未知で不確実な環境に対する認知特性の意味と応用
The Meaning and Application of Cognitive Characteristics in Unknown and Uncertain Environment

甲野 佑†, 高橋 達二‡
Taro Nagoya, Hanako Nagoya

† 東京電機大学大学院, ‡ 東京電機大学理工学部
Tokyo Denki University, Tokyo Denki University

yu.kohno.02@gmail.com

Abstract

An Loosely Symmetric model (LS) is as a subjective probability model that came from human beings' cognitive characteristics. We developed value function LSX which expanded LS, to investigate an adaptive meaning of human beings' cognitive characteristics in the decision-making. Furthermore, we derived a conditional expression to switch exploration and exploitation in LSX. Therefore, relationships between human beings' cognitive characteristics and decision-making strategy were able to be described quantitatively.

Keywords — Symmetric reasoning, Decision-making, Multi-armd bandit problem, speed-accuracy trade-off.

1. はじめに

未知の環境においてより多くの報酬を獲得するためには，選択肢となる行動を探索的に試行して情報を獲得し，選択肢を評価してより効率的な選択肢を発見していく必要がある．正確に価値を評価しようとする場合，何度も探索的試行を行い，より多くの情報を獲得する必要がある．しかし，探索してばかりでは高い報酬獲得を妨げるため，多くの報酬を得るためには探索的試行はある程度に収めて，獲得した知識を利用した利益追求行動を行わなければならない．そのため，限られた時間の中での探索行動と利益追求行動のバランシングが困難であるとされる．これを探索と知識利用のジレンマと呼び，意思決定課題における速さと正確さにはトレードオフの関係がある事を表している[1]．ϵ-greedy, softmax, UCB1 等[2][3], 数学的，統計学的に探索と知識利用を上手く使い分けるような手法が従来から考案されて来た．しかし人間は複雑な統計学的背景を持たなくても限られた時間の中で探索と知識利用を適宜調整して未知の環境に対応している．つまり数学的において非規範である人間の認知特性が未知な環境における優れた性質を有する可能性がある．

$$LS(e|a) = \frac{P(a,e) + S_p}{P(a,e) + S_p + P(a,\bar{e}) + S_n} \quad (1)$$

$$\text{Positive bias}: S_p = P(\bar{e})P(a|\bar{e})P(\bar{a}|\bar{e}) \quad (2)$$

$$\text{Negative bias}: S_n = P(e)P(a|e)P(\bar{a}|e) \quad (3)$$

本研究では人間の評価感覚と一致する数理モデルとして，篠原が確信度形成のモデルとして考案したLoosely Symmetric model(以下LS, 式1)[4]に着目する．LSは任意の原因事象Aと結果事象eの生起・不生起によって記述され，2要因間の因果帰納課題，2つの選択肢に対する意思決定課題 (2本腕バンディット問題) に対して良い成績を持つ認知的な確率モデルだとされていた．またその特性として視覚における着目対象と周辺視野との類似性 (地の不変性) が指摘されている[5]．我々はLSモデルを主観的な確率の評価関数として解釈し，複数の選択肢に対する一般化を行った (Normalized LS)．また評価基準値をパラメータ化したモデルを考案して動的な学習をする事で成績が飛躍的に上昇する事を示した(LS-VR)[6]．本研究で我々は報酬となる事象の種類を生起不生起の二種のみでなく任意の数に一般化したExtended LS(以下LSX)を考案した．またシミュレーション実験を通してLSXの非定常環境における有用性を示し，更に探索と知識利用を使い分ける条件式を導出した．この条件式を用いる事で，従来難しかった“個人” の意思決定における選択系列と直接的な比較が可能になると考えられる．

2. 意思決定課題 -多本腕バンディット問題-

本研究では単純な意思決定問題の一種である多本腕バンディット問題を例に，何も情報の無い状態からトレードオフを抱える課題，環境に対し主体的に情報を獲得して行く際の不確実な知識の扱い方や値付けを論じる．ここでの不確実な知識とは観測の不十分さに由来する曖昧な知識を意味す

る. これは強化学習課題における初期において学習を促進するためにどのような方策や価値観数を用いるかの問題に対応する[2]. 多本腕バンディット問題とは目的となる報酬を確率的に得る事の出来る幾つかの手段(腕)a_iから最適な手段を探索し, 得られる報酬eを最大化させる事を目的とする問題である. 表1はバンディット問題で扱われる変数に対する確率的な表現である.

