

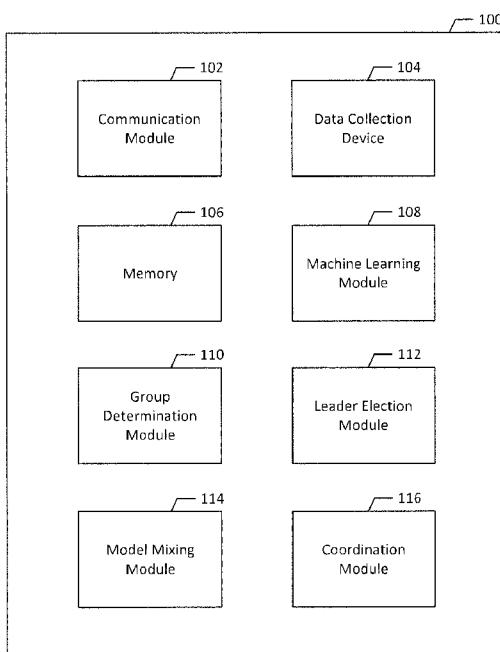
(51) International Patent Classification:
G06F 21/55 (2013.01)(21) International Application Number:
PCT/US2016/014440(22) International Filing Date:
22 January 2016 (22.01.2016)(25) Filing Language:
English(26) Publication Language:
English(30) Priority Data:
14/602,843 22 January 2015 (22.01.2015) US(71) Applicant: **PREFERRED NETWORKS, INC. [JP/JP];**
Hongo-Tokyu Building, 4F, 2-40-1 Hongo, Bunkyo-ku, Tokyo 1130033 (JP).

(72) Inventor; and

(71) Applicant : **CLAYTON, Justin, B. [US/US];** c/o Preferred Networks America, Inc., 4 West 4th Avenue, Penthouse A, San Mateo, CA 94402 (US).(72) Inventors: **OKANOHARA, Daisuke;** c/o Preferred Networks, Inc., Hongo-tokyu Building, 4f, 2-40-1 Hongo, Bunkyo-ku, 1130033 (JP). **NISHIKAWA, Toru;** c/o Preferred Networks, Inc., Hongo-tokyu Building, 4f 2-40-1Hongo, Bunkyo-ku, 1130033 (JP). **HIDO, Shohei;** c/o Preferred Networks, Inc., 4 West 4th Avenue, Penthouse A, San Mateo, CA 94402 (US). **KUBOTA, Nobuyuki;** c/o Preferred Networks, Inc., 4 West 4th Avenue, Penthouse A, San Mateo, CA 94402 (US). **OTA, Nobuyuki;** c/o Preferred Networks, Inc., 4 West 4th Avenue, Penthouse A, San Mateo, CA 94402 (US). **TOKUI, Seiya;** c/o Preferred Networks, Inc., Hongo-tokyu Building, 4f, 2-40-1 Hongo, Bunkyo-ku, 1130033 (JP).(74) Agent: **HUBBARD, Nolan R.;** K&L Gates LLP, P.O. Box 1135, Chicago, IL 80690-1135 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, 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,

[Continued on next page](54) Title: **MACHINE LEARNING WITH MODEL FILTERING AND MODEL MIXING FOR EDGE DEVICES IN A HETEROGENEOUS ENVIRONMENT**

(57) Abstract: Machine learning with model filtering and model mixing for edge devices in a heterogeneous environment is disclosed. In an example embodiment, an edge device includes a communication module, a data collection device, a memory, a machine learning module, and a model mixing module. The edge device analyzes collected data with a model for a first task, outputs a result, and updates the model to create a local model. The edge device communicates with other edge devices in a heterogeneous group, transmits a request for local models to the heterogeneous group, and receives local models from the heterogeneous group. The edge device filters the local models by structure metadata, including second local models, which relate to a second task. The edge device performs a mix operation of the second local models to generate a mixed model which relates to the second task, and transmits the mixed model to the heterogeneous group.

Fig. 1



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))*

MACHINE LEARNING WITH MODEL FILTERING AND MODEL MIXING FOR EDGE DEVICES IN A HETEROGENEOUS ENVIRONMENT

PRIORITY CLAIM

[0001] This application claims priority of U.S. patent application number 14/602,843, filed January 22, 2015, the entire disclosure of which is incorporated herein by reference.

BACKGROUND

[0002] A machine learning device uses a machine learning model to process collected data and output a result. The machine learning device may use feedback to update or improve the machine learning model. For example, a machine learning device may use a regression analysis model to make a prediction, and may then update the regression analysis model based on feedback that indicates whether the prediction was correct. A variety of types of machine learning models exists (e.g., linear regression model, naïve Bayes classifier). Machine learning models are commonly used for addressing “Big Data” problems, where the volume of data to be processed may be great and/or real time data processing may be desired.

SUMMARY

[0003] The present disclosure provides a new and innovative machine learning technique with model filtering and model mixing for edge devices in a heterogeneous environment. In an example embodiment, an edge device includes a communication module configured to communicate with a plurality of different edge devices, a data collection device configured to collect a first type of data, a memory configured to store data collected by the data collection device, a machine learning module, and a model mixing module. The edge device may be configured to analyze, using a first model relating to a first predefined task, first data collected by the data collection device, output a result including at least one of a prediction, a classification, a clustering, an anomaly detection, and a recognition, and update, based on a correctness of the result, the first model to create a first local model which relates to the first predefined task. The edge device may communicate with at least one other edge device in a heterogeneous group of edge devices, where the heterogeneous group of edge devices includes at least a first edge device and a second edge device, and the first edge device collects and analyzes the first type of data and the second edge device collects and

analyzes a different second type of data. The edge device may transmit a request for local models to the heterogeneous group of edge devices and receive a first plurality of local models from the heterogeneous group of edge devices. The edge device may filter the first plurality of local models by structure metadata, where the first plurality of local models includes a second plurality of local models, each of which relates to a second predefined task. The edge device may perform a mix operation of the second plurality of local models to generate a mixed model which relates to the second predefined task, and transmit the mixed model to the heterogeneous group of edge devices.

[0004] Additional features and advantages of the disclosed method and apparatus are described in, and will be apparent from, the following Detailed Description and the Figures.

BRIEF DESCRIPTION OF THE FIGURES

[0005] Fig. 1 is a block diagram of an example of an edge device, according to an example embodiment of the present disclosure.

[0006] Fig. 2 is a block diagram of an example machine learning module, according to an example embodiment of the present disclosure.

[0007] Fig. 3A is high level block diagram illustrating a heterogeneous group of edge devices, according to an example embodiment of the present disclosure.

[0008] Fig. 3B is block diagram illustrating a heterogeneous group of edge devices, according to an example embodiment of the present disclosure.

[0009] Fig. 4 is a flowchart illustrating an example process for machine learning in a heterogeneous group of edge devices, according to an example embodiment of the present disclosure.

[0010] Fig. 5 is a flowchart illustrating an example process for machine learning, according to an example embodiment of the present disclosure.

[0011] Fig. 6 is a flowchart illustrating an example process for generating a mixed model, according to an example embodiment of the present disclosure.

[0012] Fig. 7 is a flow diagram illustrating an example process for machine learning in a heterogeneous group of edge devices, according to an example embodiment of the present disclosure.

DETAILED DESCRIPTION OF EXAMPLE EMBODIMENTS

[0013] Fig. 1 is a block diagram of an example of an edge device 100, according to an example embodiment of the present disclosure. In an example embodiment, the edge device 100 may be a video camera. In another example embodiment, the edge device 100 may be a shopping cart. The edge device 100 is a device that is capable of performing communication with other devices, performing data collection, and performing machine learning. In an example embodiment, an edge device 100 is on the edge, or outermost layer, of a large distributed network of data connected devices, including central servers, intermediate servers, data repositories, gateways, routers, and the like. Edge devices 100 may include a wide variety of devices including recording devices (e.g., digital cameras, video cameras, audio recorders), city management devices (e.g., parking sensors, traffic sensors, water quality devices), vehicles (e.g., cars, trucks, airplanes), body sensors (e.g., activity sensors, vital signs sensor, pedometers), environmental sensors (e.g., weather sensors, pollution sensors, air quality sensors), wearable computing devices (e.g., smart watch, glasses, clothes), personal computing devices (e.g., mobile phone, tablet, laptop), home devices (e.g., appliances, thermostats, light systems, security system), advertising devices (e.g., billboards, information kiosks), etc. The edge device 100 may include a communication module 102, a data collection device 104, a memory 106, a machine learning module 108, a group determination module 110, a leader election module 112, a model mixing module 114, and a coordination module 116.

[0014] The communication module 102 is configured to communicate with other devices including other edge devices 100 of the same type (e.g., multiple video cameras) or of a different type (e.g., a video camera and a shopping cart). For example, as described in further detail below, the communication module 102 may be configured to communicate with other devices via one or more networks or communications channels, including the Internet, or any suitable wide area network, local area network, gateway, or other communication channel or network. For example, the communication module 102 may be configured for wireless communications via multiple protocols employed by cellular networks (e.g., 4G, 3G, GSM), wireless local area network (e.g., Wi-Fi), satellite (e.g., VSAT), or any suitable form of wireless communication (e.g., Bluetooth, RFID, NFC, IrDA, Li-Fi). Also, for example, the communication module 102 may be configured for a wired connection to another edge device 100 (e.g., Ethernet, DSL, USB, RS-232, coaxial cable). Further, the communication module 102 may communicate with a user, for example, via a graphical user interface implemented with a touch screen display. The user may be able to request the edge device

100 to perform a specific task and/or receive information from the edge device 100. Thus, the communication module 102 may include hardware and/or software configured to communicate via one or more communication interfaces using one or more communication protocols.

[0015] A data collection device 104 may be a sensor, detector, or any device suitable for real time collection of data representative of real world characteristics (e.g., speed, acceleration, items in a shopping cart, hand movements, shapes, temperature, angles, voice recognition, word recognition, torque, slip levels). The data collection device 104 may receive a continuous data stream or collect data on a periodic basis (e.g., every millisecond, second, minute), which may generally depend on the type of data being collected and the variability of the data stream. A data collection device 104 typically includes specific hardware and/or physical structures specifically configured to collect a certain type of data (e.g., an image sensor, an accelerometer, a gyroscope sensor, a thermometer, an altimeter, a hall effect sensor, a velocimeter, a photodetector, a flow sensor, a strain gauge, a torque sensor, a tachometer, a clinometer, a microphone, a magnetometer, a voltmeter, an ammeter, an ohmmeter, a chemical sensor, a pressure sensor, a rain sensor, a hygrometer, a humistor, an anemometer, a seismometer, a geiger counter, etc.). In an example embodiment, one edge device 100 may include multiple different data collection devices 104 that collect different types of data. The data collection device 104 provides the collected data to the memory 106. In an example embodiment, the memory 106 may be specialized memory for receiving and storing large amounts of data, such as video image data (e.g., VRAM). Thus, the memory 106 may have specialized hardware which is task specific, for example, to meet high throughput and low latency specifications of the edge device 100. The memory 106 may include different tiers of memory, such as buffers, shift registers, and the like. The memory 106 may be configured to store the collected data temporarily, and may be overwritten once the collected data is no longer needed by the machine learning module 108.

