

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)公開番号
特開2023-16762
(P2023-16762A)

(43)公開日 令和5年2月2日(2023.2.2)

(51)国際特許分類

G 0 6 N 20/00 (2019.01)

F I

G 0 6 N 20/00

審査請求		未請求	請求項の数	12	O L	外国語出願	(全17頁)
(21)出願番号	特願2022-116210(P2022-116210)		(71)出願人	390023711			
(22)出願日	令和4年7月21日(2022.7.21)			ローベルト ボツシュ ゲゼルシャフト			
(31)優先権主張番号	10 2021 207 868.0			ミット ベシユレンクテル ハフツング			
(32)優先日	令和3年7月22日(2021.7.22)			ROBERT BOSCH GMBH			
(33)優先権主張国・地域又は機関	ドイツ(DE)			ドイツ連邦共和国 シュツツトガルト (番地なし)			
				Stuttgart, Germany			
			(74)代理人	100114890			
				弁理士 アインゼル・フェリックス＝ラインハルト			
			(74)代理人	100098501			
				弁理士 森田 拓			
			(74)代理人	100116403			
				弁理士 前川 純一			
			(74)代理人	100134315			
			最終頁に続く				

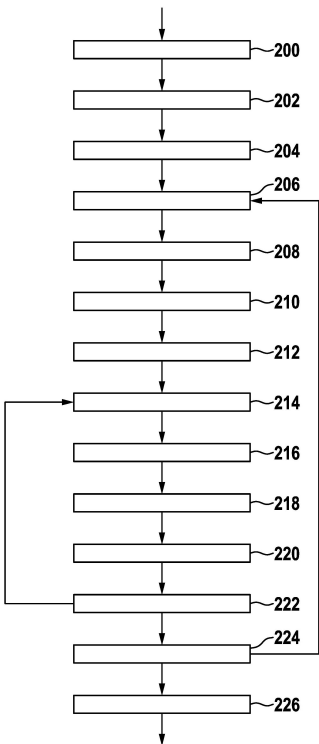
(54)【発明の名称】 機械学習のための装置、コンピュータプログラム及びコンピュータ実装された方法

(57)【要約】

【課題】機械学習のための装置、コンピュータプログラム及びコンピュータ実装された方法を提供する。

【解決手段】本方法は、多腕バンディット問題又は文脈付きバンディット問題の行動空間と、行動に対して調整された報酬にわたる分布を含むタスクを提供すること（200）と、行動空間にわたる分布であるハイパーパラメータ事前分布を提供すること（202）と、ハイパーパラメータ事前分布に関連して、行動空間にわたる分布であるハイパーパラメータ事後分布を決定すること（224）とを含み、ハイパーパラメータ事後分布に対しては、ハイパーパラメータ事後分布からサンプリングされた事前分布を使用する場合に、将来のバンディットタスクに対して予期される報酬に対する下限が、可能な限り大きい値を有する。

【選択図】図2



【特許請求の範囲】**【請求項 1】**

機械学習のためのコンピュータ実装された方法であって、

前記方法は、

多腕バンディット問題又は文脈付きバンディット問題の行動空間と、行動に対して調整された報酬にわたる分布とを含むタスクを提供すること（200）と、

前記行動空間にわたる分布であるハイパーパラメータ事前分布を提供すること（202）と、

前記ハイパーパラメータ事前分布に関連して、前記行動空間にわたる分布であるハイパーパラメータ事後分布を決定すること（224）と、

を含み、

前記ハイパーパラメータ事後分布に対しては、前記ハイパーパラメータ事後分布からサンプリングされた事前分布を使用する場合に、将来のバンディットタスクに対して予想される報酬に対する下限が、可能な限り大きい値を有する、ことを特徴とする方法。

【請求項 2】

前記ハイパーパラメータ事後分布を決定すること（224）は、前記予想される報酬に対する前記下限を最大化する前記ハイパーパラメータ事後分布を決定することを含む、請求項 1 に記載の方法。

【請求項 3】

特にセンサデータを分類するために、前記ハイパーパラメータ事後分布からサンプリングされた事前分布に関連して、前記センサデータ、特にデジタル画像データ若しくはオーディオデータを処理すること（226）、

前記センサデータ内の対象物の存在を検出すること若しくは前記センサデータに対してセマンティックセグメンテーションを実行すること、又は、

前記ハイパーパラメータ事後分布から事前分布をサンプリングする場合に、前記機械学習のロバスト性の指標、特に、次のタスクに対して予想される誤差が予め定められている値を上回らない確率を決定すること（226）、又は、

前記ハイパーパラメータ事後分布からサンプリングされた事前分布に関連してセンサデータにおける異常を検出すること（226）、又は、

物理システムをコントロールするためのポリシーを学習すること及び前記ハイパーパラメータ事後分布からサンプリングされた事前分布に関連して前記物理システムをコントロールするためのコントロール信号を決定すること

を含む、

請求項 1 又は 2 に記載の方法。

【請求項 4】

複数の反復において前記ハイパーパラメータ事後分布を決定すること（224）と、

先行する反復の前記ハイパーパラメータ事後分布から反復の前記事前分布をサンプリングすること（208）と、

を含む、

請求項 1 乃至 3 のいずれか一項に記載の方法。

【請求項 5】

タスクにわたる分布から前記反復の前記タスクをサンプリングすること（206）を含む、

請求項 4 に記載の方法。

【請求項 6】

前記ハイパーパラメータ事前分布によって前記ハイパーパラメータ事後分布を初期化すること（204）を含む、

請求項 1 乃至 5 のいずれか一項に記載の方法。

【請求項 7】

前記事前分布によって事後分布を初期化すること（２１０）と、
 前記タスクの行動ポリシーのセットから、前記タスク事後分布に関連する、確率質量を伴う行動にわたる分布を含む行動ポリシーを決定すること（２１４）と、
 前記確率質量からランダムに行動をサンプリング（２１６）又は選択することと、
 報酬にわたる前記分布から前記行動に関連して報酬をサンプリングすること（２１８）と、
 前記行動及び前記報酬を含むデータセットを決定すること（２２０）と、
 前記タスクデータセットを含むように前記事後分布を更新すること（２２２）と、
 を含む、
 請求項１乃至６のいずれか一項に記載の方法。

