

項目反応理論を用いた適応テストの WWW 上での実現性の分析

尹明, 藤田真一, 成田誠之助, 早稲田大学理工学部

1. 研究背景

遠隔教育において学習者を評価するために、Web ベース適応テストの要求は高まっている。しかし、適応テストの計算量は膨大であり、インターネット環境で利用しようとした場合、処理が追いつかない可能性がある。そのため、Web ベースの適応テストを実現するにあたり、その計算量の評価は重要である。

本論文はインターネット環境において適応テストを実現するために、計算量の分散と問題項目の選択について検討するものである。TOEIC など一部の大型の WBT においては項目反応理論をもちいたものもある。本論文では特に大学などの一般的な教育機関の教師たちが適応テストを利用することを対象として論じる。この場合、TOEIC などとは異なり、問題項目数は数千～数万程度と少なくなる。さらに、経済的な問題から必ずしも西洋の優れたマシンを持っていることは保証できない。この場合、項目反応理論に基づく Web ベース適応テストを実現するとき、主な難点として挙げられるのは反応パターンから被験者の理解度を測定することと適切な問題項目の選択をすることである。被験者の能力値をサーバ側で計算する場合とクライアント側で計算する場合のコストを比較した結果、クライアント側で推定した場合の方がシステム全体の性能がよくなることが示唆された。また、適切な問題項目の選択についてはある被験者の能力値に関する最適の項目情報をあらかじめデータベースに保存しておくという方法を採用した。テストを実施するとき、推定された被験者の能力値に基づいて、直接ルックアップ表で問題項目を検索することによって、サーバの負担を減らし、WWW 上での項目反応理論を用いた適応テストの実用化が可能になる。シミュレーションの結果からこの方法の有効性を明らかにした。

2. 項目反応理論

項目反応理論は古典的テスト理論の制限を解決するために考案された。本文はロジスティックモデルを採用する。

2.1 ロジスティック関数

項目特性関数は、問題項目に被験者が正答する確率を被験者の能力値 Q の関数として表現したものである。3パラメタロジスティックモデル (three-parameter logistic model) の場合、項目特性関数を次式のように表現される。

$$P_i(\theta) = C + \frac{1 - c_i}{1 + e^{-Da_i(\theta - b_i)}} \quad (1)$$

ただし、定数 D は 1.7、 θ は被験者の能力値 (被験者の能力)、 a_i は項目 i の識別力母数、 b_i は項目 i の困難度母数、 c_i は被験者が偶然正答する確率とする。被験者の項目 i に対する正答確率はこの項目特性関数によって示される。

2.2 被験者能力値の推定

項目 j の識別力母数、困難度母数、偶然正答する確率といった問題項目に関するパラメタの値が分かっている場合 ([5])、 N 個の項目からなるテストを受験したときの正答率から最尤推定法で被験者の能力値を推定することができる。

2.3 項目情報関数

適応テストが有効であるかどうかの評価はテスト情報関数 ([5]) を利用し、テストの測定精度をその関数の値で表すことで行なうことができる。すなわち、テスト情報量 $I(\theta)$ は各項目の情報量 $I_i(\theta)$ の総和で定義される。3母数モデルの項目 i が持つ能力値 θ に対する情報量を式 (2) のようになり、テスト情報関数の情報量が式 (3) で示される。

$$I_i(\theta) = \frac{D^2 a_i^2 (1 - c_i)}{(c + \lambda^{Da_i(\theta - b_i)}) \cdot (1 + \lambda^{-Da_i(\theta - b_i)})^2} \quad (2)$$

$$I_i(\theta) = \frac{D^2 a_i^2 \lambda^{-Da_i(\theta - b_i)}}{(1 + \lambda^{-Da_i(\theta - b_i)})^2} \quad (3)$$

3. Web ベース適応テストの流れ

適応テストではまず被験者の能力値の初期値を暫定的に与えておき、その後被験者の反応パターンにより実際の能力値へと近づけていく。また被験者の反応パターンを利用しながら、リアルタイムで次に解くべき項目を選び出し、テストを実施する。開発したテストシステムの中でクライアント側の主な役割は被験者とサーバ側の間の通信を管理することである。これに対して、サーバ側の主な役割は被験者の能力値を推定と項目プールから項目を選択する機能を持っている。また、被験者の能力値の空いてはクライアント側・サーバ側の双方にその機能をもたせ、どちらで処理したほうがより実用的かを検討する。テストの流れを以下に示す。