[0016] A machine learning module 108 executes a machine learning model using the data collected by the data collection device 104 and stored in memory 106. The machine learning module 108 receives the collected data as inputs and executes the machine learning model using the collected data to make a prediction, a classification, a clustering, an anomaly detection, and/or a recognition, which is then output as a result. The machine learning model may iteratively update the result. For example, the machine learning model may continuously execute using all available collected data stored in memory, and may produce a continuous result or a periodic result. If the volume and velocity of data collected is

relatively low (sparse and/or slow), for example, the machine learning model may only periodically execute, and may be dormant for a period of time after each result is output, while new data is collected. Each machine learning model relates to a predefined task (e.g., prediction of an item, recognition of sweetheating theft, recognition of a suspect for a be-on-the-lookout (BOLO) alert, classification of cornering speeds). For example, a machine learning model for a shopping cart edge device 100, the predefined task may be suggesting an item to a consumer. As the consumer pushes the shopping cart around, while walking throughout a grocery store and shopping, the machine learning module 108 of the shopping cart may output a prediction on which item the consumer is most likely to purchase if that item is suggested to the consumer. The prediction may be based on other items in the consumer's cart. Further, additional information, such as the consumer's age, sex, race and/or ethnicity may also be used to make the prediction of which item is most likely to be purchased. In an example embodiment, a machine learning module 108 may include several machine learning models, where each machine learning model may be specific to a subset or category (e.g., category of male consumers over 40 years of age), in which case two, four, eight, or sixteen different models for a single predefined task may be selected from, depending on the situation and the number of variables used to specify a machine learning model.

[0017] Fig. 2 is a block diagram of an example machine learning module 108, according to an example embodiment of the present disclosure. The machine learning module 108 may include an interface module 202, a model selection module 204, and one or more machine learning models 206. As described herein, a machine learning model 206 may simply be referred to as a model 206 for brevity. The interface module 202 is an interface between the machine learning module 108 and other modules or components within the edge device 100 (e.g., memory 106, communication module 102). The model selection module 204 receives information including which predefined task is to be analyzed, and may include information for choosing from multiple models 206 to select the appropriate model 206 (e.g., male consumers over 40 years of age) for the predefined task. As shown in Fig. 2, Model 1 relates to a predefined Task A, while Models 2, 3, and 4 relate to a different predefined Task B, and Model 5 relates to another different predefined Task C. In an example embodiment, an edge device 100 may switch tasks periodically. For example, a first predefined task may be handled during business hours, and after hours, a second different predefined task may be handled. A model 206 may be provided to an edge device 100, for example, at the time of manufacture, or may be loaded onto the edge device 100 at a later time from another device,

either as a new model 206 or as an updated version of a model 206. The machine learning model 206 includes an executable component 210, which is executed to process the collected data as inputs to generate results as outputs of the model 206. The machine learning model 206 includes structural metadata 212, such as the type of the model (e.g., linear classifier, linear or logistic regression model, decision tree, clustering model, bayesian model, neural network) and structure of the model (e.g., parametric fixed structure or non-parametric, number and structure of parameters, input variable(s), output variable(s), data type, feature definitions). The structural metadata 212 may include any necessary information for identifying what the model 206 is and what the model 206 does, so that the model 206 can be used for distributed machine learning, as will be discussed in greater detail below. The machine learning model 206 may also include context metadata 214 (e.g., data collection environment, model history, model stability), which may describe the contextual situation in which the model has been or is operating in. In an example embodiment, context metadata may indicate a geographic location of a video camera edge device 100, and may indicate that the video camera is operating in a dark environment or a bright environment, and/or in an indoor environment or an outdoor environment. The machine learning model 206 includes a results log 216, which may store results, for example, for a period of time or a specific number of results. The results log 216 may also include summary statistics or other information, for example, which may be useful for analysis of the model 206. The results log 216 may include a summary of collected data that was previously stored in the memory 106, but has been overwritten by new collected data. The machine learning model 206 includes an update module 218, which determines whether the model 206 should be updated based on feedback on the results output from the model 206. For example, if the model 206 makes predictions that are correct, this does not indicate that the model should be updated, but if the model 206 makes incorrect predictions, this tends to indicate that the model may be improved. Thus, the update module 218 may update the machine learning model 206 based on internal analysis, and may also update based on an outside instruction to update. When a model 206 is updated by the update module 218, the model 206 is then local to the specific edge device 100 that the model 206 resides on, and may thus be referred to as a local model 206, as opposed to a global model 206. A manufacturer may provide new edge devices 100 with global models 206, and then each edge device 100 may update their models 206, resulting in local models 206 that are different from the global model 206. A local model 206 may be trained based on a specific environment to provide better results in that specific environment than the global model 206 provides.

[0018] A group determination module 110 determines which edge devices 100 will form a group that will participate in an analysis of the machine learning models 206 on the edge devices 100 that are members of the group. Group membership may be based on geographic proximity between edge devices 100. A leader election module 112 determines which edge device 100 should act as a leader for a group of edge devices. A leader may be elected by using a consensus protocol which may receive input from all of the members of a group. A model mixing module 114 is used by a leader edge device 100 to mix local models 206 from different edge devices 100 in the group to create a mixed model 206. A model mixing module 114 may mix models 206 using several different techniques depending on the models 206 to be mixed. A coordination module 116 may coordinate the processes of the communication module 102, the data collection device 104, the memory 106, the machine learning module 108, the group determination module 110, the leader election module 112, and the model mixing module 114. An edge device 100 may have a central processing unit, and may also have one or more additional processors dedicated to various specific tasks. Each edge device 100 may use one or more processors, memories, buses, and the like. Also, each one of the communication module 102, the data collection device 104, the memory 106, the machine learning module 108, the group determination module 110, the leader election module 112, the model mixing module 114, and the coordination module 116 may use one or more processors, memories, buses, and the like. A processor, memory, and any other component may be shared by one or more of the communication module 102, the data collection device 104, the memory 106, the machine learning module 108, the group determination module 110, the leader election module 112, the model mixing module 114, and the coordination module 116. In an example embodiment, each one of the communication module 102, the data collection device 104, the memory 106, the machine learning module 108, the group determination module 110, the leader election module 112, the model mixing module 114, and the coordination module 116 uses some dedicated hardware and software that is distinct from each other module or component.

[0019] An edge device 100 as shown in Fig. 2 may interface with the physical world using a sensor, then using the data collected by the sensor, autonomously execute a machine learning model 206, and communicate with other devices regarding the machine learning performed using the machine learning model 206. With regard to the example of a shopping cart, the edge device 100 may be a shopping cart that integrally includes all the components and functionality of an edge device 100, or alternatively, a shopping cart could include an installed edge device 100 which is configured to be secured to a standard pre-existing

shopping cart. Thus, for example, a grocery store could install the edge device 100 onto a regular shopping cart. In another example embodiment, an ordinary automobile may include a traction control system that is adapted to form an edge device 100 by downloading software modules which use existing data collection devices in the automobile. Moreover, a wide variety of different types of devices may be adapted to be edge devices 100 (e.g., a shopping cart, an automobile, a thermostat, a GPS device), and thus, a heterogeneous group of edge devices 100 may be formed when an edge device(s) 100 communicates with another different type of edge device(s) 100. A heterogeneous group of edge devices differs from a homogeneous group of edge devices 100, which includes only edge devices 100 that are the same type of device and perform the same type of data collection and the same type of machine learning.

[0020] Fig. 3A is high level block diagram illustrating a heterogeneous group of edge devices, according to an example embodiment of the present disclosure. The heterogeneous group 300 includes two edge devices 302a and 302b, which are both the same type of edge device 100 (e.g., both video cameras 302). The heterogeneous group 300 also includes three more edge devices 304, 306, and 308, which are each different types of devices 100 (e.g., one thermostat, one automatic teller machine, and one shopping cart). A server 310 may communicate with one or more of the edge devices 302a, 302b, 304, 306, 308 via a network 312. The server 310 may communicate with one or more edge devices 100. In an example embodiment, the server 310 is an intermediate server, which communicates with one or more edge devices 100 and with one or more different servers (e.g., a central server). In an example embodiment, server 310 may be part of a “cloud” which performs cloud computing. The server 310 may provide information and/or commands to edge devices 100, and may receive information, for example, regarding history and/or accuracy of local models 206. For example, the server 310 may provide an updated model 206 based on a manufacturer update. The server 310 may perform many of the same functions as an edge device, but unlike an edge device 100, a server 310 does not perform data collection using a data collection device 104. The network 312 may include one or more networks and/or communication paths, and may be configured as a peer-to-peer network. In an example embodiment, the heterogeneous group 300 may be configured in any type of network 312 (e.g., LAN, WAN, Wi-Fi, BT, Z-wave, satellite, terrestrial, etc.) and may be configured in any suitable network topology (e.g., mesh, bus, grid, ring, star, tree, line).

[0021] Fig. 3B is block diagram illustrating a heterogeneous group of edge devices, according to an example embodiment of the present disclosure. As shown in Fig. 3B, the

heterogeneous group 300 is provided as a semi-connected mesh network of edge devices 302a, 302b, 304, 306, 308a, 308b, 308c that includes four different types of edge devices (e.g., video cameras 302, thermostat 304, automatic teller machine 306, and shopping carts 308). As described herein, when referencing edge devices 100, two different types of edge devices 302, 304 may be indicated by the different reference numerals 302 and 304 (e.g., a video camera and a thermostat), or two of the same type of edge device may be indicated by the same reference numeral 302a and 302b with the letters “a” and “b” indicating multiple different edge devices of the same type are referenced (e.g., two video cameras 302). Also, edge devices 100 may be referred to herein with reference numeral 100 without specifying whether the edge devices 100 are the same or different types of edge devices 100.

[0022] In an example embodiment, multiple different types of network communications (e.g., Wi-Fi, 4G, BT, NFC, Li-Fi, IrDA) may be used within a heterogeneous group 300 so that an edge device 100 which is only capable of one type of communication may be in a heterogeneous group 300 with another edge device 100, which is only capable of a different type of communication, through an edge device 100 acting as an intermediary which is capable of both forms of communication. In an example embodiment, the network 312 may be a wireless mobile ad hoc mesh network (e.g., MANET, VANET, SPAN). In an example embodiment, the network 312 may be a scatternet. As described herein, a heterogeneous group 300 of edge devices 100 may simply be referred to as a group 300 for brevity. A group 300 is often susceptible to variations in the group 300. For example, edge devices 100 in the group 300 may unexpectedly lose communication with the group 300. For example, an edge device 100 may lose power (e.g., unplugged and/or batteries die), may be moved to an area with limited or no connectivity due to interference (e.g., mountains, rain), or may be turned off by a user. Thus, a group 300 is typically operating in heterogeneous environment that is not static, but rather, typically changes dynamically with different edge devices 100 spontaneously and/or unexpectedly entering and leaving the group 300. In an example embodiment, the heterogeneous group 300 may have edge devices 100 including one or more of shopping carts, automobiles, surveillance cameras, automatic teller machines, GPS devices, medical devices, robots, remote controls, smoke detectors, head mounted displays, or any other edge device 100, as discussed in the present application. In an example embodiment, the heterogeneous group 300 includes a plurality of edge devices 100 which are part of the Internet of Things (IoT), which is rapidly growing and already includes billions of devices. As shown in Fig. 3B, edge device 302a has

been designated as a leader edge device 302a, which is a device that functions as a leader for a group 300, which will be discussed in further detail below.

[0023] Fig. 4 is a flowchart illustrating an example process for machine learning in a heterogeneous group of edge devices, according to an example embodiment of the present disclosure. Although the process 400 is described with reference to the flowchart illustrated in Fig. 4, it will be appreciated that many other methods of performing the acts associated with the process 400 may be used. For example, the order of some of the blocks may be changed, certain blocks may be combined with other blocks, and some of the blocks described are optional.

