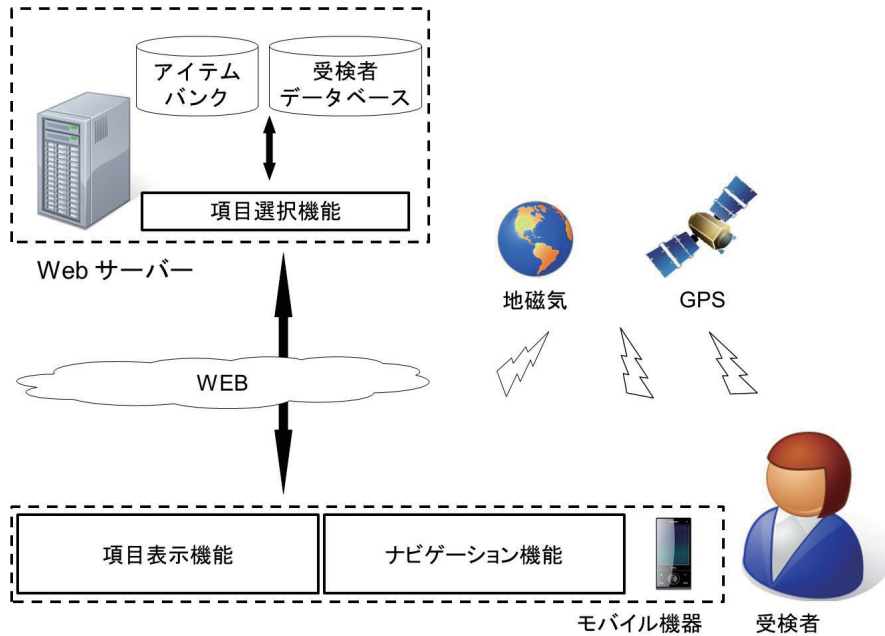


図 2.7 システム構成図

展示物を対象にするため、項目の出題に制約を加えることが必要になる。例えば、星座を対象にしたとき、期間や時間により見えない星座が存在する。出題する時間や時期を管理することにより、現実の状況における特有の制約を項目選択に組み込むことができる。5つ目は、項目に対応している展示物についての情報を管理する属性である。具体的には、展示物の名称や、その位置、項目が出題される場所から見た方向といった情報が管理される。この属性が管理している情報により、テスト出題者が任意に選択した展示物に関する項目を出題するといった制約を加えることができる。

図 2.10 に示した属性を管理するため、各属性を具体的な情報として整理した。整理した属性を図 2.11 に示す。図 2.11 から各属性が構造化されていることがわかる。例えば、項目の出題位置は、緯度・経度、または RFID といっ



図 2.8 ナビゲーション機能の画面例

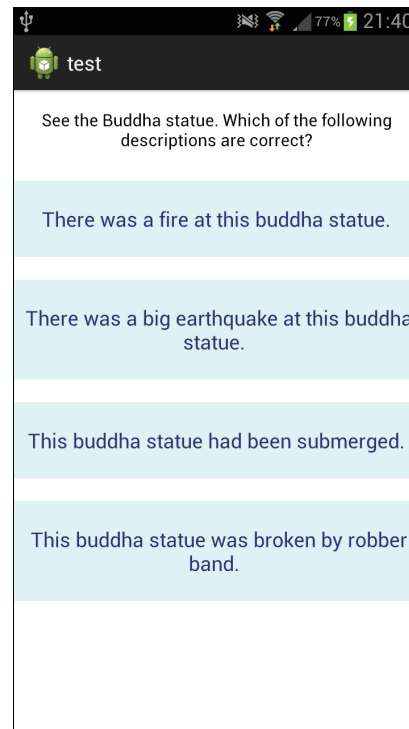


図 2.9 項目表示機能の画面例

た ID 情報で管理されている。そこで、本アイテムバンクは、XML(Extensible Markup Language) を用いて階層的に情報を管理する。

ここでは以下の 2 つの活用例を示す。1 つ目の活用例は、歴史的建造物が設置されている公園において、その展示物に関する知識を評価するテストである。具体的に、この活用例では、ご当地検定のような、その地域に設置されている施設や歴史に関する知識を評価するテストが挙げられる。この活用例では、屋内・屋外施設に仏像といった歴史的建造物が点在し、その展示物が設置されている場所まで受検者が移動し、観察や探索を通して知識を評価することを想定している。この活用例で用いられるアイテムバンクの記述例を図 2.12

要素	内容	例	活用例
出題位置	緯度・経度, RFIDやQRコードのIDによって管理	35.656163,139.544402(緯度, 経度)	公園で受検者の現在位置に応じたテスト項目を出題
期間	開始の年月日と終了の年月日によって管理	6月1日～8月1日	星座のように季節によって変化するものを時期に応じて出題
時間帯	開始の時間と終了の時間によって管理	13時00分～15時00分	アサガオのように時間によって姿が変わる草花を時間に応じて出題
天候	晴れ, 曇り, 雨, 雪の中から選ぶ	晴れ	星座のように曇りでは出題不可能なものを制御
対象物	対象物に関する情報(名称や位置, 方向)を管理	北極星, 北	誤答時に北極星といった星座の方向を示す事ができる。

図 2.10 アイテムバンクの属性

に示す。ここでは、項目の出題位置として緯度・経度の情報が管理されている。また、対象物として展示物の情報が管理されている。

2つ目の活用例は、屋内での特定の状況における知識を評価するテストである。具体的には、災害時に特定の状況において受検者の適切な判断で避難場所まで移動することができるかを評価するテストを想定している。この活用例で使われるアイテムバンクの記述例を図 2.13 に示す。ここでは、項目の出題位置として RFID の ID 情報、また、対象物として評価を行う施設に関する情報を管理する。



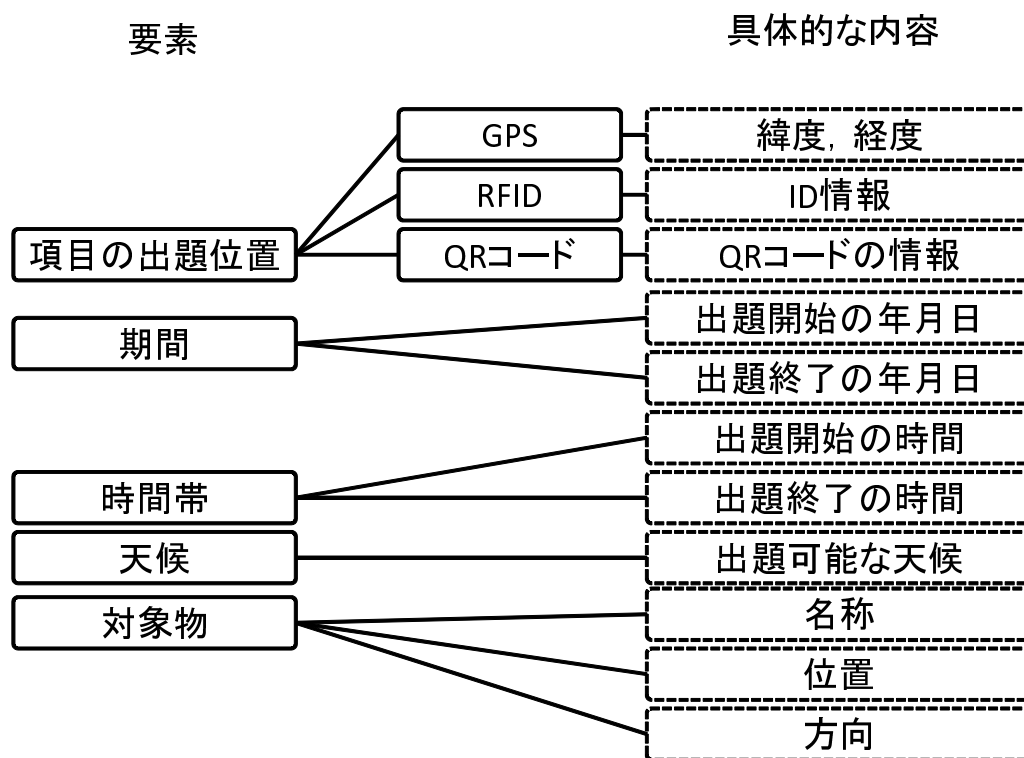


図 2.11 アイテムバンクのスキーマ

## 2.5.2 対象の寺院

本研究では、歴史的建造物が数多く設置されている東京都調布市の深大寺を実験の対象とした。深大寺は、東京で台東区の浅草寺につぐ第二の古刹であり、重要文化財である金銅造釈迦如来倚像が安置されている。深大寺に設置されている歴史的建造物の写真を図 2.15 に示す。深大寺には歴史的建造物が数多くあるため、多くの項目を作成することができる。

```
<?xml version="1.0" encoding="SHIFT-JIS" ?>
<itemBank>
  <item>
    <question>
      この仏像が遭った災害は何か？
      仏像に痕跡があります
    </question>
    (一部, 省略)
    <location>
      <gps>
        <latitude>35.655287</latitude>
        <longitude>139.544381</longitude>
      </gps>
    </location>
    <object>
      <name>〇〇仏</name>
      <location>
        <gps>
          <latitude>35.656096</latitude>
          <longitude>139.54407</longitude>
        </gps>
      </location>
      <angle>315</angle>
    </object>
    <area> 〇〇寺</area>
  </item>
</itemBank>
```

図 2.12 展示物の知識を評価するアイテム・バンクの記述例

### 2.5.3 アイテムバンクの開発

アイテムバンクには、深大寺に関する知識を評価する項目が蓄積されている。項目数は85項目である。また、状況に埋め込まれた学習を評価するため、

```
<?xml version="1.0" encoding="SHIFT-JIS" ?>
<itemBank>
  <item>
    <question>
      この標識の意味は？
    </question>
    (一部, 省略)
    <location>
      <RFID>
        <ID>2002</ID>
      </RFID>
    </location>
    <area> ○○大学</area>
  </item>
</itemBank>
```

図 2.13 特定の状況における知識を評価するアイテム・バンクの記述例

全ての項目は展示物への探索や観察が必要になるように作成されている。

項目の例は以下の通りである。

1. 白鳳仏の頭の盛り上がりは何か。
2. 大師堂の上部にある丸い造形物は、本来、何に使われたか。
3. 大師堂の中にある「植」の文字をかたどった造形物は何を置く物であるか。

受検者は、探索や観察を通して項目に解答する。これにより、受検者が事

前に持っている知識のみでなく、展示物への探索や観察を通じた学習そのものを評価することができる。

適応型テストは、能力値と項目情報量を計算するため、項目の難易度のパラメータ  $b_j$  を事前に推定する必要がある。そこで、本研究では、事前に項目の反応データを収集し、難易度のパラメータ  $b_j$  を推定する。しかし、深大寺で 100 人以上の受検者に 85 項目を解答させることは困難であった。そこで、項目反応理論の普遍尺度の性質を用いて、以下の反応データから難易度のパラメータ  $b_j$  を推定する。推定に用いた反応データは、深大寺と深大寺を再現した仮想環境の 2 か所で解答させた 9 名、仮想環境でのみ解答させた 103 名から取得した。実施期間は、仮想環境のみの反応データを取得するため、2009 年 7 月から 8 月に 1 人 2 時間で 22 日間、深大寺と仮想環境の 2 ヶ所の反応データを取得するため、2010 年 10 月に 1 人 3 時間で 5 日間であった。取得した反応データは、深大寺での反応データと仮想環境での反応データを異なる項目の反応データとみなし、横軸に項目、縦軸に受検者とした (図 2.14)。すなわち、ここでは 103 人分の深大寺での反応データを欠測値とみなして難易度のパラメータ  $b_j$  を推定する。このように欠測値が含まれたデータであってもパラメータの推定ができることが項目反応理論の利点でもある。

難易度のパラメータ  $b_i$  の推定は、能力値のパラメータ  $\theta$  の推定と同様、ベイズ推定法を用いた [23]。難易度のパラメータ  $b_i$  の行列は  $\mathbf{b}$ 、能力値の行列は  $\theta$ 、パラメータのハイパーパラメータは  $\tau_\theta, \tau_b$  とする。また、 $g(\theta|\tau_\theta)$  は能力値のパラメータ  $\theta$  の事前分布であり、 $g(\tau_\theta)$  はハイパーパラメータの分布である。同様に、 $g(\mathbf{b}|\tau_b)$  は、難易度のパラメータ  $\mathbf{b}$  の事前分布であり、 $g(\tau_b)$  はハイパーパラメータの分布である。未知のパラメータの事前同時分布  $g(\theta, \mathbf{b}, \tau_b)$  が仮定され、反応データの行列  $\mathbf{U}$  が得られたとき、事後分布は、以下の通りで

ある.

$$\begin{aligned} g(\theta, \mathbf{b}, \tau_b | \mathbf{U}) &\propto L(\mathbf{U} | \theta, \mathbf{b}) g(\theta, \mathbf{b}, \tau_b) \\ &\propto L(\mathbf{U} | \theta, \mathbf{b}) g(\theta | \tau_\theta) g(\tau_\theta) g(\mathbf{b} | \tau_b) g(\tau_b) \end{aligned} \quad (2.9)$$

ただし, 反応データは欠測値を持つため, 尤度関数は以下の通りである.

$$L(\mathbf{U} | \theta, \mathbf{b}) = \prod_{i=1}^I \prod_{j=1}^J [P_{ij}^{u_{ij}} Q_{ij}^{(1-u_{ij})}]^{D_{ij}} \quad (2.10)$$

$P_{ij}$  は受検者  $j$  が項目  $i$  に正答する確率 (式 (2.1)),  $Q_{ij}$  は誤答確率を示す.  $D_{ij}$  は反応データの有無であり, 1 ならば有り, 0 ならば無しを表す. 難易度のパラメータ  $b_i$  の推定は, 能力値のパラメータ  $\theta$  が積分消去された以下の式を用いる [23].

$$\begin{aligned} g(\tau_\theta, \mathbf{b}, \tau_b | \mathbf{U}) &\propto \int L(\mathbf{U} | \theta, \mathbf{b}) g(\theta, \mathbf{b}, \tau_b) d\theta \\ &\propto L(\mathbf{U} | \mathbf{b}, \tau_\theta) g(\tau_\theta) g(\mathbf{b} | \tau_b) g(\tau_b) \end{aligned} \quad (2.11)$$

上記の式を EM アルゴリズムを用いて計算する [23]. 推定した項目のパラメータを付録 1 に添付する.

## 2.6 評価実験

本節では, 本システムの利点を評価する. 利点は以下の通りである.

1. テスティング機能は, 受検者の能力値を効率的に測定することができる.

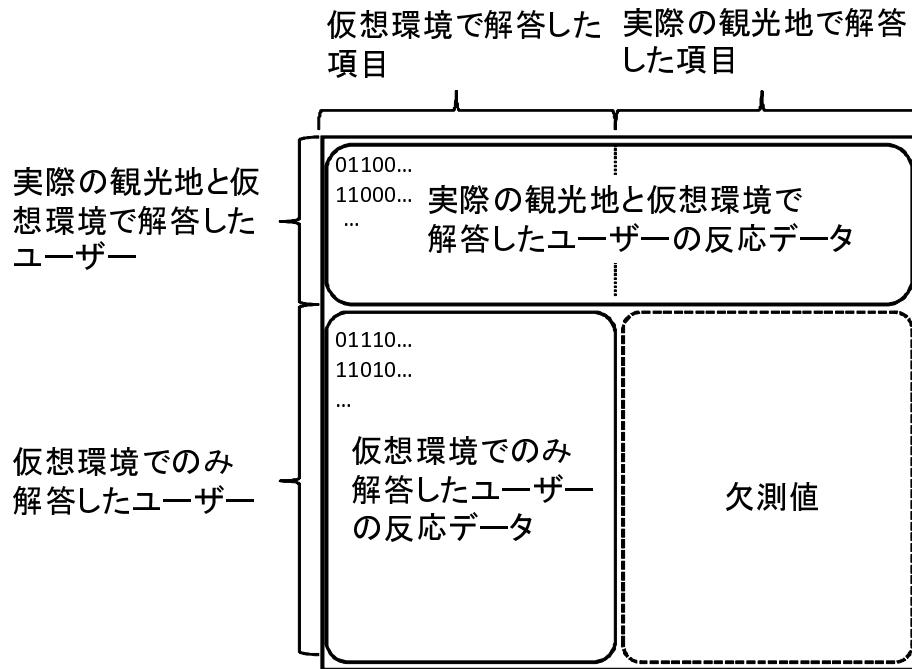


図 2.14 難易度のパラメータ  $b_i$  の推定に用いた反応データ

2. ナビゲーション機能は、受検者を展示物へ短時間で誘導することができる。
3. 状況に埋め込まれた学習そのものを評価することができる。具体的には、受検者が現実の状況において観察や探索を通してテストをすることができる。



図 2.15 深大寺の写真

### 2.6.1 テスティング機能の有効性評価

本節では、テスト機能を用いた能力推定を評価するため、シミュレーション実験を行う。

シミュレーション実験では、真の能力値を 1.0, 0.6, 0.2, -0.2, -0.6, -1.0 とし、それぞれ 200 人分の 30 項目の反応データを式 (2.1) から生成した。ここでは、真の能力値ごとに能力推定値と累積項目情報量の平均値を求めた。能力推定値と累積項目情報量の遷移を図 2.16 図 2.17 図 2.18 図 2.19 図 2.20 図 2.21 図 2.22 図 2.23 図 2.24 図 2.25 に示す。各図では、横軸は受検者が解答した項目数、左の縦軸は能力値、右の縦軸は累積項目情報量とした。各図において、能力値の推定値は、初期では真の能力値と乖離している。しかし、

受検者が解答した項目数が増えるにしたがい、真の能力の値に収束している。また、累積項目情報量が増加していることから、推定値の誤差が減少していることがわかる。

### 2.6.2 ナビゲーション機能の有効性評価

本節では、ナビゲーション機能により受検者を展示物へ短時間で誘導することができたかを評価する。具体的には、本システム、または紙の地図を被験者に利用させ、それぞれの被験者の移動時間を比較する。紙の地図として、深大寺において配布されている地図を用いた。図 2.26 に深大寺の地図の概略を示す。本実験での出発地が点線の丸であり、到着地が実線の丸である。被験者は、深大寺に訪れたことがない 12 名の大学生とした。内訳は、本システムを利用した被験者が 6 名、紙の地図を利用した被験者が 6 名である。

実験の結果、本システムを利用した被験者は移動時間が平均 85.1 秒、標準偏差 38.1 秒、紙の地図を利用した被験者は移動時間が平均 186 秒、標準偏差 85.9 秒であった。ウェルチの検定の結果、有意水準 5% で有意差があった。このため、本システムを用いることによって短時間で被験者を展示物へ誘導できることが確認できた。

### 2.6.3 観察・探索活動の分析

本節では、状況に埋め込まれた学習そのものを評価することができているかを検証する。具体的には、受検者が知識のみで項目に解答しているのではなく、現実の状況で観察や探索を通して項目に解答できているかを評価する。ここでは、眼球運動測定装置 (NacEMR-8) を用いて被験者が観察や探索を通し



