(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6145766号
(P6145766)
(45)発行日　平成29年6月14日(2017.6.14)　(24)登録日　平成29年5月26日(2017.5.26)

(51)Int.Cl.		F I		
G06Q 50/10	(2012.01)	G06Q	50/10	
G06F 19/00	(2011.01)	G06F	19/00	120

請求項の数 10　(全 16 頁)

(21)出願番号　特願2013-66768(P2013-66768)
(22)出願日　平成25年3月27日(2013.3.27)
(65)公開番号　特開2014-191598(P2014-191598A)
(43)公開日　平成26年10月6日(2014.10.6)
審査請求日　平成27年11月26日(2015.11.26)

(73)特許権者　503359821
国立研究開発法人理化学研究所
埼玉県和光市広沢２番１号
(74)代理人　100120868
弁理士　安彦　元
(72)発明者　金　成主
埼玉県和光市広沢２番１号　独立行政法人理化学研究所内
(72)発明者　青野　真士
埼玉県和光市広沢２番１号　独立行政法人理化学研究所内
(72)発明者　行田　悦資
埼玉県和光市広沢２番１号　独立行政法人理化学研究所内

最終頁に続く

(54)【発明の名称】解探索システム、解探索プログラム

(57)【特許請求の範囲】
【請求項１】
確率分布に基づいて結果を出力する２以上の被検対象のうち最良の結果の出力が期待される被検対象を探索する解探索システムにおいて、
上記出力された結果の蓄積に基づく今までの戦績を上記被検対象毎にそれぞれ求め、上記各被検対象の戦績を被検対象全体の戦績との関係においてその優劣を比較する戦績優劣比較手段と、
上記戦績優劣比較手段により比較された戦績の優劣と、上記被検対象から出力された直近の結果とに基づいて、計量変数を増加又は減少させるように制御することを当該被検対象毎に行う制御手段と、
上記計量変数が閾値を超えた上記被検対象に対して結果の出力を指示する出力指示手段とを備え、
上記出力指示手段は、上記結果の出力の指示の繰り返しを経て最終的に最も上記結果の出力の指示が行われている１以上の被検対象を、探索解として特定すること
を特徴とする解探索システム。
【請求項２】
それぞれ設定された確率分布に基づいて結果を出力する３以上の被検対象のうち最良の結果の出力が期待される被検対象の組み合わせを探索し、
上記出力指示手段は、上記結果の出力の指示の繰り返しを経て最終的に最も上記結果の出力の指示が行われている被検対象の組み合わせを、探索解として特定すること

を特徴とする請求項１記載の解探索システム。
【請求項３】
上記出力指示手段による出力指示に応じて、時系列的に確率分布が変化する被検対象の組み合わせを探索する上で、２以上の当該解探索システムを用いて解探索を行うこと
を特徴とする請求項１記載の解探索システム。
【請求項４】
上記戦績優劣比較手段は、一の被検対象がより優れた結果を出力した場合及び他の被検対象がより劣る結果を出力した場合に、上記一の被検対象における今までの戦績をより優れる側に向上させ、一の被検対象がより劣った結果を出力した場合及び他の被検対象がより優れた結果を出力した場合に、上記一の被検対象における今までの戦績をより劣る側に下降させること
を特徴とする請求項１～３のうち何れか１項記載の解探索システム。
【請求項５】
上記戦績優劣比較手段は、上記各被検対象の戦績と被検対象全体の戦績の平均との差分を内部リソース値とし、
上記制御手段は、
上記内部リソース値が正で、上記被検対象から出力された直近の結果がより優れたものである場合には、上記計量変数を増加させ、
上記内部リソース値が正で、上記被検対象から出力された直近の結果がより劣るものである場合には、上記計量変数をそのままにし、
上記内部リソース値が０で、上記被検対象から出力された直近の結果がより優れたものである場合には、上記計量変数を増加させ、
上記内部リソース値が０で、上記被検対象から出力された直近の結果がより劣るものである場合には、上記計量変数を下降させ、
上記内部リソース値が負で、上記被検対象から出力された直近の結果がより優れたものである場合には、上記計量変数をそのままにし、
上記内部リソース値が負で、上記被検対象から出力された直近の結果がより劣るものである場合には、上記計量変数を下降させること
を特徴とする請求項１～４のうち何れか１項記載の解探索システム。
【請求項６】
上記制御手段は、上記被検対象毎に割り当てられる計量変数の合計が一定となるように制御すること
を特徴とする請求項１～５のうち何れか１項記載の解探索システム。
【請求項７】
確率分布に基づいて結果を出力する２以上の被検対象のうち最良の結果の出力が期待される被検対象を探索する解探索プログラムにおいて、
上記出力された結果の蓄積に基づく今までの戦績を上記被検対象毎にそれぞれ求め、上記各被検対象の戦績を被検対象全体の戦績との関係においてその優劣を比較する戦績優劣比較ステップと、
上記戦績優劣比較ステップにより比較された戦績の優劣と、上記被検対象から出力された直近の結果とに基づいて、計量変数を増加又は減少させるように制御することを当該被検対象毎に行う制御ステップと、
上記計量変数が閾値を超えた上記被検対象に対して結果の出力を指示する出力指示ステップとを有し、
上記出力指示ステップでは、上記結果の出力の指示の繰り返しを経て最終的に最も上記結果の出力の指示が行われている１以上被検対象を、探索解として特定すること
をコンピュータに実行させることを特徴とする解探索プログラム。
【請求項８】
上記戦績優劣比較ステップでは、一の被検対象がより優れた結果を出力した場合及び他の被検対象がより劣る結果を出力した場合に、上記一の被検対象における今までの戦績を

