

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
21 April 2022 (21.04.2022)



(10) International Publication Number
WO 2022/079517 A1

(51) International Patent Classification:
G06N 20/00 (2019.01)

(21) International Application Number:
PCT/IB2021/058455

(22) International Filing Date:
16 September 2021 (16.09.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
17/068,890 13 October 2020 (13.10.2020) US

(71) Applicant: **INTERNATIONAL BUSINESS
MACHINES CORPORATION** [US/US]; New Orchard
Road, Armonk, New York, 10504 (US).

(71) Applicants (*for MG only*): **IBM UNITED KING-
DOM LIMITED** [GB/GB]; PO Box 41, North Harbour,
Portsmouth Hampshire PO6 3AU (GB). **IBM (CHINA)
INVESTMENT COMPANY LIMITED** [CN/CN]; 25/
F, Pangu Plaza, No.27, Central North 4th Ring Road,
Chaoyang District, Beijing 100101 (CN).

(72) Inventors: **SATHE, Saket**; IBM Corporation, 1101
Kitchawan Road, PO Box 218, Yorktown Heights, New
York 10598-0218 (US). **BRAMBLE, Gregory**; IBM Cor-
poration, 1101 Kitchawan Road, PO Box 218, Yorktown
Heights, New York 10598-0218 (US). **VU, Long**; IBM
Corporation, 1101 Kitchawan Road, PO Box 218, York-
town Heights, New York 10598-0218 (US). **SALONIDIS,
Theodoros**; IBM Corporation, 1101 Kitchawan Road, PO
Box 218, Yorktown Heights, New York 10598-0218 (US).

(54) Title: DISTRIBUTED RESOURCE-AWARE TRAINING OF MACHINE LEARNING PIPELINES

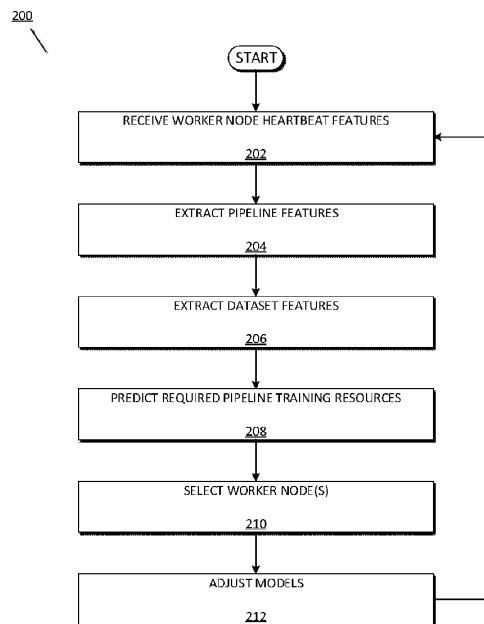


FIG. 2

(57) **Abstract:** A method, a structure, and a computer system for predicting pipeline training requirements. The exemplary embodiments may include receiving one or more worker node features from one or more worker nodes, extracting one or more pipeline features from one or more pipelines to be trained, and extracting one or more dataset features from one or more datasets used to train the one or more pipelines. The exemplary embodiments may further include predicting an amount of one or more resources required for each of the one or more worker nodes to train the one or more pipelines using the one or more datasets based on one or more models that correlate the one or more worker node features, one or more pipeline features, and one or more dataset features with the one or more resources. Lastly, the exemplary embodiments may include identifying a worker node requiring a least amount of the one or more resources of the one or more worker nodes for training the one or more pipelines.

[Continued on next page]

(74) **Agent: SHAW, Anita;** IBM United Kingdom Limited, Intellectual Property Law, Hursley Park, Hursley, Winchester Hampshire SO21 2JN (GB).

(81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))