表 1 事象A, E間の完全結合分布

	e	$\bar{e}$
a_1	$P(a_1 \cap e)$	$P(a_1 \cap \bar{e})$
a_2	$P(a_2 \cap e)$	$P(a_2 \cap \bar{e})$
$\vdots$	$\vdots$	$\vdots$
a_n	$P(a_n \cap e)$	$P(a_n \cap \bar{e})$

生き物が効率的に生きるためには, 度々このようなバンディット問題的な課題に直面する. この課題の難しさは探索と収穫のジレンマという単語で表される. 高い報酬を得るためにはどこかで探索を辞めるべきである. しかし探索しなければ高い報酬を得る事はできない. 多本腕バンディット問題はこのような知識の獲得とその利用からなる普遍的な“早さ”と“正確さ”のトレードオフを端的に表す事が出来る.

2.1 選択収束状態

本論文では議論を簡略化するために, いずれかの選択肢の選択された割合がほぼ100 %になる状態を“選択収束状態”と定義する($\exists_{a_i \in A} P(a_i) \approx 1.0$). 言い換えると, ある選択肢に執着して他の選択肢を相対的に殆ど選択していない状態を意味する. その執着している選択肢が真に最も期待値の高い正解の選択肢である場合, 期待損失の上昇が止まり, 上限が決定する. 期待値が最も高い訳では無い誤った選択肢に執着してしまっている場合, その状態から抜け出せなければ期待損失が上昇し続ける.

3. 意思決定における人間の特性と方策

人間は複雑な数学的, 統計学的知識を持たなくても, 速さと正確さのトレードオフに対応する能力を有している. それは未知の環境での行動選択を迫られた際に, 適応的, 経験的に獲得するのかもしれないし, 高度な認知能力を有する生物に先天的に備わっているのかもしれない, しかしながら, 少なくとも生物が自然環境に対して進化的に獲得した能力であると考えられる. そのような意思決定における特性は数多く存在するが, ここでは後に示すLSとの関連が深い3つの特性を挙げる. これらはHattori[7]やTenenbaum[8]の主張にも関連があり, 諸々の認知特性の中でも特に重要な特性だと考えられる.

3.1 相対評価

人間は手段a_1を試行した際に, 報酬eが得られなかった際, その他の手段であるa_2に対する評価が上昇する. 逆に, a_1で報酬が得られた際, a_2に対する評価を下げる傾向がある. このように, 一つの手段に対する試行結果が関係の無い(正確には関係あるかどうか解らない)他全ての手段の評価に影響する事は, 規範的な論理学から導出されない. しかし, 人間はよくこのような評価をしてしまう[9]. このように選択可能な手段の間に相対的な関係を想定して評価する形式は相対評価と呼ばれ, ある手段が上手くいけば其れに執着し, 上手くいかなければ他の手段を試すよう促す効果を生む. これは正に“報酬の最大化”と“探索”を毎時の個別的な試行からバランシングしているに等しい. 即ち, 対称性推論では論理的に関係があるとは限らない事象を関連づける事で, 客観的, 絶対的な評価ではなく, 主観的, 相対的な評価を可能としている.

3.2 信頼性考慮

信頼性考慮とは, 評価の期待値のみでなく, サンプル数による信頼性を評価値に考慮する事である[10]. 人間は確率的に等しい期待値と観測される選択肢に対して, サンプル数の相対的な比率で評価が異なる場合がある. また, サンプル数によって評価値の順位が逆転する事もある. サンプル数の大きさは, その選択肢の客観的に観測した期待値がどれだけ信用できるかを表している. 統計的知識を持たなくても, 相対的な比率を参照することで評価に異なりを与える事が出来る点で優れている.

3.3 規準充足化

人間は評価を連続値ではなく“良い”と“悪い”等に緩く離散化する性質がある. 連続値を離散化するには基準値(Reference value)が必要となり, 規準と個々の選択肢との間の相対評価によって評価値の二値化が行われる[11]. また, その規準そのものも全体の評価値の分布や経験等から形成される.

この評価値を離散化する性質によって，規準を満たす“良い”選択肢を見つけるまで探索するという方策が表現される．更に評価値の離散化と信頼性考慮を組み合わせる事で，“良い”評価が多い場合と“悪い”評価が多い場合によってリスク忌避とリスク追及という真逆の傾向にわかれるという反射効果が表される．この時，二値化の規準が反射効果の基準値となる[12].