より優れる側に向上させ、一の被検対象がより劣った結果を出力した場合及び他の被検対象がより優れた結果を出力した場合に、上記一の被検対象における今までの戦績をより劣る側に下降させること
を特徴とする請求項７記載の解探索プログラム。
【請求項９】
上記戦績優劣比較ステップでは、上記各被検対象の戦績と被検対象全体の戦績の平均との差分を内部リソース値とし、
上記制御ステップでは、
上記内部リソース値が正で、上記被検対象から出力された直近の結果がより優れたものである場合には、上記計量変数を増加させ、
上記内部リソース値が正で、上記被検対象から出力された直近の結果がより劣るものである場合には、上記計量変数をそのままにし、
上記内部リソース値が０で、上記被検対象から出力された直近の結果がより優れたものである場合には、上記計量変数を増加させ、
上記内部リソース値が０で、上記被検対象から出力された直近の結果がより劣るものである場合には、上記計量変数を下降させ、
上記内部リソース値が負で、上記被検対象から出力された直近の結果がより優れたものである場合には、上記計量変数をそのままにし、
上記内部リソース値が負で、上記被検対象から出力された直近の結果がより劣るものである場合には、上記計量変数を下降させること
を特徴とする請求項７又は８記載の解探索プログラム。
【請求項１０】
上記制御ステップでは、上記被検対象毎に割り当てられる計量変数の合計が一定となるように制御すること
を特徴とする請求項７～９のうち何れか１項記載の解探索プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、組み合わせ報酬最大化問題の解を高速かつ効率的に導く上で好適な解探索システム、解探索プログラムに関するものである。
【背景技術】
【０００２】
従来より、期待値を最大化する解を探索する問題の代表例として、バンディット問題がある。このバンディット問題とは、貰える合計報酬の期待値を最大化することを目的とし、プレイヤーはn種類の異なる行動選択肢から一つの選択肢を選択する動作を繰り返す。各選択の後は毎回、選択した行動に依存する確率分布から選ばれた結果がプレイヤーの報酬として与えられる。
【０００３】
仮に複数のスロットマシーンが存在し、各スロットマシーンのレバーを引くことにより、ある確率分布の下でコイン（報酬）がもらえるものとする。このコインが出る確率分布（当選確率）がスロットマシーン毎に異なる場合であって、かつプレイヤーはその当選確率が分からない場合を考えてみる。このとき、各スロットマシーンの当選確率を知る最も一般的な方法としては、とりあえず各スロットマシーンを多数回に亘り順にプレイし、実際に最も報酬が大きかったスロットマシーンが、最も当選確率が高いものと判断する。
【０００４】
しかしながら、かかる方法では、実際に最も当選確率の高いスロットマシーンを特定する上で相当の回数に亘りスロットマシーンをプレイしなければならず、結果として多くの投資が必要となる。このため、各スロットマシーンにおける当選確率を調べる上で極力投資を少なくしつつ、効率的に解を探索できるアルゴリズムを考える必要が出てくることが分かる。

【０００５】
かかる場合には、上述のような貰える合計報酬の期待値を最大化するバンディット問題に置き換えて、これを解くことができる（例えば、非特許文献１参照。）。特にこのバンディット問題の中で、n種類の異なる行動選択肢から最良の結果の出力が期待される組み合わせを選択する、いわゆる組み合わせバンディット問題も近年において注目されている。この組み合わせバンディット問題では、複数台のスロットマシーンの中からより高配当が期待できるスロットマシーンの組み合わせを選択する場合のみならず、例えばコグニティブ無線通信においてデータ伝送量を最大化できるチャネルの最適な組み合わせの選択、インターネット広告においてクリック数を最大化できる広告の最適な組み合わせ、更には最も投資リターンの大きい金融商品のポートフォリオの選択等、様々な分野においてニーズがある。このような応用例の場合には、より一般的な組み合わせ報酬最大化問題になる。つまり、プレイヤーが多数で、それぞれのプレイヤーの選択に依存して（例えばペイオフ行列によって）各プレイヤーの報酬量が決定される。しかし、本明細書では簡単のために、各スロットマシーンが独立な場合の、組み合わせ報酬最大化問題（特に２つの組み合わせ）について例を挙げて説明する。
【先行技術文献】
【非特許文献】
【０００６】
【非特許文献１】S.J.Kim,M.Aono,M.Hara, BioSystems 101,29-36 (2010)
【発明の概要】
【発明が解決しようとする課題】
【０００７】
しかしながら、従来において、組み合わせバンディット問題の解を自動的に探索して求めるためのアルゴリズムが特段提案されていなかった。情報量が増大の一途を辿る昨今において、大量の情報から高速かつ効率的に、組み合わせバンディット問題の解を求めるための社会的要請が高くなると考えられるが、これについて特段の解決策が提案されていないのが現状であった。
【０００８】
本発明は、上述した問題点に鑑みて案出されたものであり、その目的とするところは、高速かつ効率的に組み合わせバンディット問題の解を求めることが可能な解探索システム、解探索プログラムを提供することにある。
【課題を解決するための手段】
【０００９】
本発明を適用した解探索システムは、上述した課題を解決するために、確率分布に基づいて結果を出力する２以上の被検対象のうち最良の結果の出力が期待される被検対象を探索する解探索システムにおいて、上記出力された結果の蓄積に基づく今までの戦績を上記被検対象毎にそれぞれ求め、上記各被検対象の戦績を被検対象全体の戦績との関係においてその優劣を比較する戦績優劣比較手段と、上記戦績優劣比較手段により比較された戦績の優劣と、上記被検対象から出力された直近の結果とに基づいて、計量変数を増加又は減少させるように制御することを当該被検対象毎に行う制御手段と、上記計量変数が閾値を超えた上記被検対象に対して結果の出力を指示する出力指示手段とを備え、上記出力指示手段は、上記結果の出力の指示の繰り返しを経て最終的に最も上記結果の出力の指示が行われている１以上の被検対象を、探索解として特定する解探索システムことを特徴とする。
【００１０】
本発明を適用した組み合わせ探索プログラムは、確率分布に基づいて結果を出力する２以上の被検対象のうち最良の結果の出力が期待される被検対象を探索する解探索プログラムにおいて、上記出力された結果の蓄積に基づく今までの戦績を上記被検対象毎にそれぞれ求め、上記各被検対象の戦績を被検対象全体の戦績との関係においてその優劣を比較する戦績優劣比較ステップと、上記戦績優劣比較ステップにより比較された戦績の優劣と、

上記被検対象から出力された直近の結果とに基づいて、計量変数を増加又は減少させるように制御することを当該被検対象毎に行う制御ステップと、上記計量変数が閾値を超えた上記被検対象に対して結果の出力を指示する出力指示ステップとを有し、上記出力指示ステップでは、上記結果の出力の指示の繰り返しを経て最終的に最も上記結果の出力の指示が行われている１以上被検対象を、探索解として特定することをコンピュータに実行させることを特徴とする。