10

【請求項８】
 前記タスクを提供すること（２００）は、状態空間と、初期状態にわたる分布とを含む前記タスクを提供することを含み、
 前記方法はさらに、初期状態にわたる前記分布から初期状態をランダムにサンプリング（２１８）又は選択することを含み、
 報酬にわたる前記分布は、前記行動及び前記状態空間の状態に対して調整される、
 請求項７に記載の方法。

【請求項９】
 空のセットによって前記データセットを初期化すること（２１２）と、
 次いで、予め定められた数のラウンドにおいて前記事後分布を更新することと、
 を含む、
 請求項１乃至８のいずれか一項に記載の方法。

20

【請求項１０】
 前記ハイパーパラメータ事後分布を決定すること（２２４）は、前記ハイパーパラメータ事後分布及び前記ハイパーパラメータ事前分布のカルバック・ライブラ発散に関連した前記予期される報酬の決定及び近似を含む、
 請求項１乃至９のいずれか一項に記載の方法。

【請求項１１】
 請求項１乃至１０のいずれか一項に記載の方法におけるステップを実施するように構成されていることを特徴とする、機械学習のための装置（１００）。

30

【請求項１２】
 コンピュータ上において実行されるときに、前記コンピュータに、請求項１乃至１０のいずれか一項に記載の方法を実施させるためのコンピュータ可読命令を含むことを特徴とするコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

背景技術
 本発明は、機械学習のための装置、コンピュータプログラム及びコンピュータ実装された方法に関する。

40

【背景技術】

【０００２】

「A PAC-Bayesian bound for Lifelong Learning」(Anastasia Pentina, Christoph H. Lampert 著, arXiv:1311.2838) は、機械学習のための生涯学習設定の態様を開示している。

【先行技術文献】

【非特許文献】

【０００３】

【非特許文献１】「A PAC-Bayesian bound for Lifelo

50

ng Learning」(Anastasia Pentina, Christopher H. Lampert 著, arXiv:1311.2838)

【発明の概要】

【発明が解決しようとする課題】

【0004】

発明の開示

独立請求項に記載された、コンピュータ実装された方法、装置及びコンピュータプログラムは、特に、生涯学習システムをトレーニングするための目的関数に基づいて、改良された機械学習を提供する。

【課題を解決するための手段】

【0005】

機械学習のためのコンピュータ実装された方法は、多腕バンディット問題又は文脈付きバンディット問題の行動空間と、行動に対して調整された報酬にわたる分布とを含むタスクを提供することと、行動空間にわたる分布であるハイパーパラメータ事前分布(hyperprior)を提供することと、ハイパーパラメータ事前分布に関連して、行動空間にわたる分布であるハイパーパラメータ事後分布(hyperposterior)を決定することと、を含み、ハイパーパラメータ事後分布に対しては、ハイパーパラメータ事後分布からサンプリングされた事前分布を使用する場合に、将来のバンディットタスクに対して予期される報酬に対する下限が、可能な限り大きい値を有する。これは、これらのタスクが、多腕バンディット問題又は文脈付きバンディット問題のいずれかであることを意味する。下限は、PACベイズ汎化限界であり、ハイパーパラメータ事後分布からサンプリングされた事前分布を使用する場合に、将来のバンディットタスクに対して予期される報酬を下限とする。下限は、生涯学習のための目的関数として使用され、ここでは、タスクの観察の後、下限を近似的に最大化するようにハイパーパラメータ事後分布が更新される。下限は、観察可能な量を含み、その結果、観察されたタスクからのトレーニングデータを使用して計算され得る。このようにして、観察された多腕バンディットタスク又は文脈付きバンディットタスクのセットからの事前知識が、新たな多腕バンディットタスク又は文脈付きバンディットタスクに転用される。

【0006】

ハイパーパラメータ事後分布を決定することは、予期される報酬に対する下限を最大化するハイパーパラメータ事後分布を決定することを含み得る。下限を最大化するハイパーパラメータ事後分布は、結果として将来のタスクにおける誤差が少なくなる事前分布に高い確率を割り当てることが予期される。

【0007】

この方法を用いて、観察された多腕バンディットタスク又は文脈付きバンディットタスクのセットからの事前知識が、新たな多腕バンディットタスク又は文脈付きバンディットタスクに転用され得る。多くの問題を、多腕バンディット問題又は文脈付きバンディット問題のいずれかとして表すことができる。従って、この方法を使用して、多くの種類のタスクに事前知識を転用することができる。

【0008】

この方法は、

特にセンサデータを分類するために、ハイパーパラメータ事後分布からサンプリングされた事前分布に関連して、センサデータ、特にデジタル画像データ若しくはオーディオデータを処理すること、

センサデータ内の対象物の存在を検出すること若しくはセンサデータに対してセマンティックセグメンテーションを実行すること、又は、

ハイパーパラメータ事後分布から事前分布をサンプリングする場合に、機械学習のロバスト性の指標、特に、次のタスクに対して予期される誤差が予め定められた値を上回らない確率を決定すること、又は、

