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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 November 2020 (19.11.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/227983 A1

(51) International Patent Classification:
G06F 1/32 (2019.01)

(21) International Application Number:
PCT/CN2019/087083

(22) International Filing Date:
15 May 2019 (15.05.2019)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **ALIBABA GROUP HOLDING LIMITED** [—/CN]; Fourth Floor, One Capital Place, P.O. Box 847, George Town, Grand Cayman (KY).

(72) Inventors: **LI, Zhan**; 6142 147th Place SE, Bellevue, WA 98004 (CA). **REN, Zhixing**; 527 239th Ave SE, Sammamish, WA 98074 (US). **ZHANG, Yun**; Building 1, No.969 West Wen Yi Road, Yu Hang District, Hangzhou, Zhejiang 311121 (CN). **WANG, Jialong**; Building 1, No.969 West Wen Yi Road, Yu Hang District, Hangzhou, Zhejiang 311121 (CN).

(74) Agent: **BEIJING TSINGYUANHUI INTELLECTUAL PROPERTY LAW FIRM**; Room 101, 1st Floor, Building 1, No. C-18, Zhichun Road, Haidian District, Beijing, China 100190 (CN).

(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, 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,

(54) Title: HYBRID-LEARNING NEURAL NETWORK ARCHITECTURE

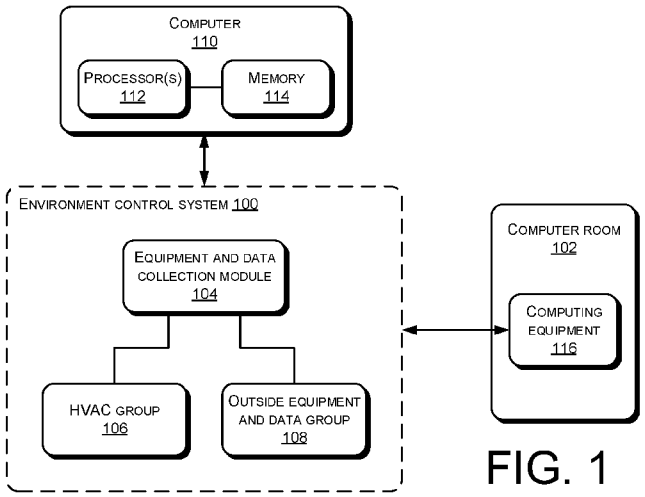

FIG. 1

(57) Abstract: Systems and methods for predicting energy efficiency of a computer room in a data center, and more specifically to predicting power usage effectiveness (PUE) of a computer room with optimized parameters using a two-tower deep learning architecture are provided. The two-tower deep learning architecture may learn embeddings from data and an ontology structure automatically, and may include training of two sub-networks, such as a first neural network that captures the domain knowledge embedded in the ontology and a second neural network that predicts the PUE from inputs. The learning of the first neural network, which may be unsupervised, and the second neural network, which may be supervised, may be simultaneous and referred to as hybrid-learning, and the two-tower deep learning architecture may also be referred to as a hybrid-learning neural network (HLNN) architecture.

WO 2020/227983 A1

[Continued on next page]

SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))*

HYBRID-LEARNING NEURAL NETWORK ARCHITECTURE

BACKGROUND

[0001] In a computer room of a data center, an environment control system, such as a heating, ventilation, and air conditioning (HVAC) system, is provided to maintain an acceptable operating environment for computing equipment that includes components such as servers, power supplies, displays, routers, network and communication modules, and the like, in the computer room. Based on the total energy consumed by the computer room and the total energy consumed by the computing equipment, power usage effectiveness (PUE), may be calculated and used to assess the energy efficiency of the computer room. The HVAC system may include many duplicative and/or similar components, such as coolers, fans, secondary pumps, air conditioners, refrigeration units, water pumps, such as cooling water pumps (CWPs), secondary chilling water pumps (SCWPs), and the like. For example, it is not unusual to equip more than fifty computer room air conditioning units (CRAC) in one computer room with tens of temperature and humidity sensors.

[0002] One method used to optimize the PUE is computational fluid dynamics (CFD), which utilizes numerical analysis and data structure to analyze and solve problems involving fluid. The CFD contains partial differential equations to provide predictions of air flows and heat distributions in computer rooms, and it is widely used in the design phase. However, CFD-based methods

are computationally intensive and are not well-suited for real-time operations. Moreover, extensive verifications and validations are needed to guarantee accuracy of the simulation, especially when there are changes involved.

[0003] Another method that may be used to predict the PUE and/or control the HVAC system is a deep learning neural network. The deep learning neural network does not depend on any of the physical models and does not distinguish various input features. Classic neural networks obtain knowledge/relations only from historical data and does not have any domain knowledge. It is more difficult for general deep learning model to apply to a system having a large number of duplicate and similar devices such as a computer room of a data center. Although these HVAC components have complex nonlinear correlations, the inputs from the sensors are treated equally from the perspective of a neural network structure, and information behind data from the inputs may be biased by the duplicative and/or similar inputs, which may result overfitting and eventual inaccuracy, causing inefficiency.

[0004] To avoid the duplicative and/or similar inputs and to improve the PUE or the energy consumption of the computer room, a popular solution is to manually aggregate the inputs based on a human expert's domain knowledge, and to set the aggregated input as the input to the neural network. However, this solution is room-specific and introduces extra manual work. Further, because this solution relies on the experience and analysis of an HVAC expert, it is difficult to fully understand the most reasonable correlations among various HVAC components to achieve an energy efficient computer room condition in

different operating conditions, such as outdoor temperature, outdoor humidity, computing load, and the like.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] The detailed description is set forth with reference to the accompanying figures. In the figures, the left-most digit(s) of a reference number identifies the figure in which the reference number first appears. The use of the same reference numbers in different figures indicates similar or identical items or features.

[0006] FIG. 1 illustrates an example block diagram of an environment control system used with a hybrid-learning neural network (HLNN) which may be utilized to predict power usage effectiveness (PUE) of a computer room.

[0007] FIG. 2 illustrates an example detailed block diagram of the environment control system of FIG. 1 with associated levels.

[0008] FIG. 3 illustrates an example block diagram of the HLNN architecture.

[0009] FIG. 4 illustrates an example flowchart describing a process of predicting the PUE by the HLNN.

DETAILED DESCRIPTION

[0010] Systems and methods discussed herein are directed to predicting energy efficiency of a computer room in a data center, and more specifically to predicting power usage effectiveness (PUE) of a computer room with optimized parameters using a two-tower deep learning architecture. The two-tower deep

learning architecture may learn embeddings from data and an ontology structure automatically, and may include simultaneous training of two sub-networks, an unsupervised Auto-encoder Net (AE-Net) that captures the domain knowledge embedded in the ontology and a supervised Prediction Net (P-Net) that predicts the PUE from inputs. The simultaneous learning of the AE-Net (unsupervised) and the P-Net (supervised) may be referred to as hybrid-learning, and the two-tower deep learning architecture may also be referred to as a hybrid-learning neural network (HLNN) architecture.

[0011] To achieve PUE optimization by ensuring a reasonable and appropriate operating environment, such as the environment of a computer room, and reducing waste in setting components of an environment control system, such as an HVAC system, machine learning methods may be used to learn from historical data to obtain complex relationships among various HVAC components and the energy efficiency of the computer room in different operating conditions.

[0012] In the HLNN architecture, the first and second neural networks, such as the AE-Net and the P-Net respectively, may share a shared structure comprising one input layer and two concept layers. Each of the AE-Net and the P-Net may have its own hidden layers and an output layer. The AE-Net, the P-Net, and the shared structure may form a hybrid-learning neural network (HLNN) architecture. The AE-Net may be an unsupervised-learning network, which may be trained to make its output copy input with a lowest possible error, while P-Net may be a deep feedforward neural network to predict the PUE.

[0013] Domain knowledge of the components associated with the HVAC system and the computing equipment of the computer room, for example, may be embedded into the HLNN architecture. By embedding the domain knowledge of the components, the number of inputs and the complexity of a search space may be reduced and the accuracy of the PUE prediction may be increased. The design of the two-tower deep learning architecture of the input layer and concept layers may be guided by the domain ontology containing multiple levels of nodes where a top level may contain a root concept and a bottom level may contain multiple instance. The instances in the bottom level of the ontology may be represented by the nodes in the input layer, and the concepts in the middle levels may also have corresponding nodes in concept layers of the shared structure. Moreover, the relations and/or connections between levels may be copied in the input and the concept layers.