【発明の効果】

【００１２】

上述した構成からなる本発明によれば、被検対象から出力された今までの戦績の優劣と、被検対象から出力された直近の結果とに基づいて、計量変数を増加又は減少させ、この計量変数の値に応じて結果の出力を指示するか否かを決定する。そして、結果の出力の指示の繰り返しを経て最終的に最も結果の出力の指示が行われている被検対象の組み合わせを、探索すべき組み合わせとして特定する。これにより、組み合わせバンディット問題の解を自動的に探索して求めることが可能となり、情報量が増大の一途を辿る昨今において、大量の情報から高速かつ効率的に、組み合わせバンディット問題の解を求めることが可能となる。

【図面の簡単な説明】

【００１３】

【図１】本発明を適用した解探索システムの全体構成を示す図である。

【図２】計量変数について説明するための図である。

【図３】本発明を適用した解探索システムの他の全体構成例を示す図である。

【図４】本発明を適用した解探索システムを、コンピュータプログラムで実現した場合における実施例を示す図である。

【図５】被検対象の当選確率が、０．２、０．５、０．８の３つのサンプルについての組み合わせを選択するシミュレーションを行う例を示す図である。

【発明を実施するための形態】

【００１４】

以下、本発明を適用した解探索システムについて図面を参照しながら詳細に説明をする。

【００１５】

図１は、本発明を適用した解探索システム解探索システム１の全体構成を示している。この解探索システム１は、２以上の被検対象５ａ～５ｄのうち最良の結果の出力が期待される被検対象５の組み合わせを探索するシステムである。解探索システム１は、戦績優劣比較部２と、この戦績優劣比較部２に接続された制御部３ａ～３ｄと、制御部３ａ～３ｄに接続された出力指示部４ａ～４ｄとを備えている。

【００１６】

ちなみに、この制御部３ａ～３ｄ、出力指示部４ａ～４ｄは、それぞれ被検対象５の数と同等となるように設けられるものであり、図１の例では４個の被検対象５からなるため、これら制御部３、出力指示部４も４個ずつで構成される。

【００１７】

被検対象５は、それぞれ設定された確率分布に基づいて結果を出力する対象物である。例えば、スロットマシーンやパチンコの台のように、設定された確率分布に基づいてコインという結果物を出力するものであってもよい。また、コグニティブ無線通信は、各チャネルのデータ伝送量の大小は、その都度変化するものであるが、これについてもある時点において設定された確率分布で表現することができる。このようなコグニティブ無線通信において任意のチャネルを選択した場合に、実際の＂データ伝送量＂という結果物を出力する。また、インターネット広告についても、掲載すべき広告のクリック数の大小は確率分布で表すことが可能となり、＂実際のクリック数＂という結果物はその確率分布に基づいて算出することが可能となる。また、金融商品については、その将来的な投資リターンも確率分布で表すことができ、＂実際の投資リターン＂という結果物も当該確率分布に基

づいて表される。

【００１８】

このように、被検対象５は、出力する結果を確率分布に変換することが可能なあらゆる事象、物、システム、プログラムやアルゴリズムを含む概念である。ちなみに、この被検対象５において出力される結果の確率分布は、通常の正規分布、ガウシアン分布のみならず、離散的な分布であってもよいし、２項分布で構成されていてもよい。ちなみに、この被検対象５の確率分布は、この解探索システム１のユーザにとって未知のものとなっている。ユーザは、これらの被検対象のうち、最良の結果の出力が期待される被検対象の組み合わせを探索するためにこの解探索システム１を使用することとなる。

【００１９】

被検対象５ａ～５ｄは、それぞれ設定された確率分布に基づいて結果ｌを出力する。このとき、被検対象５ａから出力される結果を結果l_1とし、被検対象５ｂから出力される結果を結果l_2とし、被検対象５ｃから出力される結果を結果l_3とし、被検対象５ｄから出力される結果を結果l_4とする。出力された結果l_1は、戦績優劣比較部２へと送信されると共に、制御部３ａへ送信される。出力された結果l_2は、戦績優劣比較部２へと送信されると共に、制御部３ｂへ送信される。出力された結果l_3は、戦績優劣比較部２へと送信されると共に、制御部３ｃへ送信される。出力された結果l_4は、戦績優劣比較部２へと送信されると共に、制御部３ｄへ送信される。

【００２０】

戦績優劣比較部２は、被検対象５ａ～５ｄから出力される結果l_1～l_4を受信し、これを記憶する。この戦績優劣比較部２は、被検対象５ａ～５ｄから結果l_1～l_4を受信する都度、順次記憶しておくことで、結果を蓄積する。そして、この戦績優劣比較部２は、被検対象５ａ～５ｄ毎に、出力された結果の蓄積に基づく今までの戦績をそれぞれ求める。ここでいう戦績とは、被検対象５から出力される結果がより優れているのか、或いはより劣っているのかを示すあらゆるデータを示すものである。被検対象５がスロットマシーンであれば、コインがどの程度出たかを示すものであってもよいし、被検対象５がインターネット広告であれば、クリックがどの程度行われたかを示すデータであってもよい。また、この戦績優劣比較部２は、各被検対象５の戦績を被検対象５全体の戦績との関係においてその優劣を比較する。ここで被検対象５ａについての被検対象５全体の優劣をs_1とし、被検対象５ｂについての被検対象５全体の優劣をs_2とし、被検対象５ｃについての被検対象５全体の優劣をs_3とし、被検対象５ｄについての被検対象５全体の優劣をs_4とする。これら優劣s_1は、制御部３ａに、優劣s_2は、制御部３ｂに、優劣s_3は、制御部３ｃに、優劣s_4は、制御部３ｄにそれぞれ送られる。

【００２１】

制御部３ａ～３ｄは、それぞれ戦績優劣比較部２から、戦績の優劣に関する情報s_1～s_4がそれぞれ入力されるとともに、被検対象５ａ～５ｄから直近の結果l_1～l_4がそれぞれ入力される。制御部３ａ～３ｄは、それぞれ入力された戦績の優劣に関する情報s_1～s_4と、被検対象から出力された直近の結果とに基づいて計量変数を増加又は減少させるように制御する。

【００２２】

図２は、この計量変数のイメージを説明するための図である。計量変数xは、各被検対象５に対してそれぞれ個別に割り当てられたパラメータであり、被検対象５を選択するか否かの判断する上での基準となるものである。この計量変数xが高いほど、その被検対象５を選択する可能性が高くなる。一方、この計量変数xが低いほどその被検対象５を選択しない可能性が高くなる。

