(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2019-82809
(P2019-82809A)
(43)公開日　令和1年5月30日(2019.5.30)

(51)Int.Cl.		F I			テーマコード（参考）
G08G 1/00	(2006.01)	G08G	1/00	C	5H181
G08G 1/09	(2006.01)	G08G	1/09	F	
G06N 20/00	(2019.01)	G06N	99/00	150	

審査請求 未請求　請求項の数 8　OL　（全 17 頁）

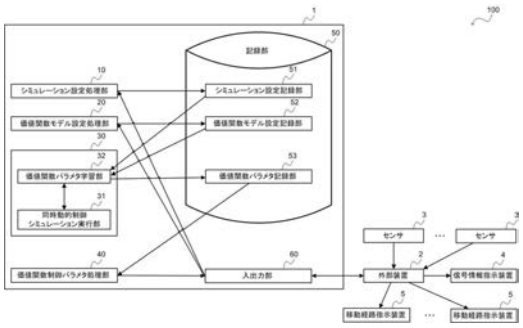
(21)出願番号　特願2017-209276 (P2017-209276)
(22)出願日　平成29年10月30日 (2017.10.30)

(71)出願人　000004226
日本電信電話株式会社
東京都千代田区大手町一丁目５番１号
(74)代理人　110001519
特許業務法人太陽国際特許事務所
(72)発明者　幸島　匡宏
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内
(72)発明者　堤田　恭太
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内
(72)発明者　松林　達史
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内

最終頁に続く

(54)【発明の名称】価値関数パラメタ学習装置、信号情報指示装置、移動経路指示装置、価値関数パラメタ学習方法、信号情報指示方法、移動経路指示方法、およびプログラム

(57)【要約】
【課題】移動体の数が増加しても、最適な交通状況を実現することができるようにする。
【解決手段】同時動的制御シミュレーション実行部３１が、交通状況を表す状態と、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動との組み合わせに対する価値関数を用いて、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動を行ったときの交通状況のシミュレーションを実行し、価値関数パラメタ学習部３２が、同時動的制御シミュレーション実行部３１によるシミュレーションの結果に基づいて、価値関数のパラメタを学習する。
【選択図】図１



【特許請求の範囲】
【請求項１】
交通状況を表す状態と、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動との組み合わせに対する価値関数を用いて、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動を行ったときの交通状況のシミュレーションを実行する同時動的制御シミュレーション実行部と、
前記同時動的制御シミュレーション実行部によるシミュレーションの結果に基づいて、前記価値関数のパラメタを学習する価値関数パラメタ学習部と、
を含む価値関数パラメタ学習装置。
【請求項２】
請求項１記載の価値関数パラメタ学習装置によって学習された前記価値関数のパラメタを用いて、入力された交通状況を表すセンサ情報に対応する前記状態について、前記行動を決定し、前記決定された行動に従って、各信号機に対して指示を行う信号情報指示装置。
【請求項３】
請求項１記載の価値関数パラメタ学習装置によって学習された前記価値関数のパラメタを用いて、入力された交通状況を表すセンサ情報に対応する前記状態について、前記行動を決定し、前記決定された行動に従って、各区間を通過する移動体に対して行う進むべき経路の指示を行う移動経路指示装置。
【請求項４】
同時動的制御シミュレーション実行部が、交通状況を表す状態と、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動との組み合わせに対する価値関数を用いて、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動を行ったときの交通状況のシミュレーションを実行するステップと、
価値関数パラメタ学習部が、前記同時動的制御シミュレーション実行部によるシミュレーションの結果に基づいて、前記価値関数のパラメタを学習するステップと、
を含む価値関数パラメタ学習方法。
【請求項５】
信号情報指示装置が、請求項４記載の価値関数パラメタ学習方法によって学習された前記価値関数のパラメタを用いて、入力された交通状況を表すセンサ情報に対応する前記状態について、前記行動を決定し、前記決定された行動に従って、各信号機に対して指示を行う信号情報指示方法。
【請求項６】
移動経路指示装置が、請求項４記載の価値関数パラメタ学習方法によって学習された前記価値関数のパラメタを用いて、入力された交通状況を表すセンサ情報に対応する前記状態について、前記行動を決定し、前記決定された行動に従って、各区間を通過する移動体に対して行う進むべき経路の指示を行う移動経路指示方法。
【請求項７】
コンピュータを、請求項１記載の価値関数パラメタ学習装置の各部として機能させるためのプログラム。
【請求項８】
コンピュータを、請求項２記載の信号情報指示装置として機能させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、価値関数パラメタ学習装置、信号情報指示装置、移動経路指示装置、価値関数パラメタ学習方法、信号情報指示方法、移動経路指示方法、およびプログラムに関し、

特に、移動体による交通を制御するための価値関数パラメタ学習装置、信号情報指示装置、移動経路指示装置、価値関数パラメタ学習方法、信号情報指示方法、移動経路指示方法、およびプログラムに関する。

【背景技術】

【０００２】

従来から、都市の交通渋滞や大規模イベントなどにおける人の混雑は社会的な課題になっている。交通渋滞は、渋滞中の車に乗車する人の時間を奪い、流通システムの遅れを生む原因にもなる。イベント会場における混雑もドミノ倒しなどの非劇的な雑踏事故を生む原因になりうる。

【０００３】

この点、強化学習（非特許文献１）によって信号機の制御を行うことで、車両の待ち時間を減少させる技術が存在する（非特許文献２）。

【先行技術文献】

【非特許文献】

【０００４】

【非特許文献１】Reinforcement learning: An introduction, Richard S Sutton and Andrew G. Barto, MIT press Cambridge, 1998.