[0014] FIG. 1 illustrates an example block diagram of an environment control system 100 used with a hybrid-learning neural network (HLNN) which may be utilized to predict power usage effectiveness (PUE) of a computer room 102.

[0015] The environment control system 100 may include a plurality of components such as an equipment and data collection module 104 communicatively coupled to an HVAC group 106 and an outside equipment and data group 108. The equipment and data collection module 104 may be configured to maintain profiles of components managed by the HVAC group 106 and the outside equipment and data group 108, receive input data from various sensors associated with those components, and transmit data to those

components to, in part, control the environment of, and calculate a predicted PUE of, the computer room 102. Some of the environment control system components may be located in the computer room 102, and other components may be located outside of a building in which the computer room 102 is located. The environment control system 100 may monitor energy consumption of components associated with the computer room 102, the equipment and data collection module 104, the HVAC group 106, and the outside equipment and data group 108. In addition, the environment control system 100 may be communicatively coupled to a computer 110. The computer 110 may comprise one or more processors 112 and memory 114 communicatively coupled to the one or more processors 112, which may store computer-readable instructions to be executed by the computer 110 to perform functions of the HLNN described below. The computer 110 may be located within the computer room 102 or may be remotely located from the computer room 102.

[0016] The computer room 102 may house computing equipment 116 including servers, power supplies, displays, routers, network and communication modules, and the like (not shown). The computing equipment 116 may be coupled to the environment control system 100 and may provide information regarding energy usage by the computing equipment 116 based on historical, current, and expected energy usage and computing loads for calculating the predicted PUE of the computer room 102.

[0017] FIG. 2 illustrates an example detailed block diagram of the environment control system 100 of FIG. 1 with associated levels (levels 1-4 shown).

[0018] The HVAC group 106 may comprise an HVAC control module 202 communicatively coupled to the equipment and data collection module 104, an air conditioning group 204, and a refrigeration group 206. The HVAC control module 202 may be configured to receive operating information from various sensors and controllers of the air conditioning group 204 and from the refrigeration group 206. The HVAC control module 202 may forward the operating information to the equipment and data collection module 104 for calculation by the HLNN. The HVAC control module 202 may also be configured to transmit control information received from the equipment and data collection module 104 to the air conditioning group 204 and the refrigeration group 206 for adjusting various parameters of the air conditioning group 204 and the refrigeration group 206 to optimize a desired parameter for predicting the PUE. The HVAC group 106 may further comprise a secondary pump group (not shown) and may similarly communicate associated operating information to and from the HVAC control module 202.

[0019] The air conditioning group 204 may comprise N air conditioners (two, AC-1 208 and AC-N 210, shown). Although not shown, each of N air conditioners may comprise several controls and sensors, such as a corresponding switch, a corresponding fan speed controller/sensor, a corresponding air conditioner output air temperature sensor, and a corresponding air conditioner

return air temperature sensor. Each of N air conditioners may be configured to receive AC operating information from the corresponding controls and sensors and forward the AC operating information to the air conditioning system 204, which, in turn, forwards the AC operating information to the HVAC control module 202. Each of N air conditioners may also be configured to transmit AC control information received from the air conditioning system 204 to the corresponding controls to optimize a desired parameter for predicting the PUE.

[0020] The refrigeration group 206 may comprise a plurality of refrigeration systems including a plurality of coolers (cooler-1 212 shown) and a plurality of cooling towers (tower-1 214 shown). Although not shown, each of the plurality of coolers may comprise associated switch, cooling mode controller, outflow cooling water temperature controller/sensor, and each of the plurality of cooling towers may comprise associated cooling tower fan speed controller/sensor, outflow cooling water temperature controller/sensor, and return cooling water temperature controller/sensor.

[0021] Each of the plurality of refrigeration systems may be configured to receive refrigeration operating information from the corresponding controls, switches, and sensors (not shown) and forward the refrigeration operating information to the HVAC control module 202 via the refrigeration group 206. Each of the plurality of refrigeration systems may also be configured to transmit refrigeration control information received from the refrigeration group 206 to the corresponding controls, switches, and sensors to optimize the desired parameter for predicting the PUE.

[0022] The outside equipment and data group 108 may comprise an outside equipment monitoring module 216 communicatively coupled to the equipment and data collection module 104, an outside humidity module 218, an outside wet bulb temperature module 220, and other modules (not shown). The outside humidity module 218 may be communicatively coupled to M humidity sensors (two humidity sensors, humidity sensor-1 222 and humidity sensor-M 224, shown). The outside wet bulb temperature module 220 may be communicatively coupled to M wet bulb temperature sensors (two wet bulb temperature sensors, wet bulb temperature sensor-1 226 and wet bulb temperature sensor-M 228, shown). The outside equipment monitoring module 216 may receive humidity and wet bulb temperature information from the corresponding sensors and forward the information to the equipment and data collection module 104 for optimizing the desired parameter for predicting the PUE.

[0023] Each block illustrated in FIG. 2 may be associated with one of a plurality of levels of the domain ontology. The domain ontology having four levels is illustrated herein as an example, however, the number of the levels of the domain ontology may not be limited to four and may be more or less than four levels. Level 1 may include the equipment and data collection module 104, which may be referred to as D1. Level 2 may include q modules including the HVAC control module 202 and the outside equipment monitoring module 216, which may be referred to as C_1, C_2, … C_q, respectively. Level 3 may include p modules including the air conditioning group 204, the refrigeration group 206, the outside humidity module 218, and the outside wet bulb temperature module

220, which may be referred to as B_1, B_2, … B_p, respectively. Level 4 may include k modules including the AC-1 208, the AC-N 210, the cooler-1 212, the tower-1 214, the humidity sensor-1 222, the humidity sensor-M 224, the wet bulb temp sensor-1 226, and the wet bulb temp sensor-M 228, which may be referred to as A_1, A_2, … A_k, respectively.

[0024] FIG. 3 illustrates an example block diagram of the HLNN architecture 300.

[0025] The HLNN structure 300 may comprise a domain ontology 302, a shared structure 304, a first neural network, such as an AE-Net 306, and a second neural network, such as a P-Net 308. There may be a plurality of levels in the ontology, and four levels corresponding to the blocks illustrated in FIG. 2 are shown in the domain ontology 302 as examples. The top level, Level 1, may contain the root concept, D_1 310, and the bottom level, Level 4, may contain a plurality of instances, of which four instances, A_1 312, A_2 314, A_n 316, and A_k 318, are shown. These four instances in Level 4 of the domain ontology 302 may be represented as nodes, A_1 320, A_2 322, A_n 324, and A_k 326, respectively, in an input layer 328 of the shared structure 304.

[0026] The second level, Level 2, and the third level, Level 3, of the domain ontology 302 may represent a plurality of instances, of which two concepts, C_1 330 and C_q 332, in Level 2 and three concepts, B_1 334, B_2 336, and B_p 338, in Level 3 are shown. These instances in Level 2 and Level 3 of the domain ontology 302 may also have corresponding nodes, C_1 340, C_q 342, B_1 344, B_2 346, and B_p 348, respectively, in concept layers 350 of the shared structure

304. Additionally, relations/connections between levels may also be copied in the input layer 328 and the concept layers 350. For example, in the domain ontology 302, the concept B _1 334 is shown to be connected to a set of instances, A_1 312, A_2 314, and A_n 316, and in the concept layer 350, the corresponding node, B_1 344 is also shown to be connected to the corresponding nodes A_1 320, A_2 322, and A_n 324 in the input layer 328.

[0027] The P-Net 308 may be a deep feedforward neural network and may comprise hidden layers 352 and a one-node output layer 354 to output PUE parameters 356, plus the input layer 328 and the concept layers 350 of the shared structure 304. An example feed-forward operation of the P-Net 308 is described below. A neutron and a node may be interchangeable used.