【００２３】

このような計量変数xの増減を制御するのが、制御部３である。つまり制御部３ａは、結果l_1と優劣s_1とに基づいて計量変数x_1の増減を制御し、制御部３ｂは、結果l_2と優劣s_2とに基づいて計量変数x_2の増減を制御し、制御部３ｃは、結果l_3と優劣s_3とに基づいて計量変数x_3の増減を制御し、制御部３ｄは、結果l_4と優劣s_4とに基づ

いて計量変数x_iの増減を制御する。これら制御部３による計量変数x_iの制御は、あくまで結果ｌと、優劣s_iに基づくものであればいかなるものであってもよい。

【００２４】

出力指示部４ａ～４ｄは、計量変数を監視し、予め設定した閾値を超えるか否かを判別する。そして、計量変数が閾値を超えた旨を判別した場合には、被検対象５に対して結果の出力を指示する。この図２の例では、閾値を超えた計量変数は、x_2とx_4である。このため、x_2を監視する出力指示部４ｂと、x_4を監視する出力指示部４ｄは、被検対象５ｂ、５ｄに対して結果の出力を指示する。

【００２５】

このようにして、被検対象５を中心として、戦績優劣比較部２、制御部３、出力指示部４の順でいわゆるフィードバック制御が行われる。このような戦績優劣比較部２、制御部３、出力指示部４からなる解探索システム１は、例えば、アナログ回路、デジタル回路を始めとしたいかなるデバイスで具現化されるものであってもよい。ちなみに、回路として具現化される場合には、FPGA(field-programmable gate array)に基づいて、構成を設定するようにしてもよい。また、本発明はプログラムで具現化されるものであってもよい。かかる場合には、戦績優劣比較部２は、これと同様の処理を実行する戦績優劣比較ステップに、制御部３は、これと同様の処理を実行する制御ステップに、出力指示部４は、これと同様の処理を実行する出力指示ステップとして具現化されることとなる。これに加えて、このようなプログラムに基づいて動作するハードウェア（例えば、パーソナルコンピュータ、各種携帯情報端末等）を介して具現化されるものであってもよい。

【００２６】

次に、本発明を適用した解探索システム１による組み合わせ探索動作について説明をする。

【００２７】

先ず、被検対象５ａ～５ｄのいくつかについて結果の出力が行われる。この結果の出力が指示されるのは上述した出力指示部４により出力指示が行われたものに限る。

【００２８】

被検対象５ａ～５ｄから出力された結果ｌは、それぞれ戦績優劣比較部２へ送られると共に、制御部３へと送られる。戦績優劣比較部２では、この送られてきた結果ｌに基づいて具体的に以下の処理動作を行う。

【００２９】

以下の式（１）におけるq_iは、それぞれの被検対象５ａ～５ｄにおける戦績を示す指数である。この戦績指数q_iのｔは、この解探索システム１におけるフィードバック回数を示している。また、ｉは、図１に示す下付きの数字に対応するものであり、それぞれの被検対象５に対応するものである。つまりｉ毎に、換言すれば被検対象５毎に、この戦績指数q_iを求めていくこととなる。

【００３０】

【数１】

$$q_i^t = \alpha \cdot q_i^{t-1} + \sum_{j \in I \setminus \{i\}} \mu \cdot \left(\rho_i^t \cdot \rho_j^t + \omega \cdot \sum_{<k,k'> \in I_2 \setminus \{<i,j>\}} \pi_k^t \cdot \pi_{k'}^t \right)$$

・・・・・・・（１）

【００３１】

戦績指数q_iは、フィードバックの回数ｔが増加するにつれて順次更新される。そして、この戦績指数q_iは、その前のフィードバック回数ｔ－１において求めた戦績指数q_i^{t-1}に加えて、Σ以降の項を足すことで表示される。また、この（１）式においてαは忘却パラメータであり、必要に応じて設定される。

【００３２】

このΣ以降の項において、μは係数である。またρ$_i$、ρ$_j$は、より優れた結果を出力

した場合を意味している。ここでρiは、これから戦績指数qiを求めようとする被検対象５が、より優れた結果を出力した場合に"１"となり、より劣る結果を出力した場合には、"０"となる。ρjは、ｊに相当する他の被検対象５が、より優れた結果を出力した場合に"１"となり、より劣る結果を出力した場合には、"０"となる。本明細では、例として、２つの組み合わせを選択するためのシステムであることからρは２変数となっている。

【００３３】

表１は、ρ、πの取り得る値を示している。

【００３４】

【表１】

	優	劣	非出力
ρ	1	0	0
π	0	1	0

【００３５】

式（１）の括弧内は、ρi・ρjのように乗算で表されることから、ρiとρjの双方が１である場合のみプラスになる。

【００３６】

また、πは、ｉ、ｊ以外の他の被検対象５が、より劣った結果を出力した場合に"１"となり、より優れた結果を出力した場合に"０"となる。

【００３７】

即ち、この（１）式の括弧内は、そのqiを求める被検対象５が、より優れた結果を出力した場合と、当該qiを求める被検対象５以外の他の被検対象５が、より劣った結果を出力した場合に、その数値が上昇することになっている。仮に被検対象５がスロットマシーンである場合には、当該qiを求めるスロットマシーンが当選した場合には、（１）式の括弧内の数値を上昇させ、当該qiを求めるスロットマシーンが落選した場合には、（１）式の括弧内の数値を変化させない。また、当該qiを求めるスロットマシーン以外が当選した場合には、（１）式の括弧内の数値を変化させず、当該qiを求めるスロットマシーン以外が落選した場合には、（１）式の括弧内の数値を上昇させる。そして、（１）式の括弧内の数値が上昇するにつれて戦績指数qiが上昇し、より優れた戦績になる。

【００３８】

このように、本発明では、qiを求める一の被検対象５が、より優れた結果を出力した場合及び他の被検対象がより劣る結果を出力した場合に、当該一の被検対象５における今までの戦績qiをより優れる側に向上させる。また、qiを求める一の被検対象５が、より劣った結果を出力した場合及び他の被検対象がより優れた結果を出力した場合に、当該一の被検対象５における今までの戦績qiを変化させない。戦績優劣比較部２は、このような制御を行うものであってもよく、上述した（１）式に基づく制御を行う場合に限定されるものではない。

