

受験者モデルに基づく出題を行なう適応型テストシステムの開発

藤原 康宏 永岡 慶三

神戸大学 発達科学部

〒657 神戸市灘区鶴甲3丁目11番

あらまし 適応型テストは、受験者の反応に応じて次に出題する項目を決定するテストシステムであり、それによる学力測定の精度の向上を主目的としている。従来の適応型テストは、受験者の理解状態を学力という1元的な数値で表わし、それを推定することを目的とした。本稿では、評価を行なう領域の内容に対する受験者の理解状態を評価することを目的として開発した適応型テストについて実施例を含めて述べる。

本システムは、意味ネットワーク表現された学習領域の構造情報とその構造情報へのオーバーレイモデルで表わす受験者モデルに基づいた出題を行なう。

システム運用の結果、従来測定が困難であった受験者個人の理解構造を評価することができ、出題方略における評価要素及び項目の難易度の選択が、効果的に機能していることが確認された。

和文キーワード 適応型テスト, 教育評価, テスト, コンピュータテスト, CMI, 意味ネットワーク, 受験者モデル

Development of Adaptive Test System on the Basis of Examinee Model

Yasuhiro Fujihara Keizo Nagaoka

Faculty of Human Development, Kobe University

3-11 Tsurukabuto, Nada-ku, Kobe 657

Abstract The adaptive testing is a computerized test system which selects test items by response of examinee on purpose to raise measuring precision of achievement. Conventional adaptive testing aims at estimating examinee's achievement defined one dimensional value. In this paper, we discuss the development of an adaptive testing which evaluates the structure of examinee's understanding.

Our system selects items on basis of structure of the domain represented by the semantic network and an examinee model described by overlay model to the domain network.

The consequence of application of our system, we realized that our system can evaluate structure of examinee's understanding, and algorithm is effective for selecting items in our system.

英文 key words Adaptive Testing, Educational Evaluation, Test, Computerized Testing, CMI, Semantic Network, Examinee Model

1. はじめに

教育評価の機能の1つとして、教師及び受験者に対する教授・学習活動を改善するための情報の提供がある。この情報は偏差値などの1元的な数値だけではなく、個人及び集団の評価を行なう学習領域の内容に対する受験者の理解状態を示す情報である必要がある。個々の受験者の理解状態を評価するためには、受験者全員に同一問題を課すよりも、個人に応じてアイテムプールより構成されるテストを実施することが望ましい。本研究での適応型テストは、教授・学習活動を改善するために有効な情報の提供を主目的として開発したものである。

適応型テストは、アメリカでは既に実用されており、我が国でも研究・開発が進んでいる^{(1)~(4)}。従来の適応型テストは、受験者に最も適した難易度の項目を出題すること及びそれによる測定精度の向上を目的としている。現在、主流となっているのは、受験者の次の項目への正答確率を推定して、受験者の能力に対して正誤反応から得られる情報量が最大となる項目をアイテムプールから選択し、出題する方式である。しかし、この方式では、受験者の理解状態を1次元の学力尺度で表わしているが、実際の受験者は、学習領域の内容について一様な理解状態ではない。よって、テストシステムは、各テスト項目が難易度を持っているだけでは不十分で、学習領域の構造情報を持つ必要がある。これにあわせて、受験者の理解状態を表すモデルを持つ必要がある。以下、このモデルを受験者モデルと呼ぶ。システムの構成は、図1のように知的CAIのそのの同等のものとなる。このことから、本テストシステムを知的適応型テストと呼ぶこととする。なお、知的CAIのシステムには評価モジュールを内包するものはあるが、それ自体独立してテストシステムとしては使用できない。

本テストシステムの目指すところは、口述試験のコンピュータ上での表現にある。すなわち、口述試験の利点である個別対応性を保ちながら、効率的に受験者の理解状態を測定し、かつ口述試験の公平性の低さ、受験者を客観的に記録することの難しさ等の欠点⁽⁵⁾を補正するものである。