[0024] The example process 400 may begin with execution of a machine learning model (block 402). For example, an edge device 100 may use a machine learning model 206 to analyze image data to make a prediction. In an example embodiment, where an edge device 308 is a shopping cart, the model 206 executed by the shopping cart may relate to the predefined task of suggesting an item to a consumer. In an example embodiment, where an edge device 302 is a video camera, the model 206 executed by the video camera may relate to the predefined task of sweetheating recognition. In an example embodiment, an edge device 100 simultaneously executes more than one machine learning model 206, for example, when two different machine learning models 206 directed to two different tasks are concurrently executing. The two different machine learning models 206 may use the same collected data or may use different data collected from different data collection devices 104 (e.g., image sensor and microphone). Also, some machine learning models 206 may use multiple different types of data collected from different data collection devices 104.

[0025] The example process 400 includes communication between heterogeneous edge devices (block 404). For example, edge devices 100 may send periodic ping messages seeking connections with various different types of edge devices 302, 304, 306, 308 within range. In an example embodiment, a video camera and a shopping cart may communicate when the shopping cart is brought into close proximity with the video camera, for example, in a check out lane when the consumer is paying for goods in the shopping cart. In an example embodiment, the range is set based on a time to live for ping messages or a number of hops through intermediate devices. In an example embodiment, the range may be limited to an effective range dictated by signal power, interference, etc. (e.g., a Bluetooth range of 30 feet based on power class and line of sight without interference). In an example embodiment, the range may be set by geographic delineations. The execution of machine learning models 206 and the communication between heterogeneous edge devices 100 may occur simultaneously,

such that the edge devices 100 may maintain communication with known edge devices 100, and make new connections with newly available edge devices 100, for example, that have been moved within range for communication, while continually executing a machine learning model 206 on an incoming stream of data collected by the data collection device 104.

[0026] The communicating edge devices 100 perform organization of a group 300 (block 406). For example, group membership and a leader of the group 300 are determined. In an example embodiment, edge devices 302a, 302b, 304, 306, 308a, 308b, 308c (e.g., video cameras, a thermostat, an automatic teller machine, and shopping carts) are determined to be members of the group 300, as illustrated in Fig. 3B, and the group 300 may collectively determine a leader edge device 302a. Members of the group 300 may be involved in distributed machine learning within the heterogeneous group 300, and the leader edge device 302a (e.g., a video camera) may lead the distributed machine learning of the group 300. For example, the leader edge device 302a may determine a structure for periodically communicating with the group 300 in order to determine whether models 206 of different edge devices (e.g., 302a and 302b) should be mixed. Also, the group 300 may be in communication with a server 310 (e.g., an intermediate server), which may monitor the group 300 (e.g., record information relating to history and mixing of models 206) and/or provide updated models 206 to the group 300. In an example embodiment, the server 310 may participate in organization of a group 300. Also, a server 310 may monitor many different groups 300 and may receive information on history and accuracy of local models 206. In an example embodiment, group membership and determination of leader edge device 302a may be based on geographic proximity and/or connection signal strength between edge devices 100.

[0027] A mixed model is generated (block 408). For example, a leader 302a requests local models 206 from edge devices 100 and mixes received local models 206. In an example embodiment, the leader 302a which is a video camera mixes two local models 206 from two shopping carts to generate a mixed model 206 relating to suggesting an item to a consumer. The leader 302a may determine which models 206 are capable of being mixed, and also determine whether it is appropriate to mix the models 206 that are capable of being mixed, as described in further detail below.

[0028] A mixed model is distributed to the group (block 410). For example, the leader 302a sends a mixed model 206 to edge devices 100 to update local models 206. The mixed model 206 may then be implemented by edge devices 100 in the group 300 (block 412). For example, an edge device 100 replaces its local model 206 with the mixed model

206. In an example embodiment, shopping carts each replace a local model 206 for suggesting an item to a consumer with a mixed model 206 that is received from the leader 302a. The example process 400 may be an ongoing process that is periodically repeated at a set interval. For example, blocks 402 through 412 may typically occur over a time period of 30 seconds, but the duration of this time period may be shorter or longer, for example, depending upon the velocity and volume of data collected (e.g., 1 second, 5 minutes), the computing power of a leader 302a, speed of communication between edge devices 100, the extent of model mixing required, etc. For example, if a time period is 1 second, the collected data may be highly variable within seconds, and it may be critical that the model 206 is as accurate as possible (e.g., automobile with self-driving features). On the other hand, if a time period is 5 minutes, the collected data may include relatively small variations over the course of minutes, and accuracy of the model 206 may not be critical (e.g., parking availability prediction). The period of time may generally include a first phase, in which machine learning models 206 are executed while communication between heterogeneous edge devices 100 is maintained (blocks 402 and 404), and a second phase, in which a group 300 is organized, models 206 are mixed, distributed, and implemented by edge devices in the group 100 (blocks 406 to 412). The first and second phases may typically overlap, and commencement of the second phase may require synchronization or a time-out period. Typically, the first phase constitutes the majority of the time period. Also, the time period may be determined dynamically, for example, based on the quantity of edge devices 100 in a group 300, computing power of a leader 302a, based on a significant change in data collected which requires a change or update to models 206, when triggered by an event (e.g., every time a consumer checks out, the consumer's shopping cart may initiate the second phase), or the like.

[0029] Fig. 5 is a flowchart illustrating an example process for machine learning, according to an example embodiment of the present disclosure. Although the process 500 is described with reference to the flowchart illustrated in Fig. 5, it will be appreciated that many other methods of performing the acts associated with the process 500 may be used. For example, the order of some of the blocks may be changed, certain blocks may be combined with other blocks, and some of the blocks described are optional.

[0030] The example process 500 may begin with receiving collected data from a data collection device (block 502). For example, a data stream including items in a shopping cart is received from a tag reader. A shopping cart may include multiple data collection devices 104, such as a tag reader (e.g., RFID reader) and a location detection device (e.g., GPS

device). For example, as a shopper moves through a grocery store, the tag reader may be used to identify items placed in the shopping cart. The memory 106 may periodically store data indicating which items are placed in the shopping cart and where the shopping cart is located in the store. For example, the memory 106 may store a current location and a list of items in the shopping cart every 0.5 seconds, thus, creating a log of the shopper's actions throughout the store. In an example embodiment, a store may track 1,000 different items on the shelves, and if one of these items is placed into the shopping cart, a vector is updated to indicate the presence or absence of the item. For example, if items 3 and 5 of the 1,000 items are placed into the shopping cart, the 1,000-term vector is updated to <0, 0, 1, 0, 1, 0, 0, ..., 0>, which may be stored with the position in the store and a timestamp. It should be appreciated that not all items in the store must have a tag (e.g., RFID tag) that is readable by the tag reader in the shopping cart.

[0031] The collected data may be analyzed with a model relating to a predefined task (block 504). For example, the data stream collected by the shopping cart is analyzed to make a prediction of an item, which may be suggested to the shopper. In an example embodiment, the model 206 may predict an item that, if suggested to the shopper, is the most likely item to be purchased from a predefined list of items that can be suggested by the model 206 for a given day. For example, the items suggested to the shopper may be high profit margin items and/or may be items that are overstocked or soon expiring. In an example embodiment, the model 206 may have a list of three items which it may suggest (e.g., a lottery ticket, a pastry tray, or a flower bouquet). The model 206 may analyze the items placed in the cart, dwell times (e.g., more than 10 seconds in one location), movement through the store (e.g., order of aisles and bypassing aisles), the day of the week, time of day, holidays or special events (e.g., the Super Bowl), etc. For example, the collected data is input as input variables of the executable component 210 of the model 206 to analyze dwell times where a consumer stops and stands for a period of time (e.g., shopper is interested in products or is making a decision), the order that items are put in cart, the sections of the store that are bypassed, etc. All of the collected data may be analyzed by the model 206 to make a prediction of which item should be suggested to the shopper. In an example embodiment, the model 206 may include a naïve Bayes classifier, which outputs a probability of purchase for each of the lottery ticket, the pastry tray, and the flower bouquet, and the item with the highest probability of purchase may be the suggested item. As noted above, not all items in the store must have a tag (e.g., RFID tag) that is readable by the tag reader in the shopping cart,

however, such items may not be considered by the model 206 for making the suggestion of an item to the consumer.

[0032] A result is output based on the analysis of the model (block 506). For example, the executable component 210 makes a prediction based on a global model 206 that recommending a lottery ticket will most likely cause a purchase of the lottery ticket. For example, a shopper may put five items in their shopping cart, and then the executable component 210 may determine the probabilities of purchase for the items (e.g., <lottery ticket, 0.847>; <pastry tray, 0.017>; <flower bouquet, 0.136>). Thus, the model 206 may output a result predicting that the lottery ticket should be suggested to the shopper. In an example embodiment, the communication module 102 may suggest the item based on the result output from the model, for example, on a digital display screen. Also, in an example embodiment, a coupon may be provided to the consumer for the suggested item. The result which is output by the model 206 may include a prediction, a classification, a clustering, an anomaly detection, and/or a recognition. As noted above, the memory 106 may store the collected data temporarily, and thus, once the collected data has been analyzed and a result is output, the memory 106 may be overwritten with new collected data. For example, the memory 106 may be overwritten every 50 milliseconds, every 500 milliseconds, every 5 seconds, or every 5 minutes.

[0033] A correctness of the result is determined by the model (block 508). For example, the model 206 uses feedback to determine whether a lottery ticket was purchased. In an example embodiment, the communication module 102 of the shopping cart may receive information from another edge device 100 (e.g., a cash register, a video surveillance camera) indicating which items were purchased, and thus, the model 206 may determine whether the suggested item was purchased or not purchased. It should be appreciated that in some cases, an edge device 100 will be unable to obtain feedback to determine a correctness of the outputted result.

[0034] The model is updated to create a local model relating to the predefined task (block 510). For example, the update module 218 changes parameter values of the global model 206 to create a local model 206. For example, if a lottery ticket was not purchased after the shopping cart suggested that the consumer purchase the lottery ticket, the update module 218 may use this feedback as training data to update parameter values in the model 206. A local model 206 may be very well trained to a specific environment after a relatively short period of time. If the environment changes, then the model 206 would generally need to be updated again, typically several times, to provide optimal results. The environment in

which an edge device 100 operates in may change because the edge device 100 is moved to a different location or because the environment changes over time, or a combination of both.

[0035] It should be appreciated that the shopping cart example discussed above is merely exemplary, and models 206 may relate to any suitable predefined task, and include any suitable type of model 206 for addressing the predefined task. In an example embodiment, a model 206 may be directed to recognition of sweetheating by cashiers in a grocery store. For example, a surveillance video camera located in a check out lane may operate as an edge device 100. The video camera may collect a stream of image data, which may be analyzed by a model 206 to detect movements by a cashier's hands which indicate that the cashier is not scanning the items properly to ring up the consumer. If the model 206 determines that sweetheating is occurring or is likely occurring, a message may be sent to a store manager, who may cause the consumer's bags and receipt to be checked at the store exit. Feedback may be provided to the model 206 based on whether the consumer had items bagged that were not scanned, and the model may be updated based on the feedback. Thus, an updated local model 206 may adapt to a cashier's style of scanning items, which may be different from another model 206 for a different cashier.

