バイオメディカル・ファジィ・システム学会誌　Vol.17，No.2，pp.67－71（2015）

[Short paper]

（2015年10月8日　Accepted）

統計的決定理論に基づくeテスティングに関する一考察

前田　康成[1]，鈴木　正清[1]

1）北見工業大学・情報システム工学科

要約：　近年の情報技術の発展にともない，教育工学分野でも多くの情報技術が応用されている．eテスティングは教育工学分野における重要なテーマの一つである．既にeテスティングに関する数多くの従来研究が存在する．しかし，従来研究では，eテスティング方法が受験者の真の理解度の判定を間違えてしまう確率である誤り率を理論的に最小化することを理論的に保証する方法は提案されていない．そこで，本研究では統計的決定理論に基づきベイズ基準のもとで誤り率を最小化するeテスティング方法を提案する．提案方法ではeテスティング問題を解くために動的計画法を利用している．提案方法の有効性を示すために数値計算例を紹介する．提案方法は適応的な仮説検定の一種でもあり，本研究の成果はヘルスケア産業においても応用可能だと考えられる．
キーワード：　eテスティング，統計的決定理論，誤り率，ベイズ基準，動的計画法

A Note on E-testing Based on Statistical Decision Theory

Yasunari MAEDA[1], Masakiyo SUZUKI[1]

1)　Department of Computer Science, Kitami Institute of Technology

Abstract: *Information technology is applied to many tasks in educational engineering. E-testing is one of important tasks in educational engineering. There are many previous studies on e-testing. An error rate in e-testing represents the probability of an event that an e-testing method computes wrong evaluations for understanding of students. In previous studies the error rate is not minimized theoretically. We propose an e-testing method based on statistical decision theory. The proposed method minimizes the error rate with respect to the Bayes criterion. Dynamic programming technique is used to solve the e-testing problem. We show the effectiveness of the proposed method by some numerical calculation examples. The proposed method is one of adaptive tests of hypotheses in statistics. The proposed method can be applied to many tasks in healthcare industry.*
Keywords: *e-testing, statistical decision theory, error rate, Bayes criterion, dynamic programming*

Yasunari MAEDA
165 Koen-cho, Kitami-shi, Hokkaido, 090-8507, Japan
Phone:　+81-157-26-9328,　E-mail: maeda@cs.kitami-it.ac.jp

1. はじめに

情報技術の発展,インターネットの普及にともない,eラーニング,eテスティングに関する研究[1][2]も数多く行われている.CBT（computer based testing）がスタンドアローン型コンピュータ上で実施されるテストであるのに対して,eテスティングはWeb上で実施されるテストであり,オンライン・テスティング,オンラインクイズ,IBT（Internet based testing）とも呼ばれている[2].eテスティングに関する従来研究[1][2]では,採用した確率モデル上での相互情報量などの情報量の最適化問題として検討している.しかし,受験者の真の理解度の判定を間違えてしまう確率である誤り率を理論的に最小化するようなeテスティング方法は未検討である.また,従来研究では誤り率による評価も行われていない.そこで,本研究ではeテスティングを統計的決定理論[3]に基づいて定式化し,ベイズ基準のもとで誤り率を最小にするという意味で最適なeテスティング方法を提案する.また,数値計算例によって提案方法の有効性を検証する.

2. 準備

本研究で使用する各種記号などの定義を行う.θ_i,$\theta_i \in \Theta$は受験者の理解度を示すi番目の状態である.Θ,$\Theta = \{\theta_1, \theta_2, \ldots, \theta_{|\Theta|}\}$は状態集合を示す.例えば,$|\Theta| = 2$で合格と不合格の2状態,$|\Theta| = 3$で良,普,否の3状態などが考えられる.$e_i$,$e_i \in E$は$i$番目のテスト問題,$E$,$E = \{e_1, e_2, \ldots, e_{|E|}\}$はテスト問題の集合を示す.$o_i$,$o_i \in O$はテスト問題に対する$i$番目の回答結果（採点結果）,$O$,$O = \{o_1, o_2, \ldots, o_{|O|}\}$は回答結果の集合を示す.例えば,$|O| = 2$であれば,正解と不正解の2種の回答結果である.

本研究で想定しているeテスティングでは,真の理解度が未知の受験者に対してテスト問題集合Eからn問の異なるテスト問題を出題し,当該受験者の回答結果（採点結果）から理解度を推定する.本研究で提案するeテスティング方法は出題するn問の選択と理解度の推定を行う方法である.

x_i,$x_i \in E$はi問目の出題問題を示す変数,y_i,$y_i \in O$はi問目の出題問題に対する受験者の回答結果を示す変数である.$p(o_k | e_i, \theta_j)$は理解度θ_jの受験者のテスト問題e_iに対する回答結果がo_kになる確率で既知である.θ^*,$\theta^* \in \Theta$は受験者の真の理解度で未知である.$p(\theta_i)$は理解度θ_iの事前確率で既知である.$x^i y^i$は$x_1 y_1 x_2 y_2 \cdots x_i y_i$という$i$問の出題問題と回答結果の系列を示す.