上述の機能を実現するために本システムでは、学習領域の構造情報を意味ネットワークで記述し、それにアイテムプールを対応させ、受験者の

理解状態は、学習領域の構造情報へのオーバーレイモデルで表わす。

2. 知的適応型テストの開発

2-1 システム概要

本システムの構成を図1に示す。システムの主な動作は次の通りである。

- (1) 教材構造と受験者モデルから情報を取り出して、次に出題する項目を選択する。
- (2) 出題する項目が決定すると項目番号をインタフェース部へ転送し、該当項目をアイテムプールから選択して、受験者に出題する。
- (3) 受験者が回答する。
- (4) 受験者の反応を受験者モデルに追加する。

なお、本システムでは、多肢選択式項目のみを扱った。回答はそれぞれの項目について正答に至るまで続ける方式をとっており、項目の出題には正答に至るまでの入力回数及び所要時間を用いる。これによって、1項目から得られる情報をより多くすることができる。

2-2 学習領域の構造情報

前述のように、今までの適応型テストは受験者の能力というものの存在を想定し、それを測定するというものである。語彙力、単語力といったアイテムプール内での1次元性が高いものには有効である。しかし、構造を持つものの場合、この能力というものを一意に決定することは、必ずしもできない。すなわち、学習領域のある部分集合についてはよくできるが、別の部分についてはほとんどできないといったことが起こりえる。よって、受験者の理解状態を測定するためには、受験者モデルが必要であり、その表現のためには学習領域の構造情報が必要である。本システムでは、構造情報の表現方法として意味ネットワークを用いた⁽⁶⁾。

テスト項目は大別すると「事実」を問うものと、「関係」を問うものがあり、前者をノードに、後者をリンクに対応させる。例をあげると、前者は「4/10を約分しなさい」が、後者は、「1970年の発表時に、SCHOLARが教材としていたものは何か」など図2にあげる項目が該当する。

意味ネットワークのノードには、それぞれの領域を構成する学習要素が対応する。例えば、算数

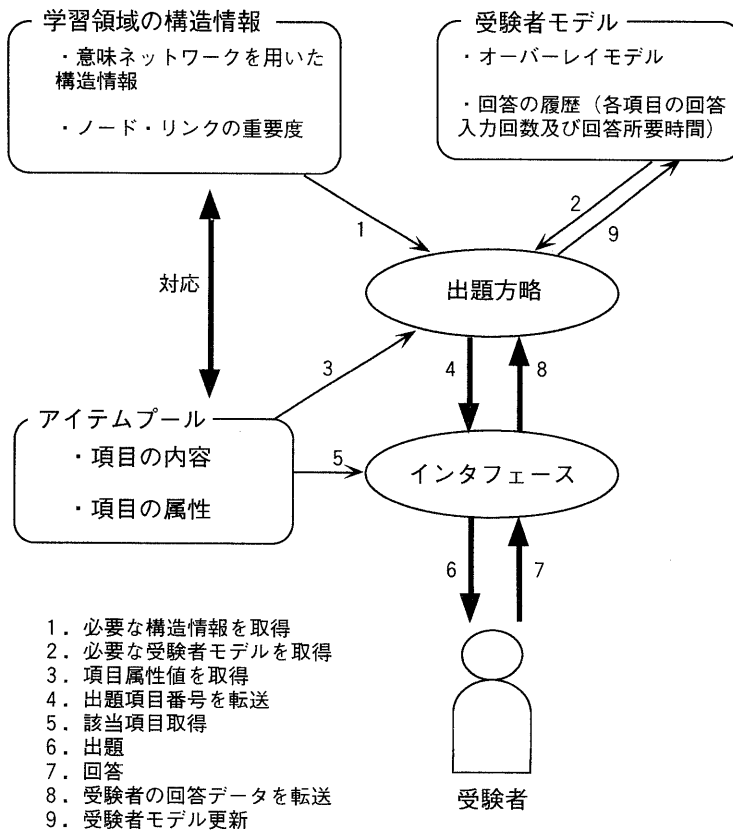


図1 システム構成

の場合は、「分数の約分」・「約分のない分数の足し算」などが、歴史の場合は、「織田信長」・「本能寺」などがノードとなる。

意味ネットワークのリンクの種類は、一般に利用される分野によって異なる。SCHOLAR⁽⁷⁾を参考にし、アイテムプールの利用形態を考慮して、ここでは表1の8種類とした。表中の1)～5)は、いわゆる縦の関係を表わし、6)