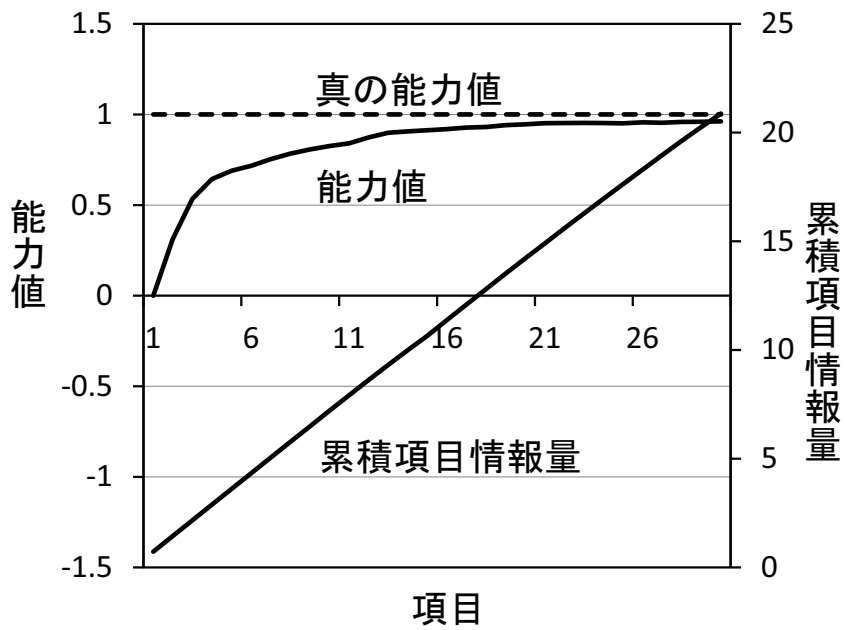


図 2.16 真の能力値が 1.0 のときの推定の過程

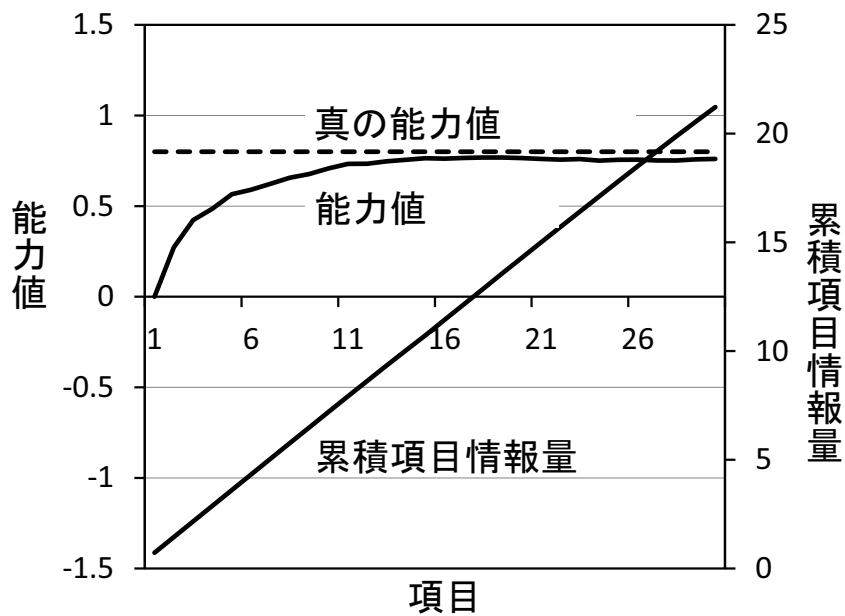


図 2.17 真の能力値が 0.8 のときの推定の過程

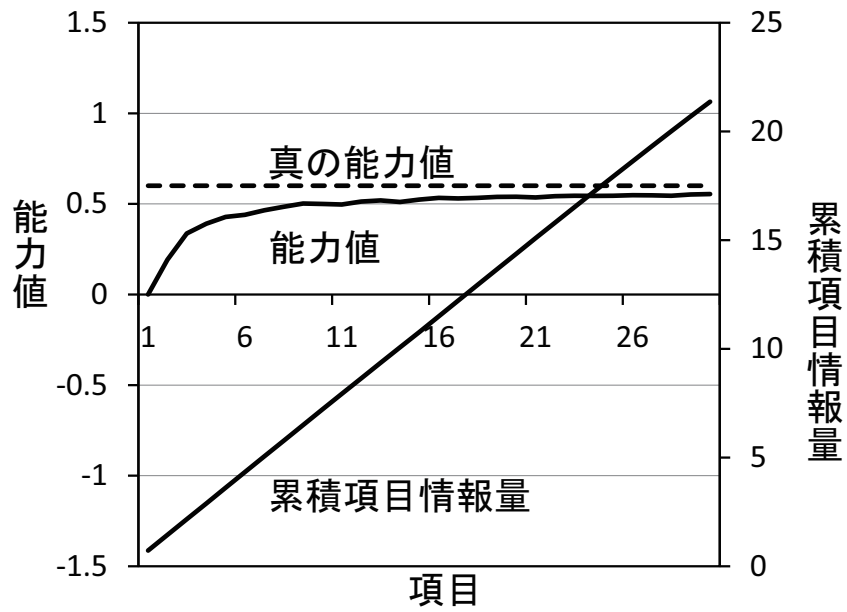


図 2.18 真の能力値が0.6のときの推定の過程

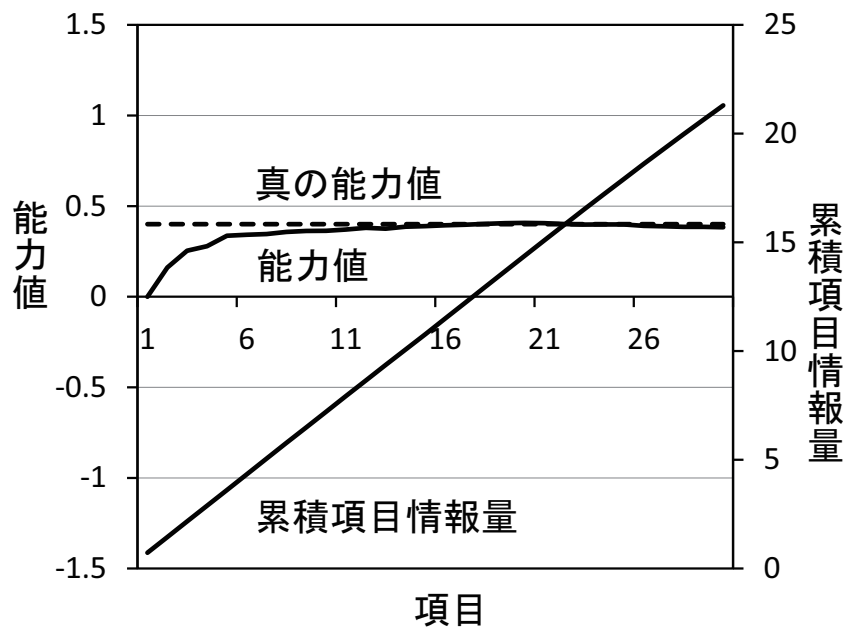


図 2.19 真の能力値が0.4のときの推定の過程

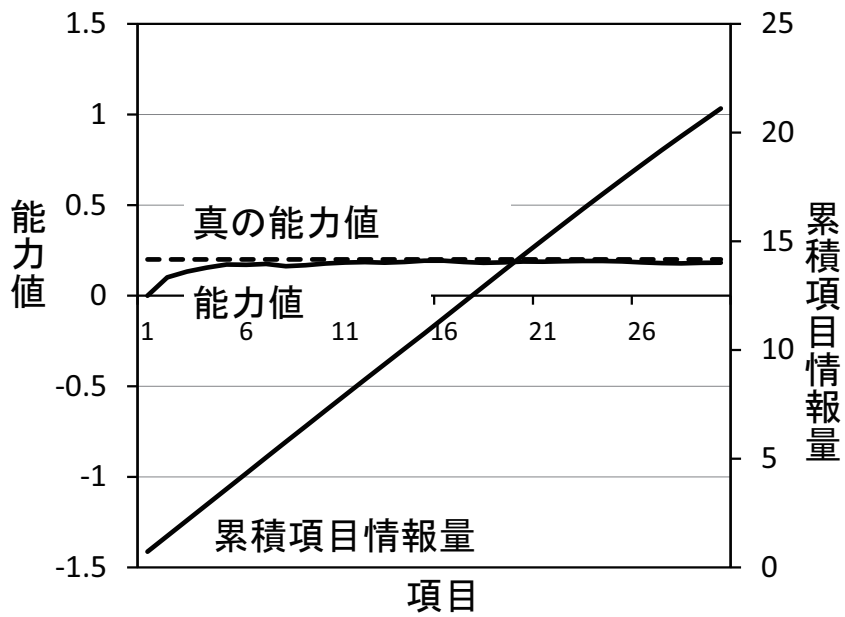


図 2.20 真の能力値が0.2のときの推定の過程

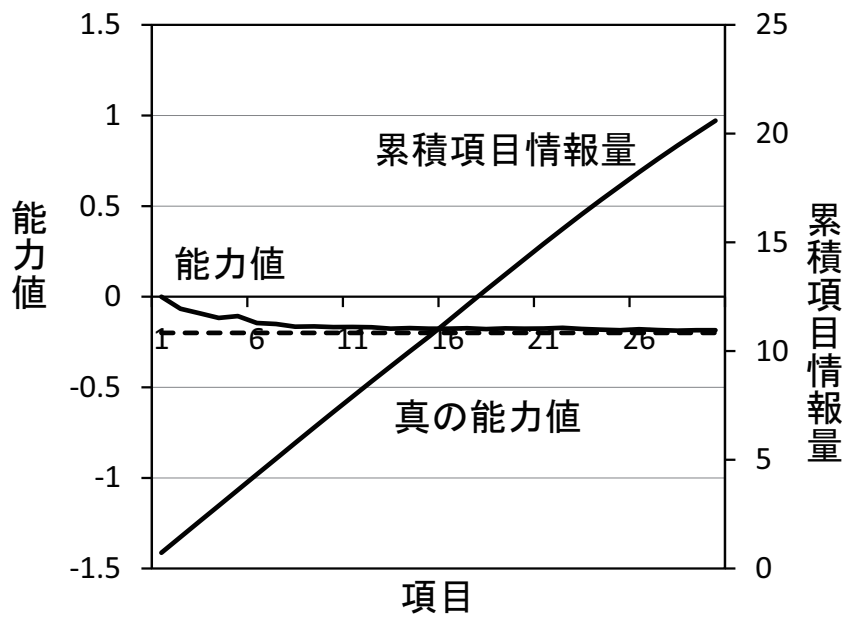


図 2.21 真の能力値が-0.2のときの推定の過程

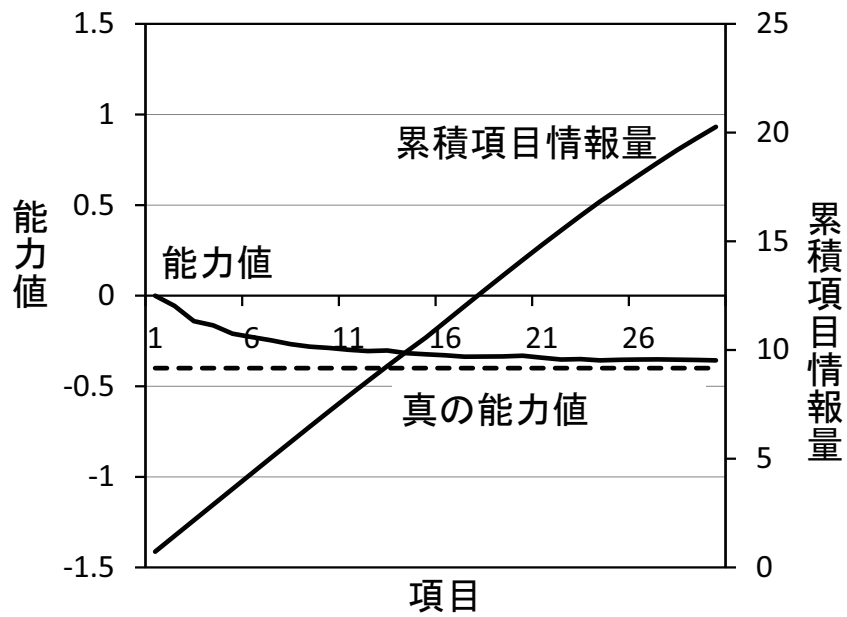


図 2.22 真の能力値が-0.4 のときの推定の過程

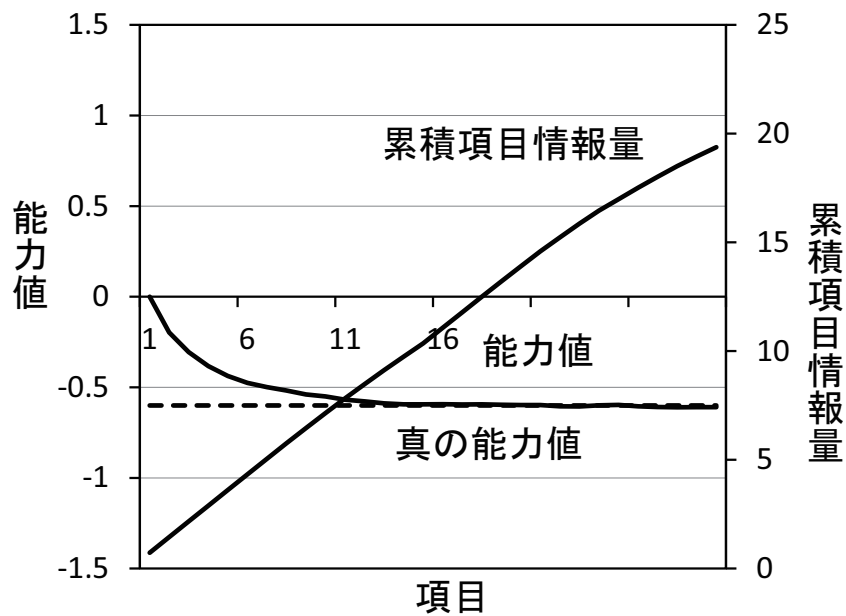


図 2.23 真の能力値が-0.6 のときの推定の過程

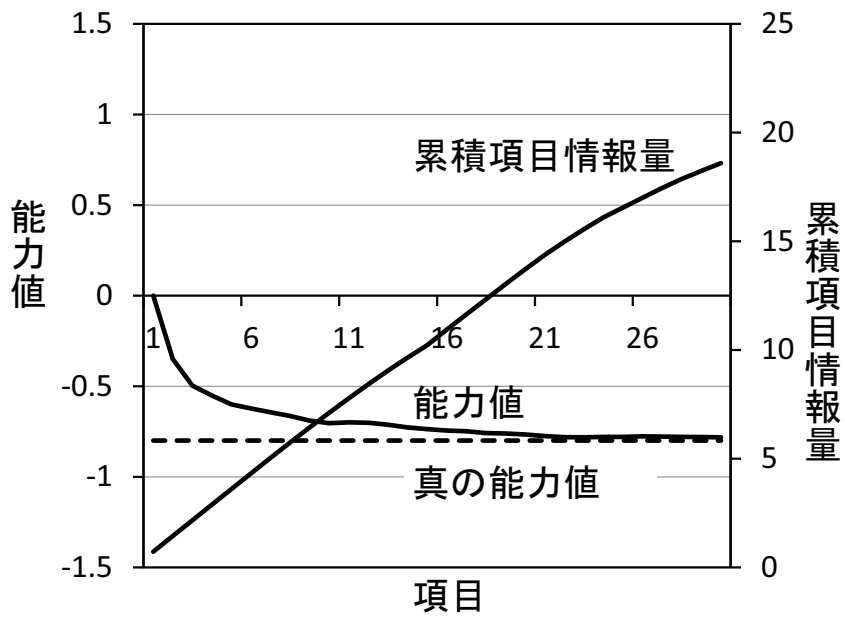


図 2.24 真の能力値が-0.8 のときの推定の過程

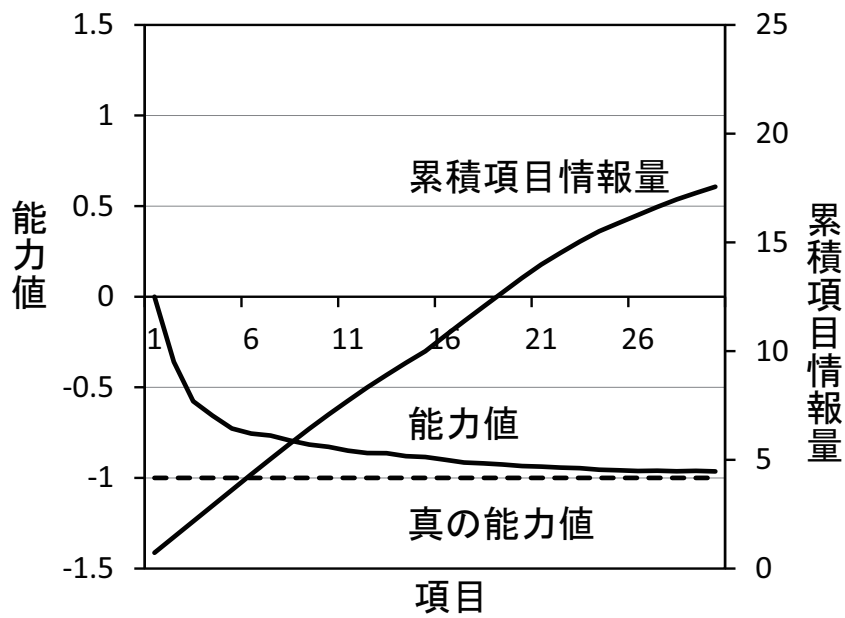


図 2.25 真の能力値が-1.0 のときの推定の過程

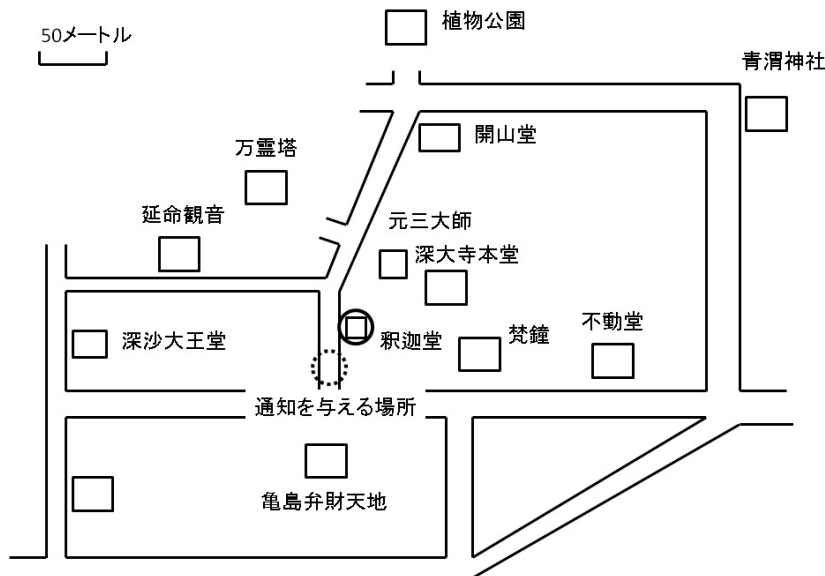


図 2.26 深大寺の地図

て項目に解答しているかを検証する。この実験は、深大寺において機器を用いた計測が困難であったため、深大寺を再現した仮想環境で実施した。仮想環境の配置図を図 2.27 に示す。仮想環境には、被験者の左側に大きな仏像の写真、右側に小さな仏像、被験者の右手前に大きな仏像の解説、小さな仏像の左側にその仏像の解説を設置している。写真の大きさと写真間の距離は、実際と同じである。被験者は、深大寺に訪れたことがない大学生 31 名とした。

