US011627165B2

(12) **United States Patent**
Silver et al.

(10) **Patent No.: US 11,627,165 B2**
(45) **Date of Patent: Apr. 11, 2023**

(54) **MULTI-AGENT REINFORCEMENT LEARNING WITH MATCHMAKING POLICIES**

(71) Applicant: **DeepMind Technologies Limited**, London (GB)

(72) Inventors: **David Silver**, Hitchin (GB); **Oriol Vinyals**, London (GB); **Maxwell Elliot Jaderberg**, London (GB)

(73) Assignee: **DeepMind Technologies Limited**, London (GB)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 589 days.

(21) Appl. No.: **16/752,496**

(22) Filed: **Jan. 24, 2020**

(65) **Prior Publication Data**

US 2020/0244707 A1 Jul. 30, 2020

Related U.S. Application Data

(60) Provisional application No. 62/894,633, filed on Aug. 30, 2019, provisional application No. 62/796,567, filed on Jan. 24, 2019.

(51) **Int. Cl.**

G06N 3/08	(2006.01)
H04L 9/40	(2022.01)
G06K 9/62	(2022.01)

(52) **U.S. Cl.**
CPC ***H04L 63/205*** (2013.01); ***G06K 9/6256*** (2013.01); ***G06N 3/08*** (2013.01)

(58) **Field of Classification Search**
CPC G06N 3/0454; G06N 3/08; G06N 20/00; G06N 3/088; G06N 3/006; G06N 3/0445; G06N 3/0472; G06N 7/005; G06N 3/082; G06N 3/126; G06N 3/04; G06N 5/043; G06N 3/02; G06N 3/0481; G06N 5/00; G06N 5/022; G06N 5/025;
(Continued)

(56) **References Cited**

U.S. PATENT DOCUMENTS

2019/0102676 A1 * 4/2019 Nazari G06N 3/0472
2020/0033868 A1 * 1/2020 Palanisamy G05D 1/0221

FOREIGN PATENT DOCUMENTS

WO WO-2018224695 A1 * 12/2018 .... G06N 3/006

OTHER PUBLICATIONS

Human-level performance in first-person multiplayer games with population-based deep reinforcement earning—Max Jaderberg et al. (Year: 2018)—IDS entry.*

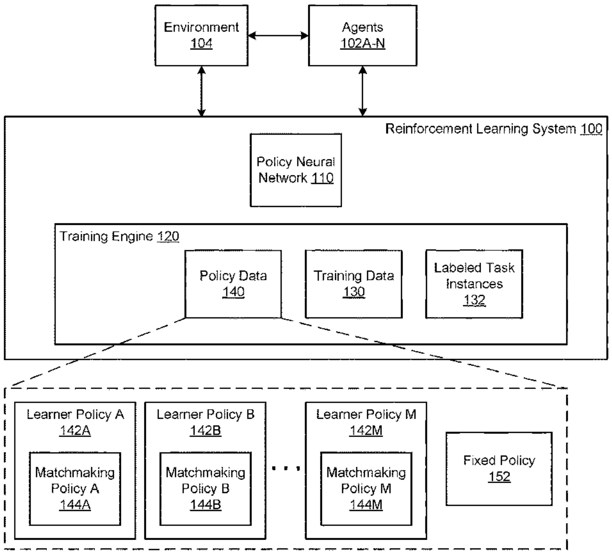
(Continued)

Primary Examiner — Quan M Hua
(74) *Attorney, Agent, or Firm* — Fish & Richardson P.C.

(57) **ABSTRACT**

Methods, systems, and apparatus, including computer programs encoded on a computer storage medium, for training a policy neural network having a plurality of policy parameters and used to select actions to be performed by an agent to control the agent to perform a particular task while interacting with one or more other agents in an environment. In one aspect, the method includes: maintaining data specifying a pool of candidate action selection policies; maintaining data specifying respective matchmaking policy; and training the policy neural network using a reinforcement learning technique to update the policy parameters. The policy parameters define policies to be used in controlling the agent to perform the particular task.

30 Claims, 3 Drawing Sheets


Environment 104
Agents 102A-N
Reinforcement Learning System 100
Policy Neural Network 110
Training Engine 120
Policy Data 140
Training Data 130
Labeled Task Instances 132
Learner Policy A 142A
Matchmaking Policy A 144A
Learner Policy B 142B
Matchmaking Policy B 144B
. . .
Learner Policy M 142M
Matchmaking Policy M 144M
Fixed Policy 152