～8)は、横の関係を表わす。「1)上位の前提」は、特に数学などで用いられ、領域によって使用する頻度にかかなりの差がでてくる。例をあげるとノード「移項」とノード「一次方程式」の関係にあたる。「2)上位集合」と「3)上位の属性」は、意味的に厳密な区分は難しい。この違いは、上位のノードに対しその下位のノードの全てを出題する必要があるれば「2)上位の集合とな

表1 リンクの種類

リンクの名称	機能
1) 上位の前提(Super Premise)	その要素を習得するために必要な前提となる要素
2) 上位集合(Super Part)	その要素を構成する要素 (AND)
3) 上位の属性(Super Attribute)	その要素を構成する要素 (OR)
4) 上位の性質・特徴(Super Property)	その要素の性質・特徴
5) 上位の例(Super Example)	その要素の具体例
6) 同義(Equal)	同一特徴関係
7) 対義(Opposition)	対立特徴関係
8) 関連(Relation)	関連特徴関係

り、そうでなければ「3）上位の属性」となる。「2）上位の集合」・「3）上位の属性」と「4）上位の性質・特徴」は、丁度、フレーム型知識表現モデルにおける、「2）上位の集合」・「3）上位の属性」がフレームとフレームの関係、「4）上位の性質・特徴」がフレームとスロットの関係にあたる。「5）上位の例」は「4）上位の性質・特徴」の特別な形と考えることができる。なお、表中のリンク2)～5)については、図2に例示する。

コンピュータ上では知識表現言語Prologを用いた。ノード・リンクの関数はそれぞれ次の形をとる。

node(Name,Importance).

link(Name1,Name2,Function,Importance,Code,Difficulty).

関数nodeでは、Nameはノード名、Importanceはノードの重要度を表わす。関数linkでは、Name1及びName2はノード名、FunctionはName1,Name2の関係、Importanceは、そのリンクの重要度、Codeはこのリンクに該当する項目番号、DifficultyはCodeそれぞれの難易度を表わす。Code（項目番号）は、一般的に複数存在するのでリストでかかれる。なお、項目が存在しない場合は空となる。Difficulty（難易度）はCodeと同様にリストでかかれ、それぞれA（最も難しい）～D（最も易しい）で表わす。このA～Dは、次節「受験者モデル」のA～Dと対応している。図2に関数の具体例を示す。項目の難易度A～Dの決定アルゴリズムは、以下の通りである。

(1) 次の5項目を、表2に示すようないくつかのカテゴリーにわけそれぞれに評価値を与える。以下それぞれの評価値を（）内の変数とする。

項目の正答率 (P)

所要時間 (T)

項目の種類 (E₁)

項目の問題文の長さ (E₂)

項目の選択肢の (E₃)

なお、E₁E₂E₃をまとめ項目の表現 $E (E = \sum_{i=1}^3 E_i)$ とする

(2) P,T,Eより、難易度の評価値D (1≤D≤4) を決定する。

$P \geq T$ のとき、PをDとする

$P < T$ のとき、 $(1-E) \times P + E \times T$ をDとする

(3) (2) で得られたDを「A」～「D」の値に変換する。

$1 \leq D < 1.5$ のとき、項目の難易度を「D」とする

$1.5 \leq D < 2.5$ のとき、項目の難易度を「C」とする

$2.5 \leq D < 3.5$ のとき、項目の難易度を「B」とする

$3.5 \leq D \leq 4$ のとき、項目の難易度を「A」とする

アイテムプール「教育情報処理法II」は約200の項目からできており、意味ネットワークの1部を示したものが図2である。このアイテムプールでは、項目は全てリンクに対応する。また、リンク6)～8)は、使用しなかった。