実験結果を表 2.1 に示す。被験者が観察や探索をした平均時間は、56.5 秒であり、その平均回数は 19.1 回であった。また、1 項目当たりの時間は、3.77 秒であり、回数は 1.27 回であった。実験の結果から、被験者が観察や探索を通して項目に解答していることが確認できた。

表 2.1 眼球運動計測装置を用いた注視の計測結果

	システム
展示物を注視した時間 (秒)	56.5
展示物を注視した回数	19.1
テストに費やした時間 (秒)	408
一項目あたりの注視時間 (秒)	3.77
一項目あたりの注視回数	1.27

加えて、被験者の行動を分析するため、深大寺において被験者をビデオカメラで撮影し、被験者の行動を分析する。被験者数は 26 名である。内訳は、本システムを利用した被験者が 16 名（深大寺に訪れたことがない被験者が 11 名、深大寺に訪れたことがある被験者が 5 名）紙媒体のガイドブックを用いた被験者が 10 名（深大寺に訪れたことがある被験者が 7 名、深大寺に訪れたことがない被験者が 3 名）である。

ここでは、被験者の顔と上半身を斜め前から撮影した後、被験者の状態を、A 観察や探索をしている状態、B テスト中の移動、C 手元のシステムを注視している状態、D その他の状態、この 4 つの状態に分類した。表 2.2 に集計した平均値、括弧内に標準偏差を示す。計測の結果、被験者が観察や探索をする平均時間は 286 秒であり、平均回数は 55 回であった。検定の結果、平均時間と平均回数に有意水準 1% の有意差があった。実験の結果、被験者は、現実の状況において観察や探索しながらテストを受けていることが確認できた。

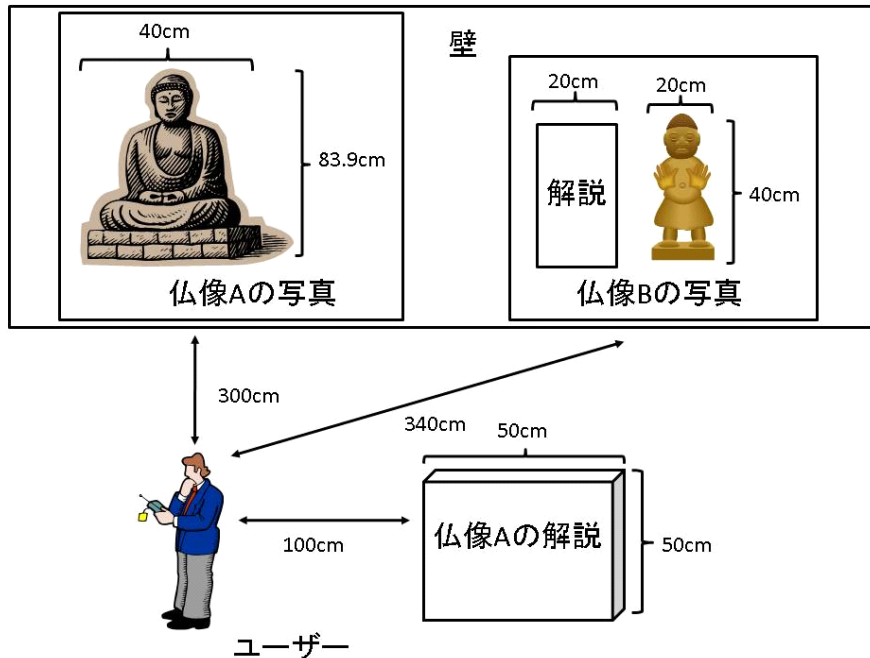


図 2.27 仮想環境の配置図

## 2.7 おわりに

本章では、状況に埋め込まれた学習を効率的に測定することを目的として、項目反応理論に基づく適応型テストを用いたモバイル・テスト・システムを開発した。適応型テストとは、受検者の解答履歴から逐次的に能力値を推定し、その能力値に対して項目情報量が最大になる項目をアイテムバンクから選択する Computer Based Testing (CBT) である。本システムは、適応型テストを用いたテスト機能と展示物がある場所まで受検者を誘導するナビゲーション機能を持つ。本システムにより期待される利点は、(1) 受検者の知



表 2.2 ビデオデータからの計測結果 (\*\*有意水準 1% で有意差あり)

	システム	紙媒体のガイドブック
展示物を注視した時間 (秒) **	286(1810)	96.3(30.0)
展示物を注視した回数**	55(17.8)	36.4(14.2)
テストに費やした時間 (秒)	1400(761)	1827(716)

識状態を効率的に測定できる, (2) 展示物がある場所まで受検者を短時間で誘導できる, (3) 状況に埋め込まれた学習そのものを評価することができる. 具体的には, フィールドにおいて観察や探索を通して項目に解答させることができる, である.

システムの有効性を評価するため, シミュレーション実験と東京都調布市の深大寺で被験者実験を実施した. まず, 受検者の能力値が効率的に測定されているかを評価するため, テスティング機能を用いたシミュレーション実験を実施した. シミュレーション実験の結果, 初期の能力推定値は受検者の真の能力値と乖離しているが, 受検者の解答数が増えるにしたがい真の能力値に収束していた. 累積項目情報量も増加しているため, 能力値の推定誤差が減少していることが確認された. 次に, 受検者を展示物へ短時間で誘導できるかを評価するため, 被験者実験を実施した. 被験者実験により, 紙媒体のマップと比較し, 被験者を短時間で展示物へ誘導できることを確認した. 最後に, 受検者が観察や探索を通して項目に解答しているか評価するため, 被験者実験を実施した. 仮想環境において眼球運動計測装置を用いた実験, 及び, 深大寺での被験者実験の結果から展示物に対して観察や探索を通して項目に解答していること

を確認した。これらの評価実験から、システムの有効性を確認した。

本システムは、観光地での学習として教育システム情報学会の論文誌に掲載されている。

## 第3章

# テスト情報量と移動距離を最適化する モバイル・アダプティブ・テストング

### 3.1 はじめに

前章では、状況に埋め込まれた学習を効率的に評価することを目的とした適応型テストを用いたモバイル・テストング・システムについて述べた。適応型テストは、受検者の能力値を逐次的に推定し、その能力値に対して情報量が最大の項目をアイテムバンクから抽出する手法である。しかし、適応型テストでは、各受検者に出題する項目が異なり、テストに費やされる解答所要時間も異なる。

同様の問題は、制限時間が設定された適応型テストにおいても報告されている。そこで、Lindenら [24, 25, 26] は、適応型テストにおける解答所要時間を最適化する項目選択手法を提案している。具体的には、2段階で項目を選択する。まず、1段階目では時間を制約として情報量が最大のテストを構成し、2段階目では構成されたテストから情報量が最大の項目を選択し、受検者に出題する。しかし、モバイル・テストングでは、項目が出題される順序に応じて移動時間が異なるため、情報量や解答所要時間のみならず、移動経路の最適を組み込んだ項目選択が必要である。

順序の最適化問題は、巡回セールスマン問題として知られている。特

に、移動距離のみならず、各地点で購入した商品の価格が最小になるパスを探索する問題は、Traveling Purchaser Problem(以下、TPP と呼ぶ)と呼ばれている。TPP とは、Traveling Salesman Problem の一般化の一つである [27, 28, 29, 30, 31, 32, 33]。TPP では、商店街に店舗が点在し、各店舗に商品が売られているとき、商品の価格と移動距離が最小になる購買経路を探索する。

本研究では、TPP における商品を項目、店舗を項目が出題される場所とみなすことで、テスト情報量と解答所要時間、出題順序を最適化できる適応型テストを提案する。TPP は、線形計画法により最適解を得ることができず、全探索が必要となり、店舗数を  $n$  とすると計算量が  $O(n!)$  になる。しかし、TPP は、動的計画法を用いることにより計算量を  $O(2n \cdot 2^n)$  に減少できることが知られている [34]。動的計画法は、計算量を減少させるため、最適化問題を複数の部分問題に分割し、その部分問題の結果を次の部分問題の計算に使う手法である [35]。しかし、TPP では目的関数が単調増加性を持たなければならず、本研究の場合、目的関数にテスト情報量の最大化と移動距離の最適化をいかに組み込むかが問題になる。そこで、本論では、移動時間に対して極力小さい重みを掛けたペナルティ項をテスト情報量に付与した目的関数を提案する。これにより、TPP の出題項目に対する単調増加関数として移動距離を最小にしなが情報量を最大化できる。さらに、TPP の店舗数  $n$  とモバイル・テストングの項目数  $I$  を  $n = I$  とし、従来  $O(I!)$  である計算量を  $O(2 \cdot I \cdot 2^I)$  に軽減できる。

提案手法の利点は以下の通りである。

1. 移動経路が最適化されるため、テストの解答所要時間に対する移動時間の割合が減少する。
2. 移動時間の減少により受検者が解答できる項目数が増加する。

3. 情報量が高い項目を出題でき、高精度な能力推定が期待できる。

さらに、シミュレーション実験、及び被験者実験を実施し、提案手法の有効性を示す。

## 3.2 制限時間を制約として組み込んだ適応型テスト

適応型テストでは、受検者ごとに出題される項目が異なり、テストに費やされる時間が異なる。モバイル・テストは、学習後の限られた時間で実施するため、受検者が十分な項目数を解答する前にテストが終了し、能力推定精度が減少する可能性がある。

このような問題は、制限時間が設定された適応型テストでも報告されており、この問題を解決するために、Linden ら [24] は、解答所要時間を最適化する適応型テストを提案している [24, 25, 26]。

この手法は、Shadow Test Approach (STA) と呼ばれ、テストの制約を満たし、かつ、テスト情報量が最大となるテストを構成し、構成したテストから項目情報量が最大の項目を選択する。構成されたテストをシャドーテストと呼ぶ。

$k$  番目の項目を出題するときのシャドーテストは、以下の制約を満たす項目により構成されている。

$$\begin{aligned} \text{Maximize } w = & \\ & \sum_{i=1}^I I_{U_1, \dots, U_{k-1}}(\hat{\theta}_{u_{i_1}, \dots, u_{i_{k-1}}}) x_i \end{aligned} \quad (3.1)$$

ただし

$$\sum_{i=1}^I \exp(\beta_i) x_i \leq t_{tot} \quad (3.2)$$

ここで,

$$x_i = \begin{cases} 1 & \text{シャドーテストに項目}i\text{が含まれる,} \\ 0 & \text{それ以外.} \end{cases} \quad (3.3)$$

を表す. また,  $\hat{\theta}_{k-1}$  は  $k-1$  番目までの項目への反応データを用いて推定した能力値である.  $t_{tot}$  は受検者がテストに費やす時間を表し, 解答速度  $\tau \in (-\infty, \infty)$  の受検者が項目  $i$  を解答する解答所要時間  $t_i$  の分布には, 対数正規分布を仮定する [36].

$$f(t_i; \tau, \alpha_i, \beta_i) = \frac{\alpha_i}{t_i \sqrt{2\pi}} \exp\left\{-\frac{1}{2}[\alpha_i(\ln t_i - (\beta_i - \tau))]^2\right\} \quad (3.4)$$

ここで,  $\beta_i \in (-\infty, \infty)$  は, 受検者が項目  $i$  の解答に費やす時間の度合を表す.  $\alpha_i \in (0, \infty)$  は, 受検者が項目  $i$  の解答に費やす時間の識別力であり,  $t_i$  の対数の分散を表している. また,  $\tau \in (-\infty, \infty)$  はスピードパラメータと呼ばれ, 以下の制約を持つ.

$$\sum_{l=1}^M \tau_l = 0 \quad (3.5)$$

式 (3.4) において  $t_i$  の対数の分布は, 期待値が  $\beta_i - \tau$  になる. スピードパラメータ  $\tau$  は, 式 (3.5) の制約を持つため, 平均は 0 になる. すなわち, パラメータ  $\beta_i$  は, 解答所要時間の対数の平均と等しくなる.

制限時間を制約とする適応型テストを用いる事で、制限時間中で情報量が最大となる項目を出題できる。本論でも、制限時間付き適応型テストを採用する。しかし、モバイル・テストングでは、項目が出題される場所に受検者を移動させるため、項目の出題順序によって移動時間が増大する可能性がある。そこで、本研究では、情報量や解答所要時間のみではなく、出題の順序を制約に組み込んだ適応型テストを提案する。

### 3.3 Traveling Purchaser Problem を用いた適応型テスト

#### 3.3.1 Traveling Purchaser Problem(TPP)

本研究では、テスト情報量最大化と移動距離最適化を同時に満たす項目を選択するために、Traveling Purchaser Problem(TPP)を用いる。TPPとは、TSPの一般形であり、移動距離のみでなく、各地点で購入した商品の価格が最小になるパスを探索する。この最適化問題では、購買者が商店街の店舗で商品を購入するとき、商品の価格と移動距離が最小になる購買経路を探索する。

TPPは、商品が売られている店舗がノード、移動距離がエッジであるグラフにおいて、各店舗での商品数と価格が既知であり、購入すべき商品が与えられたとき、移動距離と商品の価格が最小になるパスを探索する最適化問題である。TPPでは、店舗を  $S := \{v_1, \dots, v_n\}$ 、商品を  $K := \{p_1, \dots, p_m\}$  とする。全ての場所の集合は、 $S_0 := \{S \cup o\}$ 、 $o$  は出発地点である。グラフ  $G = (V, E)$  は無向グラフであり、 $V := S_0$  はノード集合、 $E := \{[v_i, v_j] : v_i, v_j \in V, i < j\}$  はエッジ集合である。商品  $p_k$  の価格は  $b_k$  とする。 $v_i, v_j$  間の移動距離は  $d_{ij}$  とする。店舗を訪れる経路はパスと呼ばれ、場所  $v_i$  の項目  $p_k$  がパスに含まれ

ているときは  $z_{ik} = 1$ , そうでなければ  $z_{ik} = 0$  とする. また, 場所  $v_i$  と場所  $v_j$  の間にエッジが含まれるときは  $x_{ij} = 1$ , そうでなければ  $x_{ij} = 0$  とする. このとき, 以下の最適化問題から最適なパスを見つける.

$$\text{Minimize } w = \sum_{v_i \in S} \sum_{p_k \in K} b_k z_{ik} + D \sum_{(i,j) \in L} d_{ij} x_{ij} \quad (3.6)$$

本研究では, モバイル・テストングに合わせて TPP を定式化し, テスト情報量と移動距離を最適化するパスを探索する.

### 3.3.2 テスト情報量と移動距離を最適化する TPP の定式化

定式化では, モバイル・テストングのため, TPP で用いた変数の名称を置き換える. 項目が出題される場所の集合を  $S := \{v_1, \dots, v_n\}$ , 項目の集合を  $K := \{p_1, \dots, p_I\}$  とする. 全ての場所の集合は  $S_0 := \{S \cup o\}$ ,  $o$  は出発地点である. 項目  $p_k$  の項目情報量は  $b_k$ , 項目  $p_k$  の解答所要時間は  $t_k$ , また,  $v_i, v_j$  間の移動時間は  $d_{ij}$  とする.  $T$  は, テストの制限時間である. このとき, 以下の最適化問題から最適なパスを見つける.

$$\text{Maximize } w = \sum_{v_i \in S} \sum_{p_k \in K} b_k z_{ik} - D \sum_{(i,j) \in L} d_{ij} x_{ij} \quad (3.7)$$

ただし,

$$\sum_{v_i \in S} \sum_{p_k \in K} t_k z_{ik} + \sum_{(i,j) \in L} d_{ij} x_{ij} < T \quad (3.8)$$

式 (3.7) はテスト情報量最大化と移動距離最適化のための目的関数, 式 (3.8) は移動時間と解答所要時間を制限時間内にするための制約条件である.



上述の最適化問題では、出題場所数  $n$  と項目数  $I$  を定義しているため、一つの場所で複数の項目が出題されることがある。ここでは、最適化問題の探索の考えを容易にするため、「一つの場所に一つの項目が出題される」として  $n = I$  とする。

TPP は線形計画法が適用できないため [37]、項目数  $I$  に対して計算量が  $O(I!)$  となる全探索法を用いなければならない。一方、近年、計算量を  $O(2I \cdot 2^I)$  に減少できる動的計画法を用いた探索法が提案され、広く利用されている [34]。動的計画法は、計算量を減少させるため、最適化問題を複数の部分問題に分割し、部分問題の結果を次の部分問題の計算に利用する [35]。しかし、TPP では目的関数が単調増加性を持たなければならず、本研究では、モバイル・テストングの目的関数にテスト情報量の最大化と移動距離の最小化をいかに組み込むかが重要となる。そこで、目的関数の移動時間に対して極力小さい重みを掛けたペナルティ項をテスト情報量に付与して単調増加性を保証する。この最適化問題により、テスト情報量最大化と移動距離最適化を同時に満たすパスを探索できる。

### 3.3.3 動的計画法を用いた最適解探索アルゴリズム

本研究では、モバイル・テストングのために定式化した TPP を探索するために、動的計画法を用いたアルゴリズム 1 を用いる。

アルゴリズム 1 は、最適解を探索する関数を呼ぶ Main 関数、最適解を探索する関数である OptCalculator 関数、探索した最適解を呼び出す SearchPath 関数から構成される。

Main 関数では、変数の初期化し、OptCalculator 関数と SearchPath 関数を呼び出している。Main 関数において、 $S$  は受検者が解答した項目集合、 $v$  は

受検者が最後に解答した項目, すなわち受検者の現在位置,  $t$  は経過時間を示している.  $0$  は, 出発地点を表すダミー項目である. OptCalculator 関数は, 引数に  $S$ ,  $v$ ,  $t$  を与え, 探索した最適解の目的関数の評価値を返す. 多次元配列 DP は,  $S$  と  $v$  に対して評価値を蓄積している. 動的計画法は, DP に計算した評価値を蓄積し, 再利用することにより, 計算量を減少させている. この多次元配列 DP を用いて計算量を減少させる仕組みを図 3.1 に示す. 例えば, ノード  $\{0, 1, 3, 4\}$  では, 受検者が解答した項目集合  $S$  に項目 1, 3, 4 が含まれ, 受検者が最後に解答した項目  $v$  は 4 である. ここでは, 受検者が項目 1, 3, 4 に解答したあと, 制限時間内で情報量が最大になる項目は 2 のみとし, 出発地点  $0$  に戻ることにする. このノード  $\{0, 1, 3, 4\}$  に対応する多次元配列 DP[S][v] には, 4 から  $0$  までのパスについて最適解の目的関数の評価値が入っている. 多次元配列 DP を用いることにより, ノード  $0, 3, 1, 4$  では格納されている評価値を呼び出すのみで計算を終えることができる. OptCalculator 関数は, 多次元配列 DP に計算結果が格納されていないとき, 計算を続ける. もし, 経過時間と出発地点に戻る時間が制限時間を超過したとき,  $-\text{INF}$  を返す.  $-\text{INF}$  は, 非常に大きな負の値である. もし, 制限時間を超過していないときは, 受検者に未出題であり, 目的関数の評価値が最大になる項目を選択する. 評価値の計算では, OptCalculator 関数を再帰関数として用いている. OptCalculator 関数では,  $S$  に出題した項目を追加し,  $v$  に受検者が最後に解答した項目,  $t$  に移動時間と項目解答時間を格納したものを引数として渡す. OptCalculator 関数は, 最適解の目的関数の評価値を返す. そのため, OptCalculator 関数の計算のあと, 多次元配列 DP から最適解であるパスを探索する必要がある. SearchPath 関数は, OptCalculator 関数とは逆の操作をし, 多次元配列 DP からパスを探索するものである. 多次元配列 DP には, 目的関数の評価値が格納されている