【非特許文献２】Using a deep reinforcement learning agent for traffic signal control, Genders, Wade and Razavi, Saiedeh, arXiv preprint arXiv:1611.01142, 2016.

【発明の概要】

【発明が解決しようとする課題】

【０００５】

実際には目的地に到達するための移動経路は多数存在しているため、車両や人の移動経路についても加味して考える必要がある。

【０００６】

しかし、非特許文献２の技術では、移動経路について考慮されていないため、移動経路を加味した渋滞緩和を行うことができなかった。

【０００７】

また、移動経路について強化学習を行ったとしても、各車両毎に移動経路を決定するため、車両数が増大すると、アクションの数が指数的に増大してしまう。このような膨大なアクション数を持つ場合、探索空間が増大し、強化学習によって正しく各車両の移動経路を推定することが極めて困難となる、という問題があった。

【０００８】

本発明は、上記の点に鑑みてなされたものであり、移動体の数が増加しても、最適な交通状況を実現するための価値関数パラメタを学習することができる価値関数パラメタ学習装置、価値関数パラメタ学習方法、およびプログラムを提供することを目的とする。

【０００９】

また、本発明は、移動体の数が増加しても、最適な交通状況を実現することができる信号情報指示装置、移動経路指示装置、信号情報指示方法、移動経路指示方法、およびプログラムを提供することを目的とする。

【課題を解決するための手段】

【００１０】

本発明に係る価値関数パラメタ学習装置は、交通状況を表す状態と、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動との組み合わせに対する価値関数を用いて、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動を行ったときの交通状況のシミュレーションを実行する同時動的制御シミュレーション実行部と、前記同時動的制御シミュレーション実行部によるシミュレーションの結果に基づいて、前記価値関数のパラメタを学習する価値関数パラメタ学習部とを備えて構成される。

【００１１】

また、本発明に係る価値関数パラメタ学習方法は、同時動的制御シミュレーション実行部が、交通状況を表す状態と、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動との組み合わせに対する価値関数を用いて、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動を行ったときの交通状況のシミュレーションを実行するステップと、価値関数パラメタ学習部が、前記同時動的制御シミュレーション実行部によるシミュレーションの結果に基づいて、前記価値関数のパラメタを学習するステップとを含む。

【００１２】

本発明に係る価値関数パラメタ学習装置及び価値関数パラメタ学習方法によれば、同時動的制御シミュレーション実行部が、交通状況を表す状態と、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動との組み合わせに対する価値関数を用いて、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動を行ったときの交通状況のシミュレーションを実行する。

【００１３】

そして、価値関数パラメタ学習部が、同時動的制御シミュレーション実行部によるシミュレーションの結果に基づいて、価値関数のパラメタを学習する。

【００１４】

このように、交通状況を表す状態と、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動との組み合わせに対する価値関数を用いて、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動を行ったときの交通状況のシミュレーションを行い、シミュレーションの結果に基づいて、価値関数パラメタを学習することにより、移動体の数が増加しても、最適な交通状況を実現するための価値関数パラメタを学習することができる。

【００１５】

本発明に係る信号情報指示装置は、価値関数パラメタ学習装置によって学習された前記価値関数のパラメタを用いて、入力された交通状況を表すセンサ情報に対応する前記状態について、前記行動を決定し、前記決定された行動に従って、各信号機に対して指示を行うことができる。

【００１６】

また、本発明に係る信号情報指示方法は、信号情報指示装置が、上記価値関数パラメタ学習方法によって学習された前記価値関数のパラメタを用いて、入力された交通状況を表すセンサ情報に対応する前記状態について、前記行動を決定し、前記決定された行動に従って、各信号機に対して指示を行うことができる。

【００１７】

本発明に係る移動経路指示装置は、上記価値関数パラメタ学習装置によって学習された前記価値関数のパラメタを用いて、入力された交通状況を表すセンサ情報に対応する前記状態について、前記行動を決定し、前記決定された行動に従って、各区間を通過する移動体に対して行う進むべき経路の指示を行うことができる。

【００１８】

また、本発明に係る移動経路指示方法は、移動経路指示装置が、上記価値関数パラメタ学習方法によって学習された前記価値関数のパラメタを用いて、入力された交通状況を表すセンサ情報に対応する前記状態について、前記行動を決定し、前記決定された行動に従って、各区間を通過する移動体に対して行う進むべき経路の指示を行うことができる。

