

ネットワーク利用型試験実施システム

3 S-6

坂田洋幸 品川和弘 丹羽いつ子 前川啓介

NTT データ通信(株) 技術開発本部 マルチメディア技術センタ

1 はじめに

情報通信網の社会への急速な浸透に伴い、教育分野においてもネットワークを利用する形態の教育支援システムの研究並びに実用化が進められている^[1]。

教育支援システムにおける重要な要素の一つとして、試験実施等による学習者能力の測定機構があげられるが、その実現に関して被験者の能力を客観的に測定するための方法論の採用が課題としてあげられる。

従来からの試験に関する理論を大別すると、以下の2通りのアプローチに集約される。

- 試験実施の結果得られる正答数（素得点）に基づき、記述統計学に立脚した分析を行なうもの
- 被験者の「能力」と試験項目への応答特性との間に数学的関数による関連を仮定し、試験の実施から結果の分析までを行なうもの

後者に分類される代表的な理論として「項目応答理論」が注目されており、最近では米国における一部の資格試験実施に適用される事例も報告されている^[2]。

本稿では、我々が現在実装を進めている、項目応答理論をベースとしたネットワーク利用型の試験実施システムの概要について報告する。

2 項目応答理論による試験モデル

項目応答理論 (Item Response Theory, 以下 IRT)^{[3][4]} は、ある試験項目が測定対象とする「能力」と被験者の当該項目への応答特性との間に確率的関数（項目応答関数）による対応を仮定する、確率モデルの一種である。

項目応答関数としては累積正規モデル等、いくつかの形式が提案されているが、その中で今日最も代表的とされているのが下記(1)式に記すロジスティック分布の分布関数を用いた記述である。（ここでは3パラメータモデルについて示す）

$$\begin{aligned} P_i(\theta) &= \Pr\{u_i = 1|\theta\} \\ &= c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad (1) \end{aligned}$$

Computerized adaptive testing system based on network
Hiroyuki SAKATA, Kazuhiro SHINAGAWA, Itsuko NIWA
and Keisuke MAEKAWA
Multimedia Technology Center, NTT DATA Corporation

ただし、 u_i は試験項目 i に対する被験者の応答を表す Bernoulli 変数であり、以下の値をとる。

$$u_i = \begin{cases} 1 & (\text{正答する場合}) \\ 0 & (\text{誤答する場合}) \end{cases} \quad (2)$$

同式中 D は累積正規分布関数との近似を計るための尺度因子 (= 1.7) であり、各項目に付与される項目パラメータ b_i, a_i, c_i はそれぞれ「項目困難度」(difficulty)、「識別力」(discrimination)、「疑似正当水準」(psudochance level) と解釈される。

この時、 $P_i(\theta)$ は「ある能力水準 θ を持つ受験者母集団に属する被験者が試験項目 i に対して正答する確率」を与えるものと捉えられる。

IRT モデルの導入によるメリットとして、従来より以下に示すものがあげられている。

- モデルの仮定に沿って導出された受験者の能力推定値は、特定の試験項目に依存しないため、異なる試験受験者間の比較を可能とする。
- 試験項目の統計量は特定の受験者集団（標本）とは独立に定まる。即ち絶対的評価を可能とする。
- 推定された能力の精度を表す統計量（情報量）が個人毎に与えられる。

また、個々の試験項目パラメータをベイズ推定等により予め算出したものを大量に蓄積するデータベース (ITEM-BANK) を構築しておくことで、被験者個々の能力に応じた試験項目セットの構成が可能となる。この枠組みを計算機上に展開することで、被験者の能力の推定を逐次行ないながら次に提示するべき適切な項目を選択する、適応型テスト (Computerized Adaptive Testing: CAT)^[5] が実現される。

3 ネットワーク利用型試験実施システム

本節では、我々が現在構築を進めている、ネットワーク経由にて ITEM-BANK を利用する CAT システムの概要について述べる。（今回の報告では ITEM-BANK の存在を前提とするため、項目パラメータ算出には言及しない）