3. 定式化と提案アルゴリズム

3.1 定式化

統計的決定理論[3]に基づき定式化する.損失関数は次式で定義される.

$$L(d(x^n y^n), \theta) = \begin{cases} 1, & d(x^n y^n) \neq \theta; \\ 0, & d(x^n y^n) = \theta, \end{cases} \tag{1}$$

ただし,$d(x^n y^n)$はn問の系列を受け取った際に理解度の推定結果を返す決定関数である.損失関数$L(d(x^n y^n), \theta)$は真の理解度がθのもとで,n問の系列$x^n y^n$に対する理解度の推定結果$d(x^n y^n)$が正しいかどうかを示す$0-1$損失である.

リスク関数は次式で示される.

$$R(d(\cdot), \theta) = \sum_{y^n \in O^n} p(y^n | x^n, \theta) L(d(x^n y^n), \theta), \tag{2}$$

ただし,決定関数は次式のようにn問未満の系列を受け取った際には次に出題すべき問題を返す.

$$d(x^{i-1} y^{i-1}) = x_i, \quad 1 \leq i \leq n. \tag{3}$$

リスク関数は真の理解度がθのもとで,n問の出題と最終的な理解度の推定に決定関数$d(\cdot)$を用いた場合の損失の期待値であり,理解度の推定を間違えてしまう確率である誤り率に相当する.

ベイズリスクは次式で定義される.

$$BR(d(\cdot), p(\theta)) = \sum_{\theta \in \Theta} p(\theta) R(d(\cdot), \theta), \tag{4}$$

ただし,$p(\theta)$は理解度θの事前確率である.

ベイズ基準のもとで最適な決定関数は次式で定義される.

$$d^*(\cdot)=\arg\min_{d(\cdot)} BR(d(\cdot),p(\theta)), \tag{5}$$

ただし，右辺のminは決定関数$d(\cdot)$の全ての候補についてベイズリスク$BR(d(\cdot),p(\theta))$を計算してベイズリスクの最小値を算出し，argはベイズリスクを最小化する決定関数を左辺の$d^*(\cdot)$に返す．次節以降で出てくるmaxは最大化を意味する．式(5)の決定関数$d^*(\cdot)$は，ベイズ基準のもとで誤り率を最小にするという意味で最適な出題方法と理解度の推定方法に相当する．この方法がベイズ最適なeテスティング方法である．

3.2　提案アルゴリズム

式(5)で定義されたベイズ最適な決定関数を具体的に算出するアルゴリズムを提案する．式(4)のベイズリスクを書き下すと以下のようになる．

$$\begin{aligned}BR(d(\cdot),p(\theta)) &= \sum_{\theta\in\Theta} p(\theta)R(d(\cdot),\theta)\\ &= \sum_{\theta\in\Theta} p(\theta)\sum_{y^n\in O^n}\prod_{i=1}^{n} p(y_i|x_i,\theta)L(d(x^ny^n),\theta)\\ &= \prod_{i=1}^{n}\sum_{\theta\in\Theta} p(\theta|x^{i-1}y^{i-1})\sum_{y_i\in O} p(y_i|x_i,\theta)L(d(x^ny^n),\theta),\end{aligned} \tag{6}$$

ただし,

$$p(\theta|x^iy^i)=\frac{p(\theta|x^{i-1}y^{i-1})p(y_i|x_i,\theta)}{\sum_{\theta'\in\Theta} p(\theta'|x^{i-1}y^{i-1})p(y_i|x_i,\theta')}, \tag{7}$$

$$p(\theta|x^0y^0)=p(\theta). \tag{8}$$

式(6)の入れ子構造に動的計画法[4]を適用することにより，ベイズ最適なeテスティング方法を算出できる．

以下で，動的計画法による具体的なアルゴリズムを示す．アルゴリズムは，n問の系列x^ny^nのすべての候補それぞれに対して，当該候補を受け取ったもとで理解度の推定結果を算出する時点$n+1$の処理を最初に実施する．次に，$n-1$問の系列$x^{n-1}y^{n-1}$のすべての候補に対して，当該候補を受け取ったもとでのn問目の出題問題を選択する時点nの処理を実施する．以後，時点1へ向けて1時点ずつ遡りながら，各時点における系列のすべての候補に対して，当該候補を受け取ったもとでの出題問題を選択する処理を実施する．

