バイオメディカル・ファジィ・システム学会誌 Vol.17, No.2, pp.51－57（2015）

[Original article]

（2015 年 10 月 9 日　Accepted）

マルコフ決定過程を用いた教授戦略

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：　本研究では，マルコフ決定過程を用いて講義をモデル化し，統計的決定理論に基づく 2 つの教授戦略算出方法を提案する．従来研究で算出される教授戦略は 1 回の講義の効果を最大化するのに対して，本研究で算出する教授戦略は複数の講義で構成される科目全体での効果を最大化する．提案方法では動的計画法を利用している．1 つ目の提案方法は個別指導用の方法で，2 つ目の提案方法は複数指導用の方法である．教育目的はマルコフ決定過程の利得関数によって表現される．また，教授戦略はマルコフ決定過程の政策によって表現される．提案方法によって算出される最適な教授戦略はマルコフ決定過程の期待利得を最大化する．提案方法の有効性を示すために数値計算例を紹介する．
キーワード： 講義，教授戦略，マルコフ決定過程，動的計画法

Teaching Strategies Using Markov Decision Processes

Yasunari MAEDA[1], Masakiyo SUZUKI[1]

1) Department of Computer Science, Kitami Institute of Technology

Abstract: *In this research Markov decision processes are used in order to represent lectures. The best teaching strategy computed in previous research maximizes the effectiveness of a lecture. The best teaching strategy computed in this research maximizes the effectiveness of a subject which is composed of multiple lectures. We propose two computation methods based on statistical decision theory for teaching strategies. Dynamic programming algorithm is used in the proposed methods. The first proposed method selects teaching materials for a learner. The second proposed method selects teaching materials for multiple learners. A purpose of education is represented by the reward function of Markov decision processes. A teaching strategy is represented by the policy of Markov decision processes. The best teaching strategies computed by the proposed methods maximize the expected reward of Markov decision processes. We show the effectiveness of the proposed methods by some numerical calculation examples.*
Keywords: *lecture, teaching strategy, Markov decision processes, 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

本研究では，複数回の講義によって構成される科目に対して，各回の講義においてどの教材を選択（使用）すべきか決める教授戦略を研究対象とする．各回の講義では当該回の講義で教えるべき項目は事前に確定しており，各回の教材は前回までの講義内容の復習と新たな学習項目で構成される．ただし，各回で選択可能な複数の教材（この中から1つを選択）はそれぞれ異なる特徴を有する．特徴としては，基本的な内容の重視，応用的な内容の重視などがある．教授戦略は各回の学習者の理解度に応じて適切な特徴を有する教材を選択する．

教育工学分野において，確率モデルを用いて講義（科目）をモデル化したもとで教授戦略を検討した従来研究[1][2][3]がある．

従来研究[1]ではベイジアンネットワークを用いて講義をモデル化して，大学における対面教育とe-Learningを併用し，教授戦略について検討している．しかし，教授戦略について何らかの理論的保証を与えるような検討は行っていない．

他方，従来研究[2][3]では，ベイジアンネットワークを用いてe-Learningをモデル化し，統計的決定理論[4]に基づき期待効用を最大化する最適な教授戦略の算出方法を提案している．また，効用関数の一例として確率分布で表現された学習者の理解度の変化量を導入している．しかし，従来研究[2][3]は1回の教材選択に対する期待効用の最大化問題を繰り返し解く問題設定のため，講義回数が事前に決まっている科目に対する科目全体での期待効用を最大化する教授戦略を算出することはできない．