【００３９】

また、何れの出力結果が優れており、何れの出力結果が劣っているかについては、いかなる基準の下で判断するようにしてもよい。上述したスロットマシーンの例では、コインが出るか否かで優劣を決める場合に限定されるものではなく、コインの枚数や種別に応じて優劣を決めるようにしてもよい。また、この優劣についても、優れているか、或いは劣っているかの２段階で設定される場合に限定されるものではなく、３段階以上で優劣を評価するようにしてもよい。ちなみに、３段階以上で優劣をランク分けする場合においても、一の被検対象５がより上位ランクであるほどqiを上昇させ、他の被検対象５がより下位ランクであるほどqiを下降させるように調整を行う。

【００４０】

戦績優劣比較部２は、このようにして、各被検対象５ａ～５ｄについてそれぞれ戦績指数q_1～q_4を求めた後、その戦績指数q_1～q_4を被検対象全体５の戦績との関係においてその優劣を比較する。そして、この優劣の比較結果としての内部リソース値s_iをそれぞれ出力する。

【００４１】

かかる場合に、（２）式に基づいてその優劣を比較するようにしてもよい。

【００４２】

【数２】

$$s_i^t = x_0^t + q_i^{t-1} - \mathrm{Mean}_{k \in I \setminus \{i\}} \left\{ q_k^{t-1} \right\}$$

・・・・・・・（２）

【００４３】

式（２）は、一の被検対象５の戦績q_i^{t-1}と被検対象５全体の戦績q_i^{t-1}の平均との差分を、内部リソース値s_iとしている。例えば、被検対象５ａについての内部リソース値s_1を求める場合には、その戦績q_1^{t-1}と、全被検対象の戦績q_1^{t-1}、q_2^{t-1}、q_3^{t-1}、q_4^{t-1}の平均との差分を求める。ちなみに、（２）式の右項においてx_iはあくまで調整値であり、必須のものではない。　この（２）式に基づいて、内部リソース値s_iを判断する場合に限定されるものではない。内部リソース値s_iは、それに対応する一の被検対象５の戦績q_i^{t-1}が、他の被検対象５との間で相対的に優れているか否かを示すものであれば、いかなる計算式で、或いはいかなる方法で、これを評価するようにしてもよい。戦績優劣比較部２は、これら計算したs_1～s_4をそれぞれ、制御部３ａ～３ｄへ出力する。

【００４４】

制御部３ａ～３ｄは、これらs_1～s_4並びに被検対象５ａ～５ｄから直近の結果l_1～l_4がそれぞれ入力された場合に、例えば、以下の制御を行う。

【００４５】

表２は、この制御部３による制御を行う上で参酌するテーブルの例を示している。

【００４６】

【表２】

	$s_i^t>0$	$s_i^t=0$	$s_i^t<0$
$l_i^t=-1$	1	1	0
$l_i^t=1$	0	−1	−1

【００４７】

この表２によれば、制御部３は、内部リソース値s_iと、被検対象５から出力された直近の結果l_iとより形成されるマトリクスに基づいて、制御方法を決定する。ここでl_i＝－１は、より優れた結果が出力された場合を意味している。また、l_i＝１は、より劣った結果が出力された場合を意味している。ちなみに、このl_iの優劣の基準についてもいかなるものであってもよい。

【００４８】

また、表中の数値（－１、０、１）の意味については、先ず"－１"は、計量変数ｘを成長速度を減少させるように制御する。また、"０"は、計量変数x_iの成長速度を特に増減させないことを意味し、"１"は、計量変数x_iの成長速度を増加させるように制御する。

【００４９】

ちなみに以下の例では、x_0は、x_iの値によって決まるものと仮定している。

【００５０】

制御部３は、この表２に基づいて具体的に以下の制御を行う。

【００５１】

内部リソース値ｓiが正で、被検対象５から出力された直近の結果ｌiがより優れたものである場合（＝－１）には、"１"であることから計量変数ｘiの成長速度を増加させる。内部リソース値ｓiが正で、被検対象５から出力された直近の結果ｌiがより劣るものである場合（＝１）には、"０"であることから計量変数ｘiの成長速度を増加させることなくそのままにする。内部リソース値ｓiが０で、被検対象５から出力された直近の結果ｌiがより優れたものである場合（＝－１）には、"１"であることから計量変数計量変数ｘiの成長速度を増加させる。内部リソース値ｓiが０で、被検対象５から出力された直近の結果ｌiがより劣るものである場合（＝１）には、"－１"であることから計量変数計量変数ｘiの成長速度を減少させる。内部リソース値ｓiが負で、被検対象５から出力された直近の結果ｌiがより優れたものである場合（＝－１）には、"０"であることから計量変数計量変数ｘiの成長速度をそのままにし、内部リソース値ｓiが負で、被検対象５から出力された直近の結果ｌiがより劣るものである場合（＝１）には、"－１"であることから計量変数計量変数ｘiの成長速度を減少させる。

【００５２】

このようにして制御部３は、戦績優劣比較部２により比較された戦績の優劣に基づく内部リソース値ｓiと、被検対象５から出力された直近の結果ｌiとに基づいて、計量変数ｘiを増加又は減少させるように制御する。ちなみに表２中の数値はあくまで一例であり、内部リソース値ｓiと直近の結果ｌiとに基づくものであればいかなる数値であってもよい。

【００５３】

また、上述した例では、内部リソース値ｓiを３段階で表示し、直近の結果ｌiを２段階で表示し、合計２×３のマトリックスで表示しているが、これに限定されるものではなく、ｓi、ｌiともに最低２段階であれば何段階でランク分けするようにしてもよい。

【００５４】

さらに、この制御部３は、このようなマトリックス表で制御方法を規定する場合に限定されるものではなく、内部リソース値ｓiと直近の結果ｌiとに基づくものであれば他のいかなる方法に基づいて計量変数ｘiの増減を制御するようにしてもよい。具体的には、内部リソース値ｓiと直近の結果ｌiとを変数とした所定の演算式に従って制御方法を決めるようにしてもよい。

【００５５】

このようにして制御部により計量変数ｘiの増減を制御する結果、図２に示すように、ある被検対象５に対する計量変数ｘiは閾値未満であり、ある被検対象５に対する計量変数ｘiは閾値以上となる。出力指示部４は、これら軽量パラメータｘiをそれぞれ閾値との関係で比較し、閾値を超えた計量変数ｘiに応じた被検対象５に対してのみ、結果の出力を指示する。結果の出力を指示された被検対象５は、新たなる結果をそれぞれの確率分布に基づいて出力することとなる。