n問の系列x^ny^nを受け取ったもとで理解度の推定結果を算出する時点$n+1$の処理は以下のとおりである．

$$d^*(x^ny^n)=\arg\max_{\hat\theta\in\Theta} p(\hat\theta|x^ny^n). \tag{9}$$

$$V(x^ny^n)=\max_{\hat\theta\in\Theta} p(\hat\theta|x^ny^n). \tag{10}$$

n問未満の$i-1$問の系列$x^{i-1}y^{i-1}$を受け取ったもとで次のi問目の出題問題を選択する時点iの処理は以下のとおりである．

$$d^*(x^{i-1}y^{i-1})=\arg\max_{x_i\in E-x^{i-1}}\sum_{\theta\in\Theta} p(\theta|x^{i-1}y^{i-1})\sum_{y_i\in O} p(y_i|x_i,\theta)V(x^iy^i), \tag{11}$$

$$V(x^{i-1}y^{i-1})=\max_{x_i\in E-x^{i-1}}\sum_{\theta\in\Theta} p(\theta|x^{i-1}y^{i-1})\sum_{y_i\in O} p(y_i|x_i,\theta)V(x^iy^i), \tag{12}$$

ただし，$E-x^{i-1}$は$i-1$問目までで未出題の問題集合である．

式(9)から式(12)を用いて受験者の真の理解度を間違って判断してしまう確率である誤り率をベイズ基準のもとで最小にするという意味で最適なeテスティング方法を算出できる．なお，式(9)から式(12)による計算は実際のテスト実施前に行い，実際のテスト実施時の処理は事前の計算結果から出題すべき問題や最終的な理解度を読み取るのみである．

4．数値計算例

ここでは，提案方法の有効性を検証するために実施した数値計算結果の一例を報告する．なお，正解率などの各種設定は数値計算用の架空の設定である．理解度の状態集合Θはθ_1：良，θ_2：普，θ_3：否の3段階の理解度，回答結果（採点結果）集合Oはo_1：正解，o_2：不正解の2段階，テスト問題集合Eはe_1からe_8の8問とした．受験者の理解度に関する事前知識が何もない場合を想定し，無情報事前分布を採用して各理解度の事前確率は等確率とし，出題数は$n=4$とした．各理解度における各問題に対する正解率$p(o_1|e_i,\theta_j)$（正解する確率）を表1に示す．不正解率は$1-p(o_1|e_i,\theta_j)$である．

問題e_1からe_6は理解度良での正解率が一番高く，理解度否での正解率が一番低い一般的な問題設定の複数のパターンである．他方，e_7とe_8では理解度普程度の理解度の場合に不完全な理解に基づく間違いを起こし

易い問題を想定し，理解度普での正解率が一番低く設定してある．これは，ひっかけ問題などと呼ばれているものである．

表1　正解率 $p(o_1|e_i,\theta_j)$

	θ_1：良	θ_2：普	θ_3：否
e_1	0.99	0.8	0.7
e_2	0.9	0.7	0.6
e_3	0.9	0.6	0.4
e_4	0.8	0.5	0.2
e_5	0.7	0.4	0.1
e_6	0.6	0.3	0.1
e_7	0.9	0.2	0.4
e_8	0.9	0.4	0.6

上記の条件のもとで式(9)から式(12)を用いて算出したベイズ最適な決定結果の一部分を表2に示す．表2は系列 x^iy^i を受け取ったもとでのベイズ最適な決定 $d^*(x^iy^i)$ のリストである．

表2　ベイズ最適な決定

x^iy^i	$d^*(x^iy^i)$
-	e_5
e_5o_1	e_8
$e_5o_1e_8o_1$	e_6
$e_5o_1e_8o_1e_6o_1$	e_1
$e_5o_1e_8o_1e_6o_1e_1o_1$	θ_1：良(0.898)
$e_5o_1e_8o_1e_6o_1e_1o_2$	θ_2：普(0.632)
$e_5o_1e_8o_1e_6o_2$	e_7
$e_5o_1e_8o_1e_6o_2e_7o_1$	θ_1：良(0.838)
$e_5o_1e_8o_1e_6o_2e_7o_2$	θ_2：普(0.609)

表2の2行目の決定 e_5 が1問目の出題を示す．また，理解度の推定結果の後の数値は当該理解度の事後確率の値である．表2の結果より，当該時点までの出題問題と回答結果に依存して適応的に出題問題を選択できていることがわかる．最終的な理解度の推定結果については，ベイズ基準のもとで誤り率を最小化することは理論的に保証されているが，個別の推定結果については，事後確率が十分高いと思われる部分とそうでない部分が混在する．本研究では事前に出題数を n 問に固定する問題設定で検討したが，事後確率が十分に高くならない場合には追加出題を行うような仕組の検討も必要である．