[0036] In another example embodiment, a model 206 may relate to a predefined task of determining a target cornering speed for an automobile. For example, an automobile may include self-driving features, such as automatic braking based on a predicted collision, and alerts or alarms for the driver. In an example embodiment, an automobile may be an edge device 100 that includes data collection devices 104 such as a video camera for analyzing upcoming corners (e.g., curves or turns in the road). The data collection devices 104 may provide a data stream that is used to determine a sharpness of a bend, a slope of the road, and a camber of the road, a current speed or velocity, a slip angle, a tire-pavement friction, a weight of automobile, a distribution of weight, a moisture level, a temperature, etc. The model 206 may output a target speed for each upcoming corner, which may be used by the automobile for applying the brakes, alerting the user of a dangerous condition, or the like. The model 206 may be provided with feedback (e.g., tire slip, wheel differential) that is obtained while driving through the corner from data collection devices 104 in the automobile (e.g., a traction control system). As the automobile travels, the update module 218 may use feedback to improve the model 206, which may be based on changing location (e.g., desert, mountains), changing weather (dry, raining, snowing, windy, hot, freezing), changing weight (e.g., picking up or dropping off passengers or cargo), or any other environmental changes.

[0037] Fig. 6 is a flowchart illustrating an example process for generating a mixed model, according to an example embodiment of the present disclosure. Although the process 600 is described with reference to the flowchart illustrated in Fig. 6, it will be appreciated that many other methods of performing the acts associated with the process 600 may be used. For example, the order of some of the blocks may be changed, certain blocks may be combined with other blocks, and some of the blocks described are optional.

[0038] The example process 600 may begin with transmitting a request for models to a heterogeneous group of edge devices (block 602). For example, a leader 302a broadcasts a request for local models 206 to the group 300. In an example embodiment, a request for models 206 is sent on a periodic basis (e.g., every 30 seconds). Also, in an example embodiment, a request for models 206 is sent in response to a trigger event, such as each time a shopping cart is at a check out register. Thus, the example process 600 may be initiated on a regular periodic interval or sporadically. Also, in an example embodiment, a server 310 (e.g., an intermediate server) may transmit a request for models 206 to the edge devices 100 in a group 300.

[0039] Then, models may be received from the group (block 604). For example, the leader 302a receives several different local models 206 relating to different predefined tasks (e.g., sweetheating recognition, item suggestion) from different edge devices 302b, 308a, 308b, 308c in the group 300 (e.g., a video camera, shopping carts). The models 206 may include structural metadata 212, context metadata 214, and a results log 216, which the leader 302a may use for determining whether the received models 206 can be mixed and/or should be mixed.

[0040] The models may be filtered by structural metadata (block 606). For example, the models 206 may be filtered by model type (e.g., naïve Bayes classifier, linear regression model), model structure (e.g., parametric/fixed structure, non-parametric), variables, parameters, data type, etc. For example, different models 206 of the same type and structure may relate to different tasks and have different output variables (e.g., a suggested item, a recognition of sweetheating, a speed in miles per hour). Also, different models 206 of the same type and structure may relate to the same task (e.g., manufacturing equipment malfunction detection), but have different input variables (e.g., temperature, vibration, smoke detection, video image data, audio data). Further, different models 206 with the same types of input variables and output variables may be different types of models 206 and/or have different structures or data types. Filtering by structural metadata 212 may be used determine whether it is desirable to mix models 206, and if so, which mix operations may be employed

to mix the models 206. For example, models 206 may be filtered into sub-groups, where only models 206 within a single sub-group will be mixed. For example, it may be undesirable to mix models 206 which use input variables which are different or incompatible data types (e.g., high definition video and low definition video).

[0041] The models may be filtered by context metadata (block 608). For example, the models 206 may be filtered into sub-groups by data collection environment (e.g., outdoors, indoors), model history (e.g., quantity of results in a time period, correct and incorrect predictions in the time period), and/or model stability (e.g., quantity of updates in a time period, variance of updates in the time period). Filtering models 206 into sub-groups by context metadata 214 may include clustering the models 206 (e.g., K-means, bottom-up, GMM). In an example embodiment, models 206 may be filtered and/or clustered based on whether a quantity of results output (e.g., sweethearing recognitions) by each of the models 206 was significantly lower (e.g., sparse data) over a period of time than other models 206.

[0042] The models may be filtered by data distribution (block 610). For example, the models 206 are filtered by the similarity of the collected data streams. In an example embodiment, two video cameras may be executing models 206 for sweethearing recognition, but the collected data streams may be quite different. For example, one edge device 302b may have a data stream where a cashier is actively scanning items and ringing up shoppers, while a different edge device 302a may receive a data stream of a closed register where there is no cashier currently working. In this case, the models 206 of edge devices 302a and 302b may be filtered by structural metadata 212 and context metadata 214 into one sub-group, but then filtering by data distribution may place these models 206 into different sub-groups. Thus, even if two models 206 have identical structural metadata and relate to the same predefined task, the data streams that the models 206 have been exposed to may be completely unrelated, in which case, mixing these models 206 may not improve the models 206, and may actually degrade the models 206.

[0043] A sub-group that will not be further filtered or clustered may be referred to as a mix group, which includes models 206 that are ready for mixing. There may be cases where it would not be appropriate or necessary for filtering by context metadata and/or clustering the filtered models 206. For example, if there are shopping carts all located at the same store, it may be assumed that the shopping carts are collecting data with similar data distributions, and thus, clustering may be unnecessary. On the other hand, if shopping carts in a group 300 are located at different stores, filtering and/or clustering by context metadata (e.g., store location) may be performed. Also, for example, if based on filtering by metadata

and data distribution provides only two models 206 in a sub-group, clustering those two models 206 into a mix group may not need to be performed.

[0044] An appropriate mix operation for the models is determined (block 612). For example, a mix by enumeration operation is appropriate for two local models 206 in a mix group. A variety of mix operations may be possible for mixing models, and different mix operations may be used based on the quantity of models 206 to be mixed, a level of accuracy required, a computing power of the leader 302a, etc. A model mixing module 114 may be capable of an averaging operation, a genetic algorithm operation (“GA operation”), an enumeration operation, and an ensembling operation. For example, the ensemble operation may be used to mix a model 206 that does not have a parametric structure with another model. If a size of a mix group is small, the enumeration operation which may provide optimal accuracy may be used. If a size of a mix group is large, then the GA operation or the averaging operation may be used. In an example embodiment, with a large size mix group, if the computer power of the leader 302a is high, the GA operation may be used. On the other hand, if the computing power of the leader 302a is low, the averaging operation may be used.

[0045] The mix operation of the models is performed to generate a mixed model (block 614). For example, the leader 302a mixes two local models 206 by enumeration to create a mixed model 206. In another example embodiment, a grocery store has shopping carts that can track ten common items for the predefined task of suggesting an item A, B, or C (e.g., a lottery ticket, a pastry tray, or a flower bouquet) to the shopper. For example, the model 206 used by the shopping carts may include a naïve Bayes classifier, and the store may have several hundred shopping carts executing models 206, which may regularly be updated with mixed models 206 based on, for example, the previous 15 checked out shopping carts. For example, Model X was recently generated from a group of shopping carts consisting of carts 1-15, and Model Y was recently generated from a different group of shopping carts consisting of carts 16-30. The tables below illustrate simplified exemplary data for each cart, including which of the ten tracked items were placed in each shopping cart (1) or not placed in each shopping cart (0), and which of the three suggested items A, B, or C were purchased in conjunction with the other items.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Purchased Item
Cart1	0	0	0	0	0	1	1	1	1	1	A
Cart2	0	1	0	0	0	0	1	1	0	0	A
Cart3	0	1	0	0	1	1	0	1	1	1	A
Cart4	0	1	0	1	0	1	1	1	1	1	A

Cart5	1	0	0	1	0	0	0	0	1	0	A
Cart6	1	1	1	1	1	0	0	0	0	0	B
Cart7	1	0	1	1	1	0	1	0	0	0	B
Cart8	1	1	1	1	0	0	0	1	1	0	B
Cart9	1	1	0	1	0	0	0	0	1	0	B
Cart10	1	0	1	0	1	0	1	0	1	0	B
Cart11	1	0	1	1	1	0	0	0	1	0	C
Cart12	0	0	0	1	1	1	0	0	0	0	C
Cart13	0	0	1	1	1	1	1	0	0	1	C
Cart14	0	1	0	1	1	0	0	0	1	0	C
Cart15	0	1	0	0	0	1	1	1	1	0	C

Table 1 -- Model X

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Purchased Item
Cart16	0	1	0	0	0	0	1	0	0	1	A
Cart17	0	0	0	0	1	0	0	0	0	0	A
Cart18	0	0	0	0	0	1	0	0	0	1	A
Cart19	1	0	0	1	1	0	1	0	0	0	A
Cart20	0	0	1	0	0	0	0	0	0	0	A
Cart21	1	0	1	1	1	1	0	0	0	0	B
Cart22	1	0	1	1	1	0	1	1	0	0	B
Cart23	0	1	1	1	0	1	1	1	1	0	B
Cart24	1	0	0	1	0	0	0	0	1	0	B
Cart25	1	0	1	0	1	0	0	0	0	1	B
Cart26	0	1	1	1	1	1	1	0	0	0	C
Cart27	0	0	1	1	1	0	1	0	0	0	C
Cart28	1	0	1	1	1	0	0	1	0	1	C
Cart29	0	1	0	0	0	0	1	0	1	0	C
Cart30	1	0	0	1	1	1	0	1	1	0	C

Table 2 -- Model Y

[0046] As shown in Tables 1 and 2 above, each model is based on five purchases of item A, five purchases of item B, and five purchases of item C. However, because the data streams upon which Model X and Model Y are based are different, the models 206 may produce different results for the same set of input variables. For example, if a shopper were to place Item 2, Item 6, Item 8, Item 9, and Item 10 into a shopping cart, Model X and Model Y would suggest different items to the shopper. Specifically, for example, Model X may use a naïve Bayes classifier to determine the following probabilities of purchase: <lottery ticket, 0.847>; <pastry tray, 0.017>; <flower bouquet, 0.136>, resulting in a suggestion of item A, the lottery ticket. Model Y, based on using different collected data as shown in Table 2, may

determine the following probabilities of purchase: <lottery ticket, 0.329>; <pastry tray, 0.153>; <flower bouquet, 0.518>, resulting in a suggestion of item C, the flower bouquet. It should be noted that for Model Y, the collected data for item A (e.g., a total of 11 of items 1-10) is relatively sparse compared to the collected data for each of items B and C (e.g., both with a total of 25 of items 1-10). For example, the Model X and the Model Y may be determined as appropriate for mixing using a mix by averaging operation. If the shopper were to place Item 2, Item 6, Item 8, Item 9, and Item 10 into a shopping cart, a mixed model generated from Model X and Model Y would suggest item A, the lottery ticket. Also, a mix by merging operation may be used to mix Model X and Model Y, which similarly would result in suggesting item A, the lottery ticket, as shown below.

	Probability of A	Probability of B	Probability of C
Model X	.847	.017	.136
Model Y	.329	.153	.518
Mix by averaging	.609	.076	.314
Mix by merging	.737	.086	.175

Table 3

[0047] As shown above, Table 3 illustrates exemplary data for how model mixing may affect a result output from a model 206, and illustrates how different mixing operations may produce different results for mixing the same models 206 (e.g., Model X and Model Y).

[0048] Fig. 7 is a flow diagram illustrating an example process for machine learning in a heterogeneous group of edge devices, according to an example embodiment of the present disclosure. Although the process 700 is described with reference to the flow diagram illustrated in Fig. 7, it will be appreciated that many other methods of performing the acts associated with the process 700 may be used. For example, the order of some of the blocks may be changed, certain blocks may be combined with other blocks, and some of the blocks described are optional or may be performed by different devices.