4. Extended Loosely symmetric model

著者らが新たに考案したLSXは，複数選択肢への一般化と，規準値を動的に変化させられるようなパラメータ化が施されている(式9)．変数 R は基準点であり，以下の漸化式により，選択した選択肢 a_{chose} の報酬獲得の標本平均 (試行錯誤によって獲得された報酬獲得確率) から漸進的に学習する．ここでパラメータ α は学習率であり，新しい情報をどれだけ重視するかを決定する．パラメータ μ はある種の規格化定数であり，後述する ω_i の理論的最大値の逆数から $\mu = 2$ になる．また，ここで$R = 0.5$に固定すると式12に示される一般化LS(LSN)になる．

$$a_H = \arg\max_{a_k} P(a_k), a_L = \arg\min_{a_k} P(a_k) \quad (4)$$

$$S(e) = \frac{P(a_H \cap e)P(a_L \cap e)}{P(a_H \cap e) + P(a_L \cap e)} \quad (5)$$

$$S(\bar{e}) = \frac{P(a_H \cap \bar{e})P(a_L \cap \bar{e})}{P(a_H \cap \bar{e}) + P(a_L \cap \bar{e})} \quad (6)$$

$$W_S = S(e) + S(\bar{e}) \quad (7)$$

$$V_S(e) = \frac{S(e)}{S(e) + S(\bar{e})} \quad (8)$$

$$LSX(e|a_i) = \frac{P(a_i \cap e) + \mu R W_S + (1-\mu)S(e)}{P(a_i) + W_S} \quad (9)$$

$$R_0 = 0.5 \quad (10)$$

$$R_{t+1} = (1-\alpha)R_t + \alpha P(e|a_{chose}) \quad (11)$$

$$LSN(e|a_i) = \frac{P(a_i \cap e) + W_S - S(e)}{P(a_i) + W_S} \quad (12)$$

ここで更に信頼性考慮と関連する重みを式13，規準充足化に関する項を式14，相対評価に関する項を式15と定義する事により，LSXは式16として三つの特性に分離した式として整理される．

$$\text{RC wight}: \omega_i = \frac{W_S}{P(a_i) + W_S} \quad (13)$$

$$\text{RS差分}: \sigma_i = R - P(e|a_i) \quad (14)$$

$$\text{RE差分}: \eta_i = V_S(e) - P(e|a_i) \quad (15)$$

$$LSX(e|a_i) = P(e|a_i) + \omega_i\big(\mu\sigma_i + (1-\mu)\eta_i\big) \quad (16)$$

4.1 RC weight

RC wight (式13) とは信頼性考慮(Reliability Consideration) に関係する重み係数であると解釈できる．この重みの役割は後ろのRS差分とRE差分の強さを抽象選択肢の試行回数に応じて修飾する事で信頼性を評価値に反映させる事にある．上述の式では，重みの値域は$0 < \omega_i < 1/2$ になり，着目選択肢a_iの試行割合$P(a_i)$が増える程に減少する．選択収束状態である$P(a_H) \to 1.0$のとき$\omega_H \approx 0$になり，後ろの二項の影響がなくなる．同様に$P(a_L) \to 0.0$のとき$\omega_L \approx 1/2$になり，後ろの二項の影響が最大になる．

4.2 RS差分

RS差分(式14) は規準充足化 (Reference Satisficing) に関係する項であると解釈できる．選択収束状態において満足化に寄与するための項．しかしこの項自体がやっているのは基準値へと近似する反射効果であり，“中庸化”と呼ぶべき物である．即ち，$\omega_i \approx 1/2$となる選択肢を 基準値 に近似する事によって，基準値を越える評価値を持つ選択肢が無い時は探索し，逆に基準値を越える評価値があればその選択肢に執着するという規準充足化の振る舞いを間接的に表現している．

$$\lim_{P(a_H)\to 1.0} LSX(e|a_H) = P(e|a_H) \quad (17)$$

$$\lim_{P(a_L)\to 0.0} LSX(e|a_L) = R \quad (18)$$

4.3 RE差分

相対評価 (Relative Estimation) に関するRE差分の式 (式15) は抽象的な期待値$V_S(e)$と任意の選択肢の観測報酬確率を差分する事で，相対評価的な性質を評価値に与えているのだと考えられる．係数が負になるため，この項は中庸化に抗う項であり，選択肢が2つのときは選択収束状態においてRE差分の値は0になるので基本的には不要な項だと考えられる．しかし選択肢が3つ以上の場合では．選択収束状態において殆ど選択されない選択率ほぼ0%の選択肢はRC wight とRS差分の影響でリファレンス値に収束する．そこにRE差分は抽象期待値との差分だけ値を上昇させる．つまり選択収束状態においては選択率ほぼ0%の中でも最も高い選択肢が選択され易くなる．