なお、以後、あるノードに対して、そのノードとそのノードの下位にあたる全てをまとめて評価要素と呼ぶ(図3参照)。

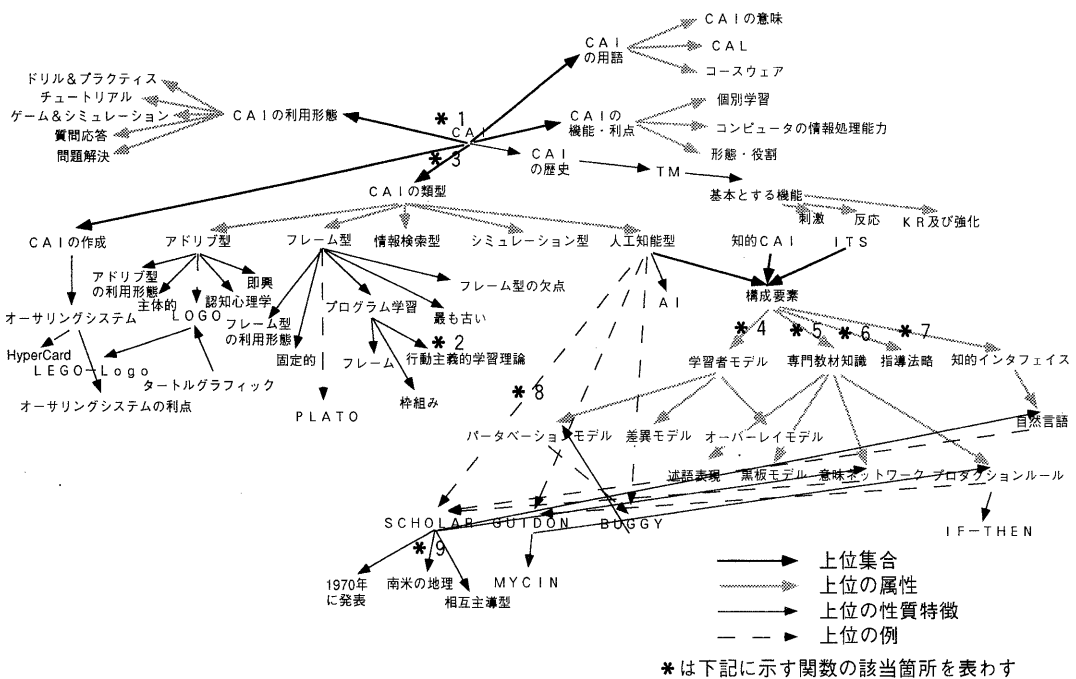
2-3 受験者モデル

受験者モデルは、受験者に応じた出題項目選択と受験者の理解状態を得ることによる有効なフィードバックを目的としている。知的CAIにおける学習者モデルとは、その利用形態が異なる。

本システムでは、受験者の理解状態を意味ネッ

表2 アイテムプール「教育情報処理法II」における項目の難易度の評価

項目	カテゴリー	評価値	
項目の正答率(P)	1.00 ~ 0.75	1	
	0.75 ~ 0.50	2	
	0.50 ~ 0.35	3	
	0.35 ~ 0	4	
項目の平均所要時間(T)	0秒 ~ 50秒	1	
	50秒 ~ 80秒	2	
	80秒 ~ 150秒	3	
	150秒 ~	4	
項目 の 表 現 (E)	不適切な選択肢を回答する項目	0.25	
	2つの命題の真偽の組み合わせ式 上記以外	0.25 0	
	問題文の長さ(E ₂)	0文字 ~ 60文字 60文字 ~	0 0.25
現 (E)	選択肢の種類(E ₃)	単語	0
		単文	0.25
		複文	0.5



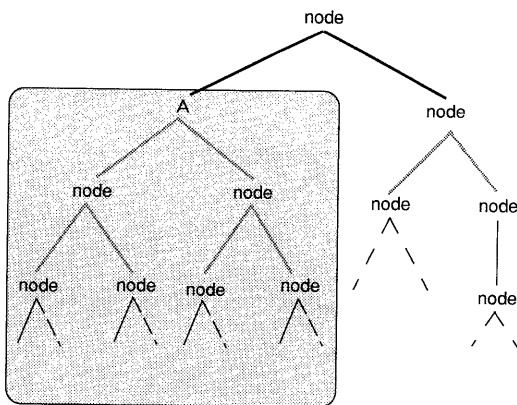
- * 1 node('CAI',10).
 - * 2 node('行動主義的学習理論',3).
 - * 3 link('CAI','CAIの類型',SuperPart,3,[],[]). ←このリンクには該当する項目がない
 - * 4 link('構成要素','学習者モデル',SuperAttribute,1,['M072AH01','M071AT60','M071AT61'],['B','C','C']).
 - * 5 link('構成要素','専門教材知識',SuperAttribute,2,['M072AH01','M071AT60','M071AT62'],['B','C','C']).
 - * 6 link('構成要素','指導方略',SuperAttribute,2,['M071AT60','M071AT62'],['C','C']).
 - * 7 link('構成要素','知的インタフェース',SuperAttribute,2,['M071AT60'],['C']).
 - * 8 link('SCHOLAR','南米の地理',SuperProperty,2,['M071AT67'],['B']).
 - * 9 link('人工知能型','SCHOLAR',SuperExample,1, ['M051AH04','M071AT65','M071AT66','M071AT71'],['B','D','C','C']).
- 1つの項目が、複数のリンクに対応している