事前に講義回数が未定の場合には，従来研究[2][3]の教授戦略が便利である．しかし，事前に講義回数が決まっている場合には，科目全体の教材選択に対して最適な教授戦略も重要である．ベイジアンネットワークは従来から主に確率推論に利用される確率モデルであるため，確率計算に関するアルゴリズムは既に多数提案されている．しかし，科目全体の教材選択に対する最適な教授戦略を算出するためには，有限回の意思決定を伴う最適化問題を解く必要がある．そこで，本研究ではマルコフ決定過程[5][6]（以下，MDPと略す）を用いて講義（科目）をモデル化する．MDPはマルコフ連鎖に制御の要素を加味して拡張された確率モデルで，有限回または無限回の意思決定を伴う最適化問題を解くのに数理工学分野でよく利用されている．MDPの利得関数の設定次第でさまざまな教育目的を表現できる．さらに，統計的決定理論に基づき科目全体に対して最適な教授戦略を算出する方法を提案する．なお，本研究は特定の教育形態に特化した研究ではない．しかし，本研究で提案する2つの方法それぞれで想定される状況がことなるため，その点については個別に後述する．2章で個別指導（学習者が1名）の場合を説明し，3章でN人（$N>1$）の学習者（N人の理解度が同一とは仮定しない）が同じ教材で学習する場合を説明する．

2．個別指導の場合の提案方法

本章では，学習者が1名の個別指導において教材を選択する問題設定を考える．本章の提案方法は個別指導の対面教育（教師と学習者が実際に相対して行う教育形態）や，e-Learning（ネットワークを介した教育形態）を想定した方法である．

2.1　準備

以下で，本研究で使用する記号などを定義する．

s_i， $s_i \in S = \{s_1, s_2, \ldots, s_{|S|}\}$は学習者の理解度を示す理解度状態でMDPの状態に相当する．Sは要素数が有限の理解度状態の集合である．例えば，状態数を$|S|=4$とした場合，s_1が優，s_2が良，s_3が可，s_4が不可で，添え字の番号が小さくなるにつれて高い理解度を示す理解度状態集合が考えられる．a_i，$a_i \in A = \{a_1, a_2, \ldots, a_{|A|}\}$は教材でMDPの行動に相当する．$A$は要素数が有限の教材の集合である．各教材は異なる特徴（基本的な内容重視，応用的な内容重視など）を有する．ただし，各回の講義で新たに学ぶ項目は事前に確定しており，各回の教材は当該回で新たに学ぶ項目と前回までの復習内容で構成されている．各回で新たに学習する項目は異なるが，教材a_iの添え字iは回が異なっても同じi番目の特徴を示し，教授戦略ではどの特徴を有する教材を学習者に提供すべきかを決める．よって，教材集合Aは厳密には教材の特徴集合であるが，本研究では単に教材集合と呼ぶ．

$p(s_k|s_i, a_j, \theta^*)$ は理解度状態 s_i の学習者に教材 a_j を提供した際に，次の期に理解度が s_k になる確率でMDPの状態遷移確率に相当する．θ^*，$\theta^* \in \Theta$ は状態遷移確率を支配する真のパラメータで既知である．Θ はパラメータ集合である．T は学習者に対して教材を選択（提供）する有限の回数（講義回数）を示す．各回で新たに学習する項目は異なるので，回によって状態遷移確率が異なるようなモデル化も可能であるが，本研究では回が異なっても状態遷移確率の値は変化しないと仮定している．ただし，各回で状態遷移確率が異なるようにモデルを拡張することは容易である．

x_t，$x_t \in S$ は学習者の t 期の理解度状態，y_t，$y_t \in A$ は t 期に選択する教材を示す変数である．なお，本研究では学習者が提供された教材で学習した後に理解度確認テストなどによって，学習者の学習後の新たな学習項目と前回までの内容に関する理解度が判定できる（次の期の理解度状態は観測可能）と仮定する．

一般的なMDPでは状態遷移に伴い利得が発生し，利得は遷移元の状態，選択された行動，遷移先の状態で定まる．本研究同様の有限期間の問題設定であれば，有限期間中の全状態遷移に伴う利得の総和を最大化することが多い．教授戦略についても全期間における利得を想定するモデルも検討できるが，本研究では T 回の学習が終了した時点での学習者の成績判定を想定した利得について検討する．具体的には，t 期（$t<T$）には利得は発生せず，学習者の最終的な理解度状態である T 期の遷移先 x_{T+1} に対してのみ利得 $r(x_{T+1})$ が発生すると仮定した．ただし，全期間の利得の総和に対応するようにモデルを拡張することは容易である．

