AI 特許紹介(5)

～Prioritized リプレイ特許～

2019 年 8 月 9 日

河野特許事務所

所長　弁理士　河野英仁

「AI 特許紹介」シリーズは、注目すべき AI 特許のポイントを紹介します。熾烈な競争となっている第 4 次産業革命下では AI 技術がキーとなり、この AI 技術・ソリューションを特許として適切に権利化しておくことが重要であることは言うまでもありません。

AI 技術は Google, Microsoft, Amazon を始めとした IT プラットフォーマ、研究機関及び大学から毎週のように新たな手法が提案されており、また AI 技術を活用した新たなソリューションも次々とリリースされています。

本稿では米国先進 IT 企業を中心に、これらの企業から出願された AI 特許に記載された AI テクノロジー・ソリューションのポイントをわかりやすく解説致します。

1.概要

特許権者　Deep Mind Technologies

出願日　2016 年 11 月 12 日

登録日　2019 年 5 月 7 日

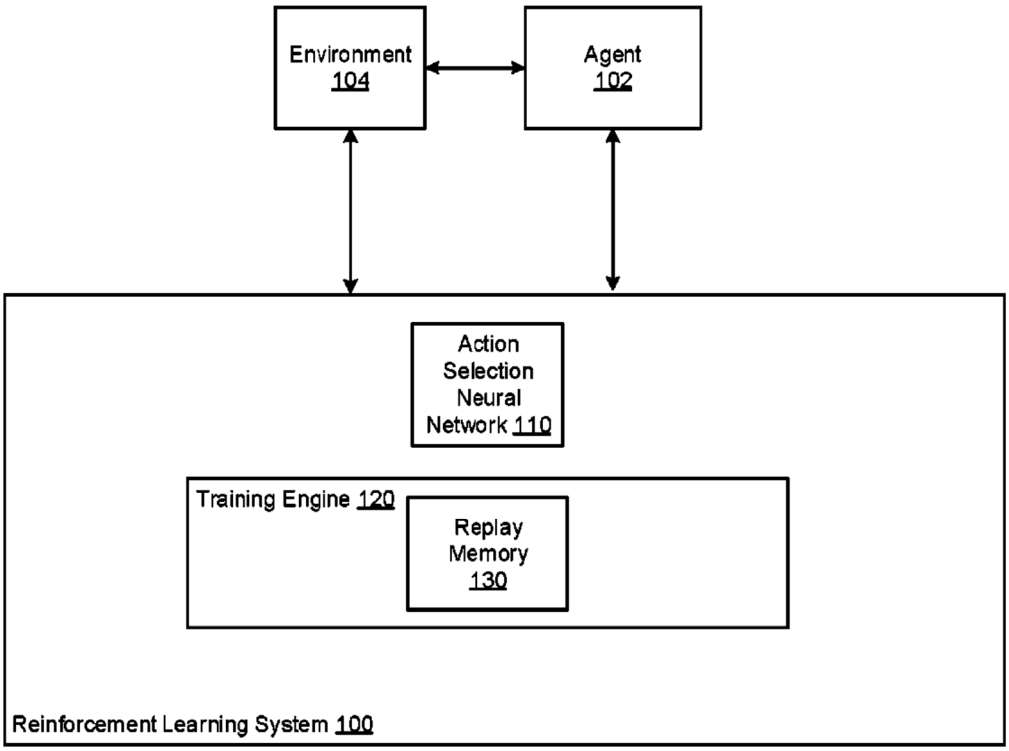
登録番号 US10,282,662

発明の名称　優先経験メモリを用いたニューラルネットワーク

662 特許は、強化学習におけるニューラルネットワークの学習の際に、期待学習進展度の高い経験データを優先的に用いるアイデアである。

2.特許内容の説明

強化学習システム 100 は、環境 104 とのインタラクションを行う強化学習エージェント 102 によって行われる行動を選択する。具体的には、強化学習システム 100 は、環境 104 の状態を特徴付ける各観測を用い、各観測に応じて、強化学習エージェント 102 によって行われる行動のセットを選択する。強化学習システム 100 は、エージェント 102 によって行われる行動に応じて報酬を受ける。



強化学習システム 100 は、行動選択ニューラルネットワーク 110 を訓練する訓練エンジン 120 を含み、行動選択ニューラルネットワーク 110 のパラメータの訓練済みの値を決定する。行動選択ニューラルネットワーク 110 の訓練支援のために、訓練エンジン 120 は、リプレイメモリ 130 を保持する。リプレイメモリ 130 は、行動選択ネットワーク 110 を訓練するために、環境 104 とのインタラクションの結果として生成された経験データを記憶する。