[0049] In the example process 700, data may flow between the edge devices 302a, 302n, 304a, 304n, for example, via the network(s) 312. The edge devices 302a, 302n, 304a, 304n may be any type of edge devices 100. Although the process for four particular edge devices 302a, 302n, 304a, 304n is shown in four separate columns, many more edge devices 100 may be involved in the process 700, and one or more servers 310 may also participate in the process 700. For example, the edge device 302n may represent a second, tenth, or fiftieth of the same type of device as edge device 302a. Likewise, edge device 304n may represent

the third, twentieth, or hundredth of the same type of device as edge device 304a. Typically, not all edge devices 302a, 302n, 304a, 304n will be able to directly communicate with each other, so communications may be routed through intermediate edge devices (e.g., 302n, 304a), and/or through other devices (e.g., a router, a gateway, a server 310). Also, the process 700 illustrates a first phase 750 and a second phase 760. In the first phase 750, edge devices 302a, 302n, 304a, 304n execute models 206 and make connections and maintain communication with the other edge devices 302a, 302n, 304a, 304n, and in the second phase 760, group organization and model mixing occurs.

[0050] The example process 700 may begin with edge device 302a executing a model (block 702a). Likewise, edge device 302n is executing a model (block 702b), edge device 304a is executing a model (block 702c), and edge device 304n is executing a model (block 702d). Any or all of the executing models may be local models 206. As described above, executing models 206 may analyze collected data, which results in outputting a result. The first phase 750 may have a duration of 45 seconds, for example, during which edge device 302a may output 90 results (e.g., 2 results per second), however, different edge devices (e.g., 304a) may output more results or less results, for example, depending on the model 206 and depending upon the data stream collected by the edge device (e.g., 304a). The edge device 302a may send ping messages 704a to edge device 302n, initially to make a connection, and then to maintain the connection. Additionally, ping messages 704b and 704c are transmitted between edge devices 302n, 304a, 304n, for example. It should be appreciated that only exemplary communications are illustrated in the process 700, and that many more communications may occur, including direction communication of ping messages from edge device 302a to 304n, for example.

[0051] An edge device 302n has a trigger event occur (block 706). For example, a trigger event could be a shopping cart in a check out lane after the sale of items in the shopping cart is completed. Also, for example, a trigger event may be a time-based event (e.g., 45 seconds after the first phase 750 begins), and may be triggered by a timer expiring. In an example embodiment, an edge device 302n may be configured to have multiple different trigger events.

[0052] Once an edge device 302n has a trigger event occur, that edge device 302n may enter the second phase 760 by initiating group organization with the other edge devices 302a, 304a, 304n, and determining group membership and a leader (block 708). Edge devices 302a, 304a, 304n may also determine group membership and the leader (blocks 710a, 710b, 710c). For example, the edge device 302n may transmit organization information

712a, 712b, and 712c to the other edge devices 302a, 304a, 304n, including a request to establish group membership and to perform a consensus protocol to elect a leader using the group determination module 110. For example, all of the edge devices 302a, 302n, 304a, 304n may determine that they should have group membership. However, for example, one or more edge devices may decline group membership, for example, if a poor wireless connection is likely to fail or cause problems in the second phase 760. Further, each of the edge devices 302a, 302n, 304a, 304n may determine a new leader 302a for the current second phase 760. For example, in a prior second phase 760, the edge device 304n may have acted as leader, but the edge device 304n may have limited processing capability at the moment, so a different leader is elected. As discussed above, determining group membership and determining a leader may both be performed in many different ways, and the particular determination methods may be dynamically employed by the edge devices 302a, 302n, 304a, 304n.

[0053] Once a leader 302a is determined, a request for models 716 may be broadcast by the leader (block 714). The request for models 716 may be routed from edge device 302n to edge device 304a and edge device 304n. Once all edge devices 302n, 304a, 304n receive the request for models 716, the edge devices 302n, 304a, 304n provide their local models 720a, 720b, 720c to the leader 302a (blocks 718a, 718b, 718c). Once the leader 302a has the local models 720a, 720b, 720c, the leader 302a will begin the model mixing process using the model mixing module 114.

[0054] The leader 302a filters and mixes the local models 720a, 720b, 720c (block 722). For example, the leader 302a may perform filtering by structural metadata, filtering by context metadata, and filtering by data distribution to create mix groups. Also, the leader 302a may filter its own local models 206 with one or more of the received local models 720a, 720b, 720c. Then, the local models 720a, 720b, 720c are then mixed using an appropriate mixing operation. For example, local models 720b and 720c are mixed, and the leader's 302a own local model 206 and a local model 720c are mixed. The leader 302a distributes the mixed models 726a, 726b, 726c to the edge devices 302n, 304a, 304n (block 724).

[0055] The edge device 302n implements the mixed models 726a (block 728a). For example, the mixed models 726a may relate to one task (e.g., models for different segments of the population) or may relate to multiple different tasks (e.g., sweetheating recognition or BOLO alert). As discussed above, an edge device 302n may switch tasks periodically (e.g., time interval or on command). For example, the edge device 302n may switch between

sweethearts recognition (e.g., when a cashier is scanning items) and a BOLO alert (e.g., when a check out lane is empty).

[0056] The edge device 304a implements the mixed models 726b (block 728b). Also, the edge device 304n implements the mixed models 726c (block 728c). Further, the leader 302a may implement a mixed model 206 if one of the leader's own local model's 206 was mixed with a local model 720c from another edge device 302n. Also, it should be appreciated that the edge devices 302a, 304a, 304n may not receive any mixed models 726a, 726b, 726c depending on the other models 720a, 720b, 720c received by the leader 302a, and the filtering process, as there may not be any suitable models for mixing. The second phase 760 may have a duration of 15 seconds, for example, after which the first phase 750 begins again.

[0057] Previously existing machine learning methods were generally useful for handling a data stream from a homogeneous environment. However, we recognize that the volume and velocity of available data in heterogeneous environments, where variation of environment is common, cannot be adequately handled using the existing methods. Attempting to address this problem with better and faster computer and/or networking hardware proves inadequate, as even the most technologically advanced computer hardware, software, and networking capabilities are not suitable for processing the volume of data at the velocity we require, while being able to handle variations in a heterogeneous environment (e.g., different types of edge devices 100 with changing connectivity). In other words, the existing machine learning techniques are generally insufficient for handling machine learning in a heterogeneous environment. On the other hand, machine learning in a heterogeneous environment with the edge devices 100 as proposed herein takes a different technological approach which was not possible using previously existing methods and systems. Accordingly, edge devices 100 and the heterogeneous environment are improved by using the methods and system as described herein.

[0058] It will be appreciated that all of the disclosed methods and procedures described herein can be implemented using one or more computer programs, modules, or components. These modules or components may be provided as a series of computer instructions on any conventional computer readable medium or machine readable medium, including volatile or non-volatile memory, such as RAM, ROM, flash memory, magnetic or optical disks, optical memory, or other storage media. The instructions may be provided as software or firmware, and/or may be implemented in whole or in part in hardware components such as ASICs, FPGAs, DSPs or any other similar devices. The instructions

may be configured to be executed by one or more processors, which when executing the series of computer instructions, performs or facilitates the performance of all or part of the disclosed methods and procedures.

[0059] Aspects of the subject matter described herein may be useful alone or in combination with one or more other aspects described herein. Without limiting the following description, in an exemplary aspect of the present disclosure, an edge device comprises: a communication module configured to communicate with a plurality of different edge devices; a data collection device configured to collect a first type of data; a memory configured to store data collected by the data collection device; a machine learning module; a model mixing module, where the edge device is configured to: analyze, using a first model relating to a first predefined task, first data collected by the data collection device; output a result including at least one of a prediction, a classification, a clustering, an anomaly detection, and a recognition; update, based on a correctness of the result, the first model to create a first local model which relates to the first predefined task; communicate with at least one other edge device in a heterogeneous group of edge devices, where the heterogeneous group of edge devices includes at least a first edge device and a second edge device, and the first edge device collects and analyzes the first type of data and the second edge device collects and analyzes a different second type of data; transmit a request for local models to the heterogeneous group of edge devices; receive a first plurality of local models from the heterogeneous group of edge devices; filter the first plurality of local models by structure metadata, where the first plurality of local models includes a second plurality of local models, each of which relates to a second predefined task; perform a mix operation of the second plurality of local models to generate a mixed model which relates to the second predefined task; and transmit the mixed model to the heterogeneous group of edge devices. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with the above aspect, the first type of data is video image data and the second type of data is velocity data. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the edge device is further configured to filter the plurality of local models by context metadata. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the edge device is further configured to filter the plurality of local models by data distribution. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the edge device is further configured to cluster

the plurality of local models. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the mix operation is determined using a size of a group of the plurality of local models and a computer power of the edge device. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the mix operation is an averaging operation. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the mix operation is a genetic algorithm operation. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the mix operation is an enumeration operation. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the mix operation is an ensemble operation. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the first predefined task is the same task as the second predefined task. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the edge device is further configured to replace the first local model with the mixed model. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the first plurality of local models includes a third plurality of local models which is different from the second plurality of local models, and the edge device is further configured to perform a mix operation of the third plurality of local models to generate a second mixed model which relates to a third predefined task which is different from the second predefined task. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the edge device is one of a shopping cart device and a surveillance camera. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the edge device is incorporated in an automobile. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the edge device is an automatic teller machine.

[0060] In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, a computer readable medium stores instructions which, when executed by an edge device which comprises a communication module configured to communicate with a plurality of different

edge devices, a data collection device configured to collect a first type of data, a memory configured to store data collected by the data collection device, a machine learning module, and a model mixing module, cause the edge device to: analyze, using a first model relating to a first predefined task, first data collected by the data collection device; output a result including at least one of a prediction, a classification, a clustering, an anomaly detection, and a recognition; update, based on a correctness of the result, the first model to create a first local model which relates to the first predefined task; communicate with at least one other edge device in a heterogeneous group of edge devices, where the heterogeneous group of edge devices includes at least a first edge device and a second edge device, and the first edge device collects and analyzes the first type of data and the second edge device collects and analyzes a different second type of data; transmit a request for local models to the heterogeneous group of edge devices; receive a first plurality of local models from the heterogeneous group of edge devices; filter the first plurality of local models by structure metadata, where the first plurality of local models includes a second plurality of local models, each of which relates to a second predefined task; perform a mix operation of the second plurality of local models to generate a mixed model which relates to the second predefined task; and transmit the mixed model to the heterogeneous group of edge devices.

[0061] In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, a method executed by an edge device comprises: communicating, by a communication module in an edge device, with a plurality of different edge devices; collecting, with a data collection device, a first type of data; storing, in a memory, data collected by the data collection device; analyzing, using a first model relating to a first predefined task, first data collected by the data collection device; outputting a result including at least one of a prediction, a classification, a clustering, an anomaly detection, and a recognition; updating, based on a correctness of the result, the first model to create a first local model which relates to the first predefined task; communicating with at least one other edge device in a heterogeneous group of edge devices, where the heterogeneous group of edge devices includes at least a first edge device and a second edge device, and the first edge device collects and analyzes the first type of data and the second edge device collects and analyzes a different second type of data; transmitting a request for local models to the heterogeneous group of edge devices; receiving a first plurality of local models from the heterogeneous group of edge devices; filtering the first plurality of local models by structure metadata, where the first plurality of local models includes a second plurality of local models, each of which relates to a second predefined task;

performing a mix operation of the second plurality of local models to generate a mixed model which relates to the second predefined task; and transmitting the mixed model to the heterogeneous group of edge devices. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the method further comprises filtering the plurality of local models by context metadata. In accordance with another exemplary aspect of the present disclosure, which may be used in combination with any one or more of the preceding aspects, the method further comprises filtering the plurality of local models by data distribution.

