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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 December 2024 (19.12.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/255036 A1

(51) International Patent Classification:
H04W 24/02 (2009.01)

(21) International Application Number:
PCT/CN2023/124985

(22) International Filing Date:
17 October 2023 (17.10.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/507,854 13 June 2023 (13.06.2023) US

(71) Applicant: **HUAWEI TECHNOLOGIES CO., LTD.** [CN/CN]; Huawei Administration Building, Bantian, Longgang, Shenzhen, Guangdong 518129 (CN).

(72) Inventors: **TANG, Hao**; Huawei Administration Building, Bantian, Longgang, Shenzhen, Guangdong 518129 (CN). **GE, Yiqun**; Huawei Administration Building, Bantian, Longgang, Shenzhen, Guangdong 518129 (CN). **MA, Jianglei**; Huawei Administration Building, Bantian, Longgang, Shenzhen, Guangdong 518129 (CN).

(74) Agent: **LONGSUN LEAD IP LTD.**; Room 801-1, Floor 8, Building 3, Block 2, No. 81, Beiqing Road, Haidian District, Beijing 100094 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))*

WO 2024/255036 A1

(54) Title: COMMUNICATION METHOD AND COMMUNICATION APPARATUS

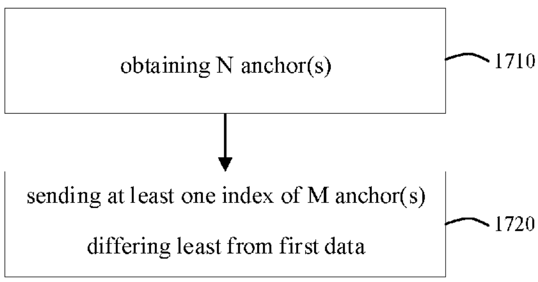

FIG. 17

(57) Abstract: Embodiments of the present application provide a communication method and a communication apparatus. The method includes: obtaining N anchor (s), one of the N anchors comprising one or multiple pieces of reference data, N≥1; and sending at least one index of M anchor (s) differing least from first data, M≤N and M≥1. In this method, a BS or UE can send the index of the M anchor (s) with the smallest difference from the first data to the receiver, and the receiver indicates the BS or the UE to realize the proactive switching of the model or the mode.

COMMUNICATION METHOD AND COMMUNICATION APPARATUS

[0001] The present application is related to, and claims priority to, United States provisional patent application Serial No. 63/507,854, entitled "AI MODEL SWITCH BY DATA ANCHOR", filed on June 13, 2023.

[0002] The disclosures of the aforementioned applications are hereby incorporated by reference in their entirety.

TECHNICAL FIELD

[0003] Embodiments of the present application relate to the field of communications, and more specifically, to a communication method and a communication apparatus.

BACKGROUND

[0004] AI-based algorithms have been introduced into modern wireless communications to solve some wireless problems such as channel estimation, scheduling, channel state information (CSI) compression (from user equipment to base-station), Multiple-in Multiple-Out (MIMO)'s beamforming, positioning and so on. As a data-driven method, AI-based algorithms inevitably suffer from low generalization. Performance of artificial intelligence (AI) models is only as good as the data they are trained on. Even if the AI model is trained on a large number of data sets, it may also not possess the necessary knowledge to perform effectively in other environments, especially in wireless communication where the channel information is changed rapidly.

[0005] Due to generalization issues, a user equipment (UE) or a base station (BS) needs to detect its generalization performance and then switch to a proper AI model. For example, there are multiple AI models for multiple scenarios, including indoor urban, outdoor urban, and rural areas, high speed train, etc. When the surrounding environment is changed, the BS or UE needs to switch to another model. In prior art when the inference performance is getting worse, the UE or BS switches its AI model or falls back to non-AI mode. It is a kind of reactive solution.

[0006] Therefore, how to perform AI model switching in advance is an urgent technical problem to be solved.

SUMMARY

[0007] Embodiments of the present application provide a communication method and a communication apparatus. In the technical solutions of the present application, the BS or UE can proactively switch models or modes after a change in the surrounding environment.

[0008] According to a first aspect, an embodiment of the present application provides a communication method including: obtaining N anchor(s), one of the N anchors including one or multiple pieces of reference data, $N \geq 1$; and sending at least one index of M anchor(s) differing least from first data, $M \leq N$ and $M \geq 1$.

[0009] In the communication method provided by the present application, the BS or UE can send the index of the M anchor(s) with the smallest difference from the first data to the receiver, and the receiver indicates the BS or the UE to realize the proactive switching of the model or the mode.