[0028] Let ω_{ji}^{l} denote the weight between the j^{th} neuron, or node, in the $(l\text{-}1)^{\text{th}}$ layer and the i^{th} neuron in the l^{th} layer, and b_i^l be the bias of the i^{th} neuron in the l^{th} layer. With these notations, the feed-forward operation may be described as

$$x_i^l = \sum_{j=1}^{R_{l-1}} \omega_{ji}^l \, o_j^{l-1} + b_i^l \tag{1}$$

where x_i^l is the weighted input of node i, o_j^{l-1} is the output of the j^{th} node in the $(l\text{-}1)^{\text{th}}$ layer, and $R_{l\text{-}1}$ is the number of neurons in the $(l\text{-}1)^{\text{th}}$ layer.

[0029] Given $o_j^l = 1$ and $\omega_0^{il} = b_i^l$, the equation (1) may be simplified to:

$$x_i^l = \sum_{j=1}^{R_{l-1}} \omega_{ji}^l \, o_j^{l-1} \tag{2}$$

[0030] Using the notation above, the activation of node i is

$$o_i^l = f_p(x_i^l) \tag{3}$$

where f_p is the activation function.

[0031] In the shared structure 304, the connections may be guided by the domain knowledge, which may not fully connect the nodes in the concept layers 350. Let γ_{ji}^{l} denote the concept relation weight between two concept nodes, i.e., node j and node i, the weighted input of node i may then be expressed as:

$$x_i^l = \sum_{j=1}^{R_{l-1}} \omega_{ji}^l \, o_j^{l-1} \tag{4}$$

[0032] Each layer of the concept layers 350 may be mapped from a corresponding level of the concepts in the domain ontology 302. Let n_{ci}^{l} denote the number of nodes (instances and concepts) in the concept hierarchy that are connected to the i^{th} node, i.e., node C_i in the domain ontology 302, in the l^{th} layer, then the concept relation weight γ_{ji}^{l} may be expressed as

$$\gamma_{ji}^{l} = \begin{cases} 0 & if\ n_{ci}^{l} = 0 \\ 1 & if\ n_{ci}^{l} \neq 0 \end{cases} \tag{5}$$

That is, if the corresponding sub-concept/instance node j in the $(l\text{-}1)^{\text{th}}$ layer is not connected to the concept of node i, then γ_{ji}^{l} is zero. If $n_{ci}^{l} \neq 0$, the concept relation weights are 1, which do not affect the learning process. The loss function, $L_{PN}(a, d_p)$, may then define the error between d_P and o_P for the input a by letting a denote the input vector, o_P denote the calculated output of the neural network, and d_P denote the desired output.

[0033] The AE-Net 306 may be an unsupervised learning model comprising hidden layers 358 and an output layer 360, plus the input layer 328 and the concept layers 350 of the shared structure 304. The AE-Net 306 may be designed to minimize the difference between the input from the input layer 328 of the shared structure 304 and the output from the output layer 360. Given that an input

vector a from the input layer 328, a representation vector c from the top concept layer of the concept layers 350, and an output vector r (R_1 362 and R_k 364 shown) from the output layer 360, a mapping that transforms a into c may be called an encoder, and a mapping that transforms c back to r may be called a decoder. The encoder may be composed of the input layer 328 and the concept layers 350, while the decoder may be composed of the hidden layers 358 and the output layer 360. The training process in the AE-Net 306 may help the encoder preserve the domain knowledge in the domain ontology 302.

[0034] Setting the input vector a as $a = \{a_1, a_2, \ldots, a_k\}$, the representation vector c may be expressed as $c = \{c_1, c_2, \ldots, c_q\}$, and the output vector r as $r = \{r_1, r_2, \ldots, r_k\}$, then the encoder function, f_θ, and the decoder function, g_θ, may be expressed as:

$$c = f_\theta(a) \tag{6}$$

$$r = g_\theta(c) \tag{7}$$

[0035] In both encoder function f_θ and decoder function g_θ, the parameter set is $\theta = \{W, b, W', d\}$, where W and W' are the encoder and decoder weight matrices, and b and d are encoder and decoder bias vectors. The encoder function f_θ and decoder function g_θ may then be expressed, respectively, as:

$$f_\theta(a) = s_f(b + Wa) \tag{8}$$

$$g_\theta(c) = s_g(d + W'c) \tag{9}$$

where s_f and s_g are the encoder and decoder activation functions. In probabilistic terms r is not an exact reconstruction of a but the parameters of a distribution

$p(A|R = r)$ that generates a with high probability. The AE-Net 306 may be trained to find a parameter set to minimize reconstruction error in the equation below:

$$E_{AE}(\theta) = \sum_{a \in A} L_{AE}(a, r) = \sum_{a \in A} L_{AE}(a, g_\theta(f_\theta(a))) \quad (10)$$

where A denotes the training set of examples, L_{AE} is the loss function or the reconstruction error. The input vector may be of real-value and the loss function L_{AE} may be squared error $L_{AE}(a, r) = ||a - r||^2$. Both s_f and s_g may be sigmoid functions.

[0036] The HLNN 300 may be trained in a manner similar to standard neural networks. The only difference may be that the loss function L_{Model} may be composed of two components: the loss of the AE-Net 306, L_{AE}, and the prediction loss L_{PN} of the P-Net 308:

$$L_{Model} = L_{PN} + \alpha L_{AE} \quad (11)$$

where α is a constant providing a bias or weight to L_{AE}. Alternatively, L_{PN} may be biased or weighted by another constant β, and L_{Model} may be expressed as $L_{Model} = \beta L_{PN} + L_{AE}$.

[0037] In the training, the derivatives of the loss may be expressed as

$$\frac{\partial L_{Model}}{\partial \omega_{ji}^l} = \frac{\partial L_{Model}}{\partial x_i^l} \frac{\partial x_i^l}{\partial \omega_{ji}^l} \quad (12)$$

and the following substitutions may be made:

$$\frac{\partial x_i^l}{\partial \omega_{ji}^l} = \frac{\partial (\sum_{j=1}^{R_{l-1}} \omega_{ji}^l o_j^{l-1} \gamma_{ji}^l)}{\partial \omega_{ji}^l} = o_j^{l-1} \gamma_{ji}^l \quad (13)$$

$$\frac{\partial L_{Model}}{\partial x_i^l} = \sum_{k=1}^{R^{l+1}} \frac{\partial L_{Model}}{\partial x_i^{l+1}} \frac{\partial x_k^{l+1}}{\partial x_i^l} \quad (14)$$

where R^{l+1} is the number of nodes in the $(l+1)^{th}$ layer. Combining equations (12), (13), and (14), yields

$$\frac{\partial L_{Model}}{\partial \omega_{ji}^{l}} = o_j^{l-1} \sum_{k=1}^{R^{l+1}} \frac{\partial L_{Model}}{\partial x_k^{l+1}} \frac{\partial x_k^{l+1}}{\partial x_i^{l}} \tag{15}$$

[0038] If the layer $l+1$ is in the shared structure 304, i.e., the input layer 328 and the concept layers 350, the equation (15) may be transformed to:

$$\frac{\partial L_{Model}}{\partial \omega_{ji}^{l}} = o_j^{l-1} \sum_{k=1}^{R^{l+1}} \frac{\partial (L_{PN} + \alpha L_{AE})}{\partial x_k^{l+1}} \frac{\partial x_k^{l+1}}{\partial x_i^{l}} \tag{16}$$

[0039] If the layer $l+1$ is in the hidden layers 352 or the output layer 354 of the P-Net 308, the equation (15) may be transformed to:

$$\frac{\partial L_{Model}}{\partial \omega_{ji}^{l}} = o_j^{l-1} \sum_{k=1}^{R^{l+1}} \frac{\partial (L_{PN} + \alpha L_{AE})}{\partial x_k^{l+1}} \frac{\partial x_k^{l+1}}{\partial x_i^{l}}$$

$$= o_j^{l-1} \sum_{k=1}^{R_{PN}^{l+1}} \frac{\partial L_{PN}}{\partial x_k^{l+1}} \frac{\partial x_k^{l+1}}{\partial x_i^{l}} \tag{17}$$

where R_{PN}^{l+1} denotes the number of nodes in the $(l+1)^{th}$ layer that are used to calculate the output of the P-Net 308.