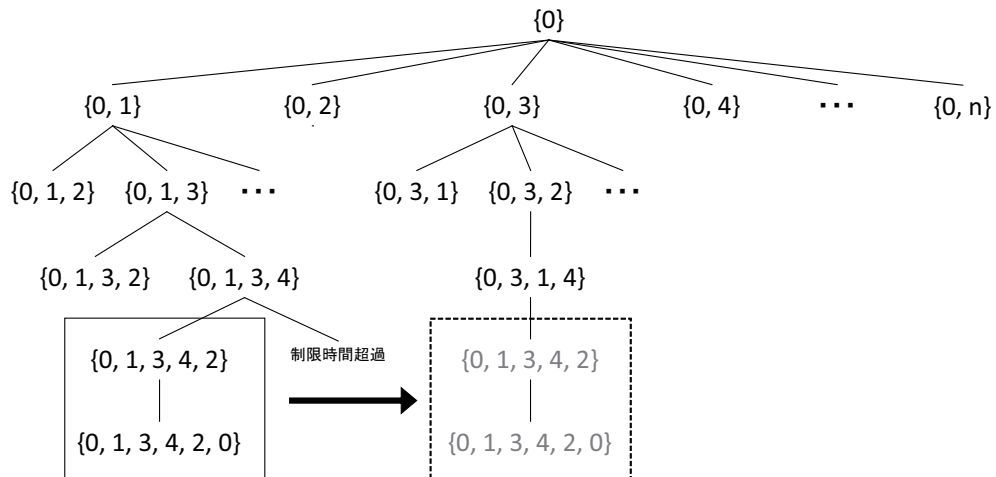


図 3.1 ダイナミックプログラミングによる最適解の探索

ため、その評価値が最大になるようにパスを探索する。SearchPath 関数では、 $S$  に解答済みの項目を追加し、 $DP[S][v]$  において評価値が最大になる  $v$  を見つける操作をする。評価値が最大になる  $v$  を発見した後、 $v$  を  $S$  に追加し、その上で評価値が最大になる  $v$  を見つける操作を繰り返す。この SearchPath 関数によって最適解のパスを探索できる。

本研究では、動的計画法を用いた上記のアルゴリズムにより TPP の最適解を探索する。

### 3.3.4 Traveling Purchaser Problem を用いた適応型テストのアルゴリズム

ここでは、TPP を組み込んだ適応型テストのアルゴリズムを詳述する。

まず、能力値の初期値を 0 とする。次に、システムは制限時間内で移動距離最小化とテスト情報量最大化を同時に満たすパスを探索し、受検者に出題する項目を選択する。受検者は選択された項目に解答する。システムは、受検者の解答を自動的に正誤判定し、それまでの解答履歴を用いて能力値を推定する。推定された能力値に応じてパスが更新される。この手続きを制限時間まで繰り返す。以上のアルゴリズムを図 3.2 に示す。

本手法は、項目数  $I$  に対して  $O(2I \cdot 2^I)$  の計算量となり、動的計画法を用いても計算量が多い。そこで、項目数  $I$  を減少させるため、項目出題エリアを複数に分割し、エリアごとの項目集合に対して提案手法を適用する。本研究では、エリア内の項目数が 20 項目以下になるようにエリア分割を行った。

本研究では、この手法を用いて、制限時間内でテスト情報量最大化と移動距離最適化を同時に満たす項目を逐次的に出題し、受検者の能力値を推定する。

### 3.4 シミュレーション実験

ここでは、提案手法の有効性を評価するため、シミュレーション実験を行う。この実験では、提案手法と時間のみ制約とした適応型テスト [24][38]、一般的な適応型テスト [12][13][14][15][16]、乱数による項目選択について受検者の真の能力値に対するテスト情報量を比較する。

シミュレーションの流れは以下の通りである。

1. アイテムバンクを、以下の条件で生成した。
  - 項目数：20 項目
  - $a_j \sim N(1.7, 0.4^2)$
  - $b_j \sim N(0, 0.5^2)$

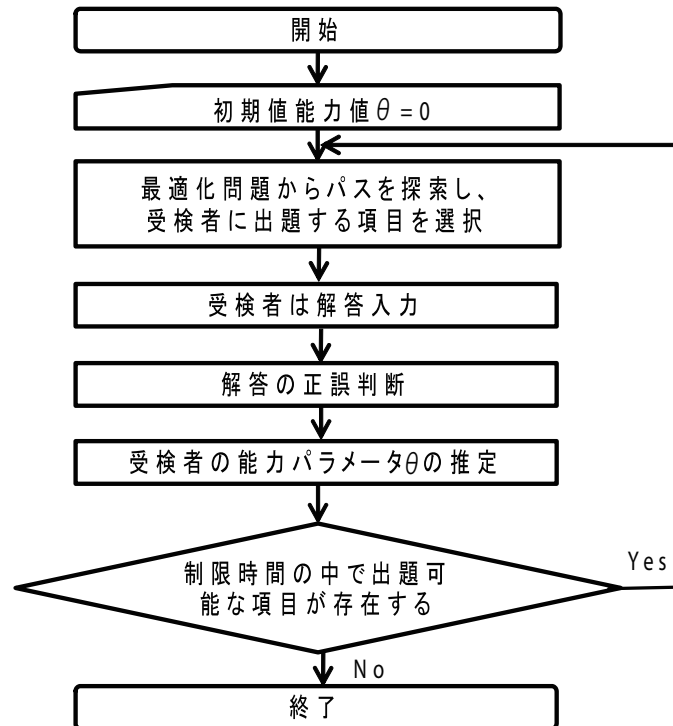


図 3.2 最適化問題を組み込んだ適応型テストのアルゴリズム

- $t_j \sim N(30, 10^2)$
  - 場所  $x \sim U(0, 300)$
  - 場所  $y \sim U(0, 300)$
2. 能力値の初期値を 0 とする.
  3. 能力推定値に応じて各手法に基づき項目を選択する.
  4. 選択した項目への反応データを, 真の能力値を所与とした式 (2.1) のロジスティック関数からランダムに発生させる.
  5. 反応データと解答履歴から受検者の能力を推定する.

6. 制限時間になるまで，4 の手順に従い項目の選択を繰り返す。
7. 真の能力値と制限時間を変えながら (2) ~ (6) を繰り返す。

シミュレーションにより生成された提案手法のパスの例を図 3.3, 図 3.4, 図 3.5, 図 3.6, 図 3.7, 図 3.8 に示す。各図は場所を表し，点は項目が出題される場所を示す。テストが開始される場所は，図の最も左上の点である。図 3.3 には，最初に探索されたパスを示す。探索されたパスは，受検者のその時の能力推定値に応じて探索されている。このパスは，次に出題する項目が現在地の右側であることを示している。図 3.4 には，図 3.3 のパスから選択された項目を受検者が解答し，その反応データから受検者の能力値を推定し，その能力値に応じて受検者の現在地から探索されたパスを示している。受検者の能力値が更新されてるため，図 3.3 のパスと図 3.4 のパスが異なっている。図 3.5, 図 3.6, 図 3.7, 図 3.8 には，この手続きが繰り返されたパスを示して

以上の実験結果を表 3.1, 表 3.2, 表 3.3 に示す。

表 3.1 の各値は，各能力値，各制限時間において 100 人のテスト情報量の平均値を表す。また，表 3.2 の各値は，受検者が解答した項目数の平均値を表す。表 3.3 の各値には，受検者が移動に費やした時間の平均値を示す。各表の(カッコ内)には，標準偏差を示す。

表 3.1 からは，制限時間が 400 秒の時を除くすべての場合で，提案手法のテスト情報量が既存手法よりも高い値を示したことがわかる。このことから，提案手法を用いることで既存手法より高精度な能力推定が可能であることが示せた。

また，表 3.3 からは，提案手法の移動時間が一般的な適応型テストやランダム移動の移動時間よりも短いことがわかる。ただし，提案手法の移動時間は，時間のみを制約とした適応型テストの移動時間と概ね同等程度である。しかし，

表 3.2 からは、提案手法の解答項目数が既存手法よりも多いことがわかる。これは、提案手法が解答所要時間のみならず、移動距離も最適化しているためである。このことから、提案手法の移動プロセスが既存手法よりも最適であることがわかる。

表 3.1 からは、制限時間が 400 秒のとき、提案手法のテスト情報量が時間のみ制約された適応型テスト [24][38] より低いことがわかる。次に、この現象の原因を明らかにする。図 3.9 と図 3.10 では、それぞれの手法を用いたテストにおいて、受検者が項目を解答したときの経路を示している。これらの図からは、提案手法の最初の項目が時間のみ制約された適応型テスト [24][38] より低いことがわかる。初期の推定では、受検者の反応データが少ないため、真の能力値と能力推定値が大きく乖離し、誤った能力推定値に基づいて項目が選択される。すなわち、コールドスタート問題が生じたといえる。特に、制限時間が短いほど、出題できる項目数が少なくなる傾向にあり、コールドスタート問題による影響が大きくなる。

この問題を解決するために、初期の項目選択のみ受検者の現在地に最も近い項目を出題するというルールを提案手法に組み込んだ。4 つの手法について比較・検討を行う。

- 提案手法 2：最初の 1 項目のみ最も近くにある項目を出題するルールを組み込んだ手法
- 提案手法 3：最初から 3 項目まで最も近くにある項目を出題するルールを組み込んだ手法
- 提案手法 4：最初から 5 項目まで最も近くにある項目を出題するルールを組み込んだ手法
- 提案手法 5：最初から 15 項目まで最も近くにある項目を出題するルー

### ルを組み込んだ手法

表 3.4 に、各手法を用いて実験を行った結果を示す。提案手法 2 は、制限時間が短いとき、テスト情報量が向上した。一方、提案手法 5 では、受検者に最も近い項目のみを出題し、項目情報量が高い項目を出題しないため、テスト情報量が増加しない。

上述の実験では、アイテムバンクを事前に生成し、共通のアイテムバンクを用いて各手法を比較した。しかし、実験結果は、アイテムバンクに依存するため、ここでは以下の 4 つのアイテムバンクを用いてシミュレーション実験を実施する。

- 前回の分析と同じ条件でアイテムバンクを受検者ごとに生成する。
- 項目の難易度のパラメータの生成では正規分布の平均を -2.0 とする。すなわち、簡単な項目が数多く蓄積されているアイテムバンクを用いる。
- 項目の難易度のパラメータの生成では正規分布の平均を 2.0 とする。すなわち、難しい項目が数多く蓄積されているアイテムバンクを用いる。
- 項目がある特定の箇所に集まっている。具体的には、任意の一点を中心とし、その場所から平均移動時間が 30 秒となるような標準偏差をもつ正規分布から発生する。(アイテムバンクの例、図 3.11)。

上記のアイテムバンクを用いてシミュレーション実験を実施する。シミュレーション実験の流れは以下の通りである。

1. 受検者の真の能力値と制限時間を設定する。
2. 受検者の能力推定値の初期値を 0 とする。
3. 上記の生成方法からアイテムバンクを生成する。



4. システムは受検者の能力推定値に応じて各手法に基づき項目を選択する。
5. 各受検者の反応データは、真の能力値に対して式 ( 2.1 ) のロジスティック関数から得られる正答確率と 0~1 で発生させた乱数を比較し、大きければ正答、そうでなければ誤答とする。
6. システムは反応パターンから受検者の能力値を推定する。
7. 受検者の能力推定値が収束していないのであれば 4 へ、そうでなければ次のステップへ。
8. 受検者が 100 人分でなければ 2 へ、そうでなければ次のステップへ。
9. 真の能力値を 5 つ、制限時間を 4 つそれぞれ計算していないなら 1 へ、そうでなければ終了する。

実験結果を表 3.5, 表 3.6, 表 3.7, 表 3.8 に示す。表 3.5 には、受検者ごとにアイテムバンクを生成したシミュレーションの結果を表す。表 3.6 には、容易な項目が多いアイテムバンクを用いたシミュレーションの結果を表す。表 3.6 には、難しい項目が多いアイテムバンクを用いたシミュレーションの結果を表す。表 3.8 には、項目がある一つの箇所に集まっているアイテムバンクを用いたシミュレーションの結果を表す。表の各値は、テスト情報量の平均を表す。(カッコ内) は標準偏差を示す。

表 3.5 の結果から、時間のみ制約された適応型テスト [24][38] と比較し、提案手法はテスト情報量が高いことがわかる。すなわち、提案手法は推定精度の高いテストを実現する。最初の項目にルートを加えた提案手法は、前回のシミュレーション結果と同様に制限時間が短いときに効果的である。すなわち、コールドスタート問題を解消し、推定精度の高いテストが実現できた。

次に表 3.6 と表 3.7 の結果から項目の特性に偏りがあるアイテムバンクを

用いた実験結果を分析する。表 3.6 の結果では、真の能力値が低い受検者のテスト情報量は高く、真の能力値が高い受検者のテスト情報量は低いことがわかる。表 3.7 の結果では、真の能力値が低い受検者のテスト情報量は低く、真の能力値が高い受検者のテスト情報量は高いことがわかる。テスト情報量は受検者の真の能力値と項目の難易度の値が等しくなるにつれて高くなる。すなわち、表 3.6 の結果では、真の能力値が低い受検者にとって能力推定に適した項目が多く、反対に真の能力値が高い受検者にとって能力推定に適した項目が少ないことがわかる。表 3.7 の結果では、真の能力値が低い受検者にとって能力推定に適した項目が少なく、反対に真の能力値が高い受検者にとって能力推定に適した項目が多いことがわかる。これらの結果から、受検者の能力推定に適した項目が少ないアイテムバンクでは、提案手法の有効性が少ないことがわかった。しかし、受検者の能力推定に適した項目が十分に多いアイテムバンクでは、提案手法の有効性が非常に高いことが分かった。初期の項目にルートを加えた提案手法は、前回のシミュレーション結果と同様に制限時間が短いときに効果的である。しかし、受検者の能力推定に適した項目が少ないアイテムバンクでは、得られる項目情報量が少なすぎるため、安定した良い効果を得ることはできなかった。

表 3.8 の結果から特定の箇所に項目が集まったアイテムバンクを用いた実験結果を分析する。表 3.8 の結果では、これまでのシミュレーションの結果と同様に、制限時間が増加すると提案手法は効果的だが、制限時間が少ない時は先行研究と推定精度の差は少ない。また、初期の項目にルートを追加することにより、制限時間が少ない時であってもテスト情報量を向上させることができる。

これらのシミュレーションの結果から、アイテムバンクに受検者の能力推

定に適した項目が数多くあることにより，提案手法が十分に効果的になることがわかった。

## 3.5 評価実験

### 3.5.1 実験の目的

TPP を組み込んだ適応型テストは，テスト情報量最大化と移動距離最適化を同時に満たす項目を受検者に出題することにより，移動プロセスの効率化と受検者の能力推定精度の向上が期待される．本章では，これらの利点を評価するために実施した被験者実験について詳述する．

ここでは，シミュレーション実験において最も高い能力推定精度を示した時間のみを制約とした適応型テスト [24][38] と提案手法を比較した．

この実験は，2章の被験者実験と同じ東京都の寺院において実施した．この寺院は，約6ヘクタール内にいくつかの歴史的建造物や仏像が点在している．各展示物は寺院の中で様々な場所に設置されているため，モバイル・テストでは移動時間が多くかかることが見込まれる．この寺院において，各適応型テストを用いてテストを実施し，受検者の能力推定精度を分析する．

項目は，2章と同じ項目であり，状況に埋め込まれた学習を評価するため，項目の解答に特定の場所での探索や観察が必要となるように作られている．項目の特性  $a_i, b_i, t_i, x_i, y_i$  は事前に推定されている．アイテムバンクのテスト項目の数は，85項目である．テストの制限時間は600秒とした．

被験者は，観察や探索を通して項目に解答させるために，この寺院を訪れたことがなく，事前知識を持っていない大学生10名とした．時間のみを制約とした適応型テストと提案手法を用いるグループとして5名ずつに分割した．