ハイパーパラメータ事後分布からサンプリングされた事前分布に関連してセンサデータ

10

20

30

40

50

における異常を検出すること、又は、

物理システムをコントロールするためのポリシーを学習すること及びハイパーパラメータ事後分布からサンプリングされた事前分布に関連して物理システムをコントロールするためのコントロール信号を決定すること

を含み得る。従って、この方法は、新たな問題、特に新たな分類問題、新たなロバスト性問題、新たな異常検出問題、及び／又は、新たなコントロール問題に事前知識を転用するために使用される。

【0009】

下限は、機械学習のロバスト性に関する指標でもある。ハイパーパラメータ事後分布は、事前分布がハイパーパラメータ事後分布からサンプリングされる場合に、次のタスクに対して予期される誤差に関する情報も提供する。予期される誤差は、例えば、事前分布が特定のハイパーパラメータ事後分布からサンプリングされる場合、特定の値を上回らない。

10

【0010】

この方法は、複数の反復においてハイパーパラメータ事後分布を決定することと、先行する反復のハイパーパラメータ事後分布から反復の事前分布をサンプリングすることと、を含み得る。従って、機械学習の反復において、事前知識が先行する反復から転用される。

【0011】

この方法は、タスクにわたる分布から反復のタスクをサンプリングすることを含み得る。特に、多数の種々異なるタスクを処理することによって、機械学習の結果が改良される。

20

【0012】

この方法は、好ましくは、ハイパーパラメータ事前分布によってハイパーパラメータ事後分布を初期化することを含む。利用可能なハイパーパラメータ事前分布はいずれも、機械学習のための良好な起点である。

【0013】

この方法は、事前分布によって事後分布を初期化することと、タスクの行動ポリシーのセットから、タスク事後分布に関連する、確率質量を伴う行動にわたる分布を含む行動ポリシーを決定することと、確率質量からランダムに行動をサンプリング又は選択することと、報酬にわたる分布から行動に関連して報酬をサンプリングすることと、行動及び報酬を含むタスクデータセットを決定することと、タスクデータセットを含むようにタスク事後分布を更新することと、を含み得る。これによって、多腕バンディットタスク又は文脈付きバンディットタスクに対して極めて効率的な機械学習が得られ、ここでは、生成された事後分布によって、先行する反復の知識としてタスクデータセットが保存される。

30

【0014】

タスクを提供することは、好ましくは、状態空間と、初期状態にわたる分布とを含むタスクを提供することを含み、この方法はさらに、初期状態にわたる分布から初期状態をランダムにサンプリング又は選択することを含み、報酬にわたる分布は、行動及び状態空間の状態に対して調整される。これによって、文脈付きバンディットタスクのための極めて効率的な機械学習が得られる。

40

【0015】

この方法は、空のセットによってタスクデータセットを初期化することと、次いで、予め定められた数のラウンドにおいてタスク事後分布を更新することと、を含み得る。こうして、観察されたタスクからのトレーニングデータから下限が計算される。

【0016】

ハイパーパラメータ事後分布を決定することは、ハイパーパラメータ事後分布及びハイパーパラメータ事前分布のカルバック・ライブラ発散に関連して下限を決定することを含み得る。観察されたタスクと新たなタスクとにより同程度の効果が得られるように事前分布が最適化される場合、目標である改良された機械学習が得られる。ハイパーパラメータ

50

事後分布及びハイパーパラメータ事前分布のカルバック・ライブラ発散は、観察されたタスクからのトレーニングデータから下限を決定する優れた方式を提供する。

【0017】

請求項1乃至10のいずれか一項に記載の方法におけるステップを実施するように構成されていることを特徴とする、機械学習のための装置を提供する。

【0018】

コンピュータ上において実行されるときに、コンピュータに、請求項1乃至10のいずれか一項に記載の方法を実施させるためのコンピュータ可読命令を含むことを特徴とするコンピュータプログラムを提供する。

【0019】

以下の説明及び図面から、さらなる有利な実施形態を導き出すことができる。

【図面の簡単な説明】

【0020】

【図1】機械学習のための装置の一部を概略的に示す図である。

【図2】機械学習のための方法のステップを示す図である。

【発明を実施するための形態】

【0021】

図1は、機械学習のための装置100の一部を概略的に示している。装置100は、少なくとも1つのプロセッサ102と、少なくとも1つの記憶装置104とを備えている。少なくとも1つの記憶装置104は、コンピュータ上において実行されるときに、コンピュータに、図2を参照して以下において説明する方法を実施させるためのコンピュータ可読命令を含むコンピュータプログラムを格納し得る。装置100は、特に、少なくとも1つのプロセッサ102がコンピュータプログラムの命令を実行するときに、方法のステップを実施するように構成されている。

【0022】

この例における機械学習の方法は、基礎学習アルゴリズム $Q = Q(D, P)$ を使用し、ステップ200から開始する。基礎学習アルゴリズムは、事後分布 Q を返す。基礎学習アルゴリズムは、データセット D と事前分布 P とを使用するので、事後分布は、 D 及び P への関連を明確にするために、この明細書においては、 $Q(D, P)$ として記載される。

【0023】

ステップ200において、この方法は、タスク T_i を提供することを含む。

【0024】

タスク T_i は、多腕バンディット問題の行動空間 A 即ち行動 a のセットと、行動 a に対して調整された報酬 r にわたる分布 $p_i(r | a)$ とを含む対

$T_i = (A, p_i)$

として表される、タスクである多腕バンディット問題であり得る。

【0025】

タスク T_i は、行動空間 A 即ち行動 a のセットと、状態空間 S と、行動 a 及び状態 s に対して調整された報酬 r にわたる分布 $p_i(r | s, a)$ と、文脈付きバンディット問題の初期状態 $\mu(s)$ にわたる分布とを含むクアドルプル（四部組：quadruple）

