

科学研究費助成事業（科学研究費補助金）研究成果報告書

平 2 4 年 6 月 1 日現在

機関番号：12612

研究種目：挑戦的萌芽研究

研究期間：2009～2011

課題番号：21650221

研究課題名（和文） 大規模型 e テスティング・システムの開発

研究課題名（英文） Development of an Large-scale Type e-Testing System

研究代表者

植野 真臣 (UENO MAOMI)

電気通信大学・大学院情報システム学研究科・准教授

研究者番号：50262316

研究成果の概要（和文）：本研究では、e テスティングの利点をすべて統合的に有し、さらに実用的な e テスティング・システムの開発を目的とする。具体的なシステムは、1. 項目作成支援システム、2. アイテム・バンク、3. テスト実施システム、4. テスト構成支援システム、5. テスト・データベース、6. データ分析システム、7. 適応型テストシステム、によって構成されている。統合的システムを開発することにより、1)過去に蓄積されたテスト・データが、各機能に一貫して自動的に配分され、テスト分析、項目分析、テスト構成支援、適応型テストなど様々な機能に利用することができる、2) 多様な機能を持ち、一つのシステムで多様な目的のテストを実施することができる、という利点を持つ。

研究成果の概要（英文）：The purpose of this study is to develop a practical e-testing system which is consistently designed to unify various functions of the traditional computer based testing systems. The system is consists of Item Authoring System, Item Bank, Test Delivery System, e-Testing Construction Support System, Test Database, Data Analysis System, and Adaptive Testing System. The advantage features of the integrative system are 1. The test data stored in the server is automatically divided into each function and utilized for test analysis, item analysis, test construction, and adaptive testing, and 2. The system has various functions, therefore is used for various test purposes.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,600,000	0	1,600,000
2010年度	700,000	0	700,000
2011年度	900,000	270,000	1,170,000
年度			
年度			
総計	3,200,000	270,000	3,470,000

研究分野：総合領域

科研費の分科・細目：科学教育・教育工学・教育工学

キーワード：e-ラーニング、e テスティング

1. 研究開始当初の背景

テスト理論研究では e テスティングで必要な完全情報量関数や制約条件付き最適化の研究が多くなされてきているが、すべて計算時間や実効性を無視した理論研究であり、実

際には未だに原始的な手法を用いざるを得なかった。しかしテスト理論のほとんどは良いテストを構成するための最適解探索問題として捉えることができ、本研究では問題によらない最適化探索問題の高速アルゴリズム

ムを適用し、実際に開発を行うことにより、これまで理論研究に留まっていたテスト理論の優れた多くの手法を実用化にまで至らせ、高機能な大規模 e テスティングの実用化を目指す。

e テスティングは項目反応理論を基礎とした適応型テストや等質複数テスト構成により、異なるテストを受検した時にも受検者を同一尺度上で評価できる利点が注目され普及してきた。しかし、すでに海外で e テスティング化されている語学テストや計算テスト、論理力テストなどとは異なり、一般のテストでは広範囲領域を持つことが多く、項目反応理論における一次元性を満たさず、e テスティングにおけるテスト情報量を最大化させる適応型テストや自動テスト構成を行った場合、類似の領域からのみ項目が選択されてしまうという問題がある。これらを防ぐためには理論的には「各領域から…問以上出題する」という制約を自動出題アルゴリズムに付与すればよく、数理計画法を用いて制約条件付きの最適化問題を解けばよいし、その他の制約条件も簡単に加えることができる。しかし、現実にはこの手法は計算時間がかかりすぎ適応型テストには向いていないし、一般のテスト構成でも項目数に対して計算量が指数爆発するために大規模な e テスティングでは実用化が難しい。さらに、e テスティングでは、テスト情報量として EVSI や大域情報量、完全ベイズ情報量などが最適な情報量関数（完全情報量と呼ばれる）であることが既に知られているが、これらの情報量はいまだに実用化には至らず、最適でない原始的なフィッシャー情報量を用いている場合が多い。これら計算時間の問題が大規模な国家試験などでの e テスティング化の大きな障害になっている。