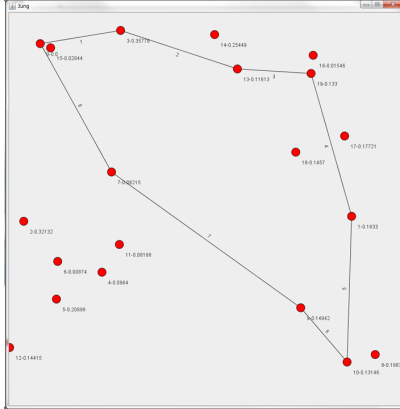


図 3.3 1 項目目のパス例

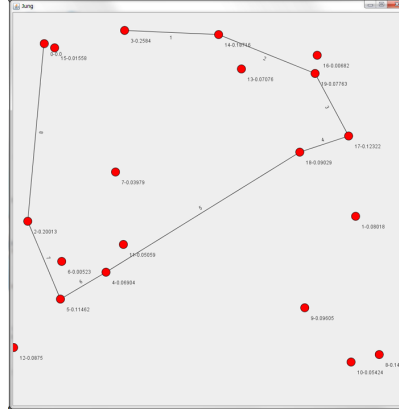


図 3.4 2 項目目のパス例

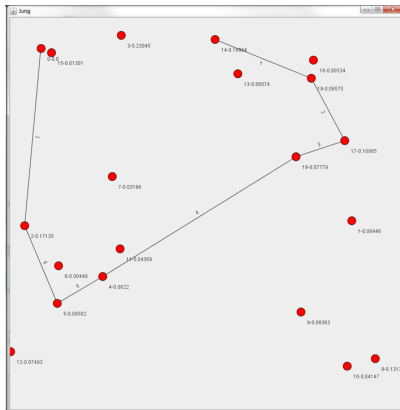


図 3.5 3 項目目のパス例

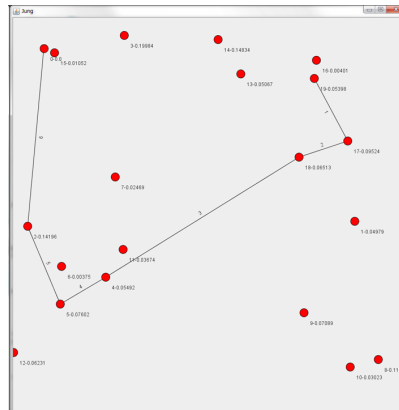


図 3.6 4 項目目のパス例

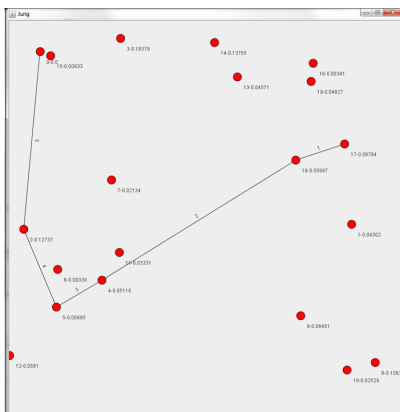


図 3.7 5 項目目のパス例

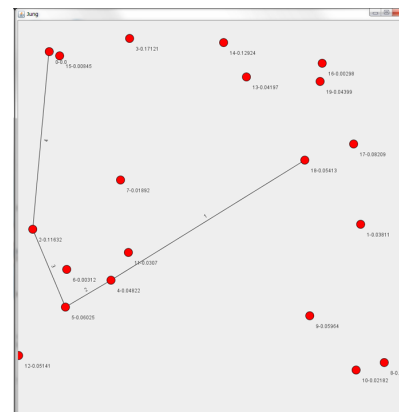


図 3.8 6 項目目のパス例

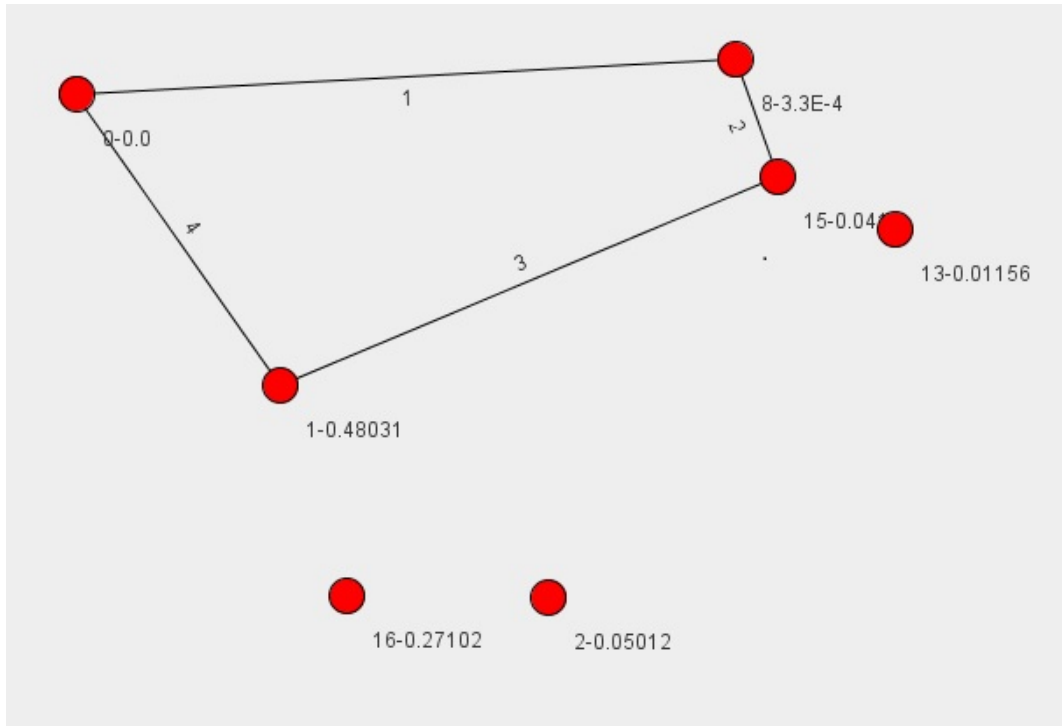


図 3.9 提案手法のパス

寺院での実験プロセスは以下の通りである。

1. 被験者は、実験の概要とシステムの使い方に関する説明を受ける。
2. システムは、項目を選択し、その項目が出題される場所を地図上に表示し、その場所まで受検者を誘導する。
3. 受検者は項目に解答する。
4. 制限時間を 10 分とし、出題できる項目があれば、2 へ、そうでなければテストを終了する。
5. テスト終了後にアンケートに回答する。

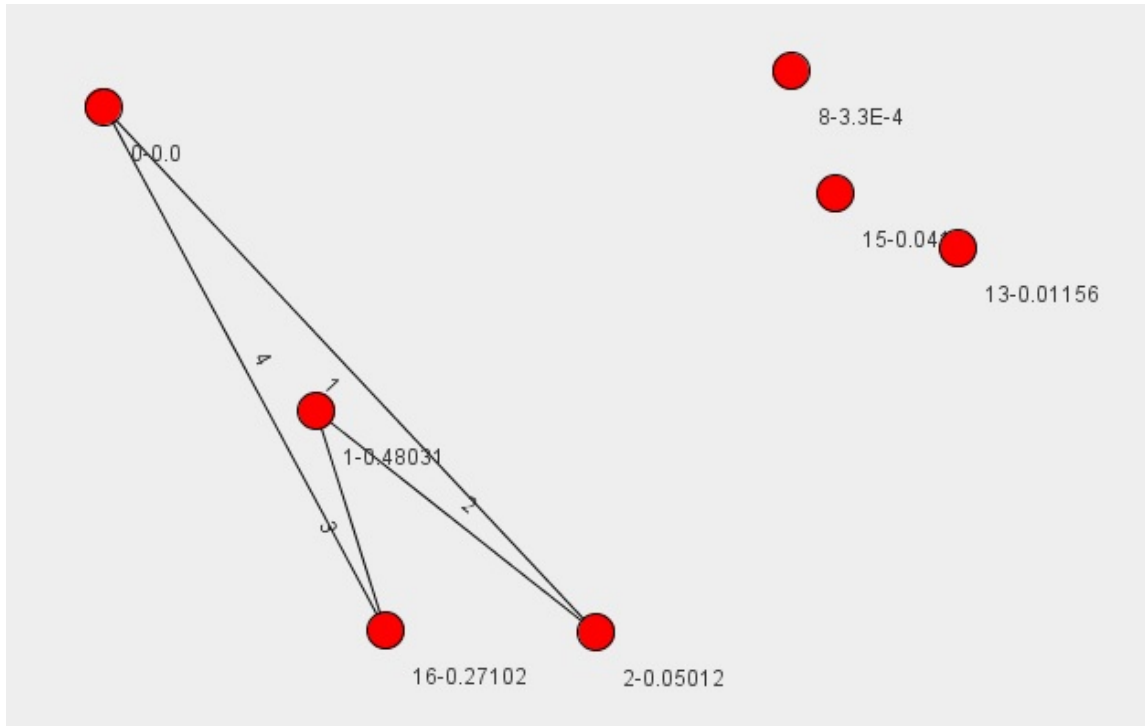


図 3.10 時間のみ制約とした適応型テストのパス

本システムでは、各適応型テストの能力推定精度と移動経路を分析するため、受検者が解答した項目、項目の解答所要時間、受検者が移動した経路、移動時間、移動回数をデータベースに蓄積している。

また、実験の最後にアンケートを実施した。アンケートは、以下の二つの質問であり、リッカート尺度を用いた。(6:とても当てはまる, 5:当てはまる, 4:どちらかといえば当てはまる, 3:どちらかといえば当てはまらない, 2:当てはまらない, 1:全く当てはまらない)。

1. 項目解答時間に対して移動時間は長すぎず適度であった。
2. 移動プロセスは無駄がなく最適な順序であった。

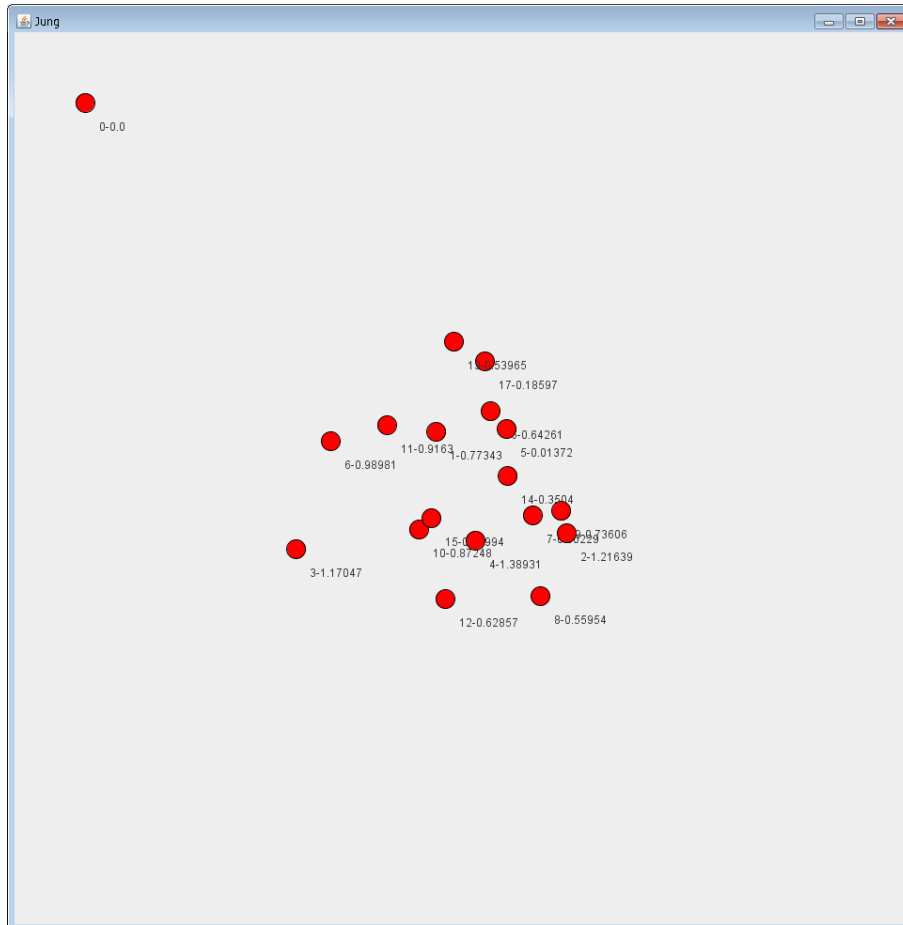


図 3.11 項目が特定の場所に偏ったアイテムバンクの例

実験の結果，提案手法は，時間のみ制約とした適応型テストと比較し，移動回数が有意に少ない．これは，提案手法では，TPP を組み込んだことにより，移動経路が最適化されたためといえる．

また，移動回数の減少により，提案手法は移動時間も有意に短い結果となった．実験を実施した寺院では，展示物間の距離が比較的長いため移動回数の減少により移動時間が大幅に減少した．

受検者が解答した項目数は，提案手法が有意に多い．寺院の各展示物には

複数の項目が存在し、同じ展示物に関する項目を連続して出題できるため、1項目ごとに移動する必要がなくなり、出題項目数が大幅に増加している。展示物は一般的に複数の項目が準備されることが考えられるため、この性質は非常に有用である。

以上により、出題項目数が増加し、移動経路が最適化されたため、テスト情報量も有意に高い値となった。

また、アンケート (1) の結果、提案手法の評価値が、有意に高いことがわかる。このことから、提案手法を用いることで、項目解答時間に対する移動時間を短縮できたといえる。同様に、アンケート (2) の結果も提案手法の評価値が有意に高い。このことから、提案手法を用いることで、移動経路を最適化できたことが示せた。

以上の実験から、提案手法の有効性を示せた。

### 3.6 おわりに

本研究では、テスト情報量最大化と移動距離最適化を同時に満たす最適解を探索するため、モバイル・テストのための TPP を提案した。さらに、TPP を組み込んだ適応型テストを用いたモバイル・テスト・システムを開発した。TPP では目的関数に単調増加性を持たせなければならないため、本研究では、移動時間に対して極力小さい重みを掛けたペナルティ項をテスト情報量に付与した目的関数を提案した。これにより、目的関数は項目数に対して単調増加性を保持しながらテスト情報量最大化と移動距離最適化を同時に満たす。さらに、項目数  $I$  に対して計算量が  $O(I!)$  である計算量を  $O(2I \cdot 2^I)$  に減少させることができる。本研究では、シミュレーション実験、及び被験者実験を実施し、提案手法の利点である、(1) 移動経路が最適化されるため、テス



トの解答所要時間に対する移動時間の割合が減少する (2) 移動時間の減少により受検者が解答できる項目数が増加する (3) 情報量が高い項目を出題でき、高精度な能力推定が期待できる、を確認し、提案手法の有効性を示した。今後、出題順序の制約を一般的な形式として適応型テストに組み込む予定である。

---

### アルゴリズム 1 提案手法の疑似コード

---

**Require:** アイテムバンク, テスト条件

**Ensure:** 情報量と移動時間を最適化するパス

```

function MAIN
  初期化
   $S = 0$ 
   $v = start$ 
   $t = 0$ 
  OPTMOBILETESTING( $S, v, t$ )
  SearchPath()
end function

function OPTCALCULATOR( $S, v, t$ )
  評価値を返す関数
  DP[S][v] 上に計算結果がキャッシュされてる場合
  if  $t > MaxTime$  then return -INF then
  end if
  ルート探索の時間が超えている場合
  if  $t + DistanceTime[v][0] > MaxTime$  then return DP[S][v] = -INF then
  end if
  DP[S][v] 上に計算結果がない場合.
  OptimumScore = 0
  for all  $i$  in Itembank do
    if  $i \notin S$  then then
       $Time_i = t + DistanceTime[v][i] + ResponceTime[i]$ 
       $S_i := S \cup i$ 
       $TmpEvalu = OPTMOBILETESTING(S_i, i, Time_i)$ 
       $+Info[i] - (DistanceTime[v][i] + ResponceTime[i])$ 
       $*D(D \ll 1)$ 
      if  $OptimumScore < TmpEvalu$  then then
         $OptimumScore := TmpEvalu$ 
      end if
    end if
  end for return OptimumScore
end function

function SEARCHPATH
  DP 表からパスを探索
  OptimumPath[0] = 0
   $S = 0$ 
  for all  $i$  in PathLength do
    OptimumScore = 0
    OptimumItem = 0
    for all  $j$  in ItemNum do
       $S_j := S \cup \{j\}$ 
       $TmpEvalu = dp[S_j][j] + Info[i]$ 
       $-(DistanceTime[v][i] + ResponceTime[i])$ 
       $*D(D \ll 1)$ 
      if  $OptimumScore < TmpEvalu$  then then
         $OptimumScore = TmpEvalu$ 
         $OptimumItem = j$ 
      end if
       $S := S \cup \{j\}$ 
       $OptimumPath[i] = OptimumItem$ 
    end for
  end for return OptimumPath
end function

```

---

表 3.1 各適応型テストの平均テスト情報量（標準偏差）の例

時間 [s]	真の能力値	提案手法	時間のみ制約とした 適応型テスト [24][38]	適応型テスト [12][13][14][15][16]	ランダム
400	-2.0	0.031 (0.0 <sup>2</sup> )	0.031 (0.0 <sup>2</sup> )	0.001 (0.0 <sup>2</sup> )	0.0 (0.0 <sup>2</sup> )
	-1.0	0.535 (0.069 <sup>2</sup> )	0.548 (0.0 <sup>2</sup> )	0.248 (0.0 <sup>2</sup> )	0.0 (0.0 <sup>2</sup> )
	0.0	1.317 (0.043 <sup>2</sup> )	1.47 (0.188 <sup>2</sup> )	1.275 (0.0 <sup>2</sup> )	0.0 (0.0 <sup>2</sup> )
	1.0	0.502 (0.059 <sup>2</sup> )	0.533 (0.061 <sup>2</sup> )	0.008 (0.0 <sup>2</sup> )	0.0 (0.0 <sup>2</sup> )
	2.0	0.046 (0.004 <sup>2</sup> )	0.048 (0.004 <sup>2</sup> )	0.0 (0.0 <sup>2</sup> )	0.0 (0.0 <sup>2</sup> )
	600	-2.0	0.039 (0.0 <sup>2</sup> )	0.038 (0.0 <sup>2</sup> )	0.001 (0.0 <sup>2</sup> )
-1.0		0.634 (0.0 <sup>2</sup> )	0.623 (0.004 <sup>2</sup> )	0.248 (0.0 <sup>2</sup> )	0.235 (0.191 <sup>2</sup> )
0.0		1.926 (0.0 <sup>2</sup> )	1.867 (0.165 <sup>2</sup> )	1.275 (0.0 <sup>2</sup> )	0.594 (0.403 <sup>2</sup> )
1.0		1.186 (0.0 <sup>2</sup> )	1.117 (0.168 <sup>2</sup> )	0.008 (0.0 <sup>2</sup> )	0.283 (0.284 <sup>2</sup> )
2.0		0.184 (0.0 <sup>2</sup> )	0.092 (0.0 <sup>2</sup> )	0.0 (0.0 <sup>2</sup> )	0.04 (0.048 <sup>2</sup> )
800		-2.0	0.057 (0.037 <sup>2</sup> )	0.03 (0.0 <sup>2</sup> )	0.03 (0.0 <sup>2</sup> )
	-1.0	0.902 (0.097 <sup>2</sup> )	0.724 (0.038 <sup>2</sup> )	0.724 (0.039 <sup>2</sup> )	0.404 (0.268 <sup>2</sup> )
	0.0	2.423 (0.199 <sup>2</sup> )	2.022 (0.091 <sup>2</sup> )	2.306 (0.33 <sup>2</sup> )	0.906 (0.442 <sup>2</sup> )
	1.0	1.668 (0.369 <sup>2</sup> )	0.906 (0.0 <sup>2</sup> )	0.692 (0.106 <sup>2</sup> )	0.417 (0.34 <sup>2</sup> )
	2.0	0.295 (0.026 <sup>2</sup> )	0.081 (0.0 <sup>2</sup> )	0.044 (0.001 <sup>2</sup> )	0.061 (0.057 <sup>2</sup> )
	1000	-2.0	0.68 (0.012 <sup>2</sup> )	0.037 (0.0 <sup>2</sup> )	0.03 (0.0 <sup>2</sup> )
-1.0		2.118 (0.243 <sup>2</sup> )	0.805 (0.017 <sup>2</sup> )	0.72 (0.055 <sup>2</sup> )	0.675 (0.341 <sup>2</sup> )
0.0		4.387 (0.522 <sup>2</sup> )	3.169 (0.445 <sup>2</sup> )	2.336 (0.291 <sup>2</sup> )	1.331 (0.545 <sup>2</sup> )
1.0		2.345 (0.162 <sup>2</sup> )	1.46 (0.156 <sup>2</sup> )	1.162 (0.176 <sup>2</sup> )	0.651 (0.4 <sup>2</sup> )
2.0		0.35 (0.012 <sup>2</sup> )	0.216 (0.012 <sup>2</sup> )	0.181 (0.0 <sup>2</sup> )	0.106 (0.066 <sup>2</sup> )