5. 実験1 -定常多本腕バンディット問題-

LSXの性能を検証するため，多本腕バンディット問題におけるシミュレーション実験を行った．比

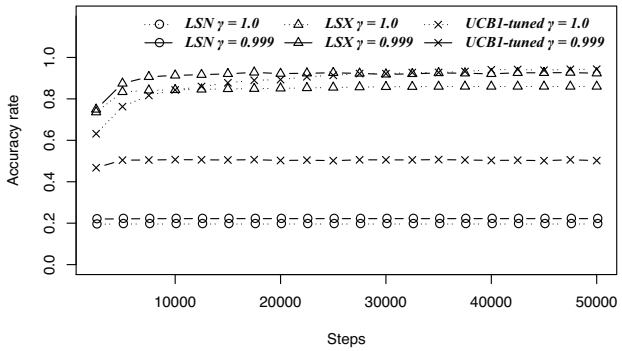

図 1 正解率の推移：定常20本腕バンディット問題

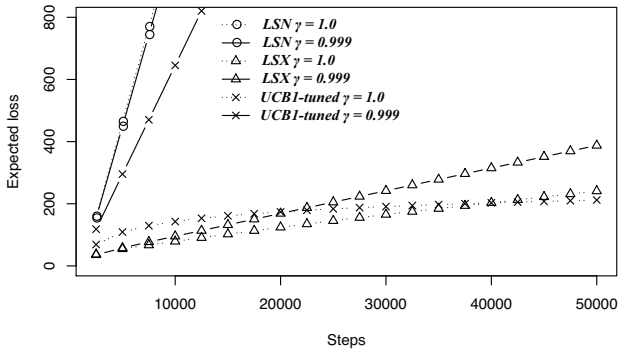

図 2 期待損失の推移：定常20本腕バンディット問題

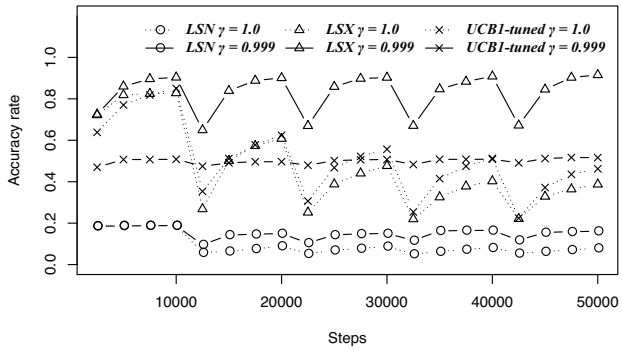

図 3 正解率の推移：非定常20本腕バンディット問題

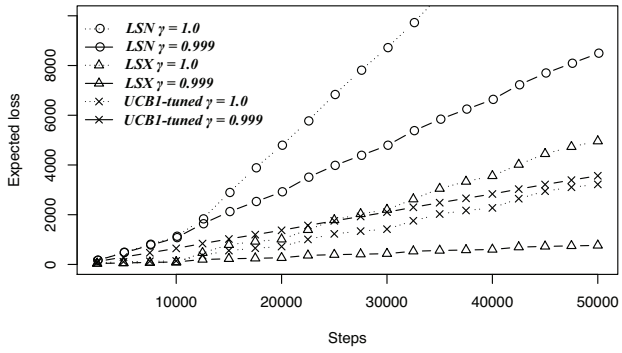

図 4 期待損失の推移：非定常20本腕バンディット問題

較モデルにはLSXの拡張前モデルである一般化LS (LSN, 式12) と，多本腕バンディット問題で最も高い成績を有するアルゴリズムの一種である UCB1-tuned アルゴリズムを用いた[13]．また，情報の更新には過去の情報を少しずつ忘れさせるパラメータである割引率 $\gamma = 0.999$ を用いた場合，割引率を用いない場合($\gamma = 1.0$)で比較した[14]．本実験では選択肢を20個とした．実際に選択して報酬の生起･不生起を観測するまでを 1 step として，実験ではそれを50,000 step 繰り返した．これをシミュレーション1回として，1,000回行い，結果を平均した．各選択肢の真の報酬確率はシミュレーション1回毎に再設定され，一様分布からランダムに設定される．本実験は二種の指標によって評価した．一つは正解率であり，それはその step に真に最も高い報酬確率を持つ選択肢を選択できていたかを評価する．もう一つは期待損失であり，始めから理想的な選択を行えていた場合からどれだけ報酬を損なったかを評価する．