2. 研究の目的

本研究では、前述の制約付き項目選択や完全情報量関数の計算について、大規模項目数でも実施可能なアルゴリズムを提案し、大規模 e テスティングの実用化を実現するだけでなく、これまで理論研究にとどまっていた多くのテスト理論を実用化に至らせることを目的とする。

3. 研究の方法

(1) テスト最適化の高速アルゴリズムを定式化する。具体的には以下のアルゴリズムを提案する。提案アルゴリズムの概要は以下の通りである。

- ① アイテム・バンクでは項目反応理論により推定された項目パラメータを持ち、それに基づきシミュレーションによって発生された受検者の能力 θ を N 個用意する。
- ② N 人の受検者について制約付き完全情報量

を最大化する項目を選択しながら、それぞれの段階で受検者の反応が正答(1)のときと誤答(0)のときの二分木を作成し、 θ の推定値が収束するまで木を作成する。一番子孫の木ノードはその受検者の θ の推定値となる。

- ③ 分枝限定法 (Branch and Bound) 法を用い、制約を満たさないことが既知の場合の木ノードの子孫ノードを候補ノードから消去する。
- ④ 構成された木構造を計算木のアルゴリズムに従って圧縮し、さらに高速化する。
- ⑤ 構築された構造をアイテム・バンクの各項目のメタ・データとして格納する。
- ⑥ 適応型テストの場合は、ルートから木と受検者の反応に従って項目出題する。自動テスト構成の場合、標準正規分布に従って θ を発生し、その θ と二分木に従って項目を選択し続けて、それを L 回繰り返して項目の出題頻度を数え上げる。その項目の出題頻度の順に構成したいテストの出題項目数まで項目を加えてテストを構成する。

(2) 提案アルゴリズムについて、エンジン・プログラムを開発し、それを実装した e テスティング・システムを開発する。

(3) さまざまな実践を行い本 e テスティング・システムの評価を行う。

4. 研究成果

(1) e テスティング・システム

本研究で開発した e テスティング・システムは、図 1 に示すモジュールによって構成されている。以下、各モジュールの概要を述べる。

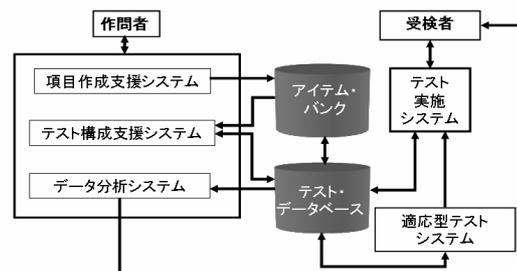


図 1 e テスティング・システム

① 項目作成支援システム

項目作成支援システムでは、動画や音声を含んだマルチメディア項目の作成が可能である。また、このシステムでは以下の項目出題形式を作成できる。

- 1) 真偽式項目
- 2) 多肢選択式項目
- 3) 並べ替え式項目
- 4) 単答式項目

5) 論述式項目

e テスティングにより、ビデオや音声を用いた質問項目を配信することができ、これまでペーパー・テストではこれまで測定することができなかった受検者の能力を測定することができるようになってきている。

採点処理について、「真偽式項目」、「多肢選択式項目」、「並べ替え式項目」は自動採点、「単答式項目」は半自動採点、「論述式項目」は手動採点によって処理する。ただし「論述式項目」では、配点は採点者が決定でき、採点はその配点の制約の中で自由な連続値を割り振ることができる。特に「単答式項目」では、アイテム・バンクに正答候補を複数入力しておき、受検者の回答がこの候補に一致していれば正答、無記入であれば誤答、いずれでもなければ作問者に手動で採点を求めるという半自動採点方式である。ただし、正答候補は以下を考慮して決定される。

- 1) 正答の別称もしくはミス・スペル
- 2) 回答の大文字や小文字の使用
- 3) 回答の同義語
- 4) 回答句読点
- 5) 否定形の使用や他言語の対応語
- 6) 数値での回答の単位の記入もしくは省略