表 3.2 各適応型テストの平均解答項目数（標準偏差）の例

時間 [s]	真の能力	提案手法	時間のみを制約とした 適応型テスト [24][38]	適応型テスト [12][13][14][15][16]	ランダム
400	-2.0	3.0 (0.0 <sup>2</sup> )	3.0 (0.0 <sup>2</sup> )	2.0 (0.0 <sup>2</sup> )	0.0 (0.0 <sup>2</sup> )
	-1.0	3.0 (0.0 <sup>2</sup> )	3.0 (0.0 <sup>2</sup> )	2.0 (0.0 <sup>2</sup> )	0.0 (0.0 <sup>2</sup> )
	0.0	3.0 (0.0 <sup>2</sup> )	3.29 (0.454 <sup>2</sup> )	2.0 (0.0 <sup>2</sup> )	0.0 (0.0 <sup>2</sup> )
	1.0	3.0 (0.0 <sup>2</sup> )	3.95 (0.218 <sup>2</sup> )	2.0 (0.0 <sup>2</sup> )	0.0 (0.0 <sup>2</sup> )
	2.0	3.0 (0.0 <sup>2</sup> )	3.99 (0.099 <sup>2</sup> )	2.0 (0.0 <sup>2</sup> )	0.0 (0.0 <sup>2</sup> )
	600	-2.0	5.0 (0.0 <sup>2</sup> )	4.0 (0.0 <sup>2</sup> )	2.0 (0.0 <sup>2</sup> )
-1.0		5.0 (0.0 <sup>2</sup> )	4.04 (0.196 <sup>2</sup> )	2.0 (0.0 <sup>2</sup> )	2.29 (0.496 <sup>2</sup> )
0.0		5.0 (0.0 <sup>2</sup> )	4.28 (0.449 <sup>2</sup> )	2.0 (0.0 <sup>2</sup> )	2.31 (0.504 <sup>2</sup> )
1.0		5.0 (0.0 <sup>2</sup> )	4.92 (0.271 <sup>2</sup> )	2.0 (0.0 <sup>2</sup> )	2.26 (0.461 <sup>2</sup> )
2.0		5.0 (0.0 <sup>2</sup> )	5.0 (0.0 <sup>2</sup> )	2.0 (0.0 <sup>2</sup> )	2.44 (0.571 <sup>2</sup> )
800		-2.0	5.09 (0.512 <sup>2</sup> )	3.0 (0.0 <sup>2</sup> )	3.0 (0.0 <sup>2</sup> )
	-1.0	5.33 (0.906 <sup>2</sup> )	3.01 (0.099 <sup>2</sup> )	3.01 (0.099 <sup>2</sup> )	3.21 (0.864 <sup>2</sup> )
	0.0	6.27 (1.232 <sup>2</sup> )	3.88 (0.325 <sup>2</sup> )	3.76 (0.427 <sup>2</sup> )	3.26 (0.702 <sup>2</sup> )
	1.0	6.48 (0.854 <sup>2</sup> )	4.0 (0.0 <sup>2</sup> )	3.27 (0.444 <sup>2</sup> )	2.96 (0.734 <sup>2</sup> )
	2.0	6.02 (0.244 <sup>2</sup> )	4.0 (0.0 <sup>2</sup> )	3.01 (0.099 <sup>2</sup> )	2.99 (0.794 <sup>2</sup> )
	1000	-2.0	9.14 (0.849 <sup>2</sup> )	4.0 (0.0 <sup>2</sup> )	3.0 (0.0 <sup>2</sup> )
-1.0		9.18 (1.108 <sup>2</sup> )	4.08 (0.392 <sup>2</sup> )	3.02 (0.14 <sup>2</sup> )	4.47 (0.921 <sup>2</sup> )
0.0		9.66 (0.552 <sup>2</sup> )	5.65 (0.684 <sup>2</sup> )	3.87 (0.336 <sup>2</sup> )	4.39 (0.835 <sup>2</sup> )
1.0		9.08 (0.271 <sup>2</sup> )	5.23 (0.421 <sup>2</sup> )	4.0 (0.0 <sup>2</sup> )	4.21 (0.898 <sup>2</sup> )
2.0		9.0 (0.0 <sup>2</sup> )	5.01 (0.099 <sup>2</sup> )	4.0 (0.0 <sup>2</sup> )	4.3 (0.964 <sup>2</sup> )

表 3.3 各適応型テストの平均移動時間（標準偏差）の例

時間 [s]	真の能力	提案手法	時間のみ制約とした 適応型テスト [24][38]	適応型テスト [12][13][14][15][16]	ランダム
400	-2.0	267.0 (0.0 <sup>2</sup> )	267.0 (0.0 <sup>2</sup> )	686.0 (0.0 <sup>2</sup> )	400.0 (0.0 <sup>2</sup> )
	-1.0	268.68 (9.553 <sup>2</sup> )	267.0 (0.0 <sup>2</sup> )	686.0 (0.0 <sup>2</sup> )	400.0 (0.0 <sup>2</sup> )
	0.0	288.28 (27.182 <sup>2</sup> )	288.75 (34.032 <sup>2</sup> )	686.0 (0.0 <sup>2</sup> )	400.0 (0.0 <sup>2</sup> )
	1.0	319.64 (13.299 <sup>2</sup> )	338.25 (16.346 <sup>2</sup> )	686.0 (0.0 <sup>2</sup> )	400.0 (0.0 <sup>2</sup> )
	2.0	322.44 (5.572 <sup>2</sup> )	341.25 (7.462 <sup>2</sup> )	686.0 (0.0 <sup>2</sup> )	400.0 (0.0 <sup>2</sup> )
	600	-2.0	472.0 (0.0 <sup>2</sup> )	426.0 (0.0 <sup>2</sup> )	686.0 (0.0 <sup>2</sup> )
-1.0		472.0 (0.0 <sup>2</sup> )	429.68 (18.028 <sup>2</sup> )	686.0 (0.0 <sup>2</sup> )	552.62 (150.27 <sup>2</sup> )
0.0		472.0 (0.0 <sup>2</sup> )	451.76 (41.308 <sup>2</sup> )	686.0 (0.0 <sup>2</sup> )	569.48 (151.854 <sup>2</sup> )
1.0		472.0 (0.0 <sup>2</sup> )	510.64 (24.959 <sup>2</sup> )	686.0 (0.0 <sup>2</sup> )	538.59 (157.905 <sup>2</sup> )
2.0		472.0 (0.0 <sup>2</sup> )	518.0 (0.0 <sup>2</sup> )	686.0 (0.0 <sup>2</sup> )	574.83 (150.051 <sup>2</sup> )
800		-2.0	707.08 (10.918 <sup>2</sup> )	707.0 (0.0 <sup>2</sup> )	707.0 (0.0 <sup>2</sup> )
	-1.0	701.46 (20.665 <sup>2</sup> )	707.06 (0.597 <sup>2</sup> )	707.41 (4.079 <sup>2</sup> )	693.48 (118.767 <sup>2</sup> )
	0.0	677.47 (31.625 <sup>2</sup> )	712.28 (1.95 <sup>2</sup> )	738.66 (16.664 <sup>2</sup> )	710.12 (108.573 <sup>2</sup> )
	1.0	680.0 (22.734 <sup>2</sup> )	713.0 (0.0 <sup>2</sup> )	721.72 (15.983 <sup>2</sup> )	697.51 (100.919 <sup>2</sup> )
	2.0	693.47 (5.44 <sup>2</sup> )	713.0 (0.0 <sup>2</sup> )	712.36 (3.582 <sup>2</sup> )	713.84 (109.183 <sup>2</sup> )
	1000	-2.0	880.18 (7.721 <sup>2</sup> )	866.28 (2.786 <sup>2</sup> )	707.0 (0.0 <sup>2</sup> )
-1.0		879.32 (12.067 <sup>2</sup> )	871.32 (12.007 <sup>2</sup> )	707.82 (5.74 <sup>2</sup> )	859.92 (136.413 <sup>2</sup> )
0.0		871.86 (9.199 <sup>2</sup> )	885.88 (39.656 <sup>2</sup> )	750.23 (29.131 <sup>2</sup> )	884.75 (149.349 <sup>2</sup> )
1.0		860.52 (11.623 <sup>2</sup> )	911.08 (46.33 <sup>2</sup> )	810.16 (36.845 <sup>2</sup> )	852.35 (131.813 <sup>2</sup> )
2.0		863.4 (4.2 <sup>2</sup> )	934.88 (11.144 <sup>2</sup> )	832.0 (0.0 <sup>2</sup> )	875.26 (134.747 <sup>2</sup> )

表 3.4 拡張した提案手法の平均テスト情報量（標準偏差）分析

時間 [s]	真の能力値	提案手法	時間のみ制約とした 適応型テスト [24][38]	提案手法 2	提案手法 3	提案手法 4	提案手法 5
400	-2	0.031 (0.0 <sup>2</sup> )	0.031 (0.0 <sup>2</sup> )	0.036 (0.0 <sup>2</sup> )	0.036 (0.0 <sup>2</sup> )	0.036 (0.0 <sup>2</sup> )	0.036 (0.0 <sup>2</sup> )
	-1	0.539 (0.057 <sup>2</sup> )	0.548 (0.007 <sup>2</sup> )	0.562 (0.021 <sup>2</sup> )	0.555 (0.0 <sup>2</sup> )	0.555 (0.0 <sup>2</sup> )	0.555 (0.0 <sup>2</sup> )
	0	1.32 (0.042 <sup>2</sup> )	1.487 (0.195 <sup>2</sup> )	1.679 (0.254 <sup>2</sup> )	0.916 (0.0 <sup>2</sup> )	0.916 (0.0 <sup>2</sup> )	0.916 (0.0 <sup>2</sup> )
	1	0.507 (0.049 <sup>2</sup> )	0.528 (0.071 <sup>2</sup> )	0.542 (0.033 <sup>2</sup> )	0.309 (0.0 <sup>2</sup> )	0.309 (0.0 <sup>2</sup> )	0.309 (0.0 <sup>2</sup> )
	2	0.047 (0.0 <sup>2</sup> )	0.048 (0.0 <sup>2</sup> )	0.048 (0.0 <sup>2</sup> )	0.038 (0.0 <sup>2</sup> )	0.038 (0.0 <sup>2</sup> )	0.038 (0.0 <sup>2</sup> )
	600	-2	0.039 (0.0 <sup>2</sup> )	0.038 (0.0 <sup>2</sup> )	0.559 (0.01 <sup>2</sup> )	0.508 (0.115 <sup>2</sup> )	0.56 (0.0 <sup>2</sup> )
-1		0.634 (0.0 <sup>2</sup> )	0.622 (0.0 <sup>2</sup> )	1.37 (0.09 <sup>2</sup> )	1.219 (0.149 <sup>2</sup> )	1.418 (0.0 <sup>2</sup> )	1.418 (0.0 <sup>2</sup> )
0		1.926 (0.0 <sup>2</sup> )	1.889 (0.174 <sup>2</sup> )	2.017 (0.131 <sup>2</sup> )	1.928 (0.181 <sup>2</sup> )	1.226 (0.0 <sup>2</sup> )	1.226 (0.0 <sup>2</sup> )
1		1.186 (0.0 <sup>2</sup> )	1.142 (0.121 <sup>2</sup> )	1.145 (0.11 <sup>2</sup> )	0.922 (0.156 <sup>2</sup> )	0.384 (0.0 <sup>2</sup> )	0.384 (0.0 <sup>2</sup> )
2		0.184 (0.0 <sup>2</sup> )	0.092 (0.0 <sup>2</sup> )	0.092 (0.0 <sup>2</sup> )	0.102 (0.003 <sup>2</sup> )	0.058 (0.0 <sup>2</sup> )	0.058 (0.0 <sup>2</sup> )
800		-2	0.057 (0.037 <sup>2</sup> )	0.03 (0.0 <sup>2</sup> )	0.618 (0.038 <sup>2</sup> )	0.582 (0.127 <sup>2</sup> )	0.622 (0.0 <sup>2</sup> )
	-1	0.897 (0.086 <sup>2</sup> )	0.713 (0.076 <sup>2</sup> )	1.638 (0.449 <sup>2</sup> )	1.49 (0.358 <sup>2</sup> )	1.925 (0.06 <sup>2</sup> )	1.504 (0.0 <sup>2</sup> )
	0	2.439 (0.14 <sup>2</sup> )	2.008 (0.105 <sup>2</sup> )	2.513 (0.142 <sup>2</sup> )	2.51 (0.236 <sup>2</sup> )	2.406 (0.116 <sup>2</sup> )	1.809 (0.0 <sup>2</sup> )
	1	1.712 (0.331 <sup>2</sup> )	0.897 (0.086 <sup>2</sup> )	1.683 (0.456 <sup>2</sup> )	1.689 (0.187 <sup>2</sup> )	1.442 (0.227 <sup>2</sup> )	1.221 (0.0 <sup>2</sup> )
	2	0.294 (0.028 <sup>2</sup> )	0.081 (0.0 <sup>2</sup> )	0.245 (0.0 <sup>2</sup> )	0.284 (0.025 <sup>2</sup> )	0.236 (0.018 <sup>2</sup> )	0.17 (0.0 <sup>2</sup> )
	1000	-2	0.679 (0.013 <sup>2</sup> )	0.037 (0.0 <sup>2</sup> )	0.67 (0.065 <sup>2</sup> )	0.638 (0.109 <sup>2</sup> )	0.682 (0.003 <sup>2</sup> )
-1		2.096 (0.237 <sup>2</sup> )	0.805 (0.018 <sup>2</sup> )	2.15 (0.207 <sup>2</sup> )	2.077 (0.225 <sup>2</sup> )	2.282 (0.15 <sup>2</sup> )	1.966 (0.0 <sup>2</sup> )
0		4.381 (0.564 <sup>2</sup> )	3.102 (0.468 <sup>2</sup> )	4.347 (0.503 <sup>2</sup> )	4.13 (0.362 <sup>2</sup> )	3.579 (0.324 <sup>2</sup> )	3.914 (0.0 <sup>2</sup> )
1		2.364 (0.131 <sup>2</sup> )	1.449 (0.162 <sup>2</sup> )	2.379 (0.127 <sup>2</sup> )	2.414 (0.152 <sup>2</sup> )	2.033 (0.353 <sup>2</sup> )	1.708 (0.0 <sup>2</sup> )
2		0.351 (0.0 <sup>2</sup> )	0.216 (0.012 <sup>2</sup> )	0.351 (0.002 <sup>2</sup> )	0.385 (0.004 <sup>2</sup> )	0.355 (0.013 <sup>2</sup> )	0.231 (0.0 <sup>2</sup> )

表 3.5 ランダム生成されたアイテムバンクを用いた各適応型テストの平均  
テスト情報量 (標準偏差) 分析

時間 [s]	真の能力値	提案手法	時間のみ制約とした 適応型テスト [24][38]	提案手法 2	提案手法 3	提案手法 4	提案手法 5
400	-2	0.077 (0.076 <sup>2</sup> )	0.069 (0.072 <sup>2</sup> )	0.096 (0.079 <sup>2</sup> )	0.095 (0.101 <sup>2</sup> )	0.089 (0.076 <sup>2</sup> )	0.088 (0.072 <sup>2</sup> )
	-1	0.835 (0.469 <sup>2</sup> )	0.617 (0.45 <sup>2</sup> )	0.911 (0.545 <sup>2</sup> )	0.792 (0.538 <sup>2</sup> )	0.778 (0.547 <sup>2</sup> )	0.737 (0.419 <sup>2</sup> )
	0	2.115 (0.876 <sup>2</sup> )	1.848 (0.675 <sup>2</sup> )	2.055 (0.828 <sup>2</sup> )	1.931 (0.931 <sup>2</sup> )	1.934 (0.777 <sup>2</sup> )	1.783 (0.714 <sup>2</sup> )
	1	0.84 (0.542 <sup>2</sup> )	0.648 (0.367 <sup>2</sup> )	0.848 (0.521 <sup>2</sup> )	0.763 (0.447 <sup>2</sup> )	0.781 (0.531 <sup>2</sup> )	0.74 (0.456 <sup>2</sup> )
	2	0.091 (0.104 <sup>2</sup> )	0.072 (0.081 <sup>2</sup> )	0.099 (0.069 <sup>2</sup> )	0.086 (0.081 <sup>2</sup> )	0.083 (0.068 <sup>2</sup> )	0.068 (0.054 <sup>2</sup> )
600	-2	0.167 (0.115 <sup>2</sup> )	0.106 (0.102 <sup>2</sup> )	0.163 (0.098 <sup>2</sup> )	0.166 (0.109 <sup>2</sup> )	0.161 (0.134 <sup>2</sup> )	0.145 (0.128 <sup>2</sup> )
	-1	1.54 (0.571 <sup>2</sup> )	1.054 (0.594 <sup>2</sup> )	1.662 (0.761 <sup>2</sup> )	1.358 (0.655 <sup>2</sup> )	1.249 (0.584 <sup>2</sup> )	1.226 (0.545 <sup>2</sup> )
	0	3.466 (1.115 <sup>2</sup> )	2.534 (0.781 <sup>2</sup> )	3.285 (1.03 <sup>2</sup> )	3.379 (0.989 <sup>2</sup> )	3.226 (1.027 <sup>2</sup> )	3.348 (1.066 <sup>2</sup> )
	1	1.364 (0.581 <sup>2</sup> )	1.023 (0.516 <sup>2</sup> )	1.513 (0.6 <sup>2</sup> )	1.377 (0.538 <sup>2</sup> )	1.377 (0.704 <sup>2</sup> )	1.258 (0.525 <sup>2</sup> )
	2	0.161 (0.118 <sup>2</sup> )	0.107 (0.073 <sup>2</sup> )	0.183 (0.142 <sup>2</sup> )	0.151 (0.105 <sup>2</sup> )	0.159 (0.125 <sup>2</sup> )	0.148 (0.103 <sup>2</sup> )
800	-2	0.262 (0.139 <sup>2</sup> )	0.172 (0.104 <sup>2</sup> )	0.293 (0.17 <sup>2</sup> )	0.258 (0.135 <sup>2</sup> )	0.293 (0.175 <sup>2</sup> )	0.207 (0.124 <sup>2</sup> )
	-1	2.233 (0.669 <sup>2</sup> )	1.502 (0.558 <sup>2</sup> )	2.219 (0.719 <sup>2</sup> )	2.22 (0.76 <sup>2</sup> )	2.085 (0.737 <sup>2</sup> )	1.705 (0.711 <sup>2</sup> )
	0	4.837 (1.044 <sup>2</sup> )	3.154 (0.898 <sup>2</sup> )	4.959 (1.105 <sup>2</sup> )	5.085 (1.222 <sup>2</sup> )	4.857 (1.129 <sup>2</sup> )	4.515 (0.987 <sup>2</sup> )
	1	2.233 (0.689 <sup>2</sup> )	1.565 (0.575 <sup>2</sup> )	2.263 (0.691 <sup>2</sup> )	2.399 (0.726 <sup>2</sup> )	2.107 (0.699 <sup>2</sup> )	1.909 (0.709 <sup>2</sup> )
	2	0.257 (0.119 <sup>2</sup> )	0.214 (0.142 <sup>2</sup> )	0.26 (0.141 <sup>2</sup> )	0.291 (0.17 <sup>2</sup> )	0.273 (0.147 <sup>2</sup> )	0.229 (0.142 <sup>2</sup> )
1000	-2	0.357 (0.166 <sup>2</sup> )	0.24 (0.134 <sup>2</sup> )	0.341 (0.142 <sup>2</sup> )	0.353 (0.162 <sup>2</sup> )	0.322 (0.138 <sup>2</sup> )	0.243 (0.115 <sup>2</sup> )
	-1	2.889 (0.771 <sup>2</sup> )	2.012 (0.661 <sup>2</sup> )	2.995 (0.783 <sup>2</sup> )	2.995 (0.869 <sup>2</sup> )	2.77 (0.829 <sup>2</sup> )	2.495 (0.745 <sup>2</sup> )
	0	6.633 (1.198 <sup>2</sup> )	4.067 (0.92 <sup>2</sup> )	6.415 (1.251 <sup>2</sup> )	6.44 (1.207 <sup>2</sup> )	6.443 (1.235 <sup>2</sup> )	5.978 (1.369 <sup>2</sup> )
	1	2.841 (0.871 <sup>2</sup> )	2.046 (0.796 <sup>2</sup> )	2.915 (0.893 <sup>2</sup> )	2.874 (0.781 <sup>2</sup> )	2.835 (0.836 <sup>2</sup> )	2.259 (0.74 <sup>2</sup> )
	2	0.358 (0.162 <sup>2</sup> )	0.251 (0.125 <sup>2</sup> )	0.334 (0.163 <sup>2</sup> )	0.354 (0.169 <sup>2</sup> )	0.351 (0.194 <sup>2</sup> )	0.308 (0.222 <sup>2</sup> )