$T_i = (S, A, \mu_i, p_i)$

として表される、タスクである文脈付きバンディット問題であり得る。

【0026】

報酬 r は、0と1との間にあると仮定され、タスク T_i は、環境 T からのサンプリングされた $i.i.d$ （独立同分布）であると仮定される。

【0027】

その後、ステップ202が実行される。

【0028】

ステップ202において、この方法は、ハイパーパラメータ事前分布 P を提供することを含む。

10

20

30

40

50

【0029】

ハイパーパラメータ事前分布 P は、可能な事前分布のセットにわたる分布である。各事前分布は、行動空間 A にわたる分布である。

【0030】

その後、ステップ 204 が実行される。

【0031】

ステップ 204 において、この方法は、ハイパーパラメータ事前分布 P によってハイパーパラメータ事後分布 Q を初期化することを含む。ハイパーパラメータ事後分布 Q は、事前分布にわたる別の分布である。

【0032】

その後、ステップ 206 が実行される。

【0033】

ステップ 206 において、この方法は、タスク T にわたる分布 T から反復 i のタスク T_i をサンプリングすることを含む。

【0034】

タスク T_i には、m_i 個の行動ポリシーを含むタスク T_i の行動ポリシーのセット

【数 1】

$$B_{ij} = (b_{i1}, \dots, b_{im_i})$$

10

20

が関連づけられる。

【0035】

行動ポリシー b_{ij} は、多腕バンディット問題の場合には、確率質量 b_{ij}(a) を伴う行動 a にわたる分布を含む。この例においては、行動ポリシー b_{ij} は、以前に観測されたトレーニングデータ

【数 2】

$$D_1, \dots, D_{1-i}, D_i^{j-1}$$

に対する反復 i に関連する行動 a にわたる分布であり、多腕バンディットタスクの場合には、

30

【数 3】

$$D_i^{j-1} = ((a_{i1}, r_{i1}), \dots, (a_{ij-1}, r_{ij-1}))$$

である。

【0036】

トレーニングデータセットの要素は、多腕バンディットタスクにおいては、

【数 4】

$$a_{ij} \sim b_{ij}(a | D_1, \dots, D_{1-i}, D_i^{j-1}), r_{ij} \sim \rho_i(r | a_{ij})$$

40

のように分布する。

【0037】

行動ポリシー b_{ij} は、文脈付きバンディット問題の場合には、確率質量 b_{ij}(a | s_{ij}) を伴う、状態 s に対して調整された行動 a にわたる分布を含む。この例においては、行動ポリシー b_{ij} は、以前に観測されたトレーニングデータ

【数 5】

50

$$D_i^{j-1} = \left((s_{i1}, a_{i1}, r_{i1}), \dots, (s, a_{ij-1}, r_{ij-1}) \right)$$

に対する反復 i に関連する行動 a にわたる分布である。

【 0 0 3 8 】

トレーニングデータセットの要素は、文脈付きバンディットタスクにおいては、

【 数 6 】

$$a_{ij} \sim b_{ij}(a|s_{ij}, D_1, \dots, D_{1-i}, D_i^{j-1}), r_{ij} \sim \rho_i(r|s_{ij}, a_{ij})$$

10

のように分布する。

【 0 0 3 9 】

その後、ステップ 2 0 8 が実行される。

【 0 0 4 0 】

ステップ 2 0 8 において、この方法は、以前の反復 $i - 1$ のハイパーパラメータ事後分布 Q から反復 i の事前分布 P をサンプリングすることを含む。

【 0 0 4 1 】

その後、ステップ 2 1 0 が実行される。

20

【 0 0 4 2 】

ステップ 2 1 0 において、この方法は、事前分布 P によって事後分布 Q を初期化することを含む。

【 0 0 4 3 】

その後、ステップ 2 1 2 が実行される。

【 0 0 4 4 】

ステップ 2 1 2 において、この方法は、空のセットによってデータセット D_i を初期化することを含む。

【 0 0 4 5 】

その後、ステップ 2 1 4 が実行される。

30

【 0 0 4 6 】

ステップ 2 1 4 において、この方法は、タスク T_i の行動ポリシー B_{ij} のセットから、タスク事後分布 Q に関連づけられた行動ポリシー b_{ij} を決定することを含む。

【 0 0 4 7 】

その後、ステップ 2 1 6 が実行される。

【 0 0 4 8 】

ステップ 2 1 6 において、この方法は、行動ポリシー b_{ij} の確率質量から行動 a_{ij} をランダムにサンプリング又は選択することを含む。

【 0 0 4 9 】

その後、ステップ 2 1 8 が実行される。

40

【 0 0 5 0 】

ステップ 2 1 8 において、この方法は、多腕バンディット問題の場合には、報酬 r にわたる分布 $p_i(r|a)$ から行動 a_{ij} に関連して、報酬 r_{ij} をサンプリングすることを含む。

【 0 0 5 1 】

ステップ 2 1 8 において、この方法は、文脈付きバンディット問題の場合には、初期状態 $\mu(s)$ にわたる分布から初期状態 s_{ij} をランダムにサンプリング又は選択することと、行動 a と状態空間 S の状態 s に対して調整された報酬 r にわたる分布 $p_i(r|s, a)$ から行動 a_{ij} に関連して報酬 r_{ij} をサンプリングすることと、を含む。