【００５６】

本発明を適用した解探索システム１では、これらの動作を繰り返し実行する。その結果、出力指示部４により出力指示が出される被検対象５の組み合わせが徐々に一定化されてくる。最終的には、この出力指示が出される被検対象５は、一定のものに収束される。この収束されてくる被検対象５の組み合わせが、解探索システム１により選択されてくる解となる。

【００５７】

ちなみに、解探索システム１により解を求める上で、被検対象５ａ～５ｄ毎に割り当てられる計量変数ｘ１～ｘ４と、ｘ０との合計が一定となるように制御することで、より探索精度を高めることが可能となる。即ち、ｘ０＋ｘ１＋ｘ２＋ｘ３＋ｘ４＝一定、としておくことにより、計量変数ｘiが全体的に大きくなってしまい、閾値との間でバランスが取れなくなるのを防止することができ、これが探索精度の向上につながることとなる。

【００５８】
x_iは、他の計量変数x_iの値により影響を受ける変数である。上述した処理動作を繰り返し進めていく結果、上述した（２）式において、当初はx_iの項が支配的になり、$q_i^{[-1]}$以後の項（一の被検対象５の戦績$q_i^{[-1]}$と被検対象５全体の戦績$q_i^{[-1]}$の平均との差分）については、あまりこのs_iを決定する上で大きな影響を与えるものではない。その結果、このs_iは、戦績$q_i^{[-1]}$と戦績$q_i^{[-1]}$の平均との差分に影響を受けることなく、自由度が高くなり、その分ランダムな値を取りやすくなる。その結果、ランダムな値を取りやすくなるs_iを介して様々な出力指示４ａ～４ｄが行われ、様々な解を探索することが可能となる。
【００５９】
これに対して、ｔが大きくなるにつれ、換言すればフィードバック回数が多くなるにつれ、徐々に$q_i^{[-1]}$以後の項（一の被検対象５の戦績$q_i^{[-1]}$と被検対象５全体の戦績$q_i^{[-1]}$の平均との差分）が大きくなる。その結果、このs_iを決定する上で、s_iよりも、戦績$q_i^{[-1]}$と戦績$q_i^{[-1]}$の平均との差分値が支配的になってくる。そして、次回に出力指示４ａ～４ｄが行われるものについては、戦績$q_i^{[-1]}$と戦績$q_i^{[-1]}$の平均との差分値がより影響を受けるものとなる。
【００６０】
このようにして最終的に出力指示が行われるものは、戦績$q_i^{[-1]}$と戦績$q_i^{[-1]}$の平均との差分値による影響を受けるものに収束されてくる。
【００６１】
上述した構成からなる本発明によれば、被検対象５から出力された今までの戦績の優劣と、被検対象５から出力された直近の結果とに基づいて、計量変数を増加又は減少させ、この計量変数の値に応じて結果の出力を指示するか否かを決定する。そして、結果の出力の指示の繰り返しを経て最終的に最も結果の出力の指示が行われている被検対象５の組み合わせを、探索すべき組み合わせとして特定する。これにより、組み合わせバンディット問題の解を自動的に探索して求めることが可能となり、情報量が増大の一途を辿る昨今において、大量の情報から高速かつ効率的に、組み合わせバンディット問題の解を求めることが可能となる。
【００６２】
なお、本発明は、上述した実施の形態に限定されるものではない。例えば、図３に示すように、２以上の解探索システム１を用いて、被検対象５を探索するようにしてもよい。図３では、２つの解探索システム１ａ、解探索システム１ｂを用いて被検対象５ａ～５ｄを探索する場合について示している。ちなみに、この２つの解探索システム１ａ、解探索システム１ｂは、互いの同一の被検対象５ａ～５ｄのグループの探索を行う。換言すれば図中において解探索システム１ｂ側の点線で示されている被検対象５ａ～５ｄは、解探索システム１ａ側の実線で示されている被検対象５ａ～５ｄと同一のものである。
【００６３】
このケースでは、被検対象５の数は、２以上であればよい。つまり最低２つの被検対象５のうち、最良の結果の出力が期待される１の被検対象５を選択するものであってもよい。かかる点において、このケースでは、２以上の被検対象５の組み合わせを探索する場合に限定されるものではなく、１の被検対象５を解として探索するものであってもよい。
【００６４】
またこのケースにおいて被検対象５は、予め確率分布が設定されているものに限定されるものではなく、確率分布が時系列的に、つまり時間の経過に応じて変化するものであってもよい。しかも、その被検対象５における時系列的な確率分布の変化が、出力指示部４ａ～４ｄによる出力指示に対応するものであってもよい。即ち、出力指示部４ｂからの出力指示があった場合、これに対応する被検対象５ｂのみが、当該出力指示に応じて自らの確率分布を変化させ、その変化させた確率分布に基づいて結果を出力する。
【００６５】
更に、各被検対象５による各確率分布の変化は、他の被検対象５による出力と独立では

なく、何らかの相関性を持たせるようにしてもよい。即ち、一の被検対象５ａがより高い確率で当たりを出す確率分布とする場合に、他の一の被検対象５ｂはこれと相関性の高い確率分布を設定するようにしてもよく、更なる他の一の被検対象５ｂは、これと負の相関をもつ確率分布を設定する等してもよい。これらの相関はいかなるもので表現されていてもよい。

【００６６】

このケースでは、２つの解探索システム１ａ、解探索システム１ｂを用いて、上述と同様に解探索を行っていく。出力指示が行われる都度、これに応じて被検対象５の確率分布が時系列的に変化し、しかもその確率分布の変化が、複数の被検対象５との間で独立ではなく互いに相関性を持っている。これらの処理を繰り返し実行することにより、上述と同様に被検対象５の選択が収束され、これが探索解となる。

【００６７】

このケースでは、例えば株等の投資対象を被検対象５に当てはめて考えることもできる。株価は、上昇、下降が、他の株価と相関性を持つことが多々ある。また、出力指示部４により出力指示された場合は、株を買った場合（或いは空売りした場合）と考えることもできるが、仮に解探索システム１ａのみならず、解探索システム１ｂが株を同じ購入した場合には、これに応じて株価が上昇する。これは、上述した出力指示に応じて被検対象５の確率分布が変化することと同様である。