【００１９】

本発明に係るプログラムは、上記の価値関数パラメタ学習装置の各部として機能させるためのプログラムである。

【００２０】

また、本発明に係るプログラムは、上記の信号情報指示装置の各部として機能させるためのプログラムである。

【発明の効果】
【００２１】
本発明の価値関数パラメタ学習装置、価値関数パラメタ学習方法、およびプログラムによれば、移動体の数が増加しても、最適な交通状況を実現するための価値関数パラメタを学習することができる。
【００２２】
本発明の信号情報指示装置、移動経路指示装置、信号情報指示方法、移動経路指示方法、およびプログラムによれば、移動体の数が増加しても、最適な交通状況を実現することができる。
【図面の簡単な説明】
【００２３】
【図１】本発明の実施の形態に係る交通制御システムの構成を示す概略図である。
【図２】本発明の実施の形態に係る価値関数パラメタ学習処理ルーチンを示すフローチャートである。
【図３】本発明の実施の形態に係る制御処理ルーチンを示すフローチャートである。
【図４】本発明の実施の形態に係る提案エージェントの一例を表す図である。
【発明を実施するための形態】
【００２４】
以下、本発明の実施の形態について図面を用いて説明する。
【００２５】
＜本発明の実施の形態の原理＞
まず、本発明の実施形態の原理について説明する。
【００２６】
本実施形態は、都市の交通渋滞や大規模イベントなどにおける人の混雑を解消するための技術である。特に本発明の実施形態は人や車などの「移動体」の経路とそれら移動体に対して進行許可・停止などの指示を与える「信号機」を同時に最適化することで混雑を解消する技術である。本発明の実施形態の新規性は、移動体の経路と信号機の最適化を個別に行うのではなく、同時に行うことで全体最適化を行うことにある。
【００２７】
この全体最適のための本実施形態の鍵となるアイディアは、移動体の経路と信号機を制御する指示主体（これをエージェント（Ａｇｅｎｔ）と呼ぶ）の定義とこの指示主体の最適な制御策の推定方法にある。本実施形態に係るエージェント（以下、提案エージェント）の定義と学習方法によって、移動体の数や指示決定に用いる観測のセンサ数が膨大となる場合であっても、最適な制御策の推定が可能になる。これによって、人の混雑を解消することを実現する。
【００２８】
本実施形態は、このような車、人などの混雑を解消する混雑緩和を行うための技術である。特に本実施形態は人や車などの「移動体の経路」とそれら移動体に対して進行許可・停止などの指示を与える「信号機」を同時に最適化することで混雑を解消する技術である。
【００２９】
なお、ここで移動体と呼んでいるものは、人や動物、バイク、車、鉄道、ヘリコプター、飛行機など、一般に移動する、動くもの全てのものを指す。川や水路を流れる水や、ネットワークを流れるパケットも移動体である。
【００３０】
また、信号機と呼んでいるものは、いわゆる道路の信号機（交通信号機）だけでなく、警察官による手信号など、人、車などの上記移動体に対して進行許可・停止などを指示する機能をもつ全てのものを指す。
【００３１】
以下、これより簡便さのために「移動体の経路」とは車の移動経路、「信号機」とは道

路の交通信号機を意味するものとして交通制御の文脈で記述を進めるが、「移動体の経路」と「信号機」は上記のいずれのものであってもよい。

【００３２】

本実施形態では、道路に設置してあるセンサや車、信号から観測される値に基づいて、動的（適応的）に信号機と車両の移動経路を制御できる、という設定を考える。

【００３３】

ここで、制御とは、信号機の場合には信号を青にするもしくは赤にすること、経路の場合には車両が移動する経路を強制し、もしくは間接的に指示することを指す。なお、センサなどの観測値に基づいて信号を変えることのできる信号は感応式信号などと呼ばれ普及している。

【００３４】

上記の設定のもと、本実施形態は、任意の観測値を入力、とるべき制御策を出力とする関数（この関数のことを最適方策と呼ぶ）をシミュレーションまたは実環境を通して学習する。一旦最適方策が学習できれば、それに従って制御を実施することで渋滞が緩和できる。

【００３５】

このような最適方策を学習するアプローチは強化学習と呼ばれる。強化学習は、本発明のような信号機と車両の移動経路のように複数のものを制御する場合ではなく、信号機単独のものを制御する場合でも利用できるアプローチである。実際、強化学習によって信号機単独の制御を行う既存技術が存在する（非特許文献２）。

【００３６】

本発明の実施形態はその既存技術の大きな発展系の一つである。そこで、まずこの既存研究について紹介する。

【００３７】

＜＜強化学習＞＞

まずはじめに強化学習について簡単に説明する。強化学習はマルコフ決定過程（Ｍａｒｋｏｖ　Ｄｅｃｉｓｉｏｎ　Ｐｒｏｃｅｓｓ、ＭＤＰ）（非特許文献１）として定義された設定で最適方策を見つける手法である。ＭＤＰは、簡単にいえば行動主体（例えばロボット）と外界の相互作用を記述したものであり、ロボットがとりうる状態の集合

$$\mathcal{S} = \{s_1, s_2, \cdots, s_S\}$$

、ロボットがとりうる行動の集合

$$\mathcal{A} = \{a_1, a_2, \cdots, a_A\}$$

、ロボットがある状態である行動を取った際の状態の遷移の仕方を定める遷移関数

$$\mathcal{P} = \{p^a_{ss'}\}_{s,s',a}(\text{ただし}\sum_{s'} p^a_{ss'} = 1)$$

、ロボットがある状態でとった行動の良さに関する情報を与える報酬関数

$$\mathcal{R} = \{r_1, r_2, \cdots, r_S\}$$

、未来に受け取る報酬の考慮度合いをコントロールする割引率γ（ただし、０≦γ＜１）の５つの組

$$(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$$

で定義される。

【００３８】

このＭＤＰの設定のもと、ロボットには各状態でどの行動を実行するかの自由度が与えられる。このロボットが各状態ｓにいる時に実行する行動ａを決定する関数を方策と呼び、πと書く。ここで、

$$\pi(s) \in \mathcal{A}$$

であり、π（ｓ）で状態ｓにいるときに実行する行動を表す。

【００３９】

強化学習では複数存在する方策のうち、最も現在から将来にいたるまで得られる報酬の期待割引和を最大化する方策、最適方策を求める。最適方策を導く際に重要な役割を果たすのが価値関数Ｑ$^{\pi}$である。