② アイテム・バンク

作成された項目は、項目のデータベースであるアイテム・バンクに蓄積される。アイテム・バンクでは具体的には以下の情報が格納される(図2)。

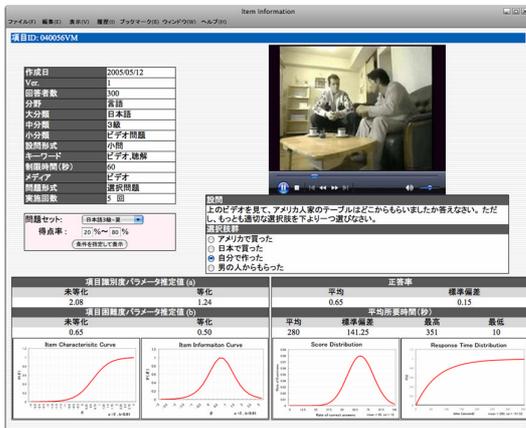


図2 アイテム・バンクにおける各質問項目の情報提示例

- 1) 細目表に対応した ID (評価目標)
- 2) 作問者
- 3) 作成日時
- 4) 項目コンテンツ
- 5) 正答 (複数可)
- 6) 解説
- 7) 正答率と標準偏差
- 8) 平均所要時間と標準偏差

9) 出題回数

10) 項目反応理論 (2 パラメータ・ロジスティック・モデル) におけるパラメータ推定値 (ベイズ推定値)

新規の項目は、1)-6)の情報のみを入力させ、テストを実施する度に 7)-9)の情報が自動的に更新、10)の情報が半自動的に更新される。

③ テスト実施システム

テスト実施システムは、テスト配信システムとも呼ばれ、指定された ID の項目を e テスティング・サーバーから各受検者の端末にインターネットを通じて配信するとともに、受検者からの回答や反応履歴 (回答、回答所要時間、回答の書き直し履歴、見直し履歴) を獲得し、サーバーに格納する機能を持つ。通常のテストでは、テスト管理ファイルに示された ID の順にアイテム・バンクから項目が検索されて配信され、適応型テストでは項目選択エンジンが指定した ID の項目を配信する。また、適応型テストでは、同一の受検者に同一項目が複数回出題されない仕組みとなっている。さらに同一受検者の複数回受検にも過去に出題された項目の出題防止機能がオプションとして用意されている。

④ テスト構成支援システム

アイテム・バンク中のすべてのテスト領域に項目が入力されるとテスト構成が可能になる。テスト構成はネットワーク上で一人もしくは複数のテスト開発者がテスト構成支援システムを用いて行う。特に、e テスティングではテスト構成がネットワーク上でブラウザを用いてどこからでも行えることが特徴である。図3は開発したテスト構成支援システムであり、構成中のテストの得点分布や所要時間分布を予測したり、現在のテスト構成状況に適した質問項目を推薦する機能を持ち、質問項目 Web 上で複数のテスト構成者が協働してテスト構成を行える環境を提供するシステムである。具体的には、システムは領域の一覧表を提示し、複数のテスト開発者がアイテム・バンクの各領域よりバランスよく出題したい項目を選択すると、システムが、領域のバランスやランダム性、信頼性を計算することにより追加すべき質問項目の候補をさらに絞り込んで推薦してくれるシステムである。あるテスト開発者が選択して、異なる開発者が選択しなかった項目などはピックアップして議論のための掲示板が表示される仕組みもある。さらに現在、構成されているテストのテスト得点、テスト所要時間の予測分布、またはテストの信頼性曲線、項目反応曲線や情報量を画面下に提示し、現在構成されているテストのモニタリングが行えるようになってきている。これにより、所要時間を無駄にせず、広範なテスト領域からバ

ランスよく選択された多数項目により信頼性の高いテストを構築することが支援できる。実際に、このシステムを用いることにより3人以上のテスト開発者で妥当性と信頼性の高い項目ができた。以上のようにeテストティングでは、ペーパー・テストでは行うことができないテスト構成支援が可能となる。このシステムにより構成されたテストは、項目ID集合としてテスト・データベースにテストID(識別子)を伴い格納され、いつでもWEB上でテスト実施ができる状態となる。