(1) テストの終了条件を定める。

(a) テスト情報量の標準誤差が一定の値以下になった場合

(b) 出題した項目数が一定数以上になった場合

(2) サーバ側はクライアント側の要求に応じて、適切な問題項目を返し、被験者に解かせる。被験者が解いた項目の反応パターンから、サーバ側かクライアント側か(後に述べる)で被験者の能力値を求める

(3) サーバ側で項目情報量の最大になる項目を検索し、被験者に出題する。

(4) 終了条件を満たしている場合、テストを終了する。そうでなければ、被験者の能力値を更新して、(2)に戻る。

4. WEB ベース適応テストにおける計算量の分析

4.1 シミュレーション用データの作成

本論文における問題項目の困難度母数は平均 0、標準偏差 1 の正規分布に従い、識別力母数は平均 1、標準偏差 0.5、下限 0.2、上限 2.0 の切断正規分布に従うと仮定する。テスト項目数は最大 200 項目までとする。被験者の問題項目に対する正誤はロジスティックモデルに従っていると仮定する。シミュレーション中の被験者の反応パターンは、0 から 1 までの一様乱数を作り、0.5 と比較して、大きければ正答、小さければ誤答とする。

表 1. 項目情報

項目番号	識別力	困難度	偶然正答確率
1001	0.936	1.046	0.219
1002	1.15	0.6194	0.177
1003	1.322	0.1439	0.1589

4.2 被験者の能力値の推定

Web サイトとの接続に伴って実行されるスクリプトやプログラムには、クライアント側で実行されるものとサーバ側で実行されるものの 2 種類がある。構築したテストシステムの中では比較実験を行うために、クライアント側、サーバ側の両方に被験者の能力値を推定する機能をもたせた。ニュートン法を利用して、クライアント側で推定の処理を行った場合と、サーバ側で推定の処理を行った場合における、システムの性能の比較を行った。異なった項目の組み合わせとそれらに対する異なった解答の組み合わせによって、被験者の能力値を推定する必要な計算の回数は異なっていることを考慮して、項目数は最大 200、計算の回数は最大 10 回までと仮定する。一般的なテストにおいて項目数が 200 を超えることはあまり起こらないと考えられるため、この条件で、十分にテストとしての要求を満たす。

サーバ側で被験者の能力値を推定するとクライアント側で被験者の能力値を推定する実験から、次のような結果が得られた

(1) 出題する項目数が増加するのに伴い、推定にかかる時間は増加する

(2) どちらの推定でも、システム全体の性能の差は小さい

(3) クライアント側で推定する場合の処理時間はコンピュータの性能によって左右される

(4) クライアント側でサーバ側で推定を処理する場合よ

りも短い

(5) サーバ側で推定する場合の処理時間は同時に接続する受験生の数が多くなる場合に長くなる可能性がある

以上の考察から、クライアント側で能力値を推定することを前提として考える

4.3 出題項目の選択に関する分析

すでに実施した項目の結果から求めた被験者の能力値を用いて各項目の項目情報量を計算する。各項目情報量が最大になる項目を次の項目として実施する。項目反応理論による測定は被験者にとって解けるか解けないか五分五分くらいの項目が多く含まれるテストを実施したときに被験者の能力値の推定精度が高くなる。推定された被験者の能力値 θ に対する、項目選択の規則が分かる。

(1) 2 母数モデルにおいては $\theta = b$ に従う

(2) 3 母数モデルにおいては式 (1) に従って困難度の値よりすこし大きい値となる項目を選択する

(3) 識別力母数の値の高い項目を識別力母数の値の低い項目より優先して選択する

(4) 偶然正解する確率の値の低い項目を優先的に選択する

しかし、条件を満たす項目が十分な数だけあることは限らず、この場合、オンラインで情報量などを計算して項目を選択することは不可能になる。実用的な時間内で計算するために推定された被験者の能力値と項目に対する正答確率に関して近似値を計算し、それに基づいて項目を選択することが必要になっている。

しかし、テスト中に項目を選択するために複雑で膨大な計算をすることはシステムのパフォーマンスの低下を招くため事前に計算を行うことにする。

(1) すべての項目に対して求めた被験者の能力値の最大値と最低値をテストの被験者の能力値の最大値と最低値と仮定する。

(2) テストの被験者の能力値の最低値より低い推定値があった場合、それを能力値の最低値とみなす

(3) テストの被験者の能力値の最低値より高い推定値があった場合、それを能力値の最大値とみなす

(4) 要求されるテストの推定の精度によって、最大値と最低値の間の値を m 分割する。つまり可能な推定値は \min から $\min+1/m$, $\min+2/m$, ..., $\min+(m-1)/m$, \max までとする。