[0040] If the layer $l+1$ is the hidden layers 358 or the output layer 360 of the AE-Net 306, the equation (15) may be transformed to:

$$\frac{\partial L_{Model}}{\partial \omega_{ji}^{l}} = o_j^{l-1} \sum_{k=1}^{R^{l+1}} \frac{\partial (L_{PN} + \alpha L_{AE})}{\partial x_k^{l+1}} \frac{\partial x_k^{l+1}}{\partial x_i^{l}}$$

$$= \alpha o_j^{l-1} \sum_{k=1}^{R_{AE}^{l+1}} \frac{\partial L_{AE}}{\partial x_k^{l+1}} \frac{\partial x_k^{l+1}}{\partial x_i^{l}} \tag{18}$$

where R_{AE}^{l+1} denotes the number of nodes in the $(l+1)^{th}$ layer that are used to calculate the output of the AE-Net 306.

[0041] Th equation (16), (17), and (18) show that the derivatives of the loss function L_{Model} are back propagated for learning both the AE-Net 306 and the P-Net 308. The solution for the PUE may be optimized by minimizing the loss calculated by the loss function L_{Model} as expressed by the equation (11), which may be accomplished by setting the derivative, such as the equations (16), (17), and (18), of the loss function L_{Model} to zero and solving for the variables. Because the solution may not always converge to zero, or may take longer than an acceptable time or a number of iterations, the value for the derivative may be set to a sufficiently small and acceptable threshold value.

[0042] FIG. 4 illustrates an example flowchart 400 describing a process of predicting the power utilization effectiveness (PUE) by the HLNN 300.

[0043] At block 402, the HLNN 300 may create an ontology having a plurality of levels, such as the domain ontology 302, of the components associated with the environment control system 100 associated with the computer room 102 as illustrated in FIGs. 1-3. The HLNN 300 may receive information of the components associated with the environment control system 100 automatically including corresponding associated historical data, locations and physical connections, and hierarchy among the components as illustrated in FIGs. 1- 3. The computing equipment 116 may include servers, power supplies, displays, routers, network and communication modules (telephone, internet, wireless devices, etc.), and the like. The relationships among components of the environment control system 100 and the computing equipment 116 may be based on loading of the computing equipment 116, such as a workload, or computing

load, of the servers and an electrical load of the servers as a function of the workload of the servers.

[0044] At block 404, the HLNN 300 may receive input feature parameters of the components associated with the environment control system 100. More specifically, the input layer 328 of the shared structure 304 may receive k instances, A_1 312, A_2 314, A_n 316, and A_k 318, from the domain ontology 302, where k is an integer. Each of k instances may have a corresponding input feature parameter in the input layer 328 (A_1 320, A_2 322, A_n 324, and A_k 326 as illustrated in FIG. 3) may belong to one or more corresponding upper concepts of a plurality of upper concepts as illustrated hierarchically in the concept layers 350.

[0045] At block 406, both the first neural network, such as the AE-Net 306, and the second neural network, such as the P-Net 308, may be trained simultaneously. As discussed above, the input vector a, or the input feature parameters, may be expressed as $a = \{a_1, a_2, \ldots, a_k\}$, the representation vector c, or the concepts, as $c = \{c_1, c_2, \ldots, c_q\}$, and the output vector r as $r = \{r_1, r_2, \ldots, r_k\}$. A mapping that transforms a into c may be called an encoder, and a mapping that transforms c back to r may be called a decoder. The encoder may be composed of the input layer 328 and the concept layers 350, while the decoder may be composed of the hidden layers 358 and the output layer 360. The training process in the AE-Net 306 may help the encoder preserve the domain knowledge in the domain ontology 302.

[0046] At block 408, the HLNN 300 may minimize the loss based on the loss function L_{Model} by utilizing the trained AE-Net 306 and the trained P-Net 308 and predict a power usage effectiveness (PUE) of the computer room 102 at block 410. The derivative of the loss function L_{Model}, such as the equations (16), (17), and (18), may be to zero for solving for the variables. Because the solution may not always converge to zero, or may take longer than an acceptable time, the value for the derivative may be set to a sufficiently small and acceptable threshold value.

[0047] The trained neural networks may be generated automatically, and the training of the trained neural networks may be performed by using a gradient descent algorithm to implement learning of the input feature parameters for corresponding concepts. An architecture of the trained neural networks may reflect deep learning of the plurality of components and associated concepts based on the relationships among the plurality of components. The trained neural networks may comprise hierarchical concept layers, such as the concept layers 350, coupled between the input layer, such as the input layer 328, and an output layer, such as the output layer 354 or 360. The concept layers 350 may be added between the input layer 328 and the hidden layers 352 and 358 as illustrated in FIG. 3. The concept layer 350 may be embedded with domain knowledge from the domain ontology 302. The concept layer 350 may construct a concept structure based on relationships among the plurality of components. The concept structure may be created manually or automatically with smart components capable of communicating with each other. The training portion of the HLNN

300 and the prediction the PUE utilizing the HLNN 300 may be performed separately and/or by different parties.

[0048] A general deep learning network may not be capable of reasonably distinguishing all duplicative and/or similar input features, and may identify the importance of each feature based entirely on historical data. In a structure, such as the computer room 102 with a large number of duplicative and similar devices, if these duplicate and/or similar input features parameters were not categorized, aggregated or abstracted, the complexity of the network and space for learning and searching would greatly increase, requiring higher quality and quantity of data. Although, it may be easy to obtain unreasonable overfitting, it would decrease prediction accuracy.

[0049] Some or all operations of the methods described above can be performed by execution of computer-readable instructions stored on a computer-readable storage medium, as defined below. The term "computer-readable instructions" as used in the description and claims, include routines, applications, application modules, program modules, programs, components, data structures, algorithms, and the like. Computer-readable instructions can be implemented on various system configurations, including single-processor or multiprocessor systems, minicomputers, mainframe computers, personal computers, hand-held computing devices, microprocessor-based, programmable consumer electronics, combinations thereof, and the like.

[0050] The computer-readable storage media may include volatile memory (such as random-access memory (RAM)) and/or non-volatile memory (such as

read-only memory (ROM), flash memory, etc.). The computer-readable storage media may also include additional removable storage and/or non-removable storage including, but not limited to, flash memory, magnetic storage, optical storage, and/or tape storage that may provide non-volatile storage of computer-readable instructions, data structures, program modules, and the like.

[0051] A non-transient computer-readable storage medium is an example of computer-readable media. Computer-readable media includes at least two types of computer-readable media, namely computer-readable storage media and communications media. Computer-readable storage media includes volatile and non-volatile, removable and non-removable media implemented in any process or technology for storage of information such as computer-readable instructions, data structures, program modules, or other data. Computer-readable storage media includes, but is not limited to, phase change memory (PRAM), static random-access memory (SRAM), dynamic random-access memory (DRAM), other types of random-access memory (RAM), read-only memory (ROM), electrically erasable programmable read-only memory (EEPROM), flash memory or other memory technology, compact disk read-only memory (CD-ROM), digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other non-transmission medium that can be used to store information for access by a computing device. In contrast, communication media may embody computer-readable instructions, data structures, program modules, or other data in a modulated data signal, such as a carrier wave, or other transmission

mechanism. As defined herein, computer-readable storage media do not include communication media.

[0052] The computer-readable instructions stored on one or more non-transitory computer-readable storage media that, when executed by one or more processors, may perform operations described above with reference to FIGs. 1-4. Generally, computer-readable instructions include routines, programs, objects, components, data structures, and the like that perform particular functions or implement particular abstract data types. The order in which the operations are described is not intended to be construed as a limitation, and any number of the described operations can be combined in any order and/or in parallel to implement the processes.

EXAMPLE CLAUSES

[0053] A. A method comprising: receiving input feature parameters of a plurality of components associated with at least one computer room; raining a first neural network and a second neural network based on the input feature parameters; and predicting the power usage effectiveness (PUE) of the at least one computer room based on an output of the first neural network and an output of the second neural network.

[0054] B. The method as paragraph A recites, wherein the first neural network is an unsupervised neural network and the second network is a supervised prediction neural network.