本章では個別指導の学習者の最終的な理解度状態に対する利得 $r(x_{T+1})$ の期待値を最大化する．具体的な教育目的はどのような利得 $r(x_{T+1})$ を想定するかに依存するが，具体的な教育目的と利得との対応例については4章で説明する．

2.2　定式化と教授戦略の算出方法

統計的決定理論に基づいて，最終的な理解度状態に対する利得の期待値の最大化を定式化する．最初に効用関数 $U(d(\cdot,\cdot), x^{T+1}, \theta^*)$ を次式で定義する．

$$U(d(\cdot,\cdot), x^{T+1}, \theta^*) = r(x_{T+1}), \tag{1}$$

ただし，$d(\cdot,\cdot)$ は当該の期の状態 x_t と，当該の期を示す整数 t を受け取って，行動（教材）y_t を返す決定関数で教授戦略に相当する．x^{T+1} は状態系列 $x_1 x_2 \cdots x_{T+1}$ を示す．効用 $U(d(\cdot,\cdot), x^{T+1}, \theta^*)$ は理解度状態の状態遷移確率を支配する真のパラメータが θ^* のもとで決定関数 $d(\cdot,\cdot)$ を用いて教材を選択し，学習者の状態が系列 x^{T+1} のように遷移した場合の利得である．

次に期待効用 $EU(d(\cdot,\cdot), x_1, \theta^*)$ を次式で定義する．

$$\begin{aligned} & EU(d(\cdot,\cdot), x_1, \theta^*) \\ &= \sum_{x_2 x_3 \cdots x_{T+1}} p(x_2 x_3 \cdots x_{T+1} | x_1, d(\cdot,\cdot), \theta^*) \\ &\quad U(d(\cdot,\cdot), x^{T+1}, \theta^*) \\ &= \sum_{x_2} p(x_2|x_1, y_1, \theta^*) \sum_{x_3} p(x_3|x_2, y_2, \theta^*) \\ &\quad \cdots \sum_{x_{T+1}} p(x_{T+1}|x_T, y_T, \theta^*) r(x_{T+1}). \end{aligned} \tag{2}$$

期待効用 $EU(d(\cdot,\cdot), x_1, \theta^*)$ は真のパラメータが θ^* で，学習者の初期の理解度状態が x_1 のもとで決定関数 $d(\cdot,\cdot)$ を用いて教材を選択する場合の利得の期待値である．

本研究では真のパラメータ θ^* が既知なので，期待効用を最大化する決定関数が最適な教授戦略であり，最適な教授戦略 $d^*(\cdot,\cdot)$ は次式で定義される．

$$d^*(\cdot,\cdot) = \arg\max_{d(\cdot,\cdot)} EU(d(\cdot,\cdot), x_1, \theta^*), \tag{3}$$

ただし，右辺のmaxは決定関数 $d(\cdot,\cdot)$ のすべての候補について期待効用 $EU(d(\cdot,\cdot), x_1, \theta^*)$ を計算して期待効用の最大値を算出する記号，argは期待効用を最大化する決定関数を左辺の $d^*(\cdot,\cdot)$ に返す記号である．

式(2)の期待効用の入れ子構造に動的計画法[5][6]（以下，DPと略す）を適用することにより，式(3)を満足する具体的な教授戦略を算出できる．以下に示すDPの処理は，T 期から1期まで遡って行う．

T 期の状態 x_T（全ての状態の候補）に対する処理を以下に示す．

$$V(x_T, T) = \max_{y_T \in A} \sum_{x_{T+1}} p(x_{T+1}|x_T, y_T, \theta^*) r(x_{T+1}), \tag{4}$$

$$d^*(x_T, T) = \arg\max_{y_T \in A} \sum_{x_{T+1}} p(x_{T+1}|x_T, y_T, \theta^*) r(x_{T+1}), \tag{5}$$