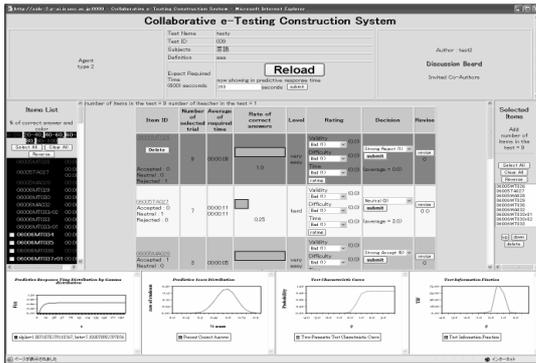


図3 協調型テスト構成支援システム

また、自動テスト構成支援システムも開発した(図4)。正答率、平均所要時間や情報量、領域ごとの最小出題数などを自動テスト構成システムに入力すると、最大クリーク抽出法に分枝限定法を用いて同一のテスト領域、テスト情報量、テスト得点分布、テスト所要時間分布、信頼性を持つ等質テストを重複率 $\alpha\%$ の制約条件でアイテムバンクから自動的に最大数生成することができる。検定テストなどでは、毎回、同一難易度で等質の異なるテストを提示しなければならない。従来のペーパーテストでは、このようなテスト構成には専門家の勘と経験に頼らねばならなかったが、eテストティングにより、労力なくより高精度にテストを自動構成できるようになった。



図4 自動テスト構成支援システム

⑤テスト・データベース

テスト・データベースは、毎回のテスト・データを管理するためのデータベースであり、以下のように構成されている。

- 1) テストID
- 2) テスト構成者
- 3) 作成日時
- 4) 出題項目IDリスト
- 5) テスト構成プロセスの履歴データ
- 6) 受検者の各項目への書き直しプロセスを含む回答データ
- 7) 受検者の各項目への正誤データ(正答のとき1、誤答のとき0)
- 8) 各受検者の得点およびテスト所要時間
- 9) 受検者の各項目への所要時間データ
- 10) 各項目の正答率と標準偏差
- 11) 各項目の平均所要時間と標準偏差

また、6)で作成される回答プロセス・データは、a)受検者ID、b)項目ID、c)回答変更数、d)日時、e)入力情報、f)所要時間、g)日時、h)入力情報、i)所要時間…となる。ただし、真偽式項目、多肢選択式項目では、選択肢が受検者に選択されるごとに回答記録される。並べ替え式項目では、選択肢の順序を回答させるが、すべての選択肢について回答したときにのみ回答が記録される。ただし、その後に変更があればそれらはすべて記録される。一方、回答書き直し時の判別が難しい単答式項目、論述式項目では、受検者がNext(次ページ)ボタンかBack(前ページ)ボタンを押したごとにそれまでの入力情報が記録されるようにしている。そのため、単答式項目、論述式項目では、NextボタンかBackボタンを押すまでに変更された回答は記録されない。

⑥データ分析システム

テスト終了後、各学習者の各項目への正誤データ、項目所要時間、回答書き直し回数、項目見直し回数などがテスト・データベースに書き込まれる。データ分析システムは、これらのデータを解析し、即時的にフィードバックを返すシステムである。フィードバックは、(1)受検者、(2)テスト実施者、(3)アイテム・バンク、の三方向へ行われる。(1)受検者には得点データや偏差値、診断結果などテスト目的に応じたフィードバック、(2)テスト実施者には(1)のフィードバックに加えて、テスト自身の信頼性などを計算した指標や項目別の情報をフィードバックし、(3)アイテム・バンクには項目に関する様々な統計解析の結果が自動的に書き込まれる。このようなデータを用いることにより、ペーパー・テストでは実現できなかった項目分析ができるようになった。

また、項目反応理論を用いた分析では、2パラメータ・ロジスティック・モデルの1)

解析的ベイズ推定、2) Bayesian Parameter Estimation、3) Marginal Maximum Likelihood、4) Gibbs Sampler の中からパラメータ推定法を選択できる。特に解析的ベイズ推定では、データ数が項目数に比較して少ない時や正答確率が0や1となるような項目が存在していても、特別な処理なしでパラメータを推定できるほか、新しく実施されたテスト・データについても等化する必要はなく、受検者が受けていない項目への反応を欠側値として簡単に計算できるという利点がある。