[0010] The first data includes monitoring data or measured data of the user equipment or the network device. Further, the first data is the monitoring data or measured data related to the AI models. The network device in this embodiment may be a base station (BS). If the first data is the data sent by the UE to the BS over the uplink, the data is the monitoring or measured data of the UE. If the first data is the data sent by the BS to the UE over the downlink, the data is the monitoring or measured data of the BS.

[0011] The M anchor(s) are the M anchor(s) among the N anchor(s) with the smallest difference from the first data. Each of the N anchor(s) corresponds to a configured AI model. M and N are integers greater than or equal to 1.

[0012] Exemplarily, the UE can report to the BS the index of the anchor(s) with the smallest difference from the first data, e.g. 1^{st}, 2^{nd}, M^{th} smallest. M can be configured by the BS. Exemplarily, the BS can indicate to the UE the index of the anchor(s) with the smallest difference from the first data, e.g. 1^{st}, 2^{nd}, M^{th} smallest. M can be reported by the UE.

[0013] The anchor among the N anchors with the smallest difference from the first data is the nearest anchor to the BS or UE. The UE or BS reports the nearest anchor index that can be periodical, semi-persistent, or aperiodic. The reporting can be performed on a physical uplink control channel (PUCCH) or a physical uplink shared channel (PUSCH).

[0014] In one possible implementation scenario, if the BS assists the UE in model switching, N anchor(s) can be configured by the BS to the UE. In one possible implementation scenario, if the UE assists the BS in model switching, N anchor(s) can be reported by the UE to the BS. A configuration signal or a reporting signal may be radio resource control (RRC), medium access control-control element (MAC-CE), or downlink control information (DCI), and may be broadcast, multicast, or unicast.

[0015] According to the UE's report, BS can assist the UE in model switching. When there are multiple AI/ML models at the UE side, the BS assists the UE in model switching or switching to non-AI mode. When there is only one AI model at the UE side, the BS assists the UE in switching between AI and non-AI modes.

[0016] In a possible implementation, M first difference value(s) corresponding to the M anchor(s) are the smallest M of N first difference value(s), and the n^{th} first difference value among the N first difference value(s) is a difference value between the first data and the n^{th} anchor among the N anchor(s), and $1 \le n \le N$.

[0017] In the communication method provided by the present application, the BS or the UE can send the index of the M anchors with the smallest difference from the first data to the receiver, and the receiver indicates the BS or the UE to realize the proactive switching of the model or the mode.

[0018] In a possible implementation, the first difference value corresponding to the n^{th} anchor is a minimum value or an average value of K second difference value(s) corresponding to the n^{th} anchor, and the n^{th} anchor includes K pieces of reference data, and the j^{th} second difference value among the K second difference value(s) is a difference value between the first data and the j^{th} reference data among the K pieces of reference data, $K \ge 1$, $1 \le j \le K$ and $1 \le n \le N$.

[0019] The n^{th} anchor is a set of reference data, e.g. reference coefficients ($\hat{c}$). A reference data ($\hat{c}$) can be a vector, e.g. a one-dimensional array, where the size of the vector is r, r is pre-defined or configured. The size of the set is K ($\hat{c}_j$, j=1, 2, … K), where K is pre-defined or configured. The n^{th} anchor includes K reference coefficients.

[0020] The second difference value(s) can be calculated by either of the equations: $d_{user,j} = \sum_i^r \left\| \hat{c}_{user}(i) - \hat{c}_j(i) \right\|^2$, $d_{user,j} = \left\langle \hat{c}_{user}, \hat{c}_j \right\rangle$, and $d_{user,j} = f\left(\hat{c}_{user}, \hat{c}_j\right)$. $d_{user,j}$ is the difference between the first data and the reference data $\hat{c}_j$ in the n^{th} anchor. $\hat{c}_{user}$ is the first data. $\hat{c}_j$ is the j^{th} reference data in the n^{th} anchor. < > represents the inner product. $\| \ \|$ represents a norm, and the norm is a way to measure the size of a vector, a matrix, a tensor, or a function. f represents other custom functions, $1 \le j \le K$ and $1 \le i \le r$. The difference(s) between the first data and the reference data can also be calculated by a dot product, Euclidean distance, or DNN-based algorithm, etc. The specific calculation should not be construed as a limitation of this application.

[0021] The first difference value can be calculated by either of the equations: $d_{user,anchor} = \min_{j=1,\dots,K}(d_{user,j})$ and $d_{user,anchor} = Avg_{j=1,\dots,K}(d_{user,j})$. $d_{user,anchor}$ is the difference between the first data and the n^{