ただし，$V(x_T, T)$ は T 期の状態が x_T という条件のもとでの期待利得の最大値，$d^*(x_T, T)$ は T 期の状態が x_T という条件のもとでの最適な行動を示す．

t 期（$1 \le t \le T-1$）の状態 x_t（全ての状態の候補）に対する処理を以下に示す．

$$V(x_t, t) = \max_{y_t \in A} \sum_{x_{t+1}} p(x_{t+1}|x_t, y_t, \theta^*) V(x_{t+1}, t+1), \tag{6}$$

$$d^*(x_t,t)=\arg\max_{y_t\in A}\sum_{x_{t+1}}p(x_{t+1}|x_t,y_t,\theta^*)V(x_{t+1},t+1),\quad(7)$$

ただし，$V(x_t,t)$はt期の状態がx_tという条件のもとでの期待利得の最大値，$d^*(x_t,t)$はt期の状態がx_tという条件のもとでの最適な行動を示す．

上記の提案方法（以下，提案1と呼ぶ）により，学習者の最終的な理解度状態に対する利得の期待値を最大にするという点で最適な教授戦略を算出できる．

3．N人が同じ教材で学習する際の提案方法

本章では，N人（$N>1$）の学習者に対して同じ教材を選択する問題設定について考える．この問題設定は対面教育や，e-Learningの中でも全学習者に同じ教材を提供する方式の遠隔教育を想定した設定である．

3.1　準備

本節では，前章と異なる部分について定義する．x_t，$x_t=(x_{t,1},x_{t,2},\ldots,x_{t,|S|})$は$t$期のMDPの状態を示し，$x_{t,i}$は状態$x_t$において理解度状態が$s_i$の学習者の人数で，$\sum_{i=1}^{|S|}x_{t,i}=N$である．つまり，$N$人の学習者の理解度はたまたま同一の場合もあるが，多くの場合はN人が$|S|$個の理解度にわかれている．$r(x_{T+1})$は最終的なN人の学習結果であるT期の遷移先x_{T+1}に対する利得で，$r(x_{T+1})=\sum_{i=1}^{|S|}x_{T+1,i}r(s_i)$である．

$p(x_{t+1}|x_t,y_t,\theta^*)$は教材$y_t$が提供されたもとで，$t$期の$N$人の理解度状態$x_t=(x_{t,1},x_{t,2},\ldots,x_{t,|S|})$が$x_{t+1}=(x_{t+1,1},x_{t+1,2},\ldots,x_{t+1,|S|})$に遷移するMDPの状態遷移確率である．$N$人の状態遷移確率は1人の理解度の状態遷移確率を利用して計算できる．t期に理解度状態がs_iの$x_{t,i}$人の学習者のうちで$t+1$期に理解度状態がs_jになる学習者の人数は0以上$x_{t,i}$以下の可能性がある．よって，N人の状態遷移確率$p(x_{t+1}|x_t,y_t,\theta^*)$は学習者のいろいろな遷移パターンに該当する確率の総和になる．例えば，t期に理解度状態がs_iの$x_{t,i}$人の学習者のうちで$t+1$期に理解度状態がs_jになる学習者の人数が$n_{i,j}$で表現される遷移パターンに該当する確率は，以下のとおりである．

$$\prod_{i=1}^{|S|}\frac{x_{t,i}!}{(x_{t,i}-n_{i,1})!\,n_{i,1}!}p(s_1|s_i,y_t,\theta^*)^{n_{i,1}}$$
$$\prod_{j=2}^{|S|}\frac{\left(x_{t,i}-\sum_{k=1}^{j-1}n_{i,k}\right)!}{\left(x_{t,i}-\sum_{k=1}^{j-1}n_{i,k}-n_{i,j}\right)!\,n_{i,j}!}p(s_j|s_i,y_t,\theta^*)^{n_{i,j}}$$