[0055] C. The method as paragraph A recites, wherein training the first neural network and the second neural network based on the input feature parameters includes simultaneously training the first neural network and the second neural network based on the input feature parameters.

[0056] D. The method as paragraph A recites, wherein receiving the input feature parameters of the plurality of components associated with the at least one computer room includes: creating an ontology having a plurality of levels associated with the plurality of components; and receiving information of the plurality of components based on the on the ontology including corresponding associated concept, historical data, locations, physical connections, and hierarchy among the plurality of components.

[0057] E. The method as paragraph D recites, wherein the relationships among the plurality of components are based, at least in part, on loading of computing equipment in the computer room.

[0058] F. The method as paragraph E recites, wherein the loading of the computing equipment includes a workload of the computing equipment and an electrical load used by the computing equipment.

[0059] G. The method as paragraph F recites, wherein the computing equipment includes a server and a power supply for the server.

[0060] H. The method as paragraph D recites, wherein training the first neural network and the second neural network based on the input feature parameters includes using a gradient descent algorithm to implement learning of the input feature parameters for corresponding concepts.

[0061] I. The method as paragraph A recites, wherein predicting the PUE of the at least one computer room based on the output of the first neural network and the output of the second neural network includes biasing a loss associated with the first neural network with a constant value.

[0062] J. The method as paragraph I recites, wherein predicting the PUE of the at least one computer room based on the output of the first neural network and the output of the second neural network includes minimizing a total loss calculated based on the biased loss associated with the first neural network and an unbiased loss associated with the second neural network.

[0063] K. The method as paragraph J recites wherein minimizing the loss calculated based on the loss function by utilizing the first neural network and the second neural network includes solving for a derivative of the loss function equaling zero.

[0064] L. The method as paragraph J recites, wherein minimizing the loss calculated based on the loss function by utilizing the trained first neural network and the trained second neural network includes solving for a derivative of the loss function being less than or equal to a threshold value.

[0065] M. A system comprising: one or more processors; and memory communicatively coupled to the one or more processors, the memory storing computer-readable instructions executable by one or more processors, that when executed by the one or more processors, cause the one or more processors to perform operations comprising: receiving input feature parameters of a plurality of components associated with at least one computer room; training a first neural

network and a second neural network based on the input feature parameters; and predicting the power usage effectiveness (PUE) of the at least one computer room based on an output of the first neural network and an output of the second neural network.

[0066] N. The system as paragraph M recites, wherein the first neural network is an unsupervised neural network and the second network is a supervised prediction neural network.

[0067] O. The system as paragraph M recites, wherein training the first neural network and the second neural network based on the input feature parameters includes simultaneously training the first neural network and the second neural network based on the input feature parameters.

[0068] P. The system as paragraph M recites, wherein receiving the input feature parameters of the plurality of components associated with the at least one computer room includes: creating an ontology having a plurality of levels associated with the plurality of components; and receiving information of the plurality of components based on the on the ontology including corresponding associated concept, historical data, locations, physical connections, and hierarchy among the plurality of components.

[0069] Q. The system as paragraph P recites, wherein the relationships among the plurality of components are based, at least in part, on loading of the computing equipment.

[0070] R. The system as paragraph Q recites, wherein the loading of the computing equipment includes a workload of the computing equipment and an electrical load used by the computing equipment.

[0071] S. The system as paragraph R recites, wherein the computing equipment includes a server and a power supply for the server.

[0072] T. The system as paragraph P recites, wherein training the first neural network and the second neural network based on the input feature parameters includes using a gradient descent algorithm to implement learning of the input feature parameters for corresponding concepts.

[0073] U. The system as paragraph M recites, wherein predicting the PUE of the at least one computer room based on the output of the first neural network and the output of the second neural network includes biasing a loss associated with the first neural network with a constant value.

[0074] V. The system as paragraph U recites, wherein predicting the PUE of the at least one computer room based on the output of the first neural network and the output of the second neural network includes minimizing a total loss calculated based on the biased loss associated with the first neural network and an unbiased loss associated with the second neural network.

[0075] W. The system as paragraph V recites, wherein minimizing the loss calculated based on the loss function by utilizing the first neural network and the second neural network includes solving for a derivative of the loss function equaling zero.

[0076] X. The system as paragraph V recites, wherein minimizing the loss calculated based on the loss function by utilizing the trained first neural network and the trained second neural network includes solving for a derivative of the loss function being less than or equal to a threshold value.

[0077] Y. A non-transitory computer-readable storage medium storing computer-readable instructions executable by one or more processors, that when executed by the one or more processors, cause the one or more processors to perform operations comprising: receiving input feature parameters of a plurality of components associated with at least one computer room; training a first neural network and a second neural network based on the input feature parameters; and predicting the power usage effectiveness (PUE) of the at least one computer room based on an output of the first neural network and an output of the second neural network.

[0078] Z. The non-transitory computer-readable storage medium as paragraph Y recites, wherein the first neural network is an unsupervised neural network and the second network is a supervised prediction neural network.

[0079] AA. The non-transitory computer-readable storage medium as paragraph Y recites, wherein training the first neural network and the second neural network based on the input feature parameters includes simultaneously training the first neural network and the second neural network based on the input feature parameters.

[0080] AB. The non-transitory computer-readable storage medium as paragraph Y recites, wherein receiving the input feature parameters

of the plurality of components associated with the at least one computer room includes: creating an ontology having a plurality of levels associated with the plurality of components; and receiving information of the plurality of components based on the on the ontology including corresponding associated concept, historical data, locations, physical connections, and hierarchy among the plurality of components.

[0081] AC. The non-transitory computer-readable storage medium as paragraph AB recites, wherein the relationships among the plurality of components are based, at least in part, on loading of computing equipment in the computer room.

[0082] AD. The non-transitory computer-readable storage medium as paragraph AC recites, wherein the loading of the computing equipment includes a workload of the computing equipment and an electrical load used by the computing equipment.

[0083] AE. The non-transitory computer-readable storage medium as paragraph AD recites, wherein the computing equipment includes a server and a power supply for the server.

[0084] AF. The non-transitory computer-readable storage medium as paragraph AB recites, wherein training the first neural network and the second neural network based on the input feature parameters includes using a gradient descent algorithm to implement learning of the input feature parameters for corresponding concepts.

[0085] AG. The non-transitory computer-readable storage medium as paragraph Y recites, wherein predicting the PUE of the at least one computer room based on the output of the first neural network and the output of the second neural network includes biasing a loss associated with the first neural network with a constant value.

[0086] AH. The non-transitory computer-readable storage medium as paragraph AG recites, wherein predicting the PUE of the at least one computer room based on the output of the first neural network and the output of the second neural network t includes minimizing a total loss calculated based on the biased loss associated with the first neural network and an unbiased loss associated with the second neural network.

[0087] AI. The non-transitory computer-readable storage medium as paragraph AH recites, wherein minimizing the loss calculated based on the loss function by utilizing the first neural network and the second neural network includes solving for a derivative of the loss function equaling zero.

[0088] AJ.The non-transitory computer-readable storage medium as paragraph AH recites, wherein minimizing the loss calculated based on the loss function by utilizing the trained first neural network and the trained second neural network includes solving for a derivative of the loss function being less than or equal to a threshold value.

CONCLUSION

[0089] Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter defined in the appended claims is not necessarily limited to the specific features or acts described. Rather, the specific features and acts are disclosed as exemplary forms of implementing the claims.

[0090]

CLAIMS

WHAT IS CLAIMED IS:

1. A method comprising:

receiving input feature parameters of a plurality of components associated with at least one computer room;

training a first neural network and a second neural network based on the input feature parameters; and

predicting a power usage effectiveness (PUE) of the at least one computer room based on an output of the first neural network and an output of the second neural network.

2. The method of claim 1, wherein the first neural network is an unsupervised neural network and the second network is a supervised prediction neural network.

3. The method of claim 1, wherein training the first neural network and the second neural network based on the input feature parameters includes simultaneously training the first neural network and the second neural network based on the input feature parameters.

4. The method of claim 1, wherein receiving the input feature parameters of the plurality of components associated with the at least one computer room includes:

creating an ontology having a plurality of levels associated with the plurality of components; and

receiving information of the plurality of components based on the on the ontology including corresponding associated concept, historical data, locations, physical connections, and hierarchy among the plurality of components.