⑦適応型テストシステム

適応型テストとは、テストへのそれまでの回答履歴を用いて逐次、受検者の能力を推定しながら、その能力に最も適応した項目をアイテム・バンクより抽出しながら出題するというものであり、以下の利点を持つ。

1. 受検者の能力に最適な項目を出題することにより、能力の推定効率を上げる。
2. 難しすぎたり易しすぎる項目が出題されないために、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. 推定効率が向上するために、ペーパー・テストと同じ精度を保って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。

一般に適応型テストシステムは単体で開発されることが多く、統合型 e テスティング・システムに含まれていることは多くない。本システムでは、共通のアイテム・バンクから項目選択基準を最大化する項目を選択し受検者に適応した項目出題を行う機能を提供する。本システムでは、完全情報量を最大化する木生成アルゴリズムを提案し、テスト領域の制約を満たしながらフィッシャー情報量パラドックスを解決した完全情報量を最大化する項目を高速に選択し、出題できる。

(2)実践と評価

本研究は単に理論研究を行うためではなく、実際に役に立つシステム開発研究として行ってきた。ここでは、20人の大学教員に実際に1年間使用してもらい、その後、アンケート調査を行うことによって、本システムが単なるプロトタイプではなく、実用的なシステムであることを示す。

講義形式の授業での評価

通常に行われている大学での講義での演習や期末テストで用いられた。大学1年生用の基礎統計学の授業の期末テストで本システムを用いて、教員一人が試験監督する中、127項目のテストを90分で受験するテストであった。ただし、受検者によるノートや参考書などの資料持ち込みは許可されている。テストの項目順序は、受検者ごとにランダムに

振り分けられており、カンニングのリスクを下げるように工夫されている。また、遠隔テストではないが、テスト採点がテスト終了と同時にわれ、教員、受検者にフィードバックされることが利点として報告された。また、本システムは、フィードバック提示機能を含め問題なく終了したと報告された。

e ラーニングでの評価

正規の e ラーニング授業で、本システムを用いて、授業の途中で理解を診断する演習問題の提示や、期末テストを遠隔で行った。ただし、e ラーニングの評価でのテストでは、試験監督は存在せず、受検者が制限時間の中、自宅のコンピュータを用いて受験している。受検者は学外の端末を用いているにも関わらず、システムは問題なく円滑に動作していたと報告された。

遠隔授業での評価

遠隔授業では、演習テスト、期末テストも遠隔テストとして実施せねばならず、本システムが用いられた。期末テストの実施には、各テスト会場には、各大学のスタッフが一人ずつ試験監督として立ち会った。ここでも、テスト実施は問題なく円滑に行われたと報告された。本実践は遠隔授業に対応した遠隔テストのモデルとしても今後重要になると考えられる。

国際遠隔授業での評価

国際遠隔授業の実験に遠隔テストとして本システムを用いた。日本側の教員とタイ側の教員が交互に授業を合計5時間行い、日本側の学生8名、タイ側の学生12名が参加した。授業終了後、30分間の最終テストを本システムを用いて実施した。テストは日本側とタイ側の教員二人によってテスト構成支援システムを用いて協調して構成された。この実践は、実験授業であったが、十分、国際間遠隔テストとして実用に耐えるシステムであることが確認された。

さらに、本システムにおいてどのような機能が有効であったかを評価するために教員20名にアンケートを実施した。その結果、マルチメディア項目作成機能、テスト構成支援システム、テスト得点および所要時間予測機能、項目分析、テスト分析はすべての教員が高い評価をしており、これらの機能が有効であったことが示された。ただし、受検者分析については反応が分散しており、一般の教員が効果的に利用するためには事前に必要な前提知識が多すぎたと解釈できる。また、適応型テストや協調型テスト構成は、20名の実践では、一度も実施されておらず、本システムの機能が十分に発揮されたわけではない。