本章ではN人の学習者の最終的な理解度状態に対する利得の期待値を最大化する．

3.2　定式化と教授戦略の算出方法

統計的決定理論に基づいて，N人の最終的な理解度状態に対する利得の期待値の最大化を定式化する．最初に効用関数$U(d(\cdot,\cdot),x^{T+1},\theta^*)$を次式で定義する．

$$U(d(\cdot,\cdot),x^{T+1},\theta^*)=r(x_{T+1})=\sum_{i=1}^{|S|}x_{T+1,i}r(s_i).\quad(8)$$

期待効用は，効用部分がN人の利得の形になる点が2章と異なるが，他は2章と同様なので式は割愛する．2章同様に期待効用の入れ子構造にDPを適用して，最適な教授戦略を算出する．

T期の状態x_T（全ての状態の候補）に対する処理を以下に示す．

$$V(x_T,T)=\max_{y_T\in A}\sum_{x_{T+1}}p(x_{T+1}|x_T,y_T,\theta^*)\sum_{i=1}^{|S|}x_{T+1,i}r(s_i).\quad(9)$$

$$d^*(x_T,T)=\arg\max_{y_T\in A}\sum_{x_{T+1}}p(x_{T+1}|x_T,y_T,\theta^*)\sum_{i=1}^{|S|}x_{T+1,i}r(s_i).\quad(10)$$

ただし，$V(x_T,T)$はT期の状態がx_Tという条件のもとでの期待利得の最大値，$d^*(x_T,T)$はT期の状態がx_Tという条件のもとでの最適な行動を示す．

t期（$1\le t\le T-1$）の状態x_t（全ての状態の候補）に対する処理は2章と同様なので式は割愛する．

上記の提案方法（以下，提案2と呼ぶ）により，N人の学習者に同じ教材を提供するもとで，N人の学習者の最終的な理解度状態に対する利得の期待値を最大にするという点で最適な教授戦略を算出できる．

4．数値計算例

提案1および提案2の有効性を確認するために行った数値計算例を報告する．状態遷移確率などは，本来であれば教育形態（対面教育，e-Learning），科目（内容）などを特定し，かつ教育現場でデータ収集して設定すべきであるが，本研究では教育形態や科目を特定せずに著者の教育経験から主観的に設定した．

学習者の理解度状態数は$|S|=4$とし，s_1が優，s_2が良，s_3が可，s_4が不可で，添え字の番号が小さくなるにつれて高い理解度を示す．教材（特徴）数は$|A|=3$で，各教材（特徴）のもとでの理解度状態の遷移確率を表1，表2，表3のように設定した．各表の2行目以降の各行の先頭の状態が遷移元s_iで，1行目の各状態が遷移先s_jである．a_1は高レベルの教材で，優の学習者をなるべく優で維持する特徴（新規学習項目のうちで高レベルの内容を重視）を有する．a_1では前回までの復習や新規学習項目のうちで中および低レベルの内容に割く時間が短いため，良や可の学習者に対しては理解度の低い状態に遷移する確率も高めである．a_2は中レベルの教材で，良と可の学習者をなるべく現状の理解度状態以上の状態へ導く特徴を有する．a_2では新規学習項目のうちで高レベルの内容に割く時間が短いため，優の学習者に対しては優未満の状態に遷移する確率も高めである．a_3は低レベルの教材で，不可の学習者をなるべく可に導く特徴（前回までの復習や新規学習項目のうちで低レベルの内容を重視）を有する．a_3では新規学習項目のうちで高および中レベルの内容に割く時間が短いため，優や良の学習者に対しては理解度の低い状態に遷移する確率も高めである．

表1　$p(s_j|s_i,a_1,\theta^*)$の値

	s_1	s_2	s_3	s_4
s_1	0.85	0.05	0.05	0.05
s_2	0.10	0.60	0.15	0.15
s_3	0.05	0.10	0.60	0.25
s_4	0.01	0.01	0.08	0.90

表2　$p(s_j|s_i,a_2,\theta^*)$の値

	s_1	s_2	s_3	s_4
s_1	0.75	0.10	0.10	0.05
s_2	0.15	0.75	0.05	0.05
s_3	0.10	0.15	0.70	0.05
s_4	0.01	0.01	0.18	0.80