【００４０】

【数１】

$$Q^{\pi}(s,a) = \lim_{T\to\infty} E^{\pi}\left[\sum_{k=0}^{T} \gamma^{k}\mathcal{R}(S_k)\middle| S_0 = s, A_0 = a\right] \qquad \cdots \quad (1)$$

【００４１】

ここで、Ｓ$_{k}$は、ある時刻ｋにおける状態であり、Ｓ$_{0}$は集合

$$\mathcal{S}$$

における最初の状態ｓを表す。

【００４２】

この関数は、状態ｓで行動ａを実行し、そのあとは方策πにしたがって無限に行動し続けた場合に得られる報酬の期待割引和を表している。方策πが最適方策であったとき、最適方策における価値関数Ｑ*（最適価値関数）は、

【数２】

$$Q^{*}(s,a) = \mathcal{R}(s) + \gamma \sum_{s'} p^{a}_{ss'} \max_{a'} Q^{*}(s',a') \qquad \cdots \quad (2)$$

を満たすことが知られ、この式のことをベルマン最適方程式と呼ぶ。

Ｑ学習に代表される強化学習の多くの手法は、上記の式の関係性を利用して、この最適価値関数をまずはじめに推定し、その結果を用いて、下記式（３）と設定することで最適方策π*を得ている。

【００４３】

【数３】

$$\pi^{*}(s) = \arg\max_{a} Q^{*}(s,a) \qquad \cdots \quad (3)$$

【００４４】

＜＜強化学習による信号制御＞＞

【００４５】
単独の信号制御を行う既存技術（非特許文献２）は前節の強化学習のアプローチにもとづき信号の制御策を発見している。信号制御の場合、ＭＤＰにおける行動は例えば「南北方向を青、東西方向を赤にする」、「南北方向を赤、東西方向を青にする」という信号の設定を切り替える操作に対応する。
【００４６】
同様に状態は、上記行動の決定の際に利用できる情報、例えば道路に設置してあるセンサや車両から送信される情報などに対応する。
【００４７】
報酬は、例えば「信号で停止している車両の台数の符号反転（－１をかける）」と設定しておけば、停止する車両の数が少なくなるような最適方策が発見されると期待できる。
【００４８】
前述した既存技術（非特許文献２）は、上記の設定に加え、Ｓｐａｃｅ　Ｉｎｖａｄｅｒなどゲーム分野で大きな成功を収めているＤｅｅｐ　Ｑ－Ｎｅｔｗｏｒｋ（ＤＱＮ）（参考文献１）のアプローチを採用している。
【００４９】
［参考文献１］
Human-level control through deep reinforcement learning, Mnih, Volodymyr and Kavukcuoglu, Koray and Silver, David and Rusu, Andrei A and Veness, Joel and Bellemare, Marc G and Graves, Alex and Riedmiller, Martin and Fidjeland, Andreas K and Ostrovski, Georg and others, Nature, 2015.
【００５０】
この研究では、最適価値関数をパラメタをもつニューラルネットワークで近似することを考える。
【００５１】
【数４】

$$Q^*(s,a) \approx Q(s,a;\theta) \qquad \cdots \quad (4)$$

【００５２】
このパラメタを、シミュレーションを通して学習することで最適価値関数と最適方策を得ている。このアプローチは、とりうる状態の数が非常に多い場合に有効な方法であることが知られている。信号制御の場合でも、センサ数やそこから送られてくる情報の種類が多い場合、一つ一つの状態が数１０～数１００次元のベクトルで表現される場合があり、このようなアプローチが採用されている。
【００５３】
このようなＤＱＮを用いた強化学習によって、既存技術（参考文献２）では最適方策によって、車両の待ち時間を減少させることができたと報告されている。
【００５４】
＜＜原理＞＞
本発明に係る実施形態は、上記既存技術を大幅に発展させ、信号機と車両の移動経路を同時に最適化することによってさらに車両の待ち時間を減少させる技術である。
【００５５】
信号機と車両の移動経路とを同時に最適化を行うために、本実施形態では、移動経路指示機、という仮想的な機械を導入する。
【００５６】
この移動経路指示機は、ある道路の一定区間を走行している車両に対して進むべき経路を指示する、というものである。
【００５７】
例えば、図４に示すように、その区間を通行した車両のその区間の通過後にとりうる経

路としてルート１またはルート２の２種類が存在するとき、この移動経路指示機は、その車両がルート１に進むか、ルート２に進むべきかを指示する。

【００５８】

なお、この指示には、車両が必ず従うとしても良いし、一定の確率に従って指示に従うか否かが決まるとしてもよい。

【００５９】

提案エージェントは、信号機と上記移動経路指示機の両方の行動を決定するエージェントである。図４の例では、提案エージェントの取りうる行動は全８種類であり、各状態においてどの行動を選択するかを定める方策を、強化学習を用いて学習する。