表 3.6 容易な項目が多いアイテムバンクを用いた各手法の平均テスト情報量 (標準偏差) 分析

時間 [s]	真の能力値	提案手法	時間のみ制約とした 適応型テスト [24][38]	提案手法 2	提案手法 3	提案手法 4	提案手法 5
400	-2	1.618 (0.864 <sup>2</sup> )	1.437 (0.777 <sup>2</sup> )	1.792 (0.758 <sup>2</sup> )	1.94 (0.76 <sup>2</sup> )	1.881 (0.739 <sup>2</sup> )	1.98 (0.789 <sup>2</sup> )
	-1	0.853 (0.465 <sup>2</sup> )	0.837 (0.483 <sup>2</sup> )	0.749 (0.45 <sup>2</sup> )	0.744 (0.475 <sup>2</sup> )	0.91 (0.549 <sup>2</sup> )	0.771 (0.463 <sup>2</sup> )
	0	0.115 (0.132 <sup>2</sup> )	0.137 (0.153 <sup>2</sup> )	0.096 (0.092 <sup>2</sup> )	0.076 (0.073 <sup>2</sup> )	0.093 (0.099 <sup>2</sup> )	0.082 (0.078 <sup>2</sup> )
	1	0.013 (0.011 <sup>2</sup> )	0.01 (0.008 <sup>2</sup> )	0.007 (0.009 <sup>2</sup> )	0.009 (0.01 <sup>2</sup> )	0.009 (0.009 <sup>2</sup> )	0.008 (0.009 <sup>2</sup> )
	2	0.002 (0.002 <sup>2</sup> )	0.002 (0.002 <sup>2</sup> )	0.001 (0.002 <sup>2</sup> )	0.001 (0.002 <sup>2</sup> )	0.001 (0.003 <sup>2</sup> )	0.001 (0.002 <sup>2</sup> )
600	-2	2.853 (1.122 <sup>2</sup> )	1.752 (1.025 <sup>2</sup> )	3.481 (1.173 <sup>2</sup> )	3.259 (0.986 <sup>2</sup> )	3.394 (1.079 <sup>2</sup> )	3.005 (0.927 <sup>2</sup> )
	-1	1.569 (0.591 <sup>2</sup> )	1.406 (0.693 <sup>2</sup> )	1.402 (0.567 <sup>2</sup> )	1.433 (0.749 <sup>2</sup> )	1.313 (0.614 <sup>2</sup> )	1.478 (0.657 <sup>2</sup> )
	0	0.199 (0.113 <sup>2</sup> )	0.202 (0.099 <sup>2</sup> )	0.175 (0.102 <sup>2</sup> )	0.153 (0.127 <sup>2</sup> )	0.153 (0.104 <sup>2</sup> )	0.144 (0.103 <sup>2</sup> )
	1	0.021 (0.014 <sup>2</sup> )	0.02 (0.014 <sup>2</sup> )	0.019 (0.015 <sup>2</sup> )	0.017 (0.014 <sup>2</sup> )	0.016 (0.011 <sup>2</sup> )	0.012 (0.01 <sup>2</sup> )
	2	0.003 (0.003 <sup>2</sup> )	0.003 (0.004 <sup>2</sup> )	0.002 (0.002 <sup>2</sup> )	0.002 (0.003 <sup>2</sup> )	0.002 (0.003 <sup>2</sup> )	0.002 (0.002 <sup>2</sup> )
800	-2	4.666 (1.328 <sup>2</sup> )	2.415 (1.015 <sup>2</sup> )	4.601 (1.168 <sup>2</sup> )	4.963 (1.265 <sup>2</sup> )	5.06 (1.179 <sup>2</sup> )	4.796 (1.164 <sup>2</sup> )
	-1	2.219 (0.723 <sup>2</sup> )	1.692 (0.684 <sup>2</sup> )	1.962 (0.686 <sup>2</sup> )	2.052 (0.738 <sup>2</sup> )	2.035 (0.714 <sup>2</sup> )	1.812 (0.609 <sup>2</sup> )
	0	0.298 (0.136 <sup>2</sup> )	0.269 (0.163 <sup>2</sup> )	0.286 (0.142 <sup>2</sup> )	0.241 (0.14 <sup>2</sup> )	0.275 (0.179 <sup>2</sup> )	0.215 (0.132 <sup>2</sup> )
	1	0.03 (0.018 <sup>2</sup> )	0.029 (0.015 <sup>2</sup> )	0.026 (0.018 <sup>2</sup> )	0.024 (0.016 <sup>2</sup> )	0.028 (0.02 <sup>2</sup> )	0.021 (0.015 <sup>2</sup> )
	2	0.004 (0.003 <sup>2</sup> )	0.004 (0.004 <sup>2</sup> )	0.004 (0.003 <sup>2</sup> )	0.003 (0.003 <sup>2</sup> )	0.004 (0.004 <sup>2</sup> )	0.003 (0.003 <sup>2</sup> )
1000	-2	6.183 (1.257 <sup>2</sup> )	3.311 (1.064 <sup>2</sup> )	6.34 (1.18 <sup>2</sup> )	6.407 (1.285 <sup>2</sup> )	6.558 (1.241 <sup>2</sup> )	5.98 (1.252 <sup>2</sup> )
	-1	2.744 (0.776 <sup>2</sup> )	2.214 (0.852 <sup>2</sup> )	2.956 (0.932 <sup>2</sup> )	2.758 (0.954 <sup>2</sup> )	2.819 (0.951 <sup>2</sup> )	2.32 (0.818 <sup>2</sup> )
	0	0.355 (0.181 <sup>2</sup> )	0.274 (0.124 <sup>2</sup> )	0.345 (0.167 <sup>2</sup> )	0.35 (0.179 <sup>2</sup> )	0.351 (0.172 <sup>2</sup> )	0.272 (0.15 <sup>2</sup> )
	1	0.033 (0.017 <sup>2</sup> )	0.033 (0.022 <sup>2</sup> )	0.034 (0.019 <sup>2</sup> )	0.037 (0.02 <sup>2</sup> )	0.033 (0.022 <sup>2</sup> )	0.03 (0.019 <sup>2</sup> )
	2	0.005 (0.004 <sup>2</sup> )	0.004 (0.003 <sup>2</sup> )	0.005 (0.004 <sup>2</sup> )	0.004 (0.003 <sup>2</sup> )	0.004 (0.004 <sup>2</sup> )	0.003 (0.003 <sup>2</sup> )



表 3.7 難しい項目が多いアイテムバンクを用いた各手法の平均テスト情報量 (標準偏差) 分析

時間 [s]	真の能力値	提案手法	時間のみ制約とした 適応型テスト [24][38]	提案手法 2	提案手法 3	提案手法 4	提案手法 5
400	-2.0	0.001 (0.002 <sup>2</sup> )	0.002 (0.003 <sup>2</sup> )	0.002 (0.003 <sup>2</sup> )	0.001 (0.001 <sup>2</sup> )	0.001 (0.003 <sup>2</sup> )	0.001 (0.002 <sup>2</sup> )
	-1.0	0.011 (0.013 <sup>2</sup> )	0.013 (0.012 <sup>2</sup> )	0.011 (0.01 <sup>2</sup> )	0.011 (0.014 <sup>2</sup> )	0.008 (0.008 <sup>2</sup> )	0.009 (0.008 <sup>2</sup> )
	0.0	0.118 (0.175 <sup>2</sup> )	0.11 (0.086 <sup>2</sup> )	0.117 (0.095 <sup>2</sup> )	0.079 (0.062 <sup>2</sup> )	0.089 (0.099 <sup>2</sup> )	0.073 (0.065 <sup>2</sup> )
	1.0	0.863 (0.462 <sup>2</sup> )	0.898 (0.511 <sup>2</sup> )	0.824 (0.511 <sup>2</sup> )	0.881 (0.525 <sup>2</sup> )	0.874 (0.545 <sup>2</sup> )	0.834 (0.523 <sup>2</sup> )
	2.0	1.555 (0.839 <sup>2</sup> )	1.221 (0.65 <sup>2</sup> )	1.889 (0.81 <sup>2</sup> )	1.945 (0.859 <sup>2</sup> )	1.816 (0.779 <sup>2</sup> )	2.002 (0.933 <sup>2</sup> )
600	-2.0	0.003 (0.003 <sup>2</sup> )	0.002 (0.003 <sup>2</sup> )	0.003 (0.003 <sup>2</sup> )	0.002 (0.002 <sup>2</sup> )	0.002 (0.004 <sup>2</sup> )	0.002 (0.003 <sup>2</sup> )
	-1.0	0.021 (0.014 <sup>2</sup> )	0.023 (0.015 <sup>2</sup> )	0.023 (0.017 <sup>2</sup> )	0.018 (0.014 <sup>2</sup> )	0.015 (0.012 <sup>2</sup> )	0.012 (0.01 <sup>2</sup> )
	0.0	0.212 (0.107 <sup>2</sup> )	0.193 (0.135 <sup>2</sup> )	0.231 (0.135 <sup>2</sup> )	0.176 (0.102 <sup>2</sup> )	0.154 (0.128 <sup>2</sup> )	0.175 (0.178 <sup>2</sup> )
	1.0	1.495 (0.577 <sup>2</sup> )	1.211 (0.505 <sup>2</sup> )	1.574 (0.793 <sup>2</sup> )	1.352 (0.614 <sup>2</sup> )	1.316 (0.684 <sup>2</sup> )	1.321 (0.692 <sup>2</sup> )
	2.0	3.066 (1.089 <sup>2</sup> )	1.779 (0.84 <sup>2</sup> )	3.277 (1.199 <sup>2</sup> )	3.263 (0.952 <sup>2</sup> )	3.122 (1.015 <sup>2</sup> )	3.151 (1.056 <sup>2</sup> )
800	-2.0	0.004 (0.004 <sup>2</sup> )	0.004 (0.004 <sup>2</sup> )	0.004 (0.003 <sup>2</sup> )	0.004 (0.003 <sup>2</sup> )	0.003 (0.003 <sup>2</sup> )	0.003 (0.003 <sup>2</sup> )
	-1.0	0.032 (0.019 <sup>2</sup> )	0.032 (0.02 <sup>2</sup> )	0.031 (0.019 <sup>2</sup> )	0.031 (0.02 <sup>2</sup> )	0.025 (0.019 <sup>2</sup> )	0.019 (0.013 <sup>2</sup> )
	0.0	0.302 (0.184 <sup>2</sup> )	0.246 (0.129 <sup>2</sup> )	0.288 (0.139 <sup>2</sup> )	0.277 (0.153 <sup>2</sup> )	0.237 (0.134 <sup>2</sup> )	0.199 (0.129 <sup>2</sup> )
	1.0	2.216 (0.752 <sup>2</sup> )	1.791 (0.694 <sup>2</sup> )	2.298 (0.815 <sup>2</sup> )	2.258 (0.79 <sup>2</sup> )	2.225 (0.861 <sup>2</sup> )	1.861 (0.669 <sup>2</sup> )
	2.0	4.426 (1.29 <sup>2</sup> )	2.474 (1.037 <sup>2</sup> )	4.804 (1.324 <sup>2</sup> )	4.888 (1.151 <sup>2</sup> )	4.843 (1.259 <sup>2</sup> )	4.549 (1.064 <sup>2</sup> )
1000	-2.0	0.005 (0.003 <sup>2</sup> )	0.004 (0.003 <sup>2</sup> )	0.005 (0.003 <sup>2</sup> )	0.005 (0.004 <sup>2</sup> )	0.004 (0.003 <sup>2</sup> )	0.003 (0.002 <sup>2</sup> )
	-1.0	0.034 (0.018 <sup>2</sup> )	0.029 (0.016 <sup>2</sup> )	0.036 (0.019 <sup>2</sup> )	0.032 (0.016 <sup>2</sup> )	0.036 (0.018 <sup>2</sup> )	0.024 (0.015 <sup>2</sup> )
	0.0	0.326 (0.111 <sup>2</sup> )	0.288 (0.142 <sup>2</sup> )	0.382 (0.17 <sup>2</sup> )	0.348 (0.152 <sup>2</sup> )	0.359 (0.181 <sup>2</sup> )	0.268 (0.133 <sup>2</sup> )
	1.0	2.806 (0.707 <sup>2</sup> )	2.221 (0.696 <sup>2</sup> )	2.894 (0.821 <sup>2</sup> )	2.745 (0.807 <sup>2</sup> )	2.817 (0.847 <sup>2</sup> )	2.485 (0.883 <sup>2</sup> )
	2.0	6.075 (1.445 <sup>2</sup> )	3.212 (1.043 <sup>2</sup> )	6.038 (1.185 <sup>2</sup> )	6.512 (0.983 <sup>2</sup> )	6.752 (1.228 <sup>2</sup> )	6.211 (1.296 <sup>2</sup> )

表 3.8 特定の一か所に項目が集まったアイテムバンクを用いた各手法の平均テスト情報量（標準偏差）分析

時間 [s]	真の能力値	提案手法	時間のみ制約とした 適応型テスト [24][38]	提案手法 2	提案手法 3	提案手法 4	提案手法 5
300	-2	0.017 (0.042 <sup>2</sup> )	0.011 (0.022 <sup>2</sup> )	0.02 (0.037 <sup>2</sup> )	0.021 (0.033 <sup>2</sup> )	0.026 (0.071 <sup>2</sup> )	0.021 (0.028 <sup>2</sup> )
	-1	0.134 (0.204 <sup>2</sup> )	0.163 (0.223 <sup>2</sup> )	0.22 (0.252 <sup>2</sup> )	0.217 (0.327 <sup>2</sup> )	0.192 (0.21 <sup>2</sup> )	0.195 (0.236 <sup>2</sup> )
	0	0.554 (0.453 <sup>2</sup> )	0.527 (0.448 <sup>2</sup> )	0.429 (0.417 <sup>2</sup> )	0.466 (0.412 <sup>2</sup> )	0.478 (0.44 <sup>2</sup> )	0.558 (0.443 <sup>2</sup> )
	1	0.17 (0.203 <sup>2</sup> )	0.137 (0.17 <sup>2</sup> )	0.193 (0.227 <sup>2</sup> )	0.176 (0.219 <sup>2</sup> )	0.22 (0.264 <sup>2</sup> )	0.237 (0.308 <sup>2</sup> )
	2	0.015 (0.026 <sup>2</sup> )	0.011 (0.017 <sup>2</sup> )	0.025 (0.05 <sup>2</sup> )	0.025 (0.058 <sup>2</sup> )	0.029 (0.05 <sup>2</sup> )	0.02 (0.035 <sup>2</sup> )
400	-2	0.086 (0.076 <sup>2</sup> )	0.061 (0.079 <sup>2</sup> )	0.098 (0.083 <sup>2</sup> )	0.074 (0.069 <sup>2</sup> )	0.075 (0.081 <sup>2</sup> )	0.088 (0.104 <sup>2</sup> )
	-1	0.815 (0.481 <sup>2</sup> )	0.68 (0.438 <sup>2</sup> )	0.853 (0.462 <sup>2</sup> )	0.713 (0.446 <sup>2</sup> )	0.821 (0.49 <sup>2</sup> )	0.768 (0.464 <sup>2</sup> )
	0	1.762 (0.725 <sup>2</sup> )	1.861 (0.588 <sup>2</sup> )	1.728 (0.716 <sup>2</sup> )	1.754 (0.738 <sup>2</sup> )	1.945 (0.724 <sup>2</sup> )	1.832 (0.711 <sup>2</sup> )
	1	0.729 (0.406 <sup>2</sup> )	0.659 (0.459 <sup>2</sup> )	0.822 (0.467 <sup>2</sup> )	0.837 (0.487 <sup>2</sup> )	0.815 (0.49 <sup>2</sup> )	0.806 (0.549 <sup>2</sup> )
	2	0.075 (0.069 <sup>2</sup> )	0.076 (0.091 <sup>2</sup> )	0.089 (0.089 <sup>2</sup> )	0.084 (0.083 <sup>2</sup> )	0.088 (0.09 <sup>2</sup> )	0.091 (0.1 <sup>2</sup> )
500	-2	0.215 (0.145 <sup>2</sup> )	0.141 (0.097 <sup>2</sup> )	0.215 (0.14 <sup>2</sup> )	0.206 (0.138 <sup>2</sup> )	0.184 (0.155 <sup>2</sup> )	0.149 (0.087 <sup>2</sup> )
	-1	1.671 (0.563 <sup>2</sup> )	1.434 (0.57 <sup>2</sup> )	1.741 (0.645 <sup>2</sup> )	1.737 (0.639 <sup>2</sup> )	1.549 (0.659 <sup>2</sup> )	1.549 (0.65 <sup>2</sup> )
	0	3.648 (0.994 <sup>2</sup> )	2.833 (0.774 <sup>2</sup> )	3.43 (0.966 <sup>2</sup> )	3.875 (0.958 <sup>2</sup> )	3.614 (1.036 <sup>2</sup> )	3.746 (0.988 <sup>2</sup> )
	1	1.685 (0.699 <sup>2</sup> )	1.475 (0.602 <sup>2</sup> )	1.658 (0.657 <sup>2</sup> )	1.794 (0.782 <sup>2</sup> )	1.608 (0.745 <sup>2</sup> )	1.545 (0.674 <sup>2</sup> )
	2	0.208 (0.141 <sup>2</sup> )	0.14 (0.104 <sup>2</sup> )	0.209 (0.129 <sup>2</sup> )	0.193 (0.119 <sup>2</sup> )	0.172 (0.118 <sup>2</sup> )	0.17 (0.116 <sup>2</sup> )
600	-2	0.34 (0.21 <sup>2</sup> )	0.229 (0.126 <sup>2</sup> )	0.338 (0.175 <sup>2</sup> )	0.3 (0.118 <sup>2</sup> )	0.296 (0.164 <sup>2</sup> )	0.221 (0.124 <sup>2</sup> )
	-1	2.534 (0.726 <sup>2</sup> )	1.918 (0.68 <sup>2</sup> )	2.548 (0.842 <sup>2</sup> )	2.382 (0.682 <sup>2</sup> )	2.516 (0.727 <sup>2</sup> )	2.195 (0.801 <sup>2</sup> )
	0	5.478 (1.185 <sup>2</sup> )	3.878 (1.064 <sup>2</sup> )	5.276 (1.091 <sup>2</sup> )	5.619 (1.168 <sup>2</sup> )	5.269 (0.98 <sup>2</sup> )	5.378 (1.195 <sup>2</sup> )
	1	2.498 (0.722 <sup>2</sup> )	1.901 (0.711 <sup>2</sup> )	2.486 (0.675 <sup>2</sup> )	2.557 (0.723 <sup>2</sup> )	2.397 (0.765 <sup>2</sup> )	2.146 (0.8 <sup>2</sup> )
	2	0.308 (0.155 <sup>2</sup> )	0.244 (0.137 <sup>2</sup> )	0.314 (0.151 <sup>2</sup> )	0.318 (0.19 <sup>2</sup> )	0.295 (0.132 <sup>2</sup> )	0.245 (0.156 <sup>2</sup> )