M071AT60 (* 4 ~ * 7)
 次の内 | CAI の構成要素でないものはどれか。
 1. 学習者モデル 2. フレーム 3. 指導方略
 4. 専門教材知識 5. 知的インタフェース (正答 2)

M071AT67 (* 8)
 1970年の発表時に、SCHLARが教材としていたものは何か。
 1. 南米の地理 2. 南米の生物 3. 南米の歴史
 4. 南米の岩石 5. 南米の言語 (正答 1)

M051AH04 (* 9)
 南米の地理を教材とし、1970年に発表された、初の本格的な人工知能型CAIといわれるシステムの名前は、次のうちどれか。
 1. SCHOLAR 2. WEST
 3. PLATO 4. LOGO
 5. EXPERT (正答 1)

図2 アイテムプールの意味ネットワーク化の例



評価要素Aは  の部分をさす

図3 評価要素の定義

トワーク表現された教材構造へのオーバーレイモデルで表わした。このオーバーレイモデルでは、各評価要素について修得済みか否かの2段階ではなく、A（完全に修得している）、B（ある程度修得しているが完全ではない）、C（全く未修得ではないが不足している）、D（全く未修得である）の4段階で表わした。

2-4 出題方略

今回開発したシステムの出題は、以下の3段階に分れる。

- (1) 主評価目標を選択する
- (2) 評価要素を選択する
- (3) 項目を選択し、出題する。

(1)における主評価目標とは、出題をする際の出発点となるノードを表わし、実施者が予め入力する。一般的に、主評価目標は複数個ある。

(2)では、主評価目標となっているノードから下位のノードを検索し、そのノードの重要度と受験者モデルに基づいて評価要素を選択する。低い評価値をとると、より重要度の高い主評価目標に近い評価要素すなわち包括的で基礎的な評価要素が選択され、高い評価値をとると、より細かい発展的な評価要素が選択されることになる。

(3)では、(2)で選択した評価要素の評価値を測定する。評価値は、前述のA（完全に修得している）～D（全く未修得である）の4段階である。(2)で選択した評価要素の中から受験者に適当な難易度の項目を出題する。それぞれの評価

要素に対する1項目目の難易度はそれまでの受験者の反応によって決定され、受験者の回答は正答に至るまでの回数、及び回答所要時間で評価され、それによって、次に出題する項目の難易度が決定される。

2-5 インタフェース

問題提示は、ディスプレイで行われ(図4)、受験者の回答の入力はマウスによって行われる。受験者は、項目が提示された後、回答し、採点ボタンを押す。その後、システムが回答を採点し、正答であれば次の項目を出題し、誤答であれば、再び、回答の入力を促す。受験者は、正答するまで項目を解くことになる。なお、1度回答した誤答選択肢には、システムによって印がつけられ、2回以上選択することはできない。各項目に正答すると、正答に至るまでの入力回数と回答所要時間が記録され、以後の項目の出題に用いられる。

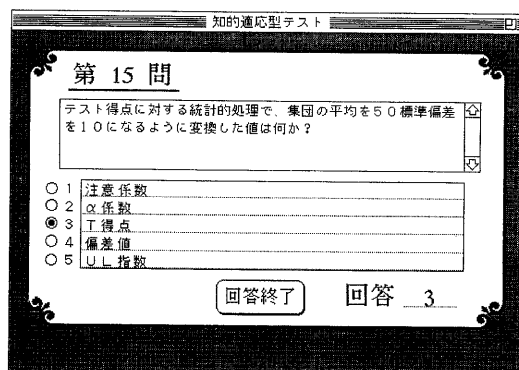


図4 項目出題画面

3. 知的適応型テストの運用

3-1 知的適応型テスト実施前の処理

本システムは今回のテストに合わせて、テストの平均所要時間が40分程度になるように、出題範囲を考慮して各評価要素の出題数(1~3)が設定されている。知的適応型テストは、受験者の理解状態の測定を目的としているので形成的評価に最も適しているが、今回のテストは総括的テストであるので、広い範囲から出題されるように出題方略が設定されている。また、テスト実施者が、教授活動に合致するように教材構造の重要度を設定し、このテストの主評価目標(「教育評価」・「S-P表」・「CAI」)をシステムに入力した。