訓練エンジン 120 は、経験データを選択する際、学習効率を高めるために期待学習進展度の高い経験データを優先的に選択する。具体的には期待学習進展度は、TD(Temporal Difference)誤差の絶対値であり、絶対値が大きい場合学習が進んでいないため優先的に用いる。一方絶対値が小さい場合、学習が進んでいるため優先度を低くする。

その他、TD 誤差がしきい値を下回った場合、リプレイメモリから当該 TD 誤差に対応する経験データを消去する。また、新たに獲得した経験データについては他の経験データよりも優先的に選択されるようにする。

3.クレーム

662 特許のシステムクレーム 1 は以下のとおりである。

1. 環境に状態を遷移させる動作を実行することによって環境と相互作用する強化学習エージェントによって実行される動作を選択するために使用されるニューラルネットワークを訓練する方法であって、

ニューラルネットワークをトレーニングする際に使用するための経験データを格納するリプレイメモリを維持し、

強化学習エージェントが環境と相互作用した結果として、それぞれの経験データが生成されており、

各経験データは、環境のそれぞれの現在の状態を特徴付けるそれぞれの現在の観察、現在の観察に応答してエージェントによって実行されるそれぞれの現在の行動、環境のそれぞれの次の状態を特徴付けるそれぞれの次の観察、および、現在のアクションを実行しているエージェントに応答して受信される報酬を含み、

複数の経験データはそれぞれ、各期待学習進展度に関連付けられ、該期待学習進展度は、（ｉ）ニューラルネットワークが１つの経験データにより訓練された場合に、ニューラルネットワークの訓練において行われるであろう進展の期待量の尺度であり、（ｉｉ）１つの経験データがニューラルネットワークを訓練するのに使用された前回の結果から導出され、

比較的高い期待学習進展度を有する経験データの部分を優先的に選択することによって、リプレイメモリから経験データを選択し、

複数の経験データについてのそれぞれの予想される期待学習進展度に基づいて、リプレイメモリ内のそれぞれの経験データについてのそれぞれの確率を決定し、

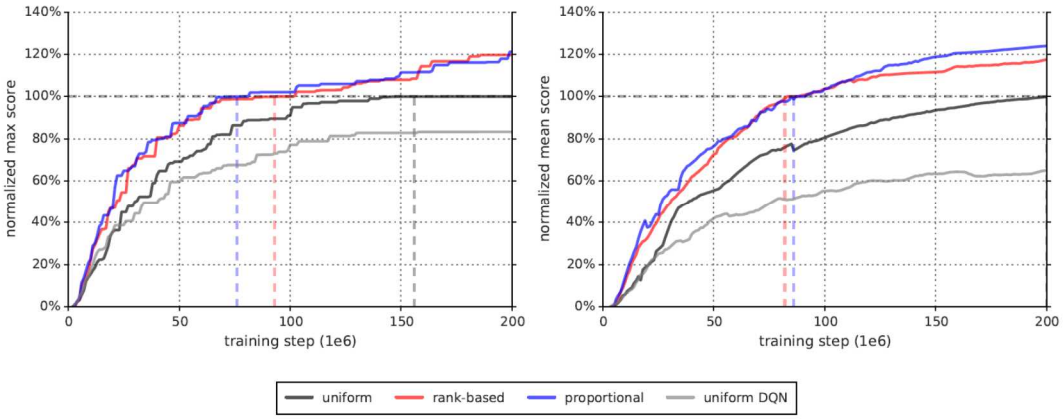
決定された確率に従ってリプレイメモリから１つの経験データをサンプリングし、

強化学習技術を使用して、選択された経験データについてニューラルネットワークを訓練し、

リプレイメモリ内で、選択された経験データを、選択された経験データについてニューラルネットワークをトレーニングした結果から導き出された新しい期待学習進展度と関連付ける。

4. Prioritized リプレイ

本特許で紹介した Prioritized リプレイは Double-DQN と共に深層強化学習の高速化に必要とされている技術である。



Prioritized リプレイに関する論文[1]には実験結果が示されている。実験では通常のDQN（uniform, uniform DQN）と2つのPrioritized リプレイ（rank-based, proportional）との比較が行われた。上のグラフに示すように青色及びオレンジ色のPrioritized リプレイによる方式が早期に学習が進んでいることが理解できる。

著者紹介

河野英仁

河野特許事務所、所長弁理士。立命館大学情報システム学博士前期課程修了、米国フランクリンピアースローセンター知的財産権法修士修了、中国清華大学法学院知的財産夏季セミナー修了、MIT(マサチューセッツ工科大学)コンピュータ科学・AI研究所 AIコース修了。

AI 特許コンサルティングの他、米国・中国特許の権利化・侵害訴訟を専門としている。著書に「世界のソフトウエア特許(共著)」、「FinTech 特許入門」、「AI/IoT 特許入門2.0」がある。

以上

[1] Tom Schaul “PRIORITIZED EXPERIENCE REPLAY” ICLR 2016