【 0 0 5 2 】

50

その後、ステップ 220 が実行される。

【0053】

ステップ 220 において、この方法は、行動 a_{ij} 及び報酬 r_{ij} を含むデータセット D_i を決定することを含む。

【0054】

データセット D_i は、多腕バンディット問題の場合に、 m_i 個の観察可能なトレーニングデータ対を含むトレーニングセットである：

【数 7】

$$D_i = ((a_{i1}, r_{i1}), \dots, (a_{im_i}, r_{im_i}))$$

10

【0055】

データセット D_i は、文脈付きバンディット問題の場合に、 m_i 個の観察可能なトレーニングデータ・トリプル（三部組：triplet）を含むトレーニングセットである：

$$D_i = ((s_{i1}, a_{i1}, r_{i1}), \dots, (s_{ij-1}, a_{ij-1}, r_{ij-1}))$$

【0056】

その後、ステップ 222 が実行される。

【0057】

ステップ 222 において、この方法は、データセット D_i を含むように事後分布 Q を更新することを含む。

20

【0058】

このために、決定的基礎学習アルゴリズムが、データセット D_i と事前分布 P とを入力として受け取り、事後分布 $Q = Q(D_i; P)$ を生じさせる。

【0059】

タスク T_i の目標は、予期される報酬を最大化する事後分布 Q を見出すことである。多腕バンディット問題の場合には、予期される報酬は、以下の通りである：

【数 8】

$$L_i(a) = E_{r \sim \rho_i(r|a)}[r]$$

30

ここで、 E は、予期値（expectancy）である。

【0060】

多腕バンディット問題の例においては、データセット D_i を用いて、予期される報酬は、以下のように推定される：

【数 9】

$$\hat{L}_i(a, D_i) = \frac{1}{m_i} \sum_{j=1}^{m_i} \frac{1}{b_{ij}(a_{ij})} r_{ij} I\{a_{ij} = a\}$$

40

$I\{a_{ij} = a\}$ は、 $a_{ij} = a$ の場合は 1 を、それ以外の場合は 0 を返す指示関数である。文脈付きバンディット問題の場合には、予期される報酬は、以下の通りである：

【数 10】

$$L_i^{(cb)}(\pi) = E_{s \sim \mu_i} E_{a \sim \pi(a|s)} E_{r \sim \rho_i(r|s, a)}[r]$$

ここで、 E は、各予期値であり、事前分布 P 及び事後分布 Q は、ポリシーの空間 Π にわ

50

たる分布であり、ここで、 $\pi \in \Pi$ は、状態 s に対して調整された行動 a にわたる分布 $\pi (a | s)$ である。

【0061】

文脈付きバンディット問題の例においては、データセット D_i を用いて、予期される報酬は、以下のように推定される：

【数11】

$$\hat{L}_i^{(cb)}(\pi, D_i) = \frac{1}{m_i} \sum_{j=1}^{m_i} \frac{\pi(a_{ij}|s_{ij})}{b_{ij}(a_{ij}|s_{ij})} r_{ij}$$

10

【0062】

この例において、この方法は、予め定められたラウンド数 m において事後分布 Q を更新することを含む。これは、この方法が、 m 回のラウンドのためにステップ 214 に続くことを意味する。

【0063】

その後、ステップ 224 が実行される。

【0064】

ステップ 224 において、この方法は、ハイパーパラメータ事前分布 P に関連して、ハイパーパラメータ事後分布 Q を決定することを含み、このハイパーパラメータ事後分布 Q に対しては、このハイパーパラメータ事後分布 Q からサンプリングされた事前情報 P を使用する場合に、将来のバンディットタスクに対して予期される報酬 L の下限が可能な限り大きい値を有する。

20

【0065】

この例においてハイパーパラメータ事後分布 Q を決定することは、予期される報酬 L に対する下限を最大化するハイパーパラメータ事後分布 Q を決定することを含む。

【数12】

$$Q \leftarrow \operatorname{argmax}_{Q'} \text{lower bound}(L(Q'))$$

30

【0066】

この例においては、多腕バンディット問題に対する、関数である下限 ($L(Q')$) は、以下の目的関数のうちの1つである：

【数13】

$$LIKLa(Q) = -\frac{1}{\sqrt{n}} (D_{KL}(Q||P) + \frac{1}{8} - \ln(\frac{\delta}{2})) + \frac{1}{\kappa_{max}} k l^{-1} \left(\frac{1}{n} \sum_{i=1}^n \kappa_i \hat{L}_i(Q(D_i, P_i), D_i) \right) \left\| \frac{D_{KL}(Q||P) + \sum_{i=1}^n D_{KL}(Q(D_i, P_i)||P_i) + n \ln(\xi(m)) - \ln(\delta/2)}{nm} \right\|$$

40

【数14】

$$LIKLb(Q) = -\frac{1}{\sqrt{8n}} \left(\frac{D_{KL}(Q||P)}{\sqrt{-\ln(\delta/2)}} + \sqrt{8} + 2\sqrt{-\ln(\delta/2)} \right) + \frac{1}{\kappa_{max}} k l^{-1} \left(\frac{1}{n} \sum_{i=1}^n \kappa_i \hat{L}_i(Q(D_i, P_i), D_i) \right) \left\| \frac{D_{KL}(Q||P) + \sum_{i=1}^n D_{KL}(Q(D_i, P_i)||P_i) + n \ln(\xi(m)) - \ln(\delta/2)}{nm} \right\|$$

【数15】

50

$$LP_a(Q) = \frac{1}{n} \sum_{i=1}^n \frac{\kappa_i}{\kappa_{max}} \hat{L}_i(Q_i(D_i, P_i), D_i) - \frac{1}{\sqrt{n}} \left(D_{KL}(Q||P) + \frac{1}{8} - \ln\left(\frac{\delta}{2}\right) \right) \\ - \frac{\sqrt{D_{KL}(Q||P) + \sum_{i=1}^n D_{KL}(Q(D_i, P_i)||P_i) + n \ln(\xi(m)) - \ln(\delta/2)}}{\kappa_{max} \sqrt{2nm}}$$