5. The method of claim 4, wherein the relationships among the plurality of components are based, at least in part, on loading of computing equipment in the computer room.

6. The method of claim 5, wherein the loading of the computing equipment includes a workload of the computing equipment and an electrical load used by the computing equipment.

7. The method of claim 6, wherein the computing equipment includes a server and a power supply for the server.

8. The method of claim 4, wherein training the first neural network and the P-Net based on the input feature parameters includes using a gradient descent algorithm to implement learning of the input feature parameters for corresponding concepts.

9. The method of claim 1, wherein predicting the PUE of the at least one computer room based on the output of the first neural network and the output of the second neural network includes biasing a loss associated with the first neural network with a constant value.

10. The method of claim 9, wherein predicting the PUE of the at least one computer room based on the output of the first neural network and the output of the second neural network includes minimizing a total loss calculated based on the biased loss associated with the first neural network and an unbiased loss associated with the second neural network.

11. The method of claim 10, wherein minimizing the loss calculated based on the loss function by utilizing the first neural network and the second neural network includes solving for a derivative of the loss function equaling zero.

12. The method of claim 10, wherein minimizing the loss calculated based on the loss function by utilizing the trained first neural network and the trained second neural network includes solving for a derivative of the loss function being less than or equal to a threshold value.

13. A system comprising:

one or more processors; and

memory communicatively coupled to the one or more processors, the memory storing computer-readable instructions executable by one or more processors, that when executed by the one or more processors, cause the one or more processors to perform operations comprising:

receiving input feature parameters of a plurality of components associated with at least one computer room;

training a first neural network and a second neural network based on the input feature parameters; and

predicting the power usage effectiveness (PUE) of the at least one computer room based on an output of the first neural network and an output of the second neural network.

14. The system of claim 13, wherein the first neural network is an unsupervised neural network and the second network is a supervised prediction neural network.

15. The system of claim 13, wherein training the first neural network and the second neural network based on the input feature parameters includes simultaneously training the first neural network and the second neural network based on the input feature parameters.

16. The system of claim 13, wherein receiving the input feature parameters of the plurality of components associated with the at least one computer room includes:

creating an ontology having a plurality of levels associated with the plurality of components; and

receiving information of the plurality of components based on the on the ontology including corresponding associated concept, historical data, locations, physical connections, and hierarchy among the plurality of components.

17. The system of claim 16, wherein the relationships among the plurality of components are based, at least in part, on loading of the computing equipment.

18. The system of claim 17, wherein the loading of the computing equipment includes a workload of the computing equipment and an electrical load used by the computing equipment.

19. The system of claim 18, wherein the computing equipment includes a server and a power supply for the server.

20. The system of claim 16, wherein training the first neural network and the second neural network based on the input feature parameters includes

using a gradient descent algorithm to implement learning of the input feature parameters for corresponding concepts.

21. The system of claim 13, wherein predicting the PUE of the at least one computer room based on the output of the first neural network and the output of the second neural network includes biasing a loss associated with the first neural network with a constant value.

22. The system of claim 21, wherein predicting the PUE of the at least one computer room based on the output of the first neural network and the output of the second neural network includes minimizing a total loss calculated based on the biased loss associated with the first neural network and an unbiased loss associated with the second neural network.

23. The system of claim 22, wherein minimizing the loss calculated based on the loss function by utilizing the first neural network and the second neural network includes solving for a derivative of the loss function equaling zero.

24. The system of claim 22, wherein minimizing the loss calculated based on the loss function by utilizing the trained first neural network and the trained second neural network includes solving for a derivative of the loss function being less than or equal to a threshold value.

25. A non-transitory computer-readable storage medium storing computer-readable instructions executable by one or more processors, that when executed by the one or more processors, cause the one or more processors to perform operations comprising:

receiving input feature parameters of a plurality of components associated with at least one computer room;

training a first neural network and a second neural network based on the input feature parameters; and

predicting the power usage effectiveness (PUE) of the at least one computer room based on an output of the first neural network and an output of the second neural network.

26. The non-transitory computer-readable storage medium of claim 25, wherein the first neural network is an unsupervised neural network and the second network is a supervised prediction neural network.

27. The non-transitory computer-readable storage medium of claim 25, wherein training the first neural network and the second neural network based on the input feature parameters includes simultaneously training the first neural network and the second neural network based on the input feature parameters.

28. The non-transitory computer-readable storage medium of claim 25, wherein receiving the input feature parameters of the plurality of components associated with the at least one computer room includes:

creating an ontology having a plurality of levels associated with the plurality of components; and

receiving information of the plurality of components based on the on the ontology including corresponding associated concept, historical data, locations, physical connections, and hierarchy among the plurality of components.

29. The non-transitory computer-readable storage medium of claim 28, wherein the relationships among the plurality of components are based, at least in part, on loading of computing equipment in the computer room.

30. The non-transitory computer-readable storage medium of claim 29, wherein the loading of the computing equipment includes a workload of the computing equipment and an electrical load used by the computing equipment.

31. The non-transitory computer-readable storage medium of claim 30, wherein the computing equipment includes a server and a power supply for the server.

32. The non-transitory computer-readable storage medium of claim 28, wherein training the first neural network and the second neural network based

on the input feature parameters includes using a gradient descent algorithm to implement learning of the input feature parameters for corresponding concepts.

33. The non-transitory computer-readable storage medium of claim 25, wherein predicting the PUE of the at least one computer room based on the output of the first neural network and the output of the second neural network includes biasing a loss associated with the first neural network with a constant value.

34. The non-transitory computer-readable storage medium of claim 33, wherein predicting the PUE of the at least one computer room based on the output of the first neural network and the output of the second neural network includes minimizing a total loss calculated based on the biased loss associated with the first neural network and an unbiased loss associated with the second neural network.

35. The non-transitory computer-readable storage medium of claim 34, wherein minimizing the loss calculated based on the loss function by utilizing the first neural network and the second neural network includes solving for a derivative of the loss function equaling zero.

36. The non-transitory computer-readable storage medium of claim 34, wherein minimizing the loss calculated based on the loss function by utilizing

the trained first neural network and the trained second neural network includes solving for a derivative of the loss function being less than or equal to a threshold value.