表 3.9 実験の結果 (\*\* 有意水準 1% の有意差, \* 有意水準 5% の有意差)

	提案手法	時間のみ制約とした適応型テスト [24][38]
項目情報量**	4.24(0.486 <sup>2</sup> )	2.03(0.503 <sup>2</sup> )
移動時間 [秒]*	110(28.0 <sup>2</sup> )	238(94.0 <sup>2</sup> )
移動回数*	4.0(0.707 <sup>2</sup> )	5.8(1.30 <sup>2</sup> )
解答された項目数**	18.2(3.70 <sup>2</sup> )	9.0(1.41 <sup>2</sup> )
アンケート (1)**	4.8(0.447 <sup>2</sup> )	3.4(0.548 <sup>2</sup> )
アンケート (2)*	4.6(0.548 <sup>2</sup> )	3.2(0.837 <sup>2</sup> )

## 第4章

### 結言

本研究では、状況に埋め込まれた学習を効率的に評価するモバイル・テストイング・システムを開発した。

第2章では、テスト理論に基づいて適応型テストを用いてモバイル・テストイング・システムを開発した。本システムは、適応型テストを用いたテストイング機能と展示物がある場所まで受検者を誘導するナビゲーション機能を持つ。本システムにより期待される利点は、(1) 受検者の知識状態を効率的に測定できる、(2) 展示物がある場所まで受検者を短時間で誘導できる、(3) 状況に埋め込まれた学習そのものを評価することができる。具体的には、観察や探索を通して受検者に項目を解答させることができる、である。ここでは、受検者が観察や探索をしているかを検証するため、眼球運動計測装置を用いる。システムの有効性を評価するため、シミュレーション実験と東京都調布市の深大寺で被験者実験を実施した。まず、受検者の能力値が効率的に推定されているかを評価するため、テストイング機能を用いてシミュレーション実験を実施した。シミュレーション実験の結果、初期の能力推定値は受検者の真の能力値と乖離しているが、受検者の解答数が増えると真の能力値に収束していた。累積項目情報量も増加しているため、能力推定誤差が減少していることが確認された。次に、受検者を展示物へ短時間で誘導できるかを評価するため、被験者実験を実施した。被験者実験より、紙媒体のマップと比較し、被験者が短時間で展示物へ誘導されていることを確認した。最後に、受検者が観察や探索を通し

て項目に解答しているか評価するため、被験者実験を実施した。仮想環境において眼球運動計測装置を用いた実験、及び、深大寺での被験者実験の結果から展示物に対して観察や探索を通してテストを受けていることが確認された。これらの評価実験から、システムの有効性を確認した。

第3章では、状況に埋め込まれた学習の評価のため、移動距離とテスト情報量の最適化を組み込んだ適応型テストを提案する。具体的には、テスト情報量の最大化と移動距離の最小化を同時に満たす項目選択のため、最適化問題の一つである Traveling Purchaser Problem (TPP) を組み込んだ適応型テストを提案した。TPP は、複数の商品が売られている店舗が点在するとき、商品の価格と移動距離が最小なパスを探索する最適化問題である。本研究では、TPP における商品をテスト項目、店舗を項目が出題される場所とみなすことで、テスト情報量と解答所要時間、出題順序を最適化できる適応型テストを提案した。TPP は、線形計画法により最適解を得ることができず、全探索が必要となり、店舗数を  $n$  とすると計算量が  $O(n!)$  になる。しかし、TPP は、動的計画法を用いることにより計算量を  $O(2n \cdot 2^n)$  に減少できることが知られている [34]。動的計画法は、計算量を減少させるため、最適化問題を複数の部分問題に分割し、その部分問題の結果を次の部分問題の計算に使う手法である [35]。しかし、TPP では目的関数が単調増加性を持たなければならず、本研究の場合、目的関数にテスト情報量の最大化と移動距離の最適化をいかに組み込むかが問題になる。そこで、本論では、移動時間に対して極力小さい重みを掛けたペナルティ項をテスト情報量に付与した目的関数を提案する。これにより、TPP の出題項目に対する単調増加関数として移動距離を最小にしながら情報量を最大化できる。さらに、TPP の店舗数  $n$  とモバイル・テストの項目数  $I$  を  $n = I$  とし、従来  $O(I!)$  である計算量を  $O(2I \cdot 2^I)$  に軽減できる。提案手法の利点は

以下の通りである。(1) 移動経路が最適化されるため、テストの解答所要時間に対する移動時間の割合が減少する。(2) 移動時間の減少により受検者が解答できる項目数が増加する。(3) 情報量が高い項目を出題でき、高精度な能力推定が期待できる。提案手法の有効性を評価するため、シミュレーション実験と被験者実験を行った。シミュレーション実験の結果、提案手法は、一般的な適応型テスト、ランダム出題より、能力推定精度が高かった。しかし、制限時間が短いとき、時間を制約とした適応型テストより能力推定精度が低かった。この原因は、コールドスタート問題が生じていたからであった。そこで、本研究では、初期の項目は受検者の最も近くにある項目を出題するというルールを提案手法に加えた。その結果、コールドスタート問題の影響が軽減され、能力推定精度が向上した。被験者実験では、提案手法を時間のみを制約とした適応型テストと比較した。その結果、提案手法は、移動回数が少なく、出題された項目数が多く、移動プロセスが最適化されていることが確認できた。また、提案手法はテスト情報量が高く、能力推定精度が高いことも確認した。

今後、出題順序を制約とした適応型テストとして一般化し、適用範囲を広げていきたい。

## 参考文献

- [1] Willard Van Orman Quine. *Word and Object (Studies in Communication)*. The MIT Press, first edition edition, 3 1964.
- [2] Ludwig Wittgenstein. *Philosophical Investigations*. Wiley-Blackwell (an imprint of, 4th revised edition edition, 2009).
- [3] Chuang-Kai Chiou, Judy C. R. Tseng, Gwo-Jen Hwang, and Shelly Heller. An adaptive navigation support system for conducting context-aware ubiquitous learning in museums. *Computers & Education*, Vol. 55, No. 2, pp. 834–845, SEP 2010.
- [4] Hui-Chun Chu, Gwo-Jen Hwang, and Chin-Chung Tsai. A knowledge engineering approach to developing mindtools for context-aware ubiquitous learning. *Computers & Education*, Vol. 54, No. 1, pp. 289–297, JAN 2010.
- [5] Y.S. Chen, T.C. Kao, and J.P. Sheu. A mobile learning system for scaffolding bird watching learning. *Journal of Computer Assisted Learning*, Vol. 19, pp. 347–59, Sept. 2003.
- [6] Gwo-Jen Hwang, Chih-Hsiang Wu, Judy C. R. Tseng, and Iwen Huang. Development of a ubiquitous learning platform based on a real-time help-seeking mechanism. *British Journal of Educational Technology*, Vol. 42, No. 6, pp. 992–1002, 2011.
- [7] H. Ogata, Y. Matsuka, M.M. El-Bishouty, and Y. Yano. LORAMS: linking physical objects and videos for capturing and sharing learn-

- ing experiences towards ubiquitous learning. *International Journal of Mobile Learning and Organisation*, Vol. 3, pp. 337–50, 2009 2009.
- [8] Gwo-Jen Hwang, Tzu-Chi Yang, Chin-Chung Tsai, and Stephen JH Yang. A context-aware ubiquitous learning environment for conducting complex science experiments. *Computers & Education*, Vol. 53, No. 2, pp. 402–413, 2009.
- [9] Hui-Chun Chu, Gwo-Jen Hwang, Chin-Chung Tsai, and Judy C.R. Tseng. A two-tier test approach to developing location-aware mobile learning systems for natural science courses. *Computers & Education*, Vol. 55, No. 4, pp. 1618 – 1627, 2010.
- [10] Patricia Rodríguez Santos, Mar Pérez-Sanagustín, Davinia Hernández-Leo, and Josep Blat. Questinsitu: From tests to routes for assessment in situ activities. *Computers & Education*, Vol. 57, No. 4, pp. 2517 – 2534, 2011.
- [11] Shu-Hsien Huang, Ting-Ting Wu, Hui-Chun Chu, and Gwo-Jen Hwang. A decision tree approach to conducting dynamic assessment in a context-aware ubiquitous learning environment. In *Wireless, Mobile, and Ubiquitous Technology in Education, 2008. WMUTE 2008. Fifth IEEE International Conference on*, pp. 89 –94, march 2008.
- [12] Wim J. van der Linden and Cees A.W. Glas. *Computerized Adaptive Testing: Theory and Practice*. Springer, 1 edition, 7 2000.
- [13] Craig N. Mills, Maria T. Potenza, John J. Fremer, and William C. Ward, editors. *Computer-Based Testing: Building the Foundation for Future Assessments*. Routledge, 4 2002.



- 
- [14] Sandra A. McIntire and Leslie A. Miller. *Foundations of Psychological Testing: A Practical Approach*. Sage Publications, Inc., 2nd edition, 11 2006.
- [15] Dato N. M. de Gruijter and Leo J. Th. van der Kamp. *Statistical Test Theory for the Behavioral Sciences (Chapman & Hall CRC Statistics in the Social and Behavioral Sciences)*. Chapman and Hall CRC, 1 edition, 8 2007.
- [16] Theresa J.B. Kline. *Psychological Testing: A Practical Approach to Design and Evaluation*. Sage Publications, Inc., 1 edition, 2 2005.
- [17] Keizo Nagaoka and Maomi Ueno. An algorithm for a multidimensional computerized adaptive testing system. *Japan Journal of Educational Technology*, Vol. 15, No. 4, pp. 157–166, mar 1992.
- [18] 植野真臣, 永岡慶三. e テスティング. 培風館, 2009.
- [19] Frank B. Baker and Seock-Ho Kim, editors. *Item Response Theory: Parameter Estimation Techniques, Second Edition (Statistics: A Series of Textbooks and Monographs)*. CRC Press, 2 edition, 7 2004.
- [20] 豊田秀樹. 項目反応理論 [入門編]—テストの測定と科学. 朝倉書店, 2002.
- [21] 豊田秀樹. 項目反応理論 [理論編]—テストの数理. 朝倉書店, 2005.
- [22] A. Birnbaum. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, editors, *Statistical theories of mental test scores*, pp. 397–479. Addison-Wesley, 1968.
- [23] J Mislevy, R. Bayes modal estimation in item response models. *Psychometrika*, Vol. 51, No. 2, pp. 177–195, 1986.
- [24] Wim J. van der Linden. *Linear Models for Optimal Test Design (Statis-*

- tics for Social and Behavioral Sciences) [Hardcover]. Springer, 2005.*
- [25] Wim J. van der Linden. Test design and speededness. *Journal of Educational Measurement*, Vol. 48, No. 1, pp. 44–60, 2011.
- [26] Wim J. van der Linden and Xinhui Xiong. Speededness and adaptive testing. *Journal of Educational and Behavioral Statistics*, 2013.
- [27] T. Ramesh. Travelling purchaser problem. *Opsearch*, Vol. 18, pp. 78–91, June 1981.
- [28] WL Pearn and RC Chien. Improved solutions for the traveling purchaser problem. *Computers & Operations Research*, Vol. 25, No. 11, pp. 879–885, NOV 1998.
- [29] Hoon Liong Ong. Approximate algorithms for the travelling purchaser problem. *Operations Research Letters*, Vol. 1, pp. 201–5, Nov. 1982.
- [30] FF Boctor, G Laporte, and J Renaud. Heuristics for the traveling purchaser problem. *Computers & Operations Research*, Vol. 30, No. 4, pp. 491–504, APR 2003.
- [31] Boris Bontoux and Dominique Feillet. Ant colony optimization for the traveling purchaser problem. *Computers & Operations Research*, Vol. 35, No. 2, pp. 628–637, FEB 2008.
- [32] KN Singh and DL van Oudheusden. A branch and bound algorithm for the traveling purchaser problem. *European Journal of Operational Research*, Vol. 97, No. 3, pp. 571–579, MAR 16 1997.
- [33] G Laporte, J Riera-Ledesma, and JJ Salazar-Gonzalez. A branch-and-cut algorithm for the undirected traveling purchaser problem. *Operations Research*, Vol. 51, No. 6, pp. 940–951, NOV-DEC 2003.

- [34] Luis Gouveia, Ana Paias, and Stefan Voß. A Dynamic Programming Approach for a Travelling Purchaser Problem With Additional Constraints The Dynamic Programming Approach. In *INOC*, pp. 1–6, 2009.
- [35] Richard Bellman. *Dynamic Programming (Dover Books on Computer Science)*. Dover Publications, reprint edition, 2003.
- [36] Wim J. van der Linden. A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, Vol. 31, No. 2, pp. 181–204, 2006.
- [37] 山本芳嗣, 久保幹雄. 巡回セールスマン問題への招待 (シリーズ「現代人の数理」). 朝倉書店, 1997.
- [38] Wim J. van der Linden, David J. Scrams, and Deborah L. Schnipke. Using response-time constraints to control for differential speededness in computerized adaptive testing. *Applied Psychological Measurement*, Vol. 23, No. 3, pp. 195–210, 1999.

## 謝辞

本研究を進めるにあたり、終始懇切なる御指導を賜った、電気通信大学大学院教授の植野真臣先生に、心より感謝を申し上げます。本論文の審査過程において、数々の貴重な御助言と御指導を賜りました大須賀昭彦教授、栗原聡教授、田原康之准教授、笠井裕之准教授に、深謝申し上げます。また、本研究における議論・検討に当たって、ご教示とご激励を頂いた電気通信大学大学院植野真臣研究室の皆様に御礼申し上げます。

## 関連論文の印刷公表の方法及び時期

### 査読付き論文（本学位申請論文関連論文）

宮澤芳光、植野真臣、「適応型テストを用いた携帯型観光・学習ナビゲーションシステム」、教育システム情報学会、Vol. 29、No.2、[110-123] (2012)

### 国際会議

Yoshimitsu Miyasawa, Maomi Ueno, "Mobile testing for authentic assessment in a field", Artificial Intelligence in Education (AIED 2013), (2013)

Yoshimitsu Miyasawa, Maomi Ueno, "Mobile testing for authentic assessment in a field: Evaluation from Actual Performances", IEEE Region 10 Humanitarian Technology Conference, (2013)

### その他（研究会等）

宮澤芳光、植野真臣、「携帯電話を用いた適応型クイズによる観光ナビゲーション」、観光情報学会第1回研究発表会講演論文集、[1-6]、(2009)

宮澤芳光、植野真臣、「携帯電話を用いた適応型クイズによる観光ナビゲーション」、教育システム情報学会研究報告、[46-51]、(2010)

宮澤芳光、宇都雅輝、安藤雅洋、植野真臣、「現実的場面での評価を支援する携帯型e-テストング」、日本教育工学会 研究報告集、[99-104]、(2010)

宮澤芳光, 宇都雅輝, 安藤雅洋, 植野真臣, 「現実的場面での評価を支援する携帯型 e-テストング」, 教育システム情報学会第 35 回全国大会論文集, [107-108], (2010)

宮澤芳光, 宇都雅輝, 安藤雅洋, 植野真臣, 「IRT に基づく観光ナビゲーション・システムの開発と評価」, 日本テスト学会第 8 回発表論文抄録集, [96-99], (2010)

宮澤芳光, 宇都雅輝, 安藤雅洋, 植野真臣, 「携帯電話を用いた適応型クイズによる観光ナビゲーションシステムの開発と評価」, 日本教育工学会第 26 回全国大会論文集, [305-306], (2010)

宮澤芳光, 植野真臣, 「真正な評価のためのモバイル・テストング・システム」, 日本テスト学会第 9 回発表論文抄録集, [174-175], (2011)

宮澤芳光, 植野真臣, 「真正な評価のためのモバイル・テストング・システム」, 日本教育工学会第 27 回全国大会論文集, [233-234], (2011)

宮澤芳光, 植野真臣, 「真正な評価のためのモバイル・テストング・システム」, 情報コミュニケーション学会第 8 回研究会研究報告, [12-13], (2011)

宮澤芳光, 植野真臣, 「現実場面での評価を支援するモバイルテストングシステム」, 日本教育工学会第 28 回全国大会論文集, [289-290], (2012)

## 付録

表1 被験者実験に用いた項目

ID	a	b	c	ID	a	b	c
1	1	0.058534	0	16	1	-1.1137	0
2	1	1.494171	0	17	1	0.058534	0
3	1	-1.278424	0	18	1	0.976689	0
4	1	-0.934213	0	19	1	1.06708	0
5	1	-0.473193	0	20	1	0.586117	0
6	1	-1.41155	0	21	1	0.08288	0
7	1	0.441854	0	22	1	-1.41155	0
8	1	-0.03885	0	23	1	0.360787	0
9	1	-0.236208	0	24	1	1.167831	0
10	1	-0.364672	0	25	1	0.334369	0
11	1	0.934731	0	26	1	0.818181	0
12	1	1.282568	0	27	1	0.4144	0
13	1	-0.588707	0	28	1	-0.03885	0
14	1	0.894327	0	29	1	0.441854	0
15	1	-1.672104	0	30	1	1.116031	0

表 2 被験者実験に用いた項目

ID	a	b	c	ID	a	b	c
31	1	0.526806	0	51	1	0.256928	0
32	1	0.469567	0	52	1	-0.71484	0
33	1	0.181559	0	53	1	-0.78296	0
34	1	0.586117	0	54	1	0.747215	0
35	1	-0.71484	0	55	1	0.818181	0
36	1	0.616679	0	56	1	0.976689	0
38	1	-0.0878	0	57	1	-0.26133	0
39	1	0.680134	0	58	1	-0.68195	0
40	1	0.206423	0	59	1	0.498057	0
41	1	0.156695	0	60	1	0.976689	0
42	1	-0.55892	0	61	1	-0.85574	0
43	1	-0.97565	0	62	1	1.06708	0
44	1	-0.21109	0	63	1	0.526806	0
45	1	-0.0145	0	65	1	0.680134	0
46	1	-0.31261	0	66	1	0.498057	0
47	1	1.282568	0	67	1	0.08288	0
48	1	1.06708	0	68	1	0.855477	0
49	1	0.498057	0	69	1	0.30821	0
50	1	0.586117	0	70	1	0.058534	0



表 3 被験者実験に用いた項目

ID	a	b	c
71	1	-0.31261	0
72	1	1.282568	0
73	1	0.131831	0
74	1	0.680134	0
75	1	0.713286	0
76	1	-0.68195	0
77	1	0.058534	0
78	1	0.713286	0
79	1	-0.16136	0
81	1	0.282569	0
82	1	0.360787	0
83	1	-0.52991	0
84	1	0.78218	0
85	1	0.818181	0