【００６８】

更にこのケースでは、解探索システム１ａ、解探索システム１ｂ間が互いに連関していてもよい。つまり、解探索システム１ａが特定の被検対象５ａを選択した場合、解探索システム１ｂは、被検対象５ｃを選択するように互いに連関するように設定されていてもよい。これにより、例えば株等の投資対象を被検対象５に当てはめる場合において、一の解探索システム１ａが、他の解探索システム１ｂとの間で互いに連関して意思決定を行う場合が多い場合（例えば、Ａ氏がＴ社の株を買った場合に、Ｂ氏は、Ｔ社と業務上の関連性が高いＵ社の株を買う場合）において、その探索精度を向上させることが可能となる。なお、この解探索システム１ａによる出力指示、解探索システム１ｂの出力指示の相関性はいかなるものであってもよい。

【００６９】

また解探索システム１ａ、解探索システム１ｂ間の連関は、互いの制御部３同士が直接的に相互作用するものであってもよい。かかる場合には、上述した解探索システム１ａにおいて計量変数を決める上での演算式を、他の解探索システム１ｂ側にも直接導入することで相関性を持たせるようにしてもよい。

【００７０】

なお、上述した実施形態では、２個の解探索システム１ａ、解探索システム１ｂを用いる場合を例にとり説明をしたが、これに限定されるものではなく２以上であればいかなる数で構成されていてもよい。

【実施例１】

【００７１】

図３は、本発明を適用した解探索システム１を、コンピュータプログラムで実現した場合における実施例を示している。コンピュータプログラムで実現するため、戦績優劣比較部２は、これと同様の処理を実行する戦績優劣比較ステップに、制御部３は、これと同様の処理を実行する制御ステップに、出力指示部４は、これと同様の処理を実行する出力指示ステップとして具現化したプログラムを作成した。

【００７２】

被検対象５は、予め設定された確率の下で当選するか落選するかが決まる装置であり、例えばスロットマシーン等を想定している。この被検対象５－１の当選確率＝０．３５、被検対象５－２の当選確率＝０．４５、被検対象５－３の当選確率＝０．５５、被検対象５－４の当選確率＝０．６５とされている。また５－０は、x_0である。

【００７３】

このうち、当選確率の高い２つの被検対象５の組み合わせ（即ち、被検対象５－３、５－４）を探索解として求めることを行う。

【００７４】

図３の横軸ｔはフィードバック回数である。また、戦績指数q_iは、（１）式に基づいて被検対象５毎に求めている。優劣S_iは、（２）式に基づいて被検対象５毎に求めている。a_iは、表２から導き出される－１、０、１の何れかの数値である。

【００７５】

またv_iは、a_iに基づいて計量変数を増やす方向に推進させるのか、或いは減らす方向に推進させるのかを決めるパラメータである。仮にa_iが計量変数を増減させる際の加速度であるとすえれば、このv_iは、計量変数の増減速度である。a_iは、v_iの１回微分で表すことができる。更に、計量変数x_iは、実際に計量変数の増減速度v_iに基づいて増減させられた実際のパラメータ値である。ここで閾値は０としている。

【００７６】

このようなプログラムによる結果より、フィードバック回数が増加するにつれて、計量変数x_iは、殆ど被検対象５－３、５－４のみが大きくなってそのまま出力指示部４によって出力指示が行われることが繰り返されるのがわかる。

【００７７】

このように本発明を適用したコンピュータプログラムにより、組み合わせバンディット問題を自動的に求めることができることを検証することができた。

【００７８】

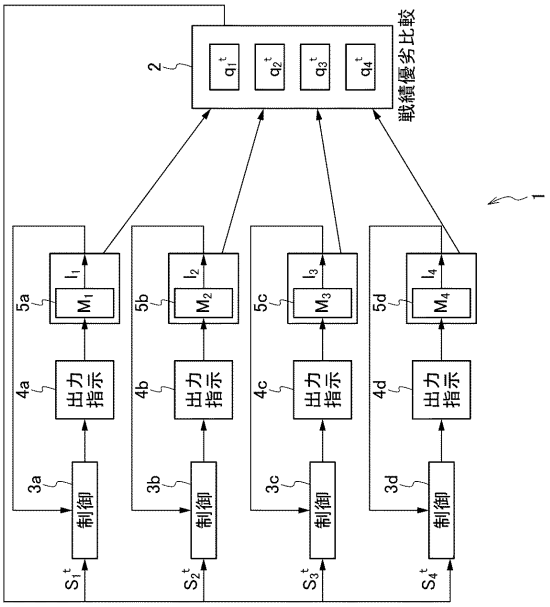
図４は、他の実施例ではあるが、それぞれ被検対象５の当選確率が、０．２、０．５、０．８の３つのサンプルについて、本発明を適用したコンピュータプログラムにより、より当選確率の高い組み合わせを選択するシミュレーションを行った結果である。実線が本発明に相当する。比較例１は、ソフトマックスアルゴリズムを組み合わせ報酬最大化問題に拡張したものである。また比較例２は、イプシロングリーディアルゴリズムを組み合わせ報酬最大化問題に拡張したものである。　その結果、本発明を適用したコンピュータプログラムは、フィードバック回数が増加するにつれて、他の比較例よりも正答率が高くなるのが分かる。