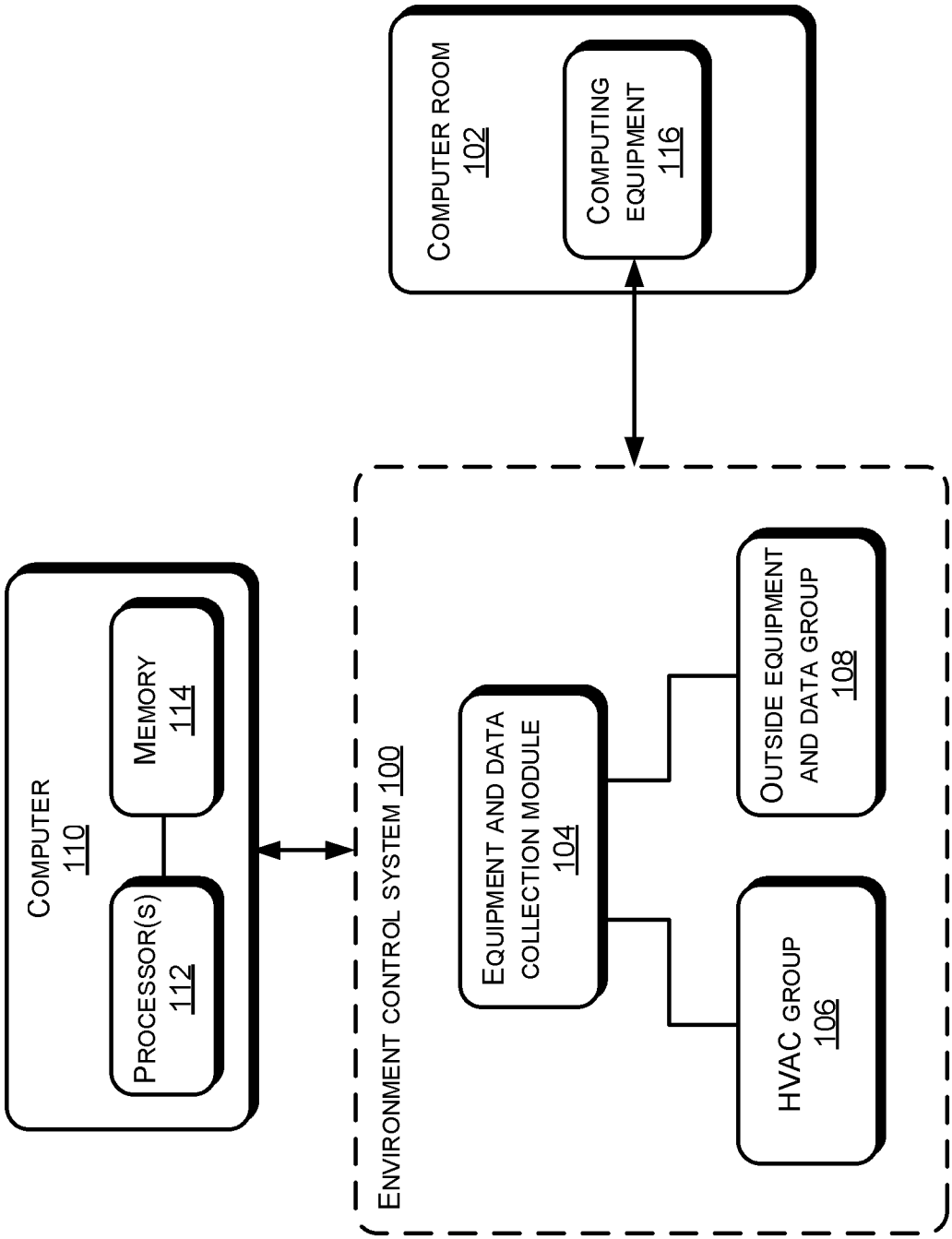
COMPUTER ROOM
102
COMPUTING EQUIPMENT
116
COMPUTER
110
MEMORY
114
PROCESSOR(S)
112
ENVIRONMENT CONTROL SYSTEM 100
EQUIPMENT AND DATA COLLECTION MODULE
104
OUTSIDE EQUIPMENT AND DATA GROUP
108
HVAC GROUP
106

FIG. 1

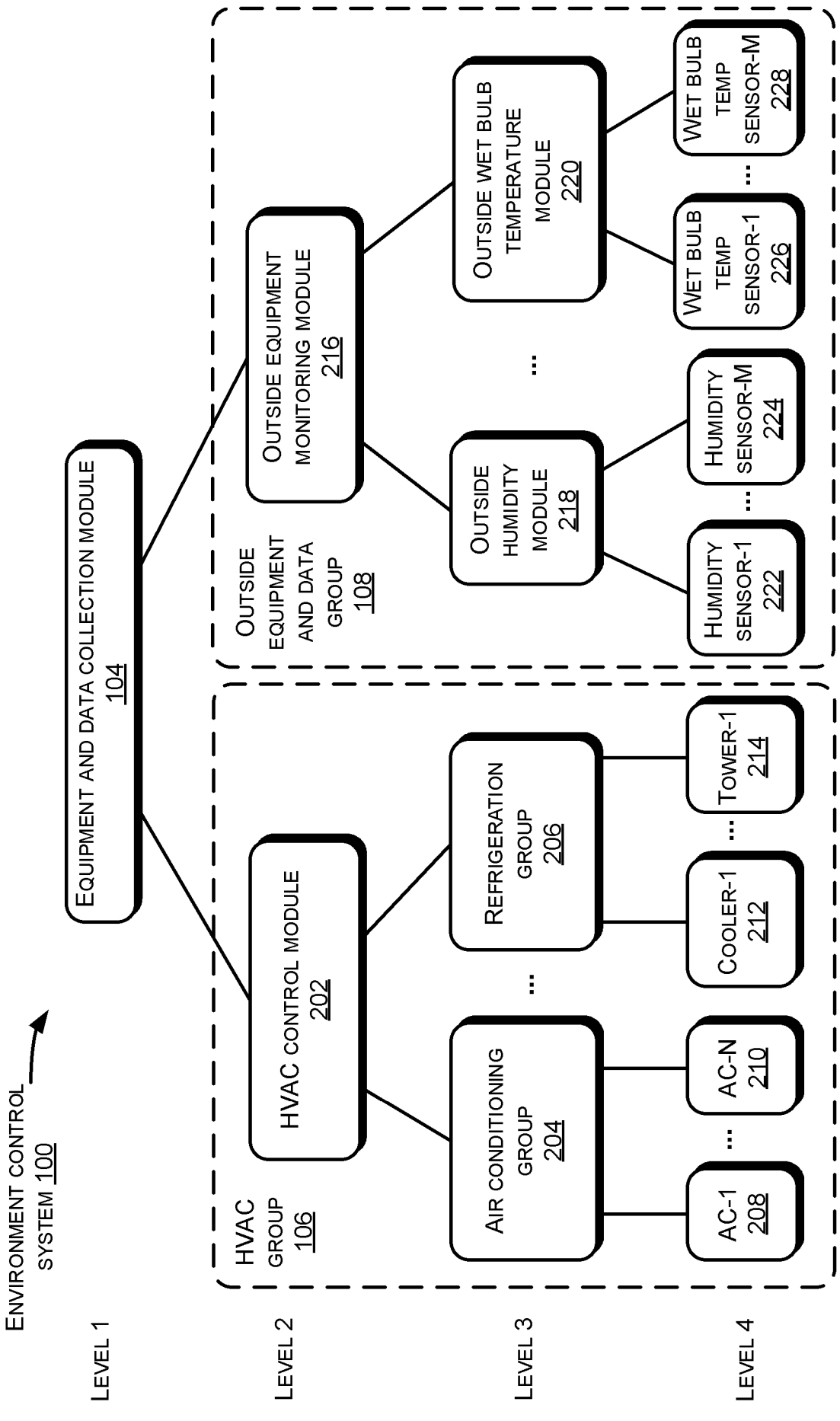
ENVIRONMENT CONTROL SYSTEM 100
LEVEL 1
LEVEL 2
LEVEL 3
LEVEL 4
EQUIPMENT AND DATA COLLECTION MODULE 104
HVAC GROUP 106
HVAC CONTROL MODULE 202
AIR CONDITIONING GROUP 204
AC-1 208
AC-N 210
REFRIGERATION GROUP 206
COOLER-1 212
TOWER-1 214
OUTSIDE EQUIPMENT AND DATA GROUP 108
OUTSIDE EQUIPMENT MONITORING MODULE 216
OUTSIDE HUMIDITY MODULE 218
HUMIDITY SENSOR-1 222
HUMIDITY SENSOR-M 224
OUTSIDE WET BULB TEMPERATURE MODULE 220
WET BULB TEMP SENSOR-1 226
WET BULB TEMP SENSOR-M 228