[0062] It should be understood that various changes and modifications to the example embodiments described herein will be apparent to those skilled in the art. Such changes and modifications can be made without departing from the spirit and scope of the present subject matter and without diminishing its intended advantages. It is therefore intended that such changes and modifications be covered by the appended claims.

CLAIMS

The invention is claimed as follows:

1. An edge device comprising:
 - a communication module configured to communicate with a plurality of different edge devices;
 - a data collection device configured to collect a first type of data;
 - a memory configured to store data collected by the data collection device;
 - a machine learning module;
 - a model mixing module, wherein the edge device is configured to:
 - analyze, using a first model relating to a first predefined task, first data collected by the data collection device;
 - output a result including at least one of a prediction, a classification, a clustering, an anomaly detection, and a recognition;
 - update, based on a correctness of the result, the first model to create a first local model which relates to the first predefined task;
 - communicate with at least one other edge device in a heterogeneous group of edge devices, wherein the heterogeneous group of edge devices includes at least a first edge device and a second edge device, and the first edge device collects and analyzes the first type of data and the second edge device collects and analyzes a different second type of data;
 - transmit a request for local models to the heterogeneous group of edge devices;
 - receive a first plurality of local models from the heterogeneous group of edge devices;
 - filter the first plurality of local models by structure metadata, wherein the first plurality of local models includes a second plurality of local models, each of which relates to a second predefined task;
 - perform a mix operation of the second plurality of local models to generate a mixed model which relates to the second predefined task; and
 - transmit the mixed model to the heterogeneous group of edge devices.
2. The edge device of Claim 1, wherein the first type of data is video image data and the second type of data is velocity data.

3. The edge device of Claim 1, wherein the edge device is further configured to filter the plurality of local models by context metadata.

4. The edge device of Claim 1, wherein the edge device is further configured to filter the plurality of local models by data distribution.

5. The edge device of Claim 1, wherein the edge device is further configured to cluster the plurality of local models.

6. The edge device of Claim 1, wherein the mix operation is determined using a size of a group of the plurality of local models and a computer power of the edge device.

7. The edge device of Claim 1, wherein the mix operation is an averaging operation.

8. The edge device of Claim 1, wherein the mix operation is a genetic algorithm operation.

9. The edge device of Claim 1, wherein the mix operation is an enumeration operation.

10. The edge device of Claim 1, wherein the mix operation is an ensemble operation.

11. The edge device of Claim 1, wherein the first predefined task is the same task as the second predefined task.

12. The edge device of Claim 1, wherein the edge device is further configured to replace the first local model with the mixed model.

13. The edge device of Claim 1, wherein the first plurality of local models includes a third plurality of local models which is different from the second plurality of local models, and wherein the edge device is further configured to perform a mix operation of the

third plurality of local models to generate a second mixed model which relates to a third predefined task which is different from the second predefined task.

14. The edge device of Claim 1, wherein the edge device is one of a shopping cart device and a surveillance camera.

15. The edge device of Claim 1, wherein the edge device is incorporated in an automobile.

16. The edge device of Claim 1, wherein the edge device is an automatic teller machine.

17. A computer readable medium storing instructions which, when executed by an edge device which comprises a communication module configured to communicate with a plurality of different edge devices, a data collection device configured to collect a first type of data, a memory configured to store data collected by the data collection device, a machine learning module, and a model mixing module, cause the edge device to:

analyze, using a first model relating to a first predefined task, first data collected by the data collection device;

output a result including at least one of a prediction, a classification, a clustering, an anomaly detection, and a recognition;

update, based on a correctness of the result, the first model to create a first local model which relates to the first predefined task;

communicate with at least one other edge device in a heterogeneous group of edge devices, wherein the heterogeneous group of edge devices includes at least a first edge device and a second edge device, and the first edge device collects and analyzes the first type of data and the second edge device collects and analyzes a different second type of data;

transmit a request for local models to the heterogeneous group of edge devices;

receive a first plurality of local models from the heterogeneous group of edge devices;

filter the first plurality of local models by structure metadata, wherein the first plurality of local models includes a second plurality of local models, each of which relates to a second predefined task;

perform a mix operation of the second plurality of local models to generate a mixed model which relates to the second predefined task; and

transmit the mixed model to the heterogeneous group of edge devices.

18. A method executed by an edge device, the method comprising:
 - communicating, by a communication module in an edge device, with a plurality of different edge devices;
 - collecting, with a data collection device, a first type of data;
 - storing, in a memory, data collected by the data collection device;
 - analyzing, using a first model relating to a first predefined task, first data collected by the data collection device;
 - outputting a result including at least one of a prediction, a classification, a clustering, an anomaly detection, and a recognition;
 - updating, based on a correctness of the result, the first model to create a first local model which relates to the first predefined task;
 - communicating with at least one other edge device in a heterogeneous group of edge devices, wherein the heterogeneous group of edge devices includes at least a first edge device and a second edge device, and the first edge device collects and analyzes the first type of data and the second edge device collects and analyzes a different second type of data;
 - transmitting a request for local models to the heterogeneous group of edge devices;
 - receiving a first plurality of local models from the heterogeneous group of edge devices;
 - filtering the first plurality of local models by structure metadata, wherein the first plurality of local models includes a second plurality of local models, each of which relates to a second predefined task;
 - performing a mix operation of the second plurality of local models to generate a mixed model which relates to the second predefined task; and
 - transmitting the mixed model to the heterogeneous group of edge devices.

