

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第3区分

【発行日】令和4年8月23日(2022.8.23)

【国際公開番号】WO2021/144963

【出願番号】特願2021-570601(P2021-570601)

【国際特許分類】

G 0 6 N 2 0 / 0 0 (2 0 1 9 . 0 1)

【 F I 】

G 0 6 N 2 0 / 0 0

10

【手続補正書】

【提出日】令和4年6月23日(2022.6.23)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

20

所定の環境における所定の状態が他の状態に遷移する際に選択される行動要素として、当該行動要素の選択肢の数が前記状態に依存しない第一の行動要素と、当該行動要素の選択肢の数が前記状態に依存する第二の行動要素と、がある場合に、

学習中のモデルを用いて前記状態に対する前記第一の行動要素の選択肢毎の選択率を算出し、当該選択率に基づいて前記第一の行動要素を選択し、

選択した前記第一の行動要素を適用し、前記第二の行動要素をさらに選択肢として適用した選択肢毎の前記他の状態について、当該他の状態に遷移させることによる報酬と当該他の状態の価値とを算出して、前記報酬と前記価値とに基づいて前記他の状態を決定し、

前記他の状態を決定する際に用いた情報に基づいて学習データを生成し、当該学習データを用いて前記モデルをさらに学習する、

30

方策学習方法。

【請求項2】

請求項1に記載の方策学習方法であって、

学習中の第二のモデルを用いて前記他の状態の前記価値を算出し、

前記学習データを用いて前記第二のモデルをさらに学習する、

方策学習方法。

【請求項3】

請求項1又は2に記載の方策学習方法であって、

前記報酬と前記価値との和が最大となる前記他の状態を決定する、

方策学習方法。

40

【請求項4】

請求項1乃至3のいずれかに記載の方策学習方法であって、

前記状態と、選択された前記第一の行動要素と、前記他の状態を決定する際に算出した前記報酬と前記価値との和の最大値と、を少なくとも関連付けた前記学習データを生成する、

方策学習方法。

【請求項5】

請求項1乃至4のいずれかに記載の方策学習方法であって、

前記環境が、前記状態であるグラフを書き換えることで前記他の状態である他のグラフに遷移するグラフ書き換え系である場合に、

50

前記第一の行動要素が前記グラフを書き換える規則を表すグラフ書き換え規則であり、前記第二の行動要素が前記グラフ内における前記グラフ書き換え規則を適用する場所を表す規則適用場所である、
方策学習方法。

【請求項 6】

請求項 5 に記載の方策学習方法であって、

前記モデルを用いて前記グラフに対する前記グラフ書き換え規則の選択肢毎の選択率を算出し、当該選択率に基づいて前記グラフ書き換え規則を選択し、

選択した前記グラフ書き換え規則を前記グラフ内における前記規則適用場所毎に適用した前記他の状態について、前記報酬と前記価値とを算出して、前記報酬と前記価値とに基づいて前記他の状態を決定する、
方策学習方法。

10

【請求項 7】

所定の環境における所定の状態が他の状態に遷移する際に選択される行動要素として、当該行動要素の選択肢の数が前記状態に依存しない第一の行動要素と、当該行動要素の選択肢の数が前記状態に依存する第二の行動要素と、がある場合に、

学習中のモデルを用いて前記状態に対する前記第一の行動要素の選択肢毎の選択率を算出し、当該選択率に基づいて前記第一の行動要素を選択する第一手段と、

選択した前記第一の行動要素を適用し、前記第二の行動要素をさらに選択肢として適用した選択肢毎の前記他の状態について、当該他の状態に遷移させることによる報酬と当該他の状態の価値とを算出して、前記報酬と前記価値とに基づいて前記他の状態を決定する第二手段と、

20

前記他の状態を決定する際に用いた情報に基づいて学習データを生成し、当該学習データを用いて前記モデルをさらに学習する第三手段と、

を備えた方策学習装置。

【請求項 8】

請求項 7 に記載の方策学習装置であって、

前記第二手段は、学習中の第二のモデルを用いて前記他の状態の前記価値を算出し、

前記第三手段は、前記学習データを用いて前記第二のモデルをさらに学習する、

方策学習装置。

30

【請求項 9】

請求項 7 又は 8 に記載の方策学習装置であって、

前記第二手段は、前記報酬と前記価値との和が最大となる前記他の状態を決定する、

方策学習装置。

【請求項 10】

所定の環境における所定の状態が他の状態に遷移する際に選択される行動要素として、当該行動要素の選択肢の数が前記状態に依存しない第一の行動要素と、当該行動要素の選択肢の数が前記状態に依存する第二の行動要素と、がある場合に、

情報処理装置に、

学習中のモデルを用いて前記状態に対する前記第一の行動要素の選択肢毎の選択率を算出し、当該選択率に基づいて前記第一の行動要素を選択する第一手段と、

40

選択した前記第一の行動要素を適用し、前記第二の行動要素をさらに選択肢として適用した選択肢毎の前記他の状態について、当該他の状態に遷移させることによる報酬と当該他の状態の価値とを算出して、前記報酬と前記価値とに基づいて前記他の状態を決定する第二手段と、

前記他の状態を決定する際に用いた情報に基づいて学習データを生成し、当該学習データを用いて前記モデルをさらに学習する第三手段と、

を実現させるためのプログラム。

50