表3　$p(s_j|s_i,a_3,\theta^*)$の値

	s_1	s_2	s_3	s_4
s_1	0.65	0.15	0.15	0.05
s_2	0.05	0.75	0.15	0.05
s_3	0.05	0.10	0.80	0.05
s_4	0.01	0.01	0.80	0.18

学習者の最終的な理解度状態に対する3パターンの利得を表4に示す．この利得の設定パターンが具体的な教育目的の例に相当する．提案1でのパターン1による利得の期待値は最終的な理解度状態が可以上になる確率である．理解度を成績判定に利用することを想定すると，この確率が合格率に相当する．提案2の場合も利得の期待値を学習者数Nで除すと，N人での合格率になる．よって，合格率の最大化がパターン1の教育目的である．パターン2およびパターン3は理解度状態が高いほど高い利得を設定しており，パターン3はパターン2よりも利得の増加傾向が緩やかな設定である．よって，学習者の理解度をなるべく高くすることがパターン2およびパターン3の教育目的である．パターン3については，厳密にはGPA（Grade Point Average）とは状態数が1つ異なるが，基本的な考え方としてはGPAの期待値最大化に相当する．パターン2およびパターン3はともになるべく高い理解度を目指す利得であるが，算出される教授戦略は必ずしも一致しないことも数値計算例で示す．教材を選択する回数は$T=10$とした．

表4　各パターンの利得

パターン	$r(s_1)$	$r(s_2)$	$r(s_3)$	$r(s_4)$
1	1	1	1	0
2	25	5	1	0
3	3	2	1	0

提案1の数値計算結果を表5，表6に示す．利得パターン1が表5, 利得パターン2が表6である. なお, 紙数の都合から結果の一部分を掲載している．利得パターン1では合格率の最大化が目的のため，可以上の理解度状態を目指す教材a_3が多く選択されている．他方，理解度状態が高いほど利得が高くなるパターン2では，各期の各理解度状態において将来的により高い理解度状態が期待できる教材が適応的に選択されている．

表5　提案1によるパターン1の教授戦略

	1期	5期	10期
s_1	a_1	a_2	a_3
s_2	a_3	a_3	a_3
s_3	a_3	a_3	a_3
s_4	a_3	a_3	a_3

表6　提案1によるパターン2の教授戦略

	1期	5期	10期
s_1	a_1	a_1	a_1
s_2	a_2	a_2	a_2
s_3	a_2	a_2	a_2
s_4	a_3	a_3	a_3

提案2の数値計算結果を表7に示す.

表7　提案2による教授戦略

パターン	t	$x_{t,1}$	$x_{t,2}$	$x_{t,3}$	$x_{t,4}$	教材
1	10	8	2	6	4	a_3
2	10	8	2	6	4	a_1
3	10	8	2	6	4	a_2
1	9	17	3	0	0	a_1
2	9	0	0	15	5	a_3
3	9	0	0	15	5	a_3
1	5	2	10	0	8	a_3
2	5	2	10	0	8	a_3
3	5	2	10	0	8	a_3

学習者数は$N=20$とした. 表7も結果の一部分である．10期の各理解度状態に学習者が適度に分散した状態では，各利得パターンに適した教材が選択されている．パターン2およびパターン3では理解度状態に応じた利得の増加傾向の違いにより，選択される教材が異なる．9期では，利得パターン1でも優の人数が多い場合には教材a_1が選択され，利得パターン2およびパターン3でも可や不可の人数が多い場合には教材a_3が選択されている．また，5期のように残り期間が長く，かつ不可の人数が多い場合には，利得パターンによらず教材a_3が選択されている.

事例数は少ないが，上記より，提案1および提案2によって，各教育目的（利得の各設定パターン）のもとで適応的に教材を選択できることを確認した.