【符号の説明】

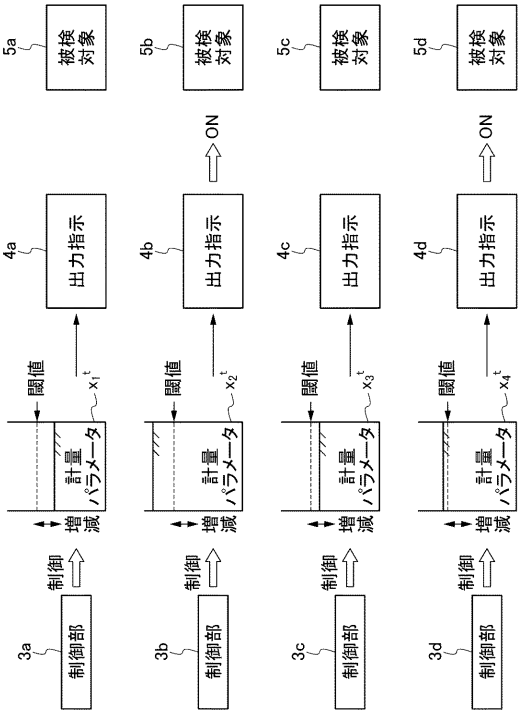
【００７９】

１　解探索システム
２　戦績優劣比較部
３　制御部
４　出力指示部
５　被検対象

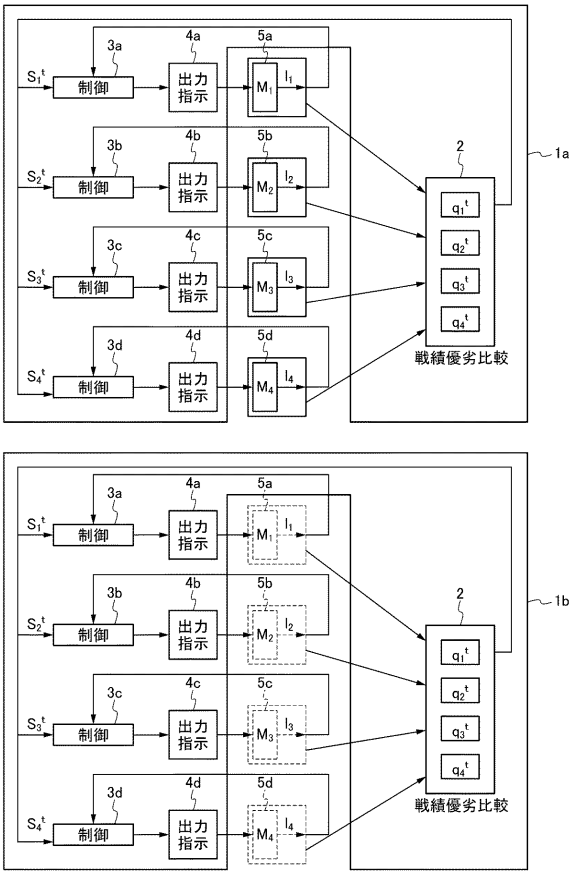
【図１】


戦績優劣比較
制御
出力指示

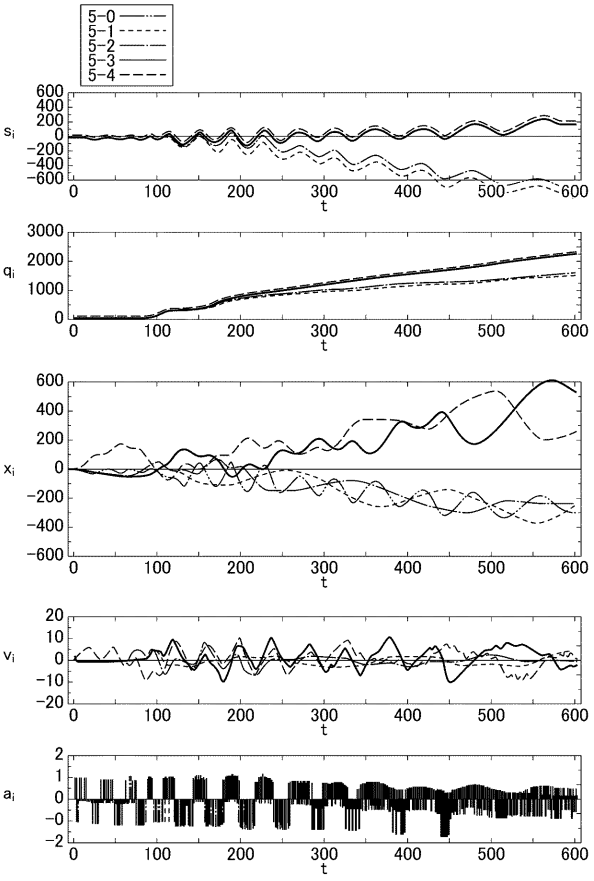
【図２】


被検対象
出力指示
ON
閾値
計量パラメータ
増減
制御
制御部

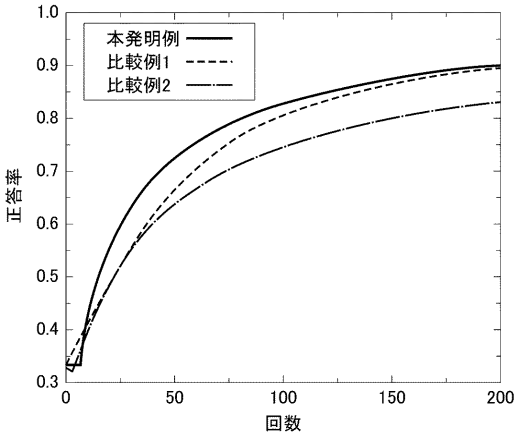
【図３】


制御
出力指示
戦績優劣比較

【図４】


5-0
5-1
5-2
5-3
5-4
t

【図５】



フロントページの続き

(72)発明者　原　正彦
埼玉県和光市広沢２番１号　独立行政法人理化学研究所内

審査官　松野　広一

(56)参考文献　米国特許出願公開第２００８／０１４０５９１（ＵＳ，Ａ１）
AUER, Peter et al, Finite-time Analysis of the Multiarmed Bandit Problem, Machine learning, ２００２年, Vol.47, pp.235-256
VERMOREL, Joanne`s, MOHRI, Mehryar, Multi-Armed Bandit Algorithms and Empirical Evaluation, In Proceedings of the 16th European Conference on Machine Learning, ２００５年１０月, Vol.3720, pp.437-448, ＵＲＬ, http://www.cs.nyu.edu/ mohri/pub/bandit.pdf
大用 庫智　外２名，非定常Ｎ本腕バンディット問題に対する人間の認知バイアスの適用，２０１１年度人工知能学会全国大会（第２５回）論文集　［ＣＤ－ＲＯＭ］　２０１１年度人工知能学会全国大，日本，２０１１年　６月　１日，pp.1-4
青野 真士　外４名，粘菌コンピュータと確率探索アルゴリズム，システム／制御／情報，日本，システム制御情報学会，２０１１年１２月１５日，Vol.55 No.12，pp.26-31

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｑ　１０／００－９９／００
Ｇ０６Ｆ　１９／００
Ｇ０６Ｎ　３／００
Ｇ０６Ｎ　５／０４