5.1 結果と考察

図1に正解率の推移を示す．約30,000 step までは割引率$\gamma = 0.999$を用いたLSXの正解率が最も高かった事がわかった．以降は UCB1-tuned アルゴリズム の正解率がLSXを勝るものの初期の成績があまり高く無い事を鑑みると単純には比較できない．また，UCB1-tunedアルゴリズムにおいて情報の更新に割引率を用いる事が出来ないのに対して，LSはむしろ割引率を用いた方が正解率が向上している．これはLSXが忘却という人間の基本的な能力を複雑なアルゴリズムを用いずに扱える事を示している．UCB1は統計的な背景を有しているからこそ，忘れる等をパラメータ的に表現できず，より複雑なメカニズムを想定する必要がある．

図2に示したのは期待損失の推移である．正解率とは異なり，割引率を用いない方がLSXの期待損失を抑える事ができた．これは割引率を用いる事で探索行動の頻度を抑えきれないことを示している定常環境において対数的に期待損失を抑えるには基準値の更新手法や割引率，あるいはそれ以外の工夫が必要になる事を示している．しかしこの性質は選択肢の真の報酬確率が常に一定ではない非定常環境下では有用に働くと考えられる．それを次に行う非定常環境かでのシミュレーション実験で検証する．

6. 実験2 -非定常多本腕バンディット問題-

前述の実験結果から，LSXの特性は長期的な正確さより，動的な環境において優れた性質を有するのだと思われる．そのため，本研究では選択肢の報酬確率が変化する非定常な環境における多本腕バンディット問題に関してもシミュレーションを行った．シミュレーションの設定，比較モデル，指標等は前述した実験1と同様にした．ただし，選択肢の真の報酬確率は 10.000 step 毎に一様分布からランダムに再設定されるようにした．

6.1 結果と考察

前述の仮説の通り，割引率を用いたLSXは非定常環境に対して優れた成績を示した．図3にその正解率の推移を示す．割引率$\gamma = 0.999$を用いたLSXは最初に正解率が変化する10,000 stepまで最も高い正解率を示しており，また，報酬確率の変化の後も高い正解率を保っており，4度環境が変わった後も高い正解率に回復している．UCB1-tunedアルゴリズムは情報の更新に割引率を用いると選択が収束せずに成績が下がるため，$\gamma = 0.999$ では環境の変化の際に成績が下がらないまでも，高い正解率を得るには至らなかった．また割引率を用いなければある程度は環境の変化に対応するが，変化に度に徐々に正解率の回復が減少していた．図3に示す期待損失の推移でも割引率を用いたLSXが最も損失を抑えられた．定常環境(図1)の結果よりも他のアルゴリズムとの差が明確に示されており，LSXが環境の変化を解り易い信号で検出できないようなより現実的な環境において優れた価値関数である事がわかった．

7. 議論

$$E \ni \{e_1, ..., e_m\} \tag{19}$$

$$A \ni \{a_1, ..., a_n\} \tag{20}$$

$$\sum_{e_k \in E} LSX(e_k|a_i) = 1 \tag{21}$$

ここまでLSXの基本的性質と，単純な意思決定課題の一種である多本腕バンディット問題において優れた成績を有する事を示した．しかしながら，著者らは以前にLS-VRというLSXとは異なるLSの拡張モデルを考案していた[6]．LS-VRはLSXと同じく基準値Rを動的に変更可能で，複数の選択肢に対してバイアス項が地の不変性を満たしている．またシミュレーション上での成績も変わらない．両モデルともLSの一般化モデルとして，多本腕バンディットモデル上での能力には違いが無い．LS-VRに無くLSXにある性質とは排中立を満たす事が挙げられる．それは報酬が複数種類存在する場合にも同様である(式21)．また，式16にある通り，LSXは相対評価，信頼性考慮，規準充足化の影響を性質を重み付き平均の形式で定量化できる．この性質を利用して意思決定における探索と知識利用を使い分ける条件を記述できる．式23はLSXにおいて知識利用を行う条件である(LSにおいても同様)．この条件式を式16を用いて変形すると，式24が導出される．式25，26，27は式24の左辺の各項を任意の性質に対応付けて分離した式である．エージェントが知識利用する際に選択される選択肢a_{max}を除く全ての選択肢に対して，これらの式の合計が負の値を取る時，エージェントは知識利用を行う．逆に一つでも正の値を取る選択肢が存在する時，エージェントは探索を行う．