【００６０】

上記提案エージェントの行動を反映することのできる交通シミュレータなどを用意することで、提案エージェントの価値関数パラメタ推定には、任意の強化学習手法が適用でき、ＤＱＮ（参考文献１）以外の手法、例えばＤｏｕｂｌｅ　ＤＱＮ（参考文献２）などを用いてもよい。

【００６１】

［参考文献２］

Deep Reinforcement Learning with Double Q-Learning, Van Hasselt, Hado and Guez, Arthur and Silver, David, AAAI, 2016.

【００６２】

上記の提案エージェントの優れた点は、エージェントのとりうる行動数が車両の数に依存しない、という点にある。

【００６３】

車両毎に移動経路を決定する、というエージェントの定義では、アクションの数が、（信号機のアクション数）×（車両１のアクション数（ルート数））×（車両２のアクション数（ルート数））…と車両数の増大にともないアクションの数が指数的に増大していく。

【００６４】

すなわち、このような膨大なアクション数を持つエージェントを定義してしまうと、探索空間が増大し、強化学習によって正しく価値関数を推定することが極めて困難となる。

【００６５】

提案エージェントでは、このような困難さを回避することで、信号機と車両の移動経路を最適化する最適方策と価値関数を推定する。

【００６６】

＜本発明の実施の形態に係る交通制御システムの構成＞

図１を参照して、本発明の実施の形態に係る交通制御システムの構成について説明する。図１は、本発明の実施の形態に係る交通制御システムの構成を示すブロック図である。

【００６７】

図１に示すように、本実施形態に交通制御システム１００は、価値関数パラメタ学習装置１と、外部装置２と、センサ３と、信号情報指示装置４と、移動経路指示装置５とを備えて構成される。

【００６８】

価値関数パラメタ学習装置１は、ＣＰＵと、ＲＡＭと、後述する行動選択学習処理ルーチンを実行するためのプログラムを記憶したＲＯＭとを備えたコンピュータで構成され、機能的には次に示すように構成されている。

【００６９】

価値関数パラメタ学習装置１は、シミュレーション設定処理部１０と、価値関数モデル設定処理部２０と、価値関数パラメタ推定部３０と、価値関数制御パラメタ処理部４０と、記録部５０と、入出力部６０とを備えて構成される。

【００７０】

シミュレーション設定処理部１０は、入出力部６０から取得したシミュレーションを行

うために必要な情報を、シミュレーション設定記録部５１に格納する。
【００７１】
具体的には、シミュレーション設定処理部１０は、シミュレーション設定に関する情報、例えば、道路ネットワーク、信号位置、センサ位置等をシミュレーション設定記録部５１に格納する。
【００７２】
価値関数モデル設定処理部２０は、入出力部６０から取得した価値関数モデルの設定に関する情報を、価値関数モデル設定記録部５２に格納する。
【００７３】
具体的には、価値関数モデル設定処理部２０は、価値関数モデルの設定に関する情報、例えばニューラルネットワークの層数、中間素子数、活性化関数等を価値関数モデル設定記録部５２に格納する。
【００７４】
価値関数パラメタ推定部３０は、シミュレーション設定記録部５１に記録されているシミュレーション設定に関する情報、価値関数モデル設定記録部５２に記録されている価値関数モデルの設定に関する情報を入力とし、同時動的制御シミュレーション実行部３１の処理を繰り返し実行することで、価値関数パラメタを学習し、学習した価値関数パラメタを価値関数パラメタ記録部５３に格納する。
【００７５】
価値関数パラメタ推定部３０は、同時動的制御シミュレーション実行部３１と、価値関数パラメタ学習部３２とを備えて構成される。
【００７６】
同時動的制御シミュレーション実行部３１は、交通状況を表す状態と、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動との組み合わせに対する価値関数を用いて、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動を行ったときの交通状況のシミュレーションを実行する。
【００７７】
具体的には、同時動的制御シミュレーション実行部３１は、価値関数パラメタ学習部３２から、シミュレーション設定に関する情報と、価値関数モデルの設定に関する情報とを取得して、シミュレーションを実行する。
【００７８】
より具体的には、同時動的制御シミュレーション実行部３１は、シミュレーション設定に関する情報に基づいて、道路に設置してあるセンサから送信される情報を、交通状況を表す状態とし、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を、行動とし、信号で停止している車両の台数の符号反転（－１をかける）を、報酬として、価値関数を用いて、行動を決定し、決定した行動を行ったときの交通状況をシミュレーションする。
【００７９】
ここで計算される報酬は、状態ｓから求められる報酬関数