また，従来研究[1]との比較も行った．従来研究と本研究ではモデル化が異なるため単純に比較はできないが従来研究の特徴として，1問ずつの出題に対して最適化するアルゴリズムを採用している．他方，本研究の提案アルゴリズムは出題数全体に対する最適化を行っている．そこで，出題数全体に対する最適化の提案アルゴリズムによる上記設定のもとでの数値計算結果と，比較用の1問ずつの最適化のアルゴリズムによる数値計算結果との比較を実施した．選択された問題は異なっており，出題数全体での最適な出題と1問ずつでの最適な出題とは異なることが確認できた．また，両アルゴリズムによる誤り率の期待値（提案アルゴリズムの期待誤り率は式(12)の1問目の値を1から減算した結果に相当）を比較したところ，提案アルゴリズムの期待誤り率が1問ずつでの最適化アルゴリズムの期待誤り率よりも約0.01小さかった．この差は小さいが，上記の数値計算例での設定に依存した偶然の結果なのか，それとも1問ずつでの最適化アルゴリズムが提案アルゴリズムに対する高精度の近似アルゴリズムに相当するかどうかを確認するためには，今後さらに検討する必要がある．

5．考察と今後の課題

従来から数多くのeテスティングに関する研究が行われているが，受験者の理解度の判定を間違えてしまう確率である誤り率を理論的に最小にするようなeテスティング方法は検討されていなかった．そこで，本研究では統計的決定理論に基づいて誤り率をベイズ基準のもとで最小にするという意味で最適なeテスティング方法を提案した．提案方法は，受験者に対してテスト問題を1問ずつ出題してその回答結果に従って適応的に次の出題問題を選択する出題問題選択方法と，最終的に受験者の理解度を推定する理解度推定方法によって構成される．

提案方法の挙動を数値計算によって確認したところ，受験者の回答結果に対して適応的に次の問題を選択していることを確認した．理解度の推定結果についてはベイズ基準のもとで誤り率は最小であるが，各推定結果にともなう事後確率が十分に高くない結果もあった．これは出題問題数を事前に決めておく問題設定のためである．よって，出題問題数も適応的に決めるような問題設定での検討も必要である．また，本研究では各理解度のもとで各回答結果が発生する確率を既知と仮定したが，実用化のためにはより現実に近い問題設定

として未知の場合を検討する必要がある．未知の場合には，何らかの事前知識や学習データ（未知の確率を推定するためのデータ）が必要である．事前知識の例としては教師の経験則，学習データの例としては専門家の口頭試問などによって理解度が判明している被験者の回答データなどが挙げられる．これらの具体的な検討については，今後の課題としたい．

数値計算例では，テスト問題集合および出題数が小さな例を対象としたが，現実には大きなテスト問題集合および出題数を対象にする必要がある．提案アルゴリズムは数値計算例よりも大きな規模のテストにも対応可能であるが，大学入試センター試験のようなとても規模の大きなテストの場合には，いろいろな効率化も必要である．例えば，各回答結果が発生する確率が同一の問題については部分集合として分類しておき，1問ずつ選択肢とするのではなく部分集合を選択肢とする．当該部分集合が選択された場合には部分集合内の問題を等確率で出題することによって，理論的な最適性は維持したまま計算量を軽減することが可能である．

本研究では教育工学の視点から e テスティングを題材に検討したが，本研究で扱った問題設定を統計学の視点から見直すと，本研究の検討内容は統計的決定理論およびベイズ統計学に基づく適応的な仮説検定に相当する．適応的な仮説検定は教育工学に限らず，ヘルスケア分野にも適用可能である．本研究における理解度状態を患者／検査対象者の健康状態，本研究におけるテスト問題を質問／検査項目，本研究における回答結果を質問への回答／検査結果として解釈し直すと，本研究の成果はヘルスケア分野における問診／健康診断項目の自動選定／選定支援ソフトなどに適用可能である．ヘルスケア分野における具体的な検討についても今後の課題としたい．

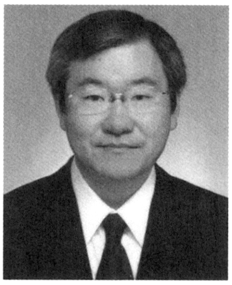
参考文献

[1] 植野真臣, 大西仁, 繁桝算男：確率ネットワークを組み込んだテスト理論の提案, 電子情報通信学会論文誌 A, Vol.J77-A, No.10, pp.1398-1408, 1994.

[2] 植野真臣：知識社会における e ラーニング, 培風館, 2007.

[3] Berger, J.O. : Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York, 1980.

[4] Bertsekas, D.P. : Dynamic Programming and Control Vol.I, Athena Scientific, Massachusetts, 1995.



前田康成（まえだやすなり）

北見工業大学准教授

統計的決定理論の学習問題への応用に関する研究に従事. 電子情報通信学会, 情報処理学会, 電気学会等各会員.