(5) すべての起こりうる推定値に基づき、すべての項目に対して正答確率、項目情報量を計算する。

(6) 以上の規則に従って計算した結果により起こりうる推定値に対する最適な項目を決定する。

(7) 最後に起こりうる全ての推定値に対応する項目番号、新しい項目番号をデータベースに保存する。

以下に、データベースのテーブルを構成しているフィールドについて説明を加え、構成を表 2 に示す。

表 2. ルックアップ表の構成

能力値	情報量	項目番号	優先度
1.1	0.9	230	1
1.1	0.9	586	1
1.1	0.865	120	2

テストの実施中に推定した被験者の能力値を θ とすると、ルックアップテーブルより項目を選択する手順は次のようになる。

(a) θ に対応するレベル 1 の最大試験問題数を数える

(b) 1 と最大項目数の間のランダム数字を作成する

(c) そのランダム数字で対応の項目番号を引き出し、すでに実施した項目ならば、上に戻す

(d) 決めた項目番号で項目をデータベースから引き出し、被験者にわたす

テストを行う時、以上の方法によって、すでに推定済の被験者の能力値を用い、サーバー側で最適のテスト項目を選択する 実験したときの時間のコストを表に示す。

(1) サーバ側で被験者の 能力値を推定する場合

同時に 接続する受験生の 数が同じ場合に 性能が異なるコンピュータで テスト項目を選択する 時間のコストは図 1 に示す。同時に 接続する受験生の 数が異なる場合に 同一のコンピュータで テスト項目を選択する 時間のコストは図 2 に示す。

同時に 接続する受験生の 数が同じ場合に 性能が異なるコンピュータで テスト項目を選択する 時間のコストは図 3 に示す。同時に 接続する受験生の 数が異なる場合に 同一のコンピュータで テスト項目を選択する 時間のコストは図 4 に示す。

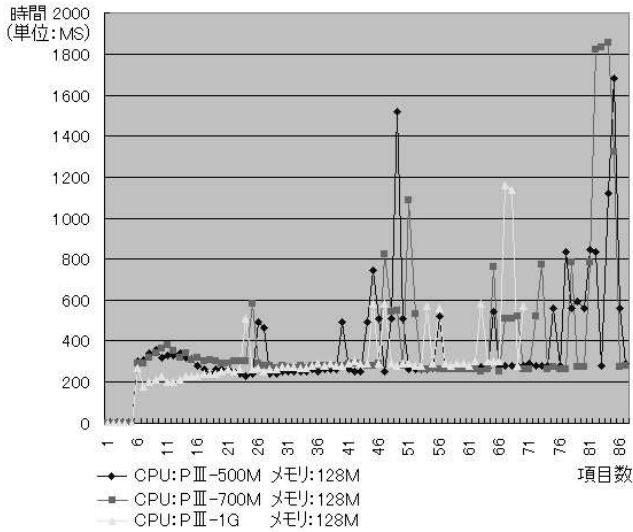


図1. 異なるコンピュータで 項目を選択する コスト (サーバ側)

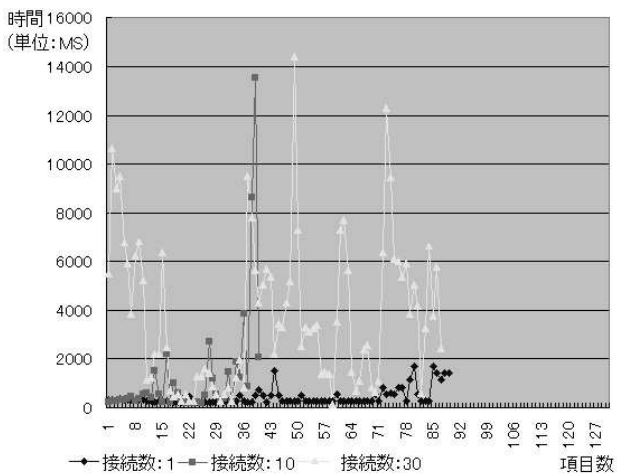


図2. 同一のコンピュータで 項目を選択する コスト (サーバ側)