$$r_S$$

であるが、状態ｓにおける行動ａも加味した報酬関数

$$r_S(a)$$

としても良い。
【００８０】
そして、同時動的制御シミュレーション実行部３１は、シミュレーションの結果を価値

関数パラメタ学習部３２に渡す。
【００８１】
価値関数パラメタ学習部３２は、同時動的制御シミュレーション実行部３１によるシミュレーションの結果に基づいて、価値関数のパラメタを学習する。
【００８２】
具体的には、まず、価値関数パラメタ学習部３２は、シミュレーション設定記録部５１からシミュレーション設定に関する情報を取得し、価値関数モデル設定記録部５２から価値関数モデルの設定に関する情報を取得する。
【００８３】
次に、価値関数パラメタ学習部３２は、シミュレーション設定に関する情報と、価値関数モデルの設定に関する情報とを、同時動的制御シミュレーション実行部３１に渡し、予め定めた反復条件を満たすまで、同時動的制御シミュレーション実行部３１にシミュレーションを繰り返し実行させる。
【００８４】
ここで、反復条件は、所定回数を繰り返す、価値関数パラメタに変化が無くなった、価値関数パラメタの学習が収束した等、様々な条件を設定することができる。
【００８５】
そして、価値関数パラメタ学習部３２は、同時動的制御シミュレーション実行部３１から取得したシミュレーション結果に基づいて、価値関数パラメタを学習する。
【００８６】
より具体的には、同時動的制御シミュレーション実行部３１によるシミュレーション結果から得られる報酬と価値関数の値とに基づいて、最適方策を得ることができる価値関数パラメタ（例えば、式（４）におけるパラメタθ）を学習する。
【００８７】
その後、価値関数パラメタ学習部３２は、学習した価値関数パラメタを、価値関数パラメタ記録部５３に記録する。
【００８８】
価値関数制御パラメタ処理部４０は、価値関数パラメタ記録部５３に記録されている価値関数パラメタを、入出力部６０に渡す。
【００８９】
記録部５０は、シミュレーション設定記録部５１と、価値関数モデル設定記録部５２と、価値関数パラメタ記録部５３とを備えて構成される。
【００９０】
シミュレーション設定記録部５１は、シミュレーション設定処理部１０から取得したシミュレーションを行うために必要な情報を記録している。
【００９１】
また、シミュレーション設定記録部５１は、予め設定されたシミュレーションを行うために必要な情報を記録している。
【００９２】
価値関数モデル設定記録部５２は、価値関数モデル設定処理部２０から取得した価値関数モデルの設定に関する情報を記録している。
【００９３】
また、価値関数モデル設定記録部５２は、予め設定された価値関数モデルの設定に関する情報を記録している。
【００９４】
価値関数パラメタ記録部５３は、価値関数パラメタ学習部３２により学習された価値関数パラメタを記録している。
【００９５】
入出力部６０は、外部装置２から、シミュレーションを行うために必要な情報と、価値関数モデルの設定に関する情報とを受け付ける。

【００９６】
入出力部６０は、シミュレーションを行うために必要な情報が入力されると、シミュレーション設定処理部１０に、シミュレーションを行うために必要な情報を渡す。
【００９７】
入出力部６０は、価値関数モデルの設定に関する情報が入力されると、価値関数モデル設定処理部２０に、価値関数モデルの設定に関する情報を渡す。
【００９８】
また、入出力部６０は、価値関数制御パラメタ処理部４０から、価値関数パラメタを受け取ると、外部装置２へ出力する。
【００９９】
外部装置２は、シミュレーションを行うために必要な情報と、価値関数モデルの設定に関する情報とを設定する装置であり、予め設定されたシミュレーションを行うために必要な情報や予め設定された価値関数モデルの設定に関する情報に修正・変更がある場合に、修正・変更を受け付ける。
【０１００】
そして、修正・変更を受け付けると、修正・変更されたシミュレーションを行うために必要な情報および／または価値関数モデルの設定に関する情報を、入出力部６０に渡す。
【０１０１】
また、外部装置２は、入力された交通状況を表すセンサ情報と、入力された価値関数パラメタとを、信号情報指示装置４と、各移動経路指示装置５とにそれぞれ渡す。
【０１０２】
具体的には、まず、外部装置２は、入出力部６０から価値関数パラメタと、複数のセンサ３の各々から、当該センサ３によって計測された交通状況を表すセンサ情報とを取得する。ここで、センサ情報は、車両の速度、車両の台数、車両が通ったか否かなどのそのセンサによって得られる車両の情報である。
【０１０３】
次に、外部装置２は、信号情報指示装置４と、各移動経路指示装置５とに対して、取得した交通状況を表すセンサ情報と、価値関数パラメタとを渡す。
【０１０４】
センサ３は、道路に複数設置されているセンサであり、各設置地点における交通の状況を計測する。例えば、設置地点の画像や設置地点を通過した車両の速度、所定時間内の車両台数、車両が通過したこと等を計測する。
【０１０５】
信号情報指示装置４は、価値関数パラメタ学習装置１によって学習された価値関数のパラメタを用いて、入力された交通状況を表すセンサ情報に対応する状態について、最適方策となる行動を決定し、決定された行動に含まれる各信号機に対する指示に従って、各信号機に対して指示を行う。
【０１０６】
具体的には、まず、信号情報指示装置４は、外部装置２から、センサ情報と、価値関数パラメタとを取得し、取得したセンサ情報に対応する状態を求める。
【０１０７】
次に、信号情報指示装置４は、取得した価値関数パラメタを用いて、求めた交通状況の状態について、最適方策となる行動を決定する。
【０１０８】
そして、信号情報指示装置４は、決定された行動に含まれる各信号機に対する指示にしたがって、各信号機に対して、「赤にする」、「青にする」等の指示を行う。
【０１０９】
移動経路指示装置５は、価値関数パラメタ学習装置１によって学習された価値関数のパラメタを用いて、入力された交通状況を表すセンサ情報に対応する状態について、最適方策となる行動を決定し、決定された行動に含まれる自装置の区間を通過する移動体に対し