5．考察と今後の課題

本研究では，講義（科目）をMDPを用いてモデル化し，学習者の合格率の最大化などの教育目的に適した教授戦略を算出する方法を提案した．統計的決定理論に基づいて最適な教授戦略を算出するという点では, 従来研究[2][3]と本研究は同じである．従来研究は1回の教材選択に対する最適化問題を解いているので，事前に全体の講義回数が未定の問題設定に適している．他方，本研究の提案方法では，事前に全体の講義回数が既知の科目について，科目全体に対して期待利得を最大化するという点で最適な教授戦略（教材選択）を算出している.

事例数は少ないが数値計算例によって，提案方法において各教育目的のもとで適応的な教材の選択が可能なことを確認した．より詳細な有効性確認には，教育現場における各種データ収集や検証実験が必要である. 具体的な教育現場との連携については今後の課題としたい．また，今回の数値計算例では提案1および提案2で教材の特徴数を同じ3としたが，提案2では学習者数の増加（マルコフ決定過程の状態数の増加）に応じて特徴数も増加させることによるさらなる教育効果（より大きな期待利得）も期待できる．これは教材作成コストと教育効果のトレードオフの問題であり，有力な今後の課題の一つである.

実用化に際しては, 教材作成にともなう課題が多い. 講義の各回の教材は特徴数に応じて用意する必要があり，すべての教材を教員一人で作成するのは難しい. よって，多数の教員の参画が必須である．また，理解度の状態遷移確率の設定については，教材作成者または他の協力教員の経験に基づいて設定する方法と，教

育現場において実際に教材を使用した際の実データに基づく設定方法の2つが考えられる．前者では主観的な設定になるという問題点があり，後者では信頼性を高めるためには多数の教育現場の協力が必要となる．これらの教材作成にともなう課題を解決するためには多数の教員が参加したもとで円滑に機能する枠組みの検討も今後必要になる．

本研究では，学習者の理解度状態の遷移確率を支配する真のパラメータが既知で，学習者の各期の理解度状態が（理解度確認テストなどによって）観測可能と仮定した．しかし，真のパラメータが未知の場合や，理解度状態が観測不可能な場合の方がより現実に近い問題設定である．これらの問題設定については，ベイズ統計学，隠れマルコフモデル，部分観測 MDP などの視点から本研究を拡張することによって対応できると考えられるが，具体的な検討は今後の課題としたい．

本研究では教育工学の視点から検討したが，本研究で解いた問題設定を数理モデルの視点から見直すと，当該問題設定は教育工学分野に限らずヘルスケア分野にも適用可能な問題設定である．MDP では期待利得の最大化が主な目的であるが，本研究のような利得を設定すると，期待利得の最大化によって目的に適した状態制御を実施できる．パターン1の利得が顕著な例であるが，合格率の最大化とは学習者の最終的な理解度を可以上の状態に滞在させる確率の最大化である．よって，本研究で扱った問題設定は教育工学に限らず，何らかの好ましい状態に滞在する確率を最大化することが目的になるような実問題に適用可能な問題設定である．ヘルスケア分野で考えると，好ましい健康状態に滞在する確率が最大になるように食事／薬／生活習慣などを選択（推薦）するような実問題に適用可能である．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] 福島潤一郎，藤原祥隆，前田康成：確率的推論を基礎とする学習成績マップを利用した対面教育適応化法，信学技報，ET2008-117，pp.139-143，2009.

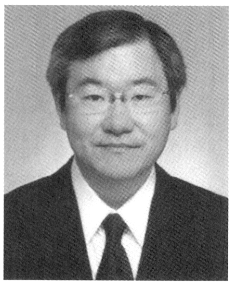
[2] 岡本敏雄，香山瑞恵：人工知能と教育工学，オーム社，2008.

[3] Ueno, M. : Intelligent Tutoring System based on Belief networks, Proc. Of IEEE, Computer Science, pp.141-142, 2000.

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

[5] Martin, L.P. : Markov Decision Processes, John Wiley & Sons, 1994.

[6] 森村英典，高橋幸雄：マルコフ解析，日科技連，東京，1979.



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．