【数 1 6】

$$LP_b(Q) = \frac{1}{n} \sum_{i=1}^n \frac{\kappa_i}{\kappa_{max}} \hat{L}_i(Q_i(D_i, P_i), D_i) - \frac{1}{\sqrt{8n}} \left(\frac{D_{KL}(Q||P)}{\sqrt{-\ln(\delta/2)}} + \sqrt{8} + 2\sqrt{-\ln(\delta/2)} \right) \\ - \frac{\sqrt{D_{KL}(Q||P) + \sum_{i=1}^n D_{KL}(Q(D_i, P_i)||P_i) + n \ln(\xi(m)) - \ln(\delta/2)}}{\kappa_{max} \sqrt{2nm}} \quad 10$$

【数 1 7】

$$LHA_a(Q) = \frac{1}{n} \sum_{i=1}^n \hat{L}_i(Q_i(D_i, P_i), D_i) - \frac{1}{\sqrt{n}} \left(D_{KL}(Q||P) + \frac{1}{8} - \ln\left(\frac{\delta}{2}\right) \right) \\ - \frac{\beta}{n} \left(\frac{D_{KL}(Q||P) + \sum_{i=1}^n D_{KL}(Q(D_i, P_i)||P_i)}{\sqrt{-\ln(\delta/2)}} + 2\sqrt{-\ln\left(\frac{\delta}{2}\right)} \right) \quad 20$$

【数 1 8】

$$LHA_b(Q) = \frac{1}{n} \sum_{i=1}^n \hat{L}_i(Q_i(D_i, P_i), D_i) - \frac{1}{\sqrt{8n}} \left(\frac{D_{KL}(Q||P)}{\sqrt{-\ln(\delta/2)}} + \sqrt{8} + 2\sqrt{-\ln(\delta/2)} \right) \\ - \frac{\beta}{n} \left(\frac{D_{KL}(Q||P) + \sum_{i=1}^n D_{KL}(Q(D_i, P_i)||P_i)}{\sqrt{-\ln(\delta/2)}} + 2\sqrt{-\ln\left(\frac{\delta}{2}\right)} \right)$$

【数 1 9】

$$CLIKL_a(Q) = -\frac{\beta_1}{\sqrt{n}} \left(D_{KL}(Q||P) + \frac{1}{8} - \ln\left(\frac{\delta}{2}\right) \right) \\ + \frac{1}{\kappa_{max}} kl_-^{-1} \left(\frac{1}{n} \sum_{i=1}^n \kappa_i \hat{L}_i(Q(D_i, P_i), D_i) \right) \left\| \beta_2 \frac{D_{KL}(Q||P) + \sum_{i=1}^n D_{KL}(Q(D_i, P_i)||P_i) + n \ln(\xi(m)) - \ln(\delta/2)}{nm} \right) \quad 30$$

【数 2 0】

$$CLIKL_b(Q) = -\frac{\beta_1}{\sqrt{8n}} \left(\frac{D_{KL}(Q||P)}{\sqrt{-\ln(\delta/2)}} + \sqrt{8} + 2\sqrt{-\ln(\delta/2)} \right) \\ + \frac{1}{\kappa_{max}} kl_-^{-1} \left(\frac{1}{n} \sum_{i=1}^n \kappa_i \hat{L}_i(Q(D_i, P_i), D_i) \right) \left\| \beta_2 \frac{D_{KL}(Q||P) + \sum_{i=1}^n D_{KL}(Q(D_i, P_i)||P_i) + n \ln(\xi(m)) - \ln(\delta/2)}{nm} \right) \quad 40$$

【数 2 1】

$$CLP_a(Q) = \frac{1}{n} \sum_{i=1}^n \frac{\kappa_i}{\kappa_{max}} \hat{L}_i(Q_i(D_i, P_i), D_i) - \frac{\beta_1}{\sqrt{n}} \left(D_{KL}(Q||P) + \frac{1}{8} - \ln\left(\frac{\delta}{2}\right) \right) \\ - \frac{\beta_2 \sqrt{D_{KL}(Q||P) + \sum_{i=1}^n D_{KL}(Q(D_i, P_i)||P_i) + n \ln(\xi(m)) - \ln(\delta/2)}}{\kappa_{max} \sqrt{2nm}}$$

【数 2 2】

$$CLP_b(Q) = \frac{1}{n} \sum_{i=1}^n \frac{\kappa_i}{\kappa_{max}} \hat{L}_i(Q_i(D_i, P_i), D_i) - \frac{\beta_1}{\sqrt{8n}} \left(\frac{D_{KL}(Q||P)}{\sqrt{-\ln(\delta/2)}} + \sqrt{8} + 2\sqrt{-\ln(\delta/2)} \right) \\ - \frac{\beta_2 \sqrt{D_{KL}(Q||P) + \sum_{i=1}^n D_{KL}(Q(D_i, P_i)||P_i) + n \ln(\xi(m)) - \ln(\delta/2)}}{\kappa_{max} \sqrt{2nm}}$$

【数 2 3】

$$CLHA_a(Q) = \frac{1}{n} \sum_{i=1}^n \hat{L}_i(Q_i(D_i, P_i), D_i) - \frac{\beta_1}{\sqrt{n}} \left(D_{KL}(Q||P) + \frac{1}{8} - \ln\left(\frac{\delta}{2}\right) \right) \\ - \frac{\beta_2 \beta}{n} \left(\frac{D_{KL}(Q||P) + \sum_{i=1}^n D_{KL}(Q(D_i, P_i)||P_i)}{\sqrt{-\ln(\delta/2)}} + 2\sqrt{-\ln\left(\frac{\delta}{2}\right)} \right) \quad 10$$