$$a_{max} = \arg\max_{a_k} \big(P(e|a_k)\big) \tag{22}$$

$$LSX(e|a_{max}) > \forall_{a_j \in (A \cap \overline{a_{max}})} LSX(e|a_j) \tag{23}$$

$$\begin{aligned} &-\tfrac{1}{\mu} P(a_j)\big(1-+\tfrac{P(a_{max})}{W_S}\big)\big(P(e|a_{max})-P(e|a_j)\big) \\ &+(1-\tfrac{1}{\mu})\big(P(a_{max})-P(a_j)\big)\big(P(e|a_{max})-V_S(e)\big) \\ &-\big(P(a_{max})-P(a_j)\big)\big(P(e|a_{max})-R\big) < 0 \end{aligned} \tag{24}$$

$$-\tfrac{1}{\mu} P(a_j)\big(1+\tfrac{P(a_{max})}{W_S}\big)\big(P(e|a_{max})-P(e|a_j)\big) \tag{25}$$

$$(1-\tfrac{1}{\mu})\big(P(a_{max})-P(a_j)\big)\big(P(e|a_{max})-V_S(e)\big) \tag{26}$$

$$-\big(P(a_{max})-P(a_j)\big)\big(P(e|a_{max})-R\big) \tag{27}$$

式25は信頼性考慮と関係のある量を意味する．ここに含まれる$\big(P(e|a_{max})-P(e|a_j)\big)$は選択肢$a_{max}$と$a_j$の観測された価値の差分を表しており$a_{max}$の定義(式22)により式25は必ず負の値を示す．またその強さはパラメータμと$P(a_{max})/W_S$によって調整されている．即ち，$P(a_{max})$の値が大きくなる程に式25の値は大きくなる．これが信頼性考慮と関連がある点である．知識利用のベースとなる値であり，探索する場合は後続の式26，27によって式27によってこの値を打ち消す必要がある．式26は相対評価と関係のある項であり，選択肢a_{max}とa_iの選択割合(P(a_{max})，$P(a_i)$)と，抽象化された価値$V_S(e)$と$P(e|a_{max})$との大小関係によって符号が決定する．式8に示されている本研究における設定では $P(e|a_{max}) > S(e)$ という関係になるため，$P(a_{max}) > P(a_i)$であれば式26は正の値になり，探索に対する促進になる．また逆であれば抑制とな

る．式27は規準充足化と関係のある項であり，式26においての$P(a_{max})$と$P(a_i)$の大小関係に符号が依存する．さらに$P(e|a_{max})$と基準値Rの大小関係にも依存する．$P(a_{max}) > P(a_i)$の際には前述の大小関係は直接的に規準充足化を示し，現状観測される中で最も高い価値$P(e|a_{max})$が基準値Rを上回っておれば探索を抑制して，逆であれば促進する．また，$P(a_{max}) < P(a_i)$であれば前述の探索と抑制が逆転する．

式25，26，27の間の量的な重みはパラメータや抽象価値$S(e)$や抽象選択割合W_sに依存する．そのため，この条件式がLSXに関する厳密な分析にはならない．しかしながら，この条件式が導出された事で，逆算的に最適なパラメータや関数を設定する事が可能だと思われる．また，パラメータを調整する事で意思決定課題(多本腕バンディット問題)における人間の選択系列を記述する事も可能になる．LSXを意思決定課題における人間の選択系列の記述に用いる利点は，人間の選択系列を記述できるとされるsoftmax [15]とは異なり，乱数を用いない事が挙げられる．また，個人の選択系列に併せてLSXのパラメータを調整することにより，LSXにおいて重要な基準点Rの推移を実際の人間に学び，その動的な学習法に新たな示唆を与えられる可能性がある．具体的には基準値Rを徐々に下げる事によって，価値に応じて選択確率を割り振り，徐々に偏って行くsoftmaxの焼き鈍し的作用をもたらすことが出来る．また初期において基準値Rを高く設定し，後期において急激に低く下げる事であるタイミングで探索と知識利用を切り替えるような方策も表現できる．