(2) クライアント側 で被験者の能力 値を推定する場合

4.4 結果と考察

Web ベースのテストを実施するとき、項目情報関数を試験中にすべて計算することは計算量の多さから実用的でないため、推定した被験者の能力値に対して近似をすることにより、実用的なテストの精度を保ちながらサーバの負荷を軽減しようと試みた。近似前にはオンラインで計算が不可能に対し、近似後にはオンラインで計算が可

能になってパフォーマンスの向 上が見られた。

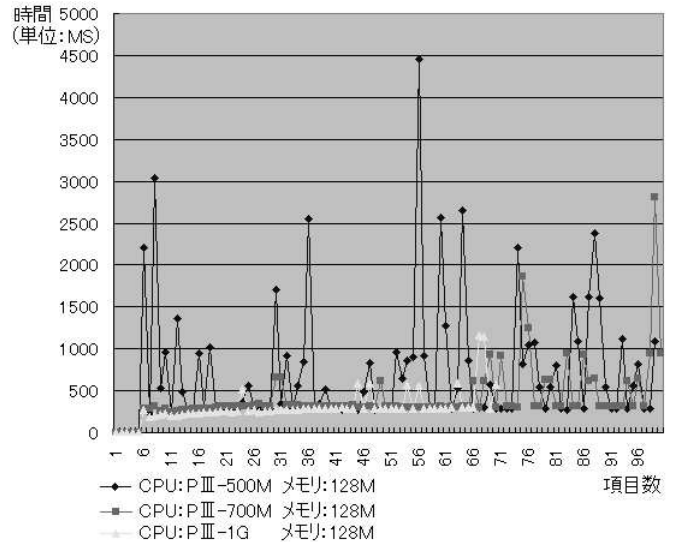


図3. 異なるコンピュータで 項目を選択する コスト (クライアント側)

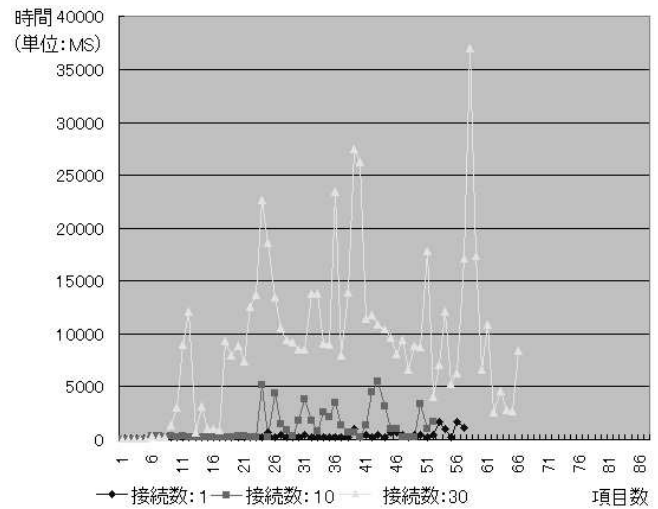


図4. 同一のコンピュータで 項目を選択する コスト (クライアント側)

5. まとめと今後の展望

Web ベースのテストは遠隔教育にとって重要である。遠隔教育に参加する学習者の能力はさまざまである。そのため Web ベースのテストには高い効率を要求され、できるだけ少ないの試験問題から、学生の実際の能力を測定することが求められる。つまり、遠隔教育に適する テストの形式は適応テストであると言える。しかし、Webベースの適応テストを実現する場合、計算に長く時間がかかりすぎると、学習者を待たすことになり学習時間が無駄になる。同時に、適切ではない試験問題を出题してしまうと、テストの学習者の能力を推定する精度が低下する。学習者がテストに対してストレスを感じることによって学習に悪い影響を与える。本論文は Web ベースの適応テストの実現に対してその実用的な構築方法を示した。

遠隔教育においては適応テストが必須となる。本論文はシミュレーションにより Web ベースの適応テストの実現可能性を実証した。現在語学教育における Web ベースの適応テストを実施しており検証後次回報告する予定である。

参考文献

- (1) 芝 佑順：“語彙理解力測定のための多層適応形テスト”、教育心理研究、第 26 卷 Vol. 4、1978
- (2) 永岡 慶三：“大学における情報教育環境としての適応型テストシステム”、信学技報 1986
- (3) 小林 修：“学習評価のための適応型逐次検定”、電子情報通信学会論文誌 A, J67-A, 1993
- (4) 芝 佑順：“項目反応理論基礎と応用”、東京大学出版社、1991
- (5) 豊田 秀樹：“項目反応理論”、朝倉書店、1999