て行う進むべき経路の指示に従って、自装置の区間を通過する移動体に対して行う進むべき経路の指示を行う。

【０１１０】

具体的には、まず、移動経路指示装置５は、外部装置２から、センサ情報と、価値関数パラメタとを取得し、取得したセンサ情報に対応する状態を求める。

【０１１１】

次に、移動経路指示装置５は、取得した価値関数パラメタを用いて、求めた交通状況の状態について、最適方策となる行動を決定する。

【０１１２】

そして、移動経路指示装置５は、決定された行動に含まれる自装置の区間を通過する移動体に対して行う進むべき経路の指示にしたがって、自装置の区間を通過する車両に対して「ルート１に進む」、「ルート２に進む」等の指示を行う。

【０１１３】

＜本発明の実施の形態に係る価値関数パラメタ学習装置の作用＞

図２は、本発明の実施の形態に係る価値関数パラメタ学習処理ルーチンを示すフローチャートである。

【０１１４】

価値関数パラメタ学習装置１に価値関数パラメタ学習処理の実行命令がなされると、価値関数パラメタ推定部３０において、図２に示す価値関数パラメタ学習処理ルーチンが実行される。

【０１１５】

まず、ステップＳ１００において、価値関数パラメタ学習部３２は、シミュレーション設定記録部５１からシミュレーション設定に関する情報を取得する。このシミュレーション設定に関する情報は、予め設定されたものでもよいし、シミュレーション設定処理部１０により格納されたものでもよい。

【０１１６】

ステップＳ１１０において、価値関数パラメタ学習部３２は、価値関数モデル設定記録部５２から価値関数モデルの設定に関する情報を取得する。この価値関数モデルの設定に関する情報は、予め設定されたものでもよいし、価値関数モデル設定処理部２０により格納されたものでもよい。

【０１１７】

ステップＳ１２０において、価値関数パラメタ学習部３２は、価値関数パラメタを初期化する。

【０１１８】

ステップＳ１３０において、同時動的制御シミュレーション実行部３１は、交通状況を表す状態と、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動との組み合わせに対する価値関数を用いて、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動を行ったときの交通状況のシミュレーションを実行する。

【０１１９】

ステップＳ１４０において、価値関数パラメタ学習部３２は、同時動的制御シミュレーション実行部３１によるシミュレーションの結果に基づいて、価値関数のパラメタを学習する。

【０１２０】

ステップＳ１５０において、価値関数パラメタ学習部３２は、予め定めた反復条件を満たすか否かを判定する。

【０１２１】

予め定めた反復条件を満たしていない場合（ステップＳ１５０のＮＯ）、ステップＳ１３０～Ｓ１４０の処理を繰り返す。

【０１２２】

予め定めた反復条件を満たしている場合（ステップＳ１５０のＹＥＳ）、ステップＳ１６０において、入出力部６０は、ステップＳ１４０により学習された価値関数パラメタを、外部装置２へ出力する。

【０１２３】

＜本発明の実施の形態に係る信号情報指示装置４及び移動経路指示装置５の作用＞

図３は、本発明の実施の形態に係る制御処理ルーチンを示すフローチャートである。

【０１２４】

外部装置２から価値関数パラメタが入力されると、信号情報指示装置４において、図３に示す制御処理ルーチンが実行される。

【０１２５】

まず、ステップＳ２００において、信号情報指示装置４は、外部装置２から入力された、価値関数パラメタ学習装置１によって学習された価値関数パラメタを取得する。

【０１２６】

ステップＳ２１０において、信号情報指示装置４は、外部装置２から、各センサ３のセンサ情報を取得し、取得したセンサ情報に対応する状態を求める。

【０１２７】

ステップＳ２２０において、信号情報指示装置４は、ステップＳ２００で取得した価値関数のパラメタを用いて、ステップＳ２１０により求められた交通状況を表すセンサ情報に対応する状態について、最適方策となる行動を決定する。

【０１２８】

ステップＳ２３０において、信号情報指示装置４は、ステップＳ２２０により決定した行動に含まれる各信号機に対する指示に従って、各信号機に対して指示を行う。

【０１２９】

また、各移動経路指示装置５においても、上記図３に示す制御処理ルーチンと同様の処理ルーチンを実行し、決定した行動に含まれる、自装置の区間を通過する移動体に対して行う進むべき経路の指示を含む行動に従って、自装置の区間を通過する移動体に対して行う進むべき経路の指示を行う。

【０１３０】

以上説明したように、本実施形態に係る価値関数パラメタ学習装置によれば、交通状況を表す状態と、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動との組み合わせに対する価値関数を用いて、各区間を通過する移動体に対して行う進むべき経路の指示、及び各信号機に対する指示を含む行動を行ったときの交通状況のシミュレーションを行い、シミュレーションの結果に基づいて、価値関数パラメタを学習するため、移動体の数が増加しても、最適な交通状況を実現するための価値関数パラメタを学習することができる。