【数 2 4】

$$CLHA_b(Q) = \frac{1}{n} \sum_{i=1}^n \hat{L}_i(Q_i(D_i, P_i), D_i) - \frac{\beta_1}{\sqrt{8n}} \left(\frac{D_{KL}(Q||P)}{\sqrt{-\ln(\delta/2)}} + \sqrt{8} + 2\sqrt{-\ln(\delta/2)} \right) \\ - \frac{\beta_2 \beta}{n} \left(\frac{D_{KL}(Q||P) + \sum_{i=1}^n D_{KL}(Q(D_i, P_i)||P_i)}{\sqrt{-\ln(\delta/2)}} + 2\sqrt{-\ln\left(\frac{\delta}{2}\right)} \right) \quad 20$$

ここで、 β_1 及び β_2 は、定数 $0 \leq \beta_1$ 、 $\beta_2 \leq 1$ であり、
ここで、

【数 2 5】

$$\varepsilon(m) = \sum_{k=0}^m \binom{m}{k} \left(\frac{k}{m}\right)^k \left(1 - \frac{k}{m}\right)^{m-k}$$

であり、ここで、 $0 < \delta \leq 1$ は、信頼度パラメータであるので、多腕バンディット問題の場合には、次の不等式はそれぞれ、確率 $1 \sim \delta$ でトレーニングデータ $D_1, \dots, D_n$ 及び事後分布 $Q_1, \dots, Q_n$ にわたって成立する： 30

【数 2 6】

$$\frac{1}{n} \sum_{i=1}^n L_i(Q_i) \geq \frac{1}{\kappa_{max}} k l^{-1} \left(\frac{1}{n} \sum_{i=1}^n \kappa_i \hat{L}_i(Q_i, D_i) \left\| \frac{D_{KL}(Q^n||P^n) + n \ln(\varepsilon(m)) - \ln(\delta)}{nm} \right\| \right)$$

【数 2 7】

$$\frac{1}{n} \sum_{i=1}^n L_i(Q_i) \geq \frac{1}{n} \sum_{i=1}^n \frac{\kappa_i}{\kappa_{max}} \hat{L}_i(Q_i, D_i) \\ - \frac{1}{\kappa_{max} \sqrt{2nm}} \sqrt{D_{KL}(Q^n||P^n) + n \ln(\varepsilon(m)) - \ln(\delta)} \quad 40$$

ここで、 Q^n は、行動 $a_1, \dots, a_n$ にわたる同時分布を指し、ここで $a_i \sim Q_i$ であり、
ここで、 P^n は、行動 $a_1, \dots, a_n$ にわたる同時分布を指し、ここで $a_i \sim P_i$ であり、
ここで、

【数 2 8】

$$\kappa_i = \min_{a \in A, j=1, \dots, m} (b_{ij}(a))$$

であり、

【数 2 9】

$$\kappa_{max} = \max_i \kappa_i$$

であり、さらに

【数 3 0】

$$\beta = \sqrt{\frac{1}{8} \sum_{i=1}^n \frac{1}{m_i^2} \sum_{j=1}^{m_i} \frac{1}{(\kappa_{i,j})^2}}$$

10

であり、ここで、

【数 3 1】

$$\kappa_{i,j} = \min_{a \in A, j=1, \dots, m_i} (b_{ij}(a))$$

20

であり、ここで、 m_i は、タスク i に対するトレーニングセットサンプルの数である。

【0 0 6 7】

文脈付きバンディット問題の場合には、これらの目的関数は、 L_i を $L_i^{(cb)}$ によって置き換え、

【数 3 2】

$$\hat{L}_i$$

を

【数 3 3】

$$\hat{L}_i^{(cb)}$$

30

によって置き換えることによって適用されるものとしてよく、ここで、 Q_i 及び P_i は、行動 a_i ではなく、ポリシー π_i にわたる分布であり、

【数 3 4】

$$\kappa_i = \min_{a \in A, j=1, \dots, m} (b_{ij}(a|s_{ij})) / \pi_{max}$$

40

にし、ここで、 π_{max} は、任意のポリシー π 及び任意の行動状態に対する可能な最大確率質量である。

【0 0 6 8】

ハイパーパラメータ事後分布 Q の性能は、例えば、事前分布 $P \sim Q$ を備えた基礎学習アルゴリズム $Q(D_{n+1}; P)$ を使用する場合に、新たなタスク T_{n+1} に対して预期される報酬である周辺転移報酬 (marginal transfer reward) によって測定される。

【0 0 6 9】

多腕バンディット問題の場合には、周辺転移報酬は以下の通りである：

【数 3 5】

50

$$L(Q) = \mathbb{E}_{P \sim Q} \left[\mathbb{E}_{(T_{n+1}, D_{n+1})} [L_{n+1}(Q(D_{n+1}, P))] \right]$$

【0070】

文脈付きバンディット問題の場合には、周辺転移報酬は以下の通りである：

【数36】

$$L^{(cb)}(Q) = \mathbb{E}_{P \sim Q} \left[\mathbb{E}_{(T_{n+1}, D_{n+1})} [L_{n+1}^{(cb)}(Q(D_{n+1}, P))] \right]$$

10

上述の関数を用いて、周辺転移報酬を最大化する代わりに、この方法は、周辺転移報酬に対するPACベイズ下限を最大化する。限界の一部は、二値のカルバック・ライブラ(KL)発散の逆元を含む。

【0071】

これは、ハイパーパラメータ事後分布Qの決定が、ハイパーパラメータ事後分布Q及びハイパーパラメータ事前分布Pのカルバック・ライブラ発散DKL(Q || P)に関連して予期される報酬Lの決定及び近似を含むことを意味する。