3.1 システム概要

システム構成を図1に、処理の流れを以下に示す。

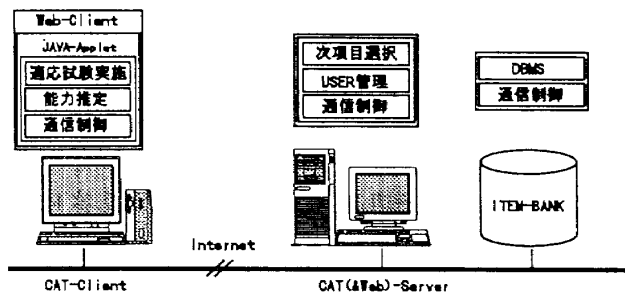


図 1: システム構成

- (1) CATクライアント (JAVA-appletとして実装) は、サーバにアクセスした被験者側端末 (JVM 対応 Web ブラウザ) 上に Down-Load される。
- (2) サーバは問題項目を ITEM-BANK から抽出し、クライアント側に転送する。
- (3) クライアントはサーバより受信した試験項目を被験者に提示し、被験者からの応答状況を基に現時点での能力推定を行なう。
- (4) クライアントは (3) にて得られた能力推定値を基に、次の試験項目をサーバに要求する。((2) へ)

以上の処理を能力推定値が安定するか、ある一定回数繰り返す、最終的に被験者の能力推定値を算出する。

上記 (2) 項目においてサーバは ITEM-BANK から次問題項目 j の選択を、現時点での被験者の能力推定値 $\hat{\theta}$ と下記 (3) 式により定まる ITEM-BANK 内各項目の情報関数値との比較により行なう。

$$I_j(\theta) = \frac{\left(\frac{\partial}{\partial \theta} P_j(\theta)\right)^2}{P_j(\theta)(1 - P_j(\theta))} \quad (3)$$

また、クライアント側における被験者の能力推定は条件付最尤推定により行なう。すなわち、潜在的な能力水準として θ を持っている被験者が現在までに n 項目の試験問題に解答している場合、下式 (4) により尤度関数 L を定め、 L の最大値を与える θ を推定値 $\hat{\theta}$ とする。 (u_k は (2) 式に従う)

$$L(u_1, \dots, u_n | \theta) = \prod_{k=1}^n P_k^{u_k} (1 - P_k)^{1 - u_k} \quad (4)$$

3.2 課題

IRT モデルの実用化に関しては以下に示す様々な議論が展開されている。

- 試験のモデルへの適合性に関するもの
- 推定より得られた結果の信頼性、妥当性に関するもの
- 安定したパラメータ推定のための方法論に関するもの
- その他

また、IRT モデルが前提とする項目間の局所独立性の改善への提案^[6]や「speed-test」など特殊な試験形態を想定した IRT モデルの拡張^[7]に関する研究報告が成されている。

これらの課題に加えて、今回示したネットワーク経由での試験実施形態の導入により、ネットワーク品質に起因する以下の様な課題が新たに発生するものと考えられる。

- (1) 輻輳による試験提示間隔の乱れ
- (2) ネットワーク断による試験中断時の扱い
- (3) 動画像、音声等を補助的に用いる試験項目の場合、マルチメディア QoS(Quality of Service) の扱い
- (4) その他 ITEM-BANK に対する負化分散など、システム設計手法に関するもの

例えば上記 (1) に対しては、項目応答関数の確率密度関数 $p(z)$ へ試験項目の提示間隔 t の確率密度関数 $g(t)$ とその条件付確率を与える $f(z|t)$ の導入によるモデルの拡張が必要になるものと考えられる。

$$p(z) = \int_0^{\infty} f(z|t)g(t)dt \quad (5)$$

4 まとめ

本稿では、我々が現在開発を行なっている項目応答理論に基づいたネットワーク型試験実施システムの概要について報告を行なった。今後は従来から議論されている IRT の実用上の課題への対応に加えて、ネットワーク化により発生する新たな問題への対応を課題とする。

参考文献

- [1] Minoli.: "DISTANCE LEARNING TECHNOLOGY AND APPLICATIONS", Artech House Publishers, Boston(1996)
- [2] 池田央.: "コンピュータテスト化の必要性和その条件", 教育工学関連学協会連合第 5 回全国大会講演論文集 (第一分冊), pp.363-366(1997)
- [3] Hambleton R.K. and Swaminathan H.: "Item Response Theory", Kluwer-Nijhoff Publishing, Boston (1985).
- [4] Linn R.L.(ed.): "Educational Measurement, 3rd ed.", Macmillan. (池田・藤田・柳井・繁樹訳 (1992), 「教育測定学」, みくに出版)
- [5] Wainer H.(ed.): "Computerized adaptive testing: a primer", Erlbaum, Hillsdale, N.J.(1990).
- [6] 植野真臣, 大西仁, 繁樹算男.: "確率ネットワークを組み込んだテスト理論の提案", 信学論 (A), J77-A, No.10, pp.1398-1408 (Oct.1994).
- [7] Linden W.J. and Hambleton R.K.(ed.): "Handbook of Modern Item Response Theory", Springer-Verlag, N.Y.(1997).