この点も本システムが一般的なユーザーには簡単には理解できない機能を持っていることを示しており、一般ユーザー向け「e テスティングの使い方」のeラーニング教材などを本システムに実装していく必要性があると考えている。

(4)まとめ

本研究では、e テスティングの利点をすべて統合的に有し、実用的なe テスティング・システムを開発した。さらに、実際に教員に実践利用してもらい、本システムが十分に実践的に役立つシステムであることを示した。

ただし、ユーザーである教員がテストの専門家でないために今回のシステムの使用法が限定され、本システムが十分に使いこなされなかった。今後、本システムを用いた様々な利用法を開発・実践し、その結果を報告していきたい。また、本研究ではある特化した機能を開発せずに統合的なe テスティング環境の構築を提案した。今後は、より特徴ある機能を本システムに追加していくことにより、理論的にも貢献したいと考える。

テスト理論研究では完全情報量関数や制約条件付き最適化の研究は多くなされてきたが、すべて計算時間を無視したシミュレーション研究であり、実用化は遠いと考えられてきた。本研究では、情報科学的アプローチより、これらの演算の高速アルゴリズムを提案し、実際に開発を行った。e テスティングにおいてこのような研究アプローチそのものがこれまで存在しておらず、新研究領域の開発に貢献できるとともにこれまで実用化されてこなかった様々な情報量関数や制約条件が実用化されることになり、大きなインパクトを与えることになる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計9件)

- ① Pokpong Songmuang, Maomi Ueno, Bees Algorithm for Construction of Multiple Test Forms in E-Testing, IEEE Transaction on Learning Technology, IEEE computer Society, 査読有、Vol. 4 no. 3, 2011, 209-221
- ② Takamitsu Hashimoto and Maomi Ueno, Latent Conditional Independence Test Using Bayesian Network Item Response Theory, IEICE Transactions on Information and Systems, 査読有、E94. D no. 4, 2011, 743-753
- ③ 吉川厚, 植野真臣, 学習評価のデザイン、人工知能学会誌、査読有、25(2)、2010、283-290
- ④ 植野真臣, e テスティング：最先端テ

ト技術、電子情報通信学会誌、査読有、92-12、2009、1017-1021

〔学会発表〕(計9件)

- ① 橋本貴充・植野真臣、ベイジアン・ネットワーク IRT によるセンター試験の分析、人工知能基本問題研究会(第83回)、2011年11月、電気通信大学
- ② Maomi Ueno, Robust learning Bayesian networks for prior belief, Association for Uncertainty in Artificial Intelligence, 2011年7月、Barcelona, Spain
- ③ Takamitsu Hashimoto and Maomi Ueno, Latent Conditional Independence Test for Bayesian Network IRT., Proceedings of the First International Workshop on Advanced Methodologies for Bayesian Networks (AMBN 2010)., 2010年11月18日、Campus Innovation Center, Tokyo
- ④ Hashimoto, T. & Ueno, M., Item Parameters Estimation for Dependent Items., The 16th International Meeting of the Psychometric Society, 2009年7月22日、St John's College, Cambridge

〔図書〕(計4件)

- ① 植野真臣、荘島宏二郎、朝倉書店、学習評価の新潮流、2010、187

6. 研究組織

(1)研究代表者

植野 真臣 (UENO MAOMI)

電気通信大学・大学院情報システム学研究科・准教授

研究者番号：50262316

(2)連携研究者

森本 康彦 (MORIMOTO YASUHIKO)

東京学芸大学・情報処理センター・准教授
研究者番号：10387532

荘島 宏二郎 (SYOUJIMA KOUJIRO)

独立行政法人大学入試センター・研究開発部・准教授

研究者番号：50360706

永森 正仁 (NAGAMORI MASAHIRO)

長岡技術科学大学・工学部・助教

研究者番号：80334697

橋本 貴充 (HASHIMOTO TAKAMITSU)

独立行政法人大学入試センター・研究開発部・助教

研究者番号：20399489

安藤 雅洋 (ANDO MASAHIRO)

長岡技術科学大学・工学部・助教

研究者番号：00345539