3-2 知的適応型テストの実施

1993年2月10日に、「教育情報処理法II」の期末試験において、このテストシステムを運用した。受験者は、14人であり、それぞれの受験者に1台のコンピュータを使用させ、一斉に行なった。今回のテストシステムは、コンピュータを用いることや正答に至るまで回答すること等、従来のテストとはかなり異なっているため、本システムに不慣れな人は、本テストシステムでの受験の練習をしてから、テストを開始するようにした。なお、テスト開始直後に、受験者に自己の学力を評価させ、それを最初の項目の難易度の決定に用いた。

キーボードからの氏名の入力を除くとシステムの練習をしてからテストを受験できることもあって、受験者に特に戸惑いはなかった。この観察結果から本システムは、従来のペーパーテストとはかなり異なっているが、実用に十分耐えうるインタフェースを持つと思われる。

受験者への出題項目数は15～30項目で、テストの回答所要時間は、5～33分であった。

3-3 運用結果

動作例として、ある受験者へ出題された項目とそれに対する反応を表3示す。この受験者の場合は、13個の評価要素の中で2個だけが「D」であり、残りは「A」、「B」であることから「D」のついた2個の評価要素は、この受験者の弱点であると解釈され、従来のテストより詳しい情報が得られた。システムの動作として特徴的なところをあげる。「S-P表」については、評価要素「S-P表の性質」について「D」の評価がついている。よって、システムは「S-P表」について、この受験者は、全く理解していないとみなし、普通は出題する評価要素「S-P表の分析指標」について出題せず、次の主評価目標「CAI」に移っている。この判断に、受験者モデルが活かされている。また、17問目以降は、この受験者は「CAI」について、システムは、良く理解していると判断し、難易度の高い項目を出題し、これに受験者も良い成績をとっているため1要素当たり

の出題数が減っている。全体を見ると、評価値が「A」及び「D」、すなわち、受験者が完全に修得している評価要素及び全く修得していない評価要素への出題項目数は、他の評価値の評価要素と比べて少なくなるよう機能している。以上、1人の受験者について述べたが、他の受験者について同様のことがいえ、本システムの受験者に対する評価要素及び出題項目の難易度の選択の適応性が、効果的に機能していると考えられる。

評価要素の評価値を「A」を4、「B」を3、「C」を2、「D」を1として、受験者ごとに平均・分散計算し、その散布図を図5に示す。この散布図は、データより4つの領域に分けた。それぞれは、「成績優秀型」・「偏り型」・「安定型」・「勉強不足型」である。以下に各領域の解釈を示す。

「成績優秀型」の受験者には、学習が十分に定

表3 動作例

項目番号	評価要素	難易度	試行数	所要時間	評価値
M022AT04	良いテスト	C	1	51 ^(秒)	
M021AH13	良いテスト	B	1	34	
M021AH06	良いテスト	A	4	68	B
M022AT09	項目類型	B	1	71	
M022AT11	項目類型	A	3	112	
M022AT07	項目類型	B	2	95	B
M012AH02	良い項目	B	4	95	
M011AT03	良い項目	D	4	81	D
M021AT07	形成的評価	D	2	21	
M021AT08	形成的評価	C	1	11	
M022AT03	形成的評価	B	1	25	B
M062AT14	S-P表の性質	B	5	60	
M062AH02	S-P表の性質	D	3	92	D
M071AT29	CAI用語	D	1	11	
M071AT06	CAI用語	B	1	12	
M071AT26	CAI用語	A	1	12	A
M051AH03	CAIの機能・利点	A	1	116	A
M071AT40	CAIの作成	A	1	59	A
M071AT34	CAIの利用形態	A	1	23	A
M071AT36	フレーム型	A	3	125	
M071AT47	フレーム型	B	1	40	B
M071AT62	ICA Iの構成要素	B	1	66	
M071AT21	ICA Iの構成要素	A	5	107	
M072AH01	ICA Iの構成要素	B	2	31	B
M071AT68	SCHOLAR	B	1	8	
M071AH04	SCHOLAR	A	1	7	A
M071AT42	アドリブ型	A	1	6	A