8. 総合考察

本研究では過去のLSの分析を元に，LSの拡張モデルであるLSXを考案した．また，シミュレーションを通してLSXが非定常環境において高い成績を有する事を示した．これらの結果は本研究で扱ったLSXが過去に著者らが考案したLS-VRと同等の能力を有する事を表している．本研究では更にLSXがLSの持つ意思決定における人間の3つの認知特性からなる方策を，式の上で定量的に分解できる事を示し，さらにLSXを用いたアルゴリズムにおいて探索と利益追求を使い分ける条件を数式によって表現した(式24)．意思決定におけるLSXの探索と利益追求のスイッチングを条件式に落とし込めた事で，価値観数による記述では不可能だった実際の人間の選択系列との比較が可能になると考えられる．個人の意思決定の系列を乱数を用いず，パラメータの動的な変化で記述する事ができれば，今後の認知的研究に新たなアプローチをもたらす事ができる．また，人間のパラメータの推移をアルゴリズム反映させる事で，LSXを用いたアルゴリズムの更なる向上，あるいはより複雑な課題への応用に新たな示唆を与えられる．本研究ではLSXの工学的な利用価値を主に示したが，前述の通り人間の意思決定に対する分析に用いるモデルとしても有益であると考えられ，今後の研究において具体的な手法を提示したいと考えている．

謝辞

本研究はJSPS科研費 25730150の助成を受けたものです．

参考文献

[1] W.A. Wickelgren, (1977) “Speed-accuracy tradeoff and information processing Dynamics,” Acta Psychologica, 41, pp. 67–85.
[2] R. S. Sutton, A. G. Barto, (2000) “強化学習,” 森北出版, (三上, 皆川 訳).
[3] P. Auer, N. Cesa-Bianchi and P. Fischer, (2002) “Finite-time analysis of the multi-armed bandit problem,” Machine Learning, 47, pp. 235–256.
[4] 篠原修二, 田口亮, 桂田浩一, 新田恒雄, (2007) “因果性に基づく信念形成モデルとN本腕バンディット問題への適用”, 人工知能学会論文誌, Vol.22, No.1, pp.58–68.
[5] T. Takahashi, M. Nakano and S. Shinohara, (2010) “Cognitive symmetry: Illogical but rational biases,” Symmetry: Culture and Science, 21, 1–3, pp. 275–294.
[6] Y. Kohno, T. Takahashi, (2012) “Loosely Symmetric Reasoning to Cope with The Speed-Accuracy Tradeoff,” SCIS-ISIS 2012, Kobe Convention Center (Kobe Portopia Hotel), pp.1166–1171.
[7] M. Hattori and M. Oaksford, (2007) “Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis,” Cognitive Science, 31, 5, pp. 765–814.
[8] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, (2011) “How to Grow a Mind: Statistics, Structure, and Abstraction,” Science, vol. 331, no. 6022, pp. 1279–1285.
[9] A. Tversky, D. Kahneman, (1974) “Judgment under uncertainty: Heuristics and biases,” Science 185 (4157), pp. 1124–1131.
[10] D. Kahneman, A. Tversky, (1984) “Choices, values and frames,” American Psychologist 39 (4), pp. 341–350.
[11] H.A. Simon, (1956) “Rational choice and the structure of the environment,” Psychological Review, 63, pp. 261–273.
[12] D. Kahneman and A. Tversky, (1979) “Prospect Theory: An Analysis of Decision under Risk,” Econometrica, 47(2), pp. 263–292.
[13] S. Gelly, Y. Wang, R. Munos and O. Teytaud, (2005) “Modification of UCT with Patterns in Monte-Carlo Go,” Technical Report, No.6062, INRIA.
[14] 甲野佑, 高橋達二, (2013)“価値推論ヒューリスティクスとしての規準学習と忘却,” 第30回日本認知科学会論文集, pp. 74–79.
[15] N.D. Daw, J.P. O’Doherty, P. Dayan, B. Seymour, R.J. Dolan, (2006) “Cortical substrates for exploratory decisions in humans,” Nature, 441(7095), 876–879.