

【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 6 年 1 月 4 日 (2024.1.4)

【国際公開番号】 WO2022/230019
【出願番号】 特願 2023-516874 (P2023-516874)
【国際特許分類】
G 0 6 N 2 0 / 0 0 (2019.01)
【 F I 】
G 0 6 N 2 0 / 0 0

10

【手続補正書】
【提出日】 令和 5 年 9 月 28 日 (2023.9.28)
【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

リブシツ連続条件を満たすように特徴量が設定された報酬関数の入力を受け付ける関数入力手段と、

熟練者の軌跡の確率分布と、前記報酬関数のパラメータに基づいて決定される軌跡の確率分布との距離を表わすワッサースタイン距離を最小にする軌跡を推定する推定手段と、
推定された軌跡に基づいて、最大エントロピーの原理で導かれるボルツマン分布の対数尤度を最大にするように前記報酬関数のパラメータを更新する更新手段とを備え、

前記更新手段は、前記対数尤度の下限として、前記ワッサースタイン距離から、前記パラメータに対する最大の報酬値から当該パラメータに対する報酬の平均値を減じた式で定義されるエントロピー正則化項を減じる式を導出し、導出された前記対数尤度の下限を最大にするように前記報酬関数のパラメータを更新する

30

ことを特徴とする学習装置。

【請求項 2】

更新手段は、エントロピー正則化項に、パラメータを更新する処理を繰り返すほど、前記エントロピー正則化項に対応する部分に対数尤度の下限の最大化に寄与する度合いを減衰させる減衰係数を設定し、設定された対数尤度の下限を最大にするように報酬関数のパラメータを更新する

請求項 1 記載の学習装置。

【請求項 3】

更新手段は、エントロピー正則化項が対数尤度の下限の最大化に寄与する度合いを減衰させる減衰係数を当該エントロピー正則化項に対応する部分に設定し、パラメータを更新する処理を繰り返す途中で、前記エントロピー正則化項に対応する部分に対数尤度の下限の最大化に寄与する度合いを減衰させるように前記減衰係数を変更する

40

請求項 1 記載の学習装置。

【請求項 4】

更新手段は、対数尤度の移動平均が一定になったと判断したときに、減衰係数を変更する

請求項 3 記載の学習装置。

【請求項 5】

更新手段は、 $\log \sum \exp$ 的上限值に基づいて、対数尤度
の下限を導出する

50

請求項 1 から請求項 4 のうちのいずれか 1 項に記載の学習装置。

【請求項 6】

関数入力手段は、線形関数になるように特徴量が設定された報酬関数の入力を受け付ける

請求項 1 から請求項 5 のうちのいずれか 1 項に記載の学習装置。

【請求項 7】

コンピュータが、リブシツ連続条件を満たすように特徴量が設定された報酬関数の入力を受け付け、

前記コンピュータが、熟練者の軌跡の確率分布と、前記報酬関数のパラメータに基づいて決定される軌跡の確率分布との距離を表わすワッサースタイン距離を最小にする軌跡を推定し、

前記コンピュータが、推定された軌跡に基づいて、最大エントロピーの原理で導かれるボルツマン分布の対数尤度を最大にするように前記報酬関数のパラメータを更新し、

前記パラメータの更新の際、前記コンピュータが、前記対数尤度の下限として、前記ワッサースタイン距離から、前記パラメータに対する最大の報酬値から当該パラメータに対する報酬の平均値を減じた式で定義されるエントロピー正則化項を減じる式を導出し、導出された前記対数尤度の下限を最大にするように前記報酬関数のパラメータを更新することを特徴とする学習方法。

【請求項 8】

コンピュータが、エントロピー正則化項に、パラメータを更新する処理を繰り返すほど、前記エントロピー正則化項に対応する部分が対数尤度の下限の最大化に寄与する度合いを減衰させる減衰係数を設定し、設定された対数尤度の下限を最大にするように報酬関数のパラメータを更新する

請求項 7 記載の学習方法。

【請求項 9】

コンピュータに、

リブシツ連続条件を満たすように特徴量が設定された報酬関数の入力を受け付ける関数入力処理、

熟練者の軌跡の確率分布と、前記報酬関数のパラメータに基づいて決定される軌跡の確率分布との距離を表わすワッサースタイン距離を最小にする軌跡を推定する推定処理、および、

推定された軌跡に基づいて、最大エントロピーの原理で導かれるボルツマン分布の対数尤度を最大にするように前記報酬関数のパラメータを更新する更新処理を実行させ、

前記更新処理で、前記対数尤度の下限として、前記ワッサースタイン距離から、前記パラメータに対する最大の報酬値から当該パラメータに対する報酬の平均値を減じた式で定義されるエントロピー正則化項を減じる式を導出させ、導出された前記対数尤度の下限を最大にするように前記報酬関数のパラメータを更新させる

ための学習プログラム。

【請求項 10】

コンピュータに、

更新処理で、エントロピー正則化項に、パラメータを更新する処理を繰り返すほど、前記エントロピー正則化項に対応する部分が対数尤度の下限の最大化に寄与する度合いを減衰させる減衰係数を設定させ、設定された対数尤度の下限を最大にするように報酬関数のパラメータを更新させる

請求項 9 記載の学習プログラム。

10

20

30

40

50