19. The method of Claim 18, further comprising filtering the plurality of local models by context metadata.

20. The method of Claim 18, further comprising filtering the plurality of local models by data distribution.

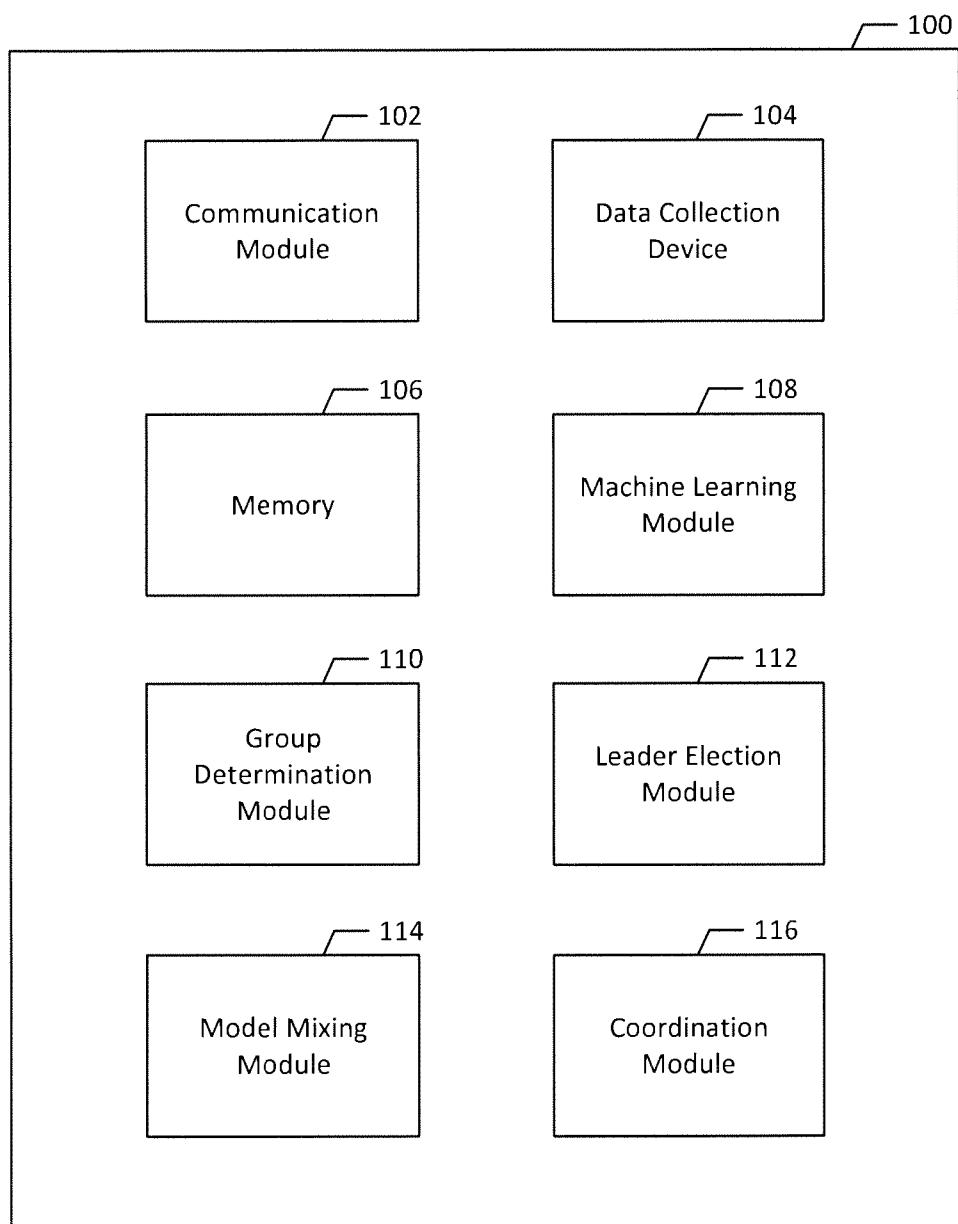


Fig. 1

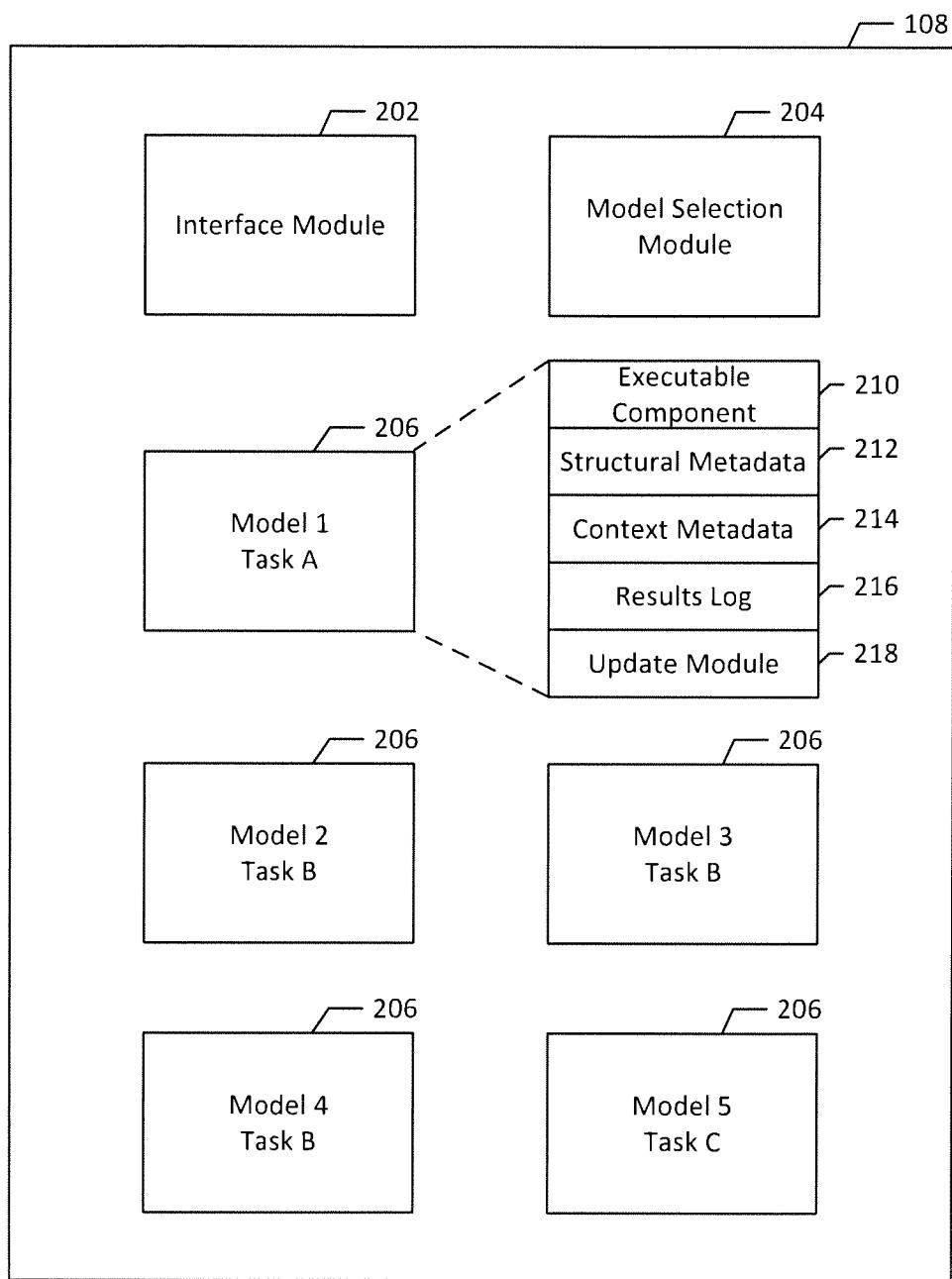


Fig. 2

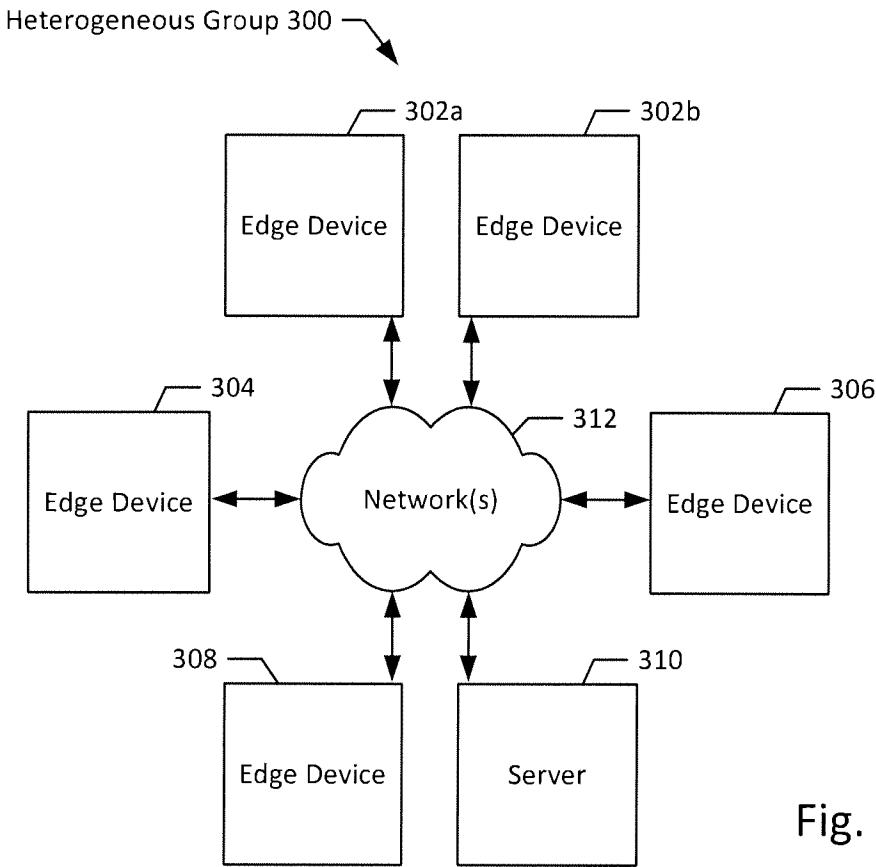


Fig. 3A

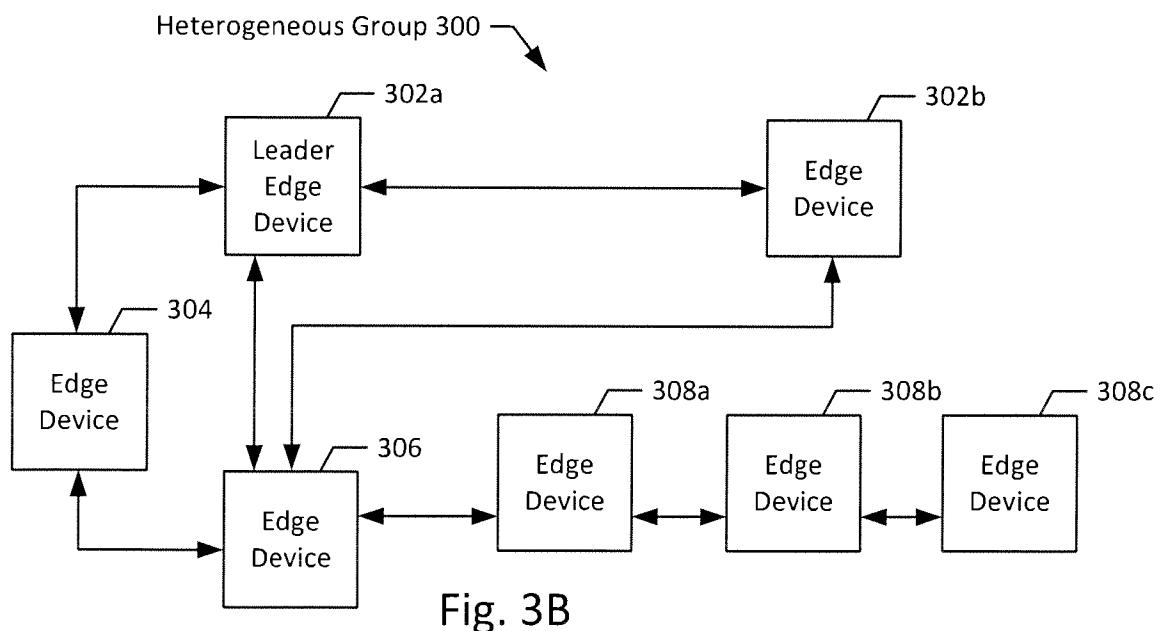


Fig. 3B

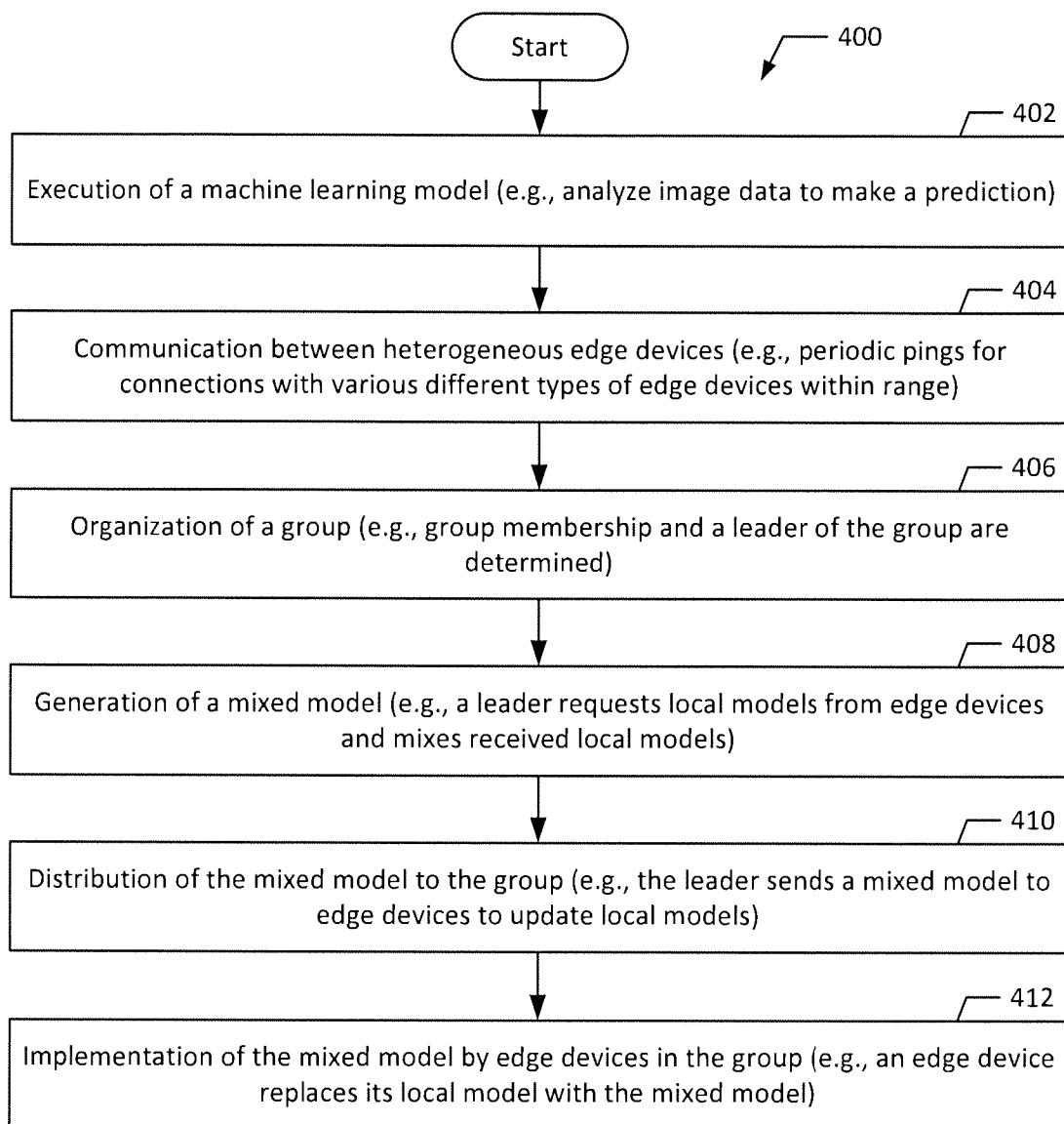


Fig. 4

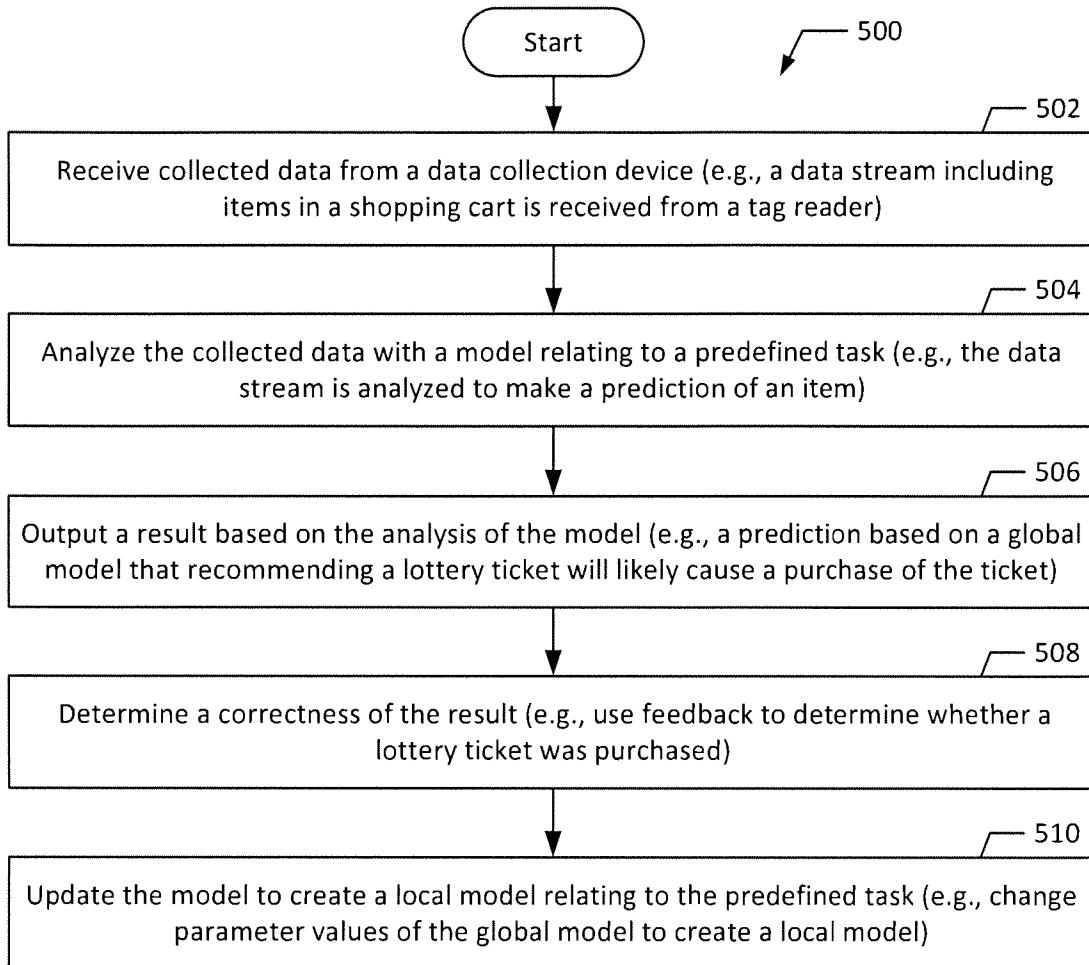


Fig. 5

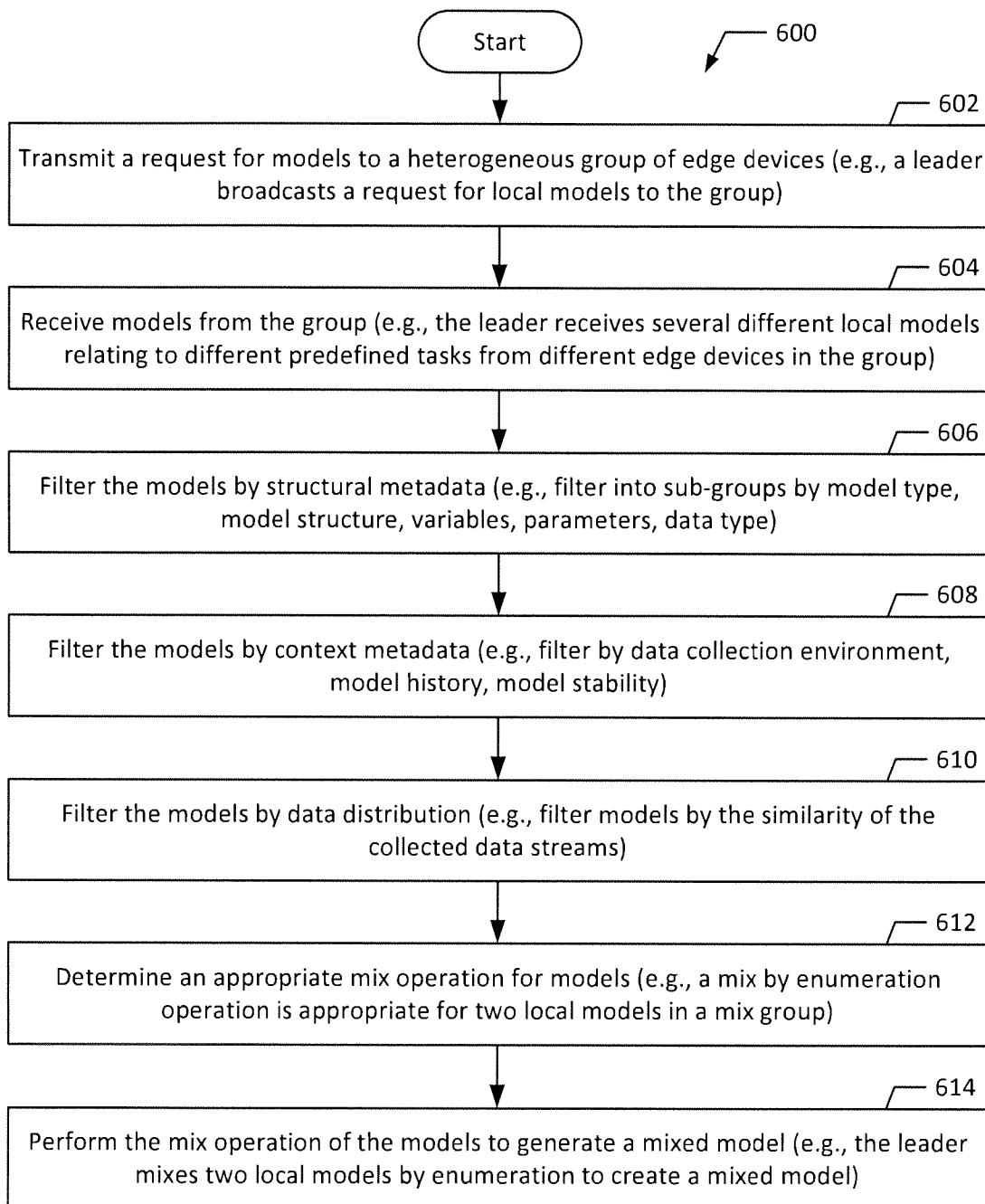


Fig. 6

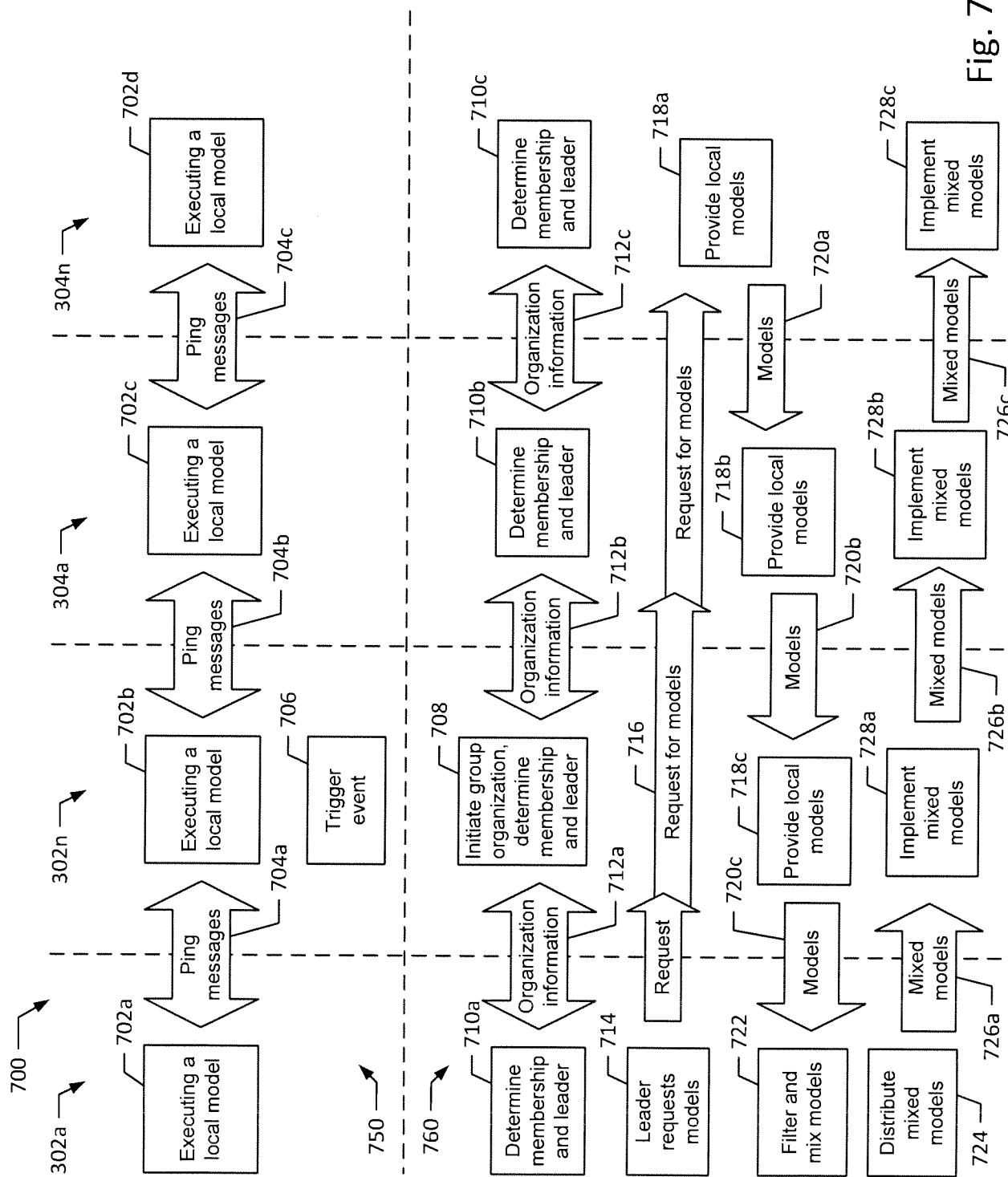


Fig. 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/014440

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 21/55 (2016.01)

CPC - G06F 21/00 (2016.01)

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)

IPC(8) - G06F 21/55; G06K 9/00; G06N 5/00 (2016.01)

CPC - G06F 21/00, 21/50, 21/552; G06K 9/209, 9/32; H04N 7/181 (2016.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 382/181; 706/12; 717/104 (keyword delimited)

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

Orbit, Google Patents, Google, Google Scholar.

Search terms used: models, mixed models, tasks, machine learning, edge devices.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2009/0220153 A1 (HALL et al) 03 September 2009 (03.09.2009) entire document	1-20
Y	US 2014/0096249 A1 (CATAPHORA, INC) 03 April 2014 (03.04.2014) entire document	1-20
Y	US 2011/0282828 A1 (PRECUP et al) 17 November 2011 (17.11.2011) entire document	6
Y	US 2003/0083936 A (MUELLER et al) 01 May 2003 (01.05.2003) entire document	8
A	US 2010/042563 A1 (LIVINGSTON et al) 18 February 2010 (18.02.2010) entire document	1-20

 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

18 March 2016

Date of mailing of the international search report

24 MAR 2016

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
 P.O. Box 1450, Alexandria, VA 22313-1450
 Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774