【0072】

この例においては、ハイパーパラメータ事後分布Qは、n回の反復iにおいて決定される。

20

【0073】

これは、この方法が、n回の反復のためにステップ206に続くことを意味する。その後、任意選択的に、ステップ226が実行される。

【0074】

この例においては、ステップ226において、この方法は、ハイパーパラメータ事後分布Qからサンプリングされた事前分布に関連して、センサデータ、特にデジタル画像データ又はオーディオデータを処理することを含む。センサは、マイクロフォン、ビデオ、レーダ、LiDAR、超音波センサ、モーションセンサ、熱画像センサ又はソナーセンサであるものとしてよい。

30

【0075】

センサデータは、例えば、センサデータを分類するために、センサデータ内の対象物の存在を検出するために、又は、センサデータに対してセマンティックセグメンテーションを実行するために、処理される。センサデータ分類は、例えば、文脈付きバンディット問題としてフレーム化される。画像は、例えば、交通標識、路面、歩行者、車両の有無又は種類に関して分類されるものとしてよい。

【0076】

代替的又は付加的に、ステップ226は、物理的システムをコントロールするためのポリシーを学習すること及びハイパーパラメータ事後分布Qからサンプリングされた事前分布に関連して物理的システムをコントロールするためのコントロール信号を決定することを含み得る。

40

【0077】

例えば、ハイパーパラメータ事後分布Qから事前分布をサンプリングする場合に、次のタスクに対して予期される誤差が予め定められた値を上回らない確率が決定されるものとしてよい。

【0078】

代替的又は付加的に、ステップ226は、機械学習のロバスト性に関する指標を決定することを含み得る。この例においては、ロバスト性の指標は、ハイパーパラメータ事後分布Qからサンプリングされた事前分布に対する予期された誤差である。

【0079】

50

事前分布が、センサデータを処理するために使用されるものとしてよい。これは、例えば、予期された誤差が閾値よりも小さい場合であり、その他の場合には、別の事前分布が、センサデータを処理するために、又は、コントロール信号を決定するために、サンプリングされるものとしてよい。

【 0 0 8 0 】

代替的又は付加的に、ステップ 2 2 6 は、ハイパーパラメータ事後分布 Q からサンプリングされた事前分布に関連してセンサデータにおける異常を検出することを含み得る。この態様においては、異常検出問題は、多腕バンディット問題としてフレーム化される。

【 0 0 8 1 】

この例においては、その後、この方法が終了する。

10

【 0 0 8 2 】

多腕バンディット問題の機械学習のための方法においてコンピュータによって実行されるステップの例を以下に示す：

【 数 3 7 】

Algorithm 1 PAC-Bayesian Lifelong MABs

```

1: Input: Task distribution  $\mathcal{T}$ , base learning algorithm  $Q = Q(D, P)$ , hyperprior  $\mathcal{P}$ 
2: Initialise hyperposterior  $Q \leftarrow \mathcal{P}$ 
3: for  $i = 1$  to  $n$  do
4:   Sample task  $T_i \sim \mathcal{T}$ 
5:   Sample prior  $P \sim Q$ 
6:   Initialise task posterior  $Q \leftarrow P$ 
7:   Initialise task dataset  $D_i \leftarrow ()$ 
8:   for  $j = 1$  to  $m$  do
9:     Get behaviour policy  $b_{ij} \leftarrow B(Q)$ 
10:    Sample  $a_{ij} \sim b_{ij}(a)$  and  $r_{ij} \sim \rho_i(r|a_{ij})$ 
11:    Append  $(a_{ij}, r_{ij})$  to  $D_i$ 
12:    Update task posterior  $Q \leftarrow Q(D_i, P)$ 
13:   end for
14:    $Q \leftarrow \operatorname{argmax}_{Q'}(\text{lower bound}(\mathcal{L}(Q')))$ 
15: end for

```

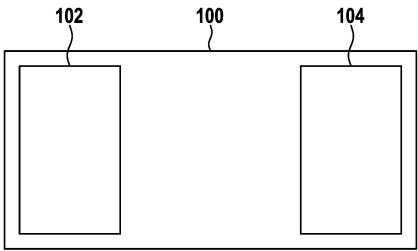
20

30

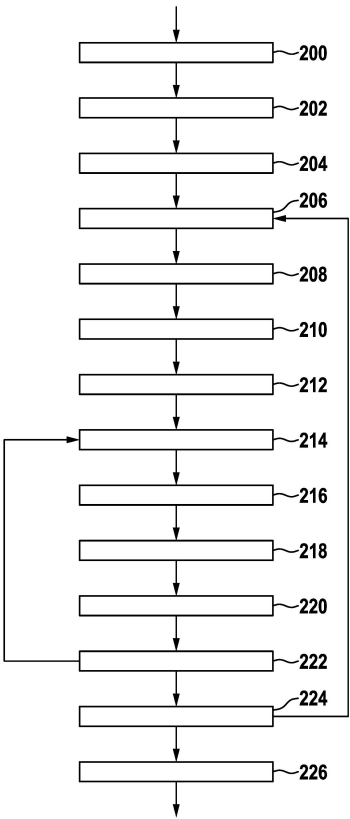
40

50

【図面】
【図 1】



【図 2】



10

20

30

40

50

フロントページの続き

弁理士 永島 秀郎
(74)代理人 100162880
弁理士 上島 類
(72)発明者 ハミッシュ フリン
ドイツ連邦共和国 シュトゥットガルト シュティッツェンブルクシュトラッセ 7
(72)発明者 ダーフィット レーブ
ドイツ連邦共和国 レニンゲン ネルケンシュトラッセ 80
(72)発明者 ヤン ペータース
ドイツ連邦共和国 ゼーハイム-ユーゲンハイム (オルツタイル マルヒェン) イム・シュタイ
ンヴィンガート 7
(72)発明者 メリフ カンデミル
デンマーク国 オーデンセ ビャークゴースレッケン 22