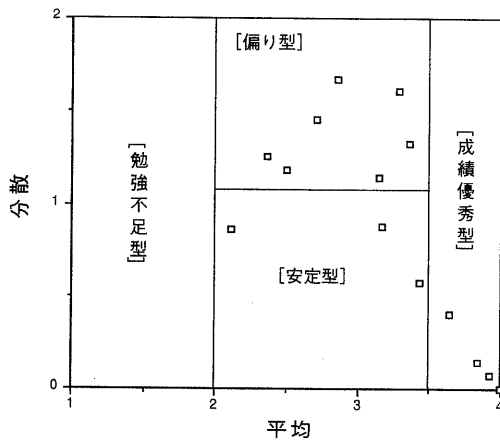


図5 評価要素の評価値の平均一分散平面

着しているのので、今後も持続するよう指示をする。

「偏り型」の受験者には、学習の定着が不十分である部分があるので、そこを重点的に指導する。

「安定型」の受験者には、領域全体的にまだ学習の余地が残されている。

「勉強不足型」は、基礎的なことから学習が必要である。

本システムでは、この中でも特に「偏り型」の受験者の理解状態を測定することができ、その指導のために必要な情報を得ることできた。

今回の運用の場合は、総括的テストであったので、テスト全体から1つの評価値を受験者につける必要があった。その方法としては、各主評価目標ごとに、その評価要素の値から評価値を計算し(今回は、平均をとった)、各主評価目標の評価値の加重平均で求めた。

4. おわりに

本研究では、テストは教授・学習活動を改善できる情報を得るための教育活動という視点で、単に受験者を一元的な尺度だけではなく、受験者の理解状態を教材構造と照らし合わせて測定することを目的とした。そのため、本システムは、教材構造及び受験者モデルを持ち、個々の受験者に評価要素及び項目の難易度を適応させた。システム運用の結果、従来測定が困難であった個人の理解状態を評価することができ、評価要素及び項目の難易度の選択が、効果的に機能していることが確

認された。

今回開発したシステムは、アイテムプールの項目数が少なく、教材構造は作成者の主観で作成されており、出題方略の各モジュールは単純であるが、今回の運用では特に問題はなかった。今後これらが充実することによって、より効果をあげられると思われる。

今回は、「教育情報処理法II」の期末試験に合わせて開発を行なった。しかし、知的適応型テストの考え方は、他の学習領域や形成的テストにもあてはまる。従来のテスト実施前のテスト構成の作業の変わりに、実施する知的適応型テストの情報を入力し、それに応じて出題することが可能であり、これによって汎用性の高いシステムの構築が可能となる。実施する知的適応型テストの情報とは、テストを行なう教科名、評価の種類、実施する項目数、時間、出題範囲、特に強調して評価したい点等である。

《謝辞》

HyperCard (インタフェース部) と Prolog (出題方略部) の間の通信技術について、オーガス情報システム総研の協力を得た。

《文献》

- (1) Wainer, H: "Computerized Adaptive Testing", Lawrence Erlbaum Associates Publishers, (1990).
- (2) 池田央: "テストの科学", 日本文化科学社, (1992).
- (3) 永岡慶三, 植野真臣: "多元的適応型テストシステムのアルゴリズム", 日本教育工学雑誌, 15, 4, pp.157-165, (1992).
- (4) 芝祐順, 野口裕之, 南風原朝和: "語彙理解力測定のための多層適応型テスト", 教育心理学研究, 26, 4, pp.11-20, (1978).
- (5) 印東太郎: "心理学的測定", 金子書房, (1950).
- (6) 永岡慶三, 藤原康宏: "意味ネットワークを用いたアイテムプールの構造化", C A I 学会研究報告, 92, 4, pp.41-44, (1992).
- (7) Carbonell, J. R.: "AI in CAI: An artificial intelligence approach to computer aided instruction", IEEE Transactions on Man-Machine Systems, MMS-11, 4, pp.190-202, (1970).