

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
19 August 2021 (19.08.2021)



(10) International Publication Number
WO 2021/161136 A1

(51) International Patent Classification:
G06N 3/04 (2006.01)

(21) International Application Number:
PCT/IB2021/050895

(22) International Filing Date:
04 February 2021 (04.02.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
16/787,443 11 February 2020 (11.02.2020) US

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

(71) Applicants (for MG only): **IBM UNITED KINGDOM 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, Cental North 4th Ring Road, Chaoyang District, Beijing 100101 (CN).

(72) Inventor: **MORIMURA, Tetsuro**; IBM Japan Limited, 19-21 Nihonbashi-hakozakicho, Chuo-ku, Tokyo, 13 103-8510 (JP).

(74) Agent: **GRAHAM, Timothy**; 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,

(54) Title: LEARNING WITH MOMENT ESTIMATION USING DIFFERENT TIME CONSTANTS

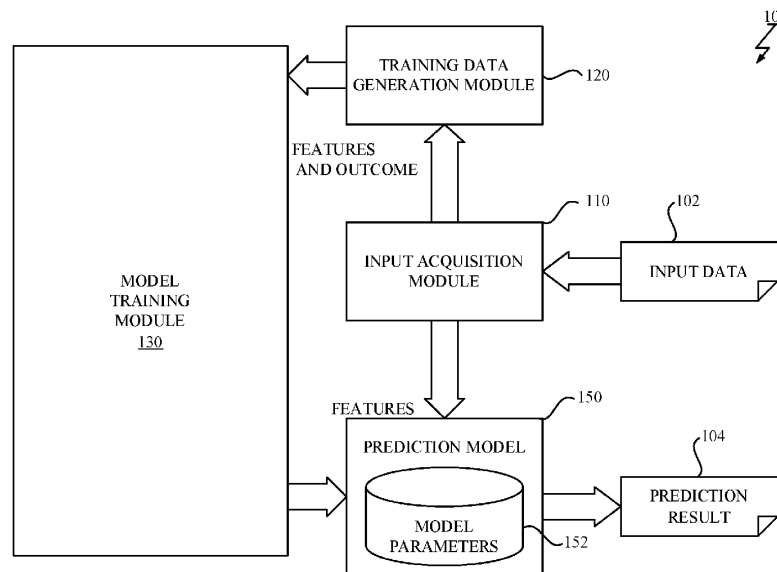


FIG. 1

(57) Abstract: A technique for training a model includes obtaining a training example for a model having model parameters stored on one or more computer readable storage mediums operably coupled to the hardware processor. The training example includes an outcome and features to explain the outcome. A gradient is calculated with respect to the model parameters of the model using the training example. Two estimates of a moment of the gradient with two different time constants are computed for the same type of the moment using the gradient. Using a hardware processor, the model parameters of the model are updated using the two estimates of the moment with the two different time constants to reduce errors while calculating the at least two estimates of the moment of the gradient.

[Continued on next page]

WO 2021/161136 A1

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))