FIG. 2

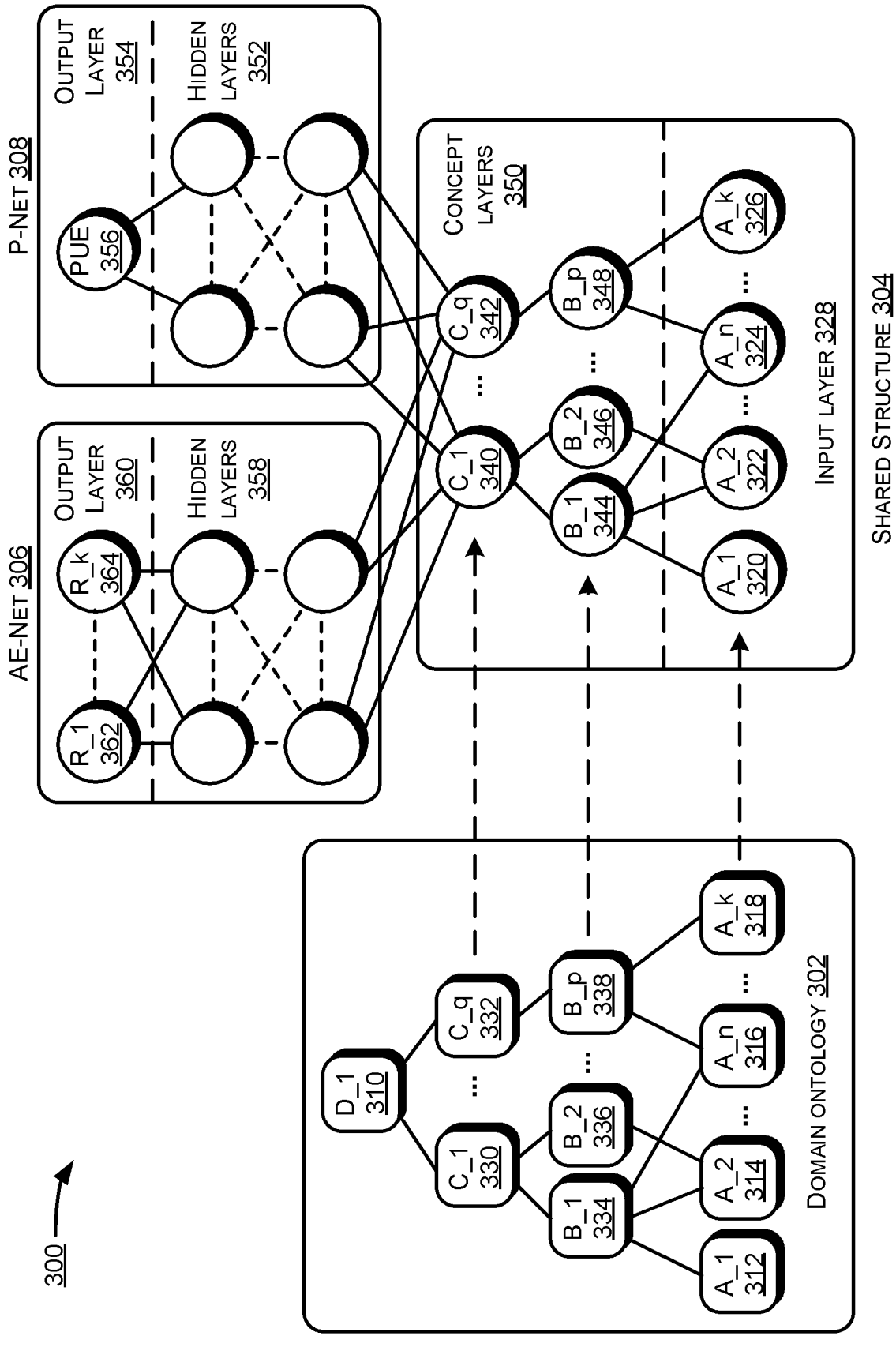
300
P-NET 308
OUTPUT LAYER 354
HIDDEN LAYERS 352
PUE 356
AE-NET 306
OUTPUT LAYER 360
HIDDEN LAYERS 358
R_k 364
R_1 362
CONCEPT LAYERS 350
C_q 342
C_1 340
B_p 348
B_2 346
B_1 344
A_k 326
A_n 324
A_2 322
A_1 320
INPUT LAYER 328
SHARED STRUCTURE 304
D_1 310
C_q 332
C_1 330
B_p 338
B_2 336
B_1 334
A_k 318
A_n 316
A_2 314
A_1 312
DOMAIN ONTOLOGY 302
LEVEL 1
LEVEL 2
LEVEL 3
LEVEL 4

FIG. 3

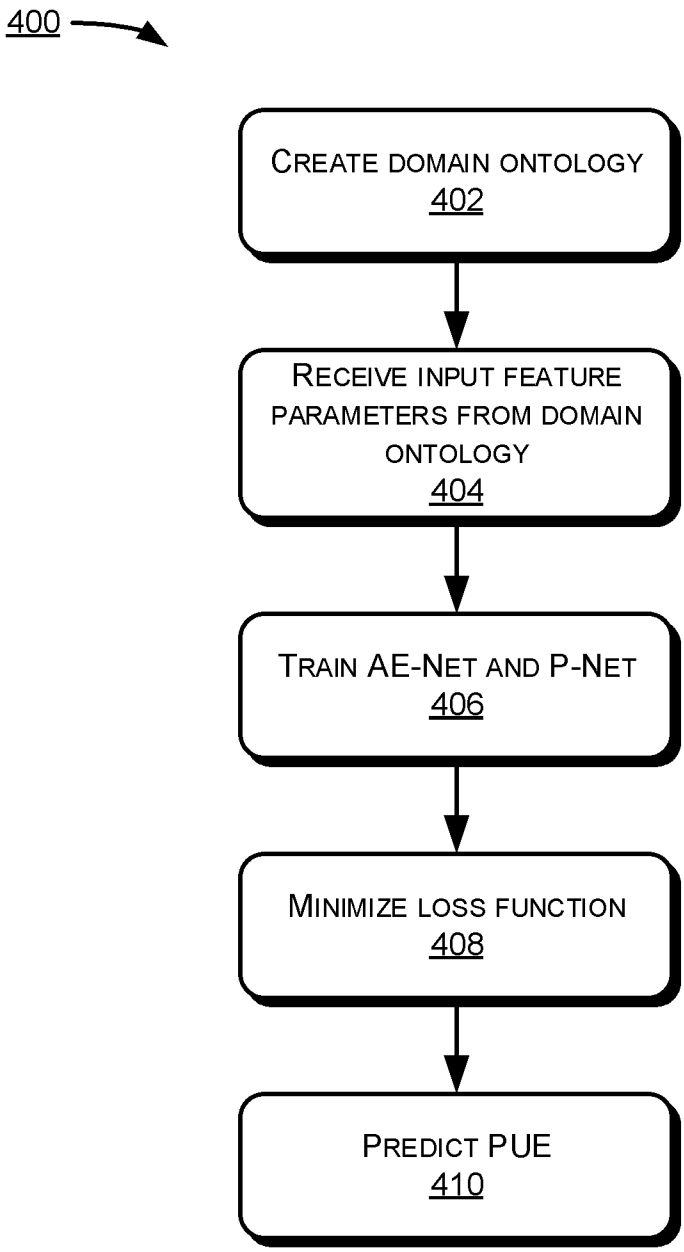
400
CREATE DOMAIN ONTOLOGY
402
RECEIVE INPUT FEATURE PARAMETERS FROM DOMAIN ONTOLOGY
404
TRAIN AE-NET AND P-NET
406
MINIMIZE LOSS FUNCTION
408
PREDICT PUE
410

FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/087083

A. CLASSIFICATION OF SUBJECT MATTER

G06F 1/32(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT; CNKI; WPI; EPODOC; GOOGLE: PUE, HVAC, power w usage w effectiveness, optimiz+, parameter?, neural w network?, HLNN, Pnet, AEnet, DSSM

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2010076607 A1 (AHMED, Osman et al.) 25 March 2010 (2010-03-25) description, paragraphs [0047]-[0055], figures 1, 1A	1-36
A	CN 103645795 A (INSPUR ELECTRONIC INFORMATION INDUSTRY CO., LTD.) 19 March 2014 (2014-03-19) the whole document	1-36
A	US 2009201293 A1 (ACCENTURE GLOBAL SERVICES GMBH) 13 August 2009 (2009-08-13) the whole document	1-36
Y	LIU, Wei et al. "Research of Mutual Learning Neural Network Training Method" *Chinese journal of computers,* Vol. 40, No. 6, 30 June 2017 (2017-06-30), ISSN: 0254-4164, abstract	1-36

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 February 2020	**27 February 2020**
Name and mailing address of the ISA/CN **National Intellectual Property Administration, PRC** **6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088** **China**	Authorized officer **LI,Di**
Facsimile No. **(86-10)62019451**	Telephone No. **86-(10)-53961300**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/087083

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
US 2010076607 A1	25 March 2010	WO 2010017429 A2	11 February 2010
CN 103645795 A	19 March 2014	None	
US 2009201293 A1	13 August 2009	CA 2652513 A1	12 August 2009
		CN 101510114 A	19 August 2009
		EP 2120128 A2	18 November 2009
		IN 200900225 I3	15 October 2010