【０１３１】

また、本実施形態に係る外部装置によれば、価値関数パラメタ学習装置によって学習された価値関数のパラメタを用いて、入力された交通状況を表すセンサ情報に対応する状態について、行動を決定し、決定された行動に従って、各信号機に対する指示、及び各区間を通過する移動体に対して行う進むべき経路の指示を行うため、移動体の数が増加しても、最適な交通状況を実現することができる。

【０１３２】

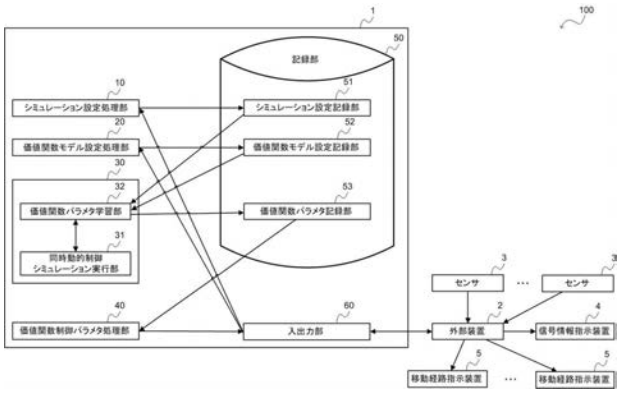
なお、本発明は、上述した実施の形態に限定されるものではなく、この発明の要旨を逸脱しない範囲内で様々な変形や応用が可能である。

【０１３３】

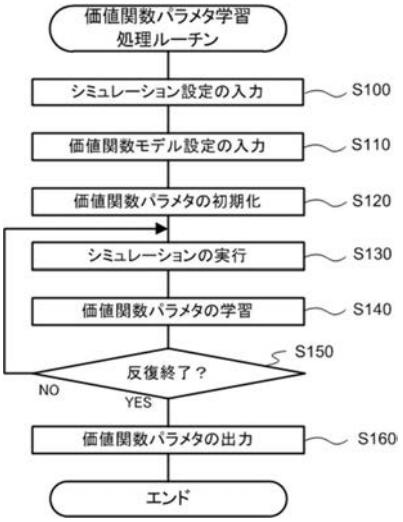
本実施形態では、信号情報指示装置４と、各移動経路指示装置５とが、各センサ３から取得したセンサ情報に基づいて状態を求め、行動を決定する例を説明したが、外部装置２が、各センサ３からセンサ情報を取得し、それぞれセンサ情報に対応する状態を求めて、行動を決定する構成としても良い。この場合、外部装置２が、信号情報指示装置４と、各移動経路指示装置５に対して、各信号機に対する指示、及び経路の指示を出す。

【０１３４】
また、本願明細書中において、プログラムが予めインストールされている実施形態として説明したが、当該プログラムを、コンピュータ読み取り可能な記録媒体に格納し、またはネットワークを介して提供することも可能である。
【符号の説明】
【０１３５】
１　　価値関数パラメタ学習装置
２　　外部装置
３　　センサ
４　　信号情報指示装置
５　　移動経路指示装置
１０　シミュレーション設定処理部
２０　価値関数モデル設定処理部
３０　価値関数パラメタ推定部
３１　同時動的制御シミュレーション実行部
３２　価値関数パラメタ学習部
４０　価値関数制御パラメタ処理部
５０　記録部
５１　シミュレーション設定記録部
５２　価値関数モデル設定記録部
５３　価値関数パラメタ記録部
６０　入出力部
１００ 交通制御システム

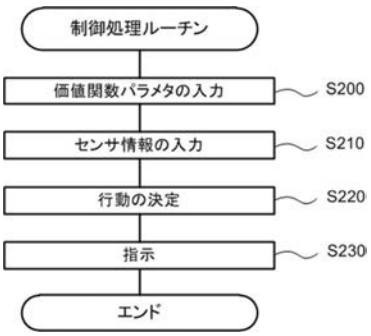
【図１】



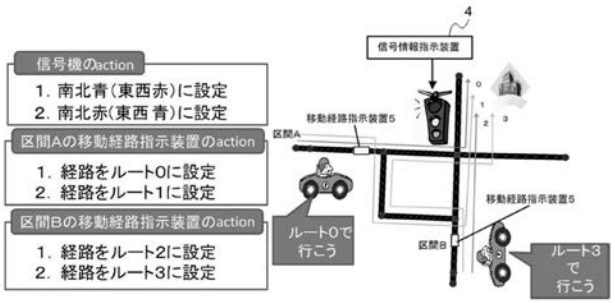
【図２】



【図３】



【図４】


信号機のaction
1. 南北青(東西赤)に設定
2. 南北赤(東西青)に設定
区間Aの移動経路指示装置のaction
1. 経路をルート0に設定
2. 経路をルート1に設定
区間Bの移動経路指示装置のaction
1. 経路をルート2に設定
2. 経路をルート3に設定
4
信号情報指示装置
移動経路指示装置5
区間A
区間B
移動経路指示装置5
ルート0で行こう
ルート3で行こう

フロントページの続き

(72)発明者　戸田　浩之
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内

Ｆターム(参考) 5H181 AA05　AA21　AA26　BB04　DD02　DD03　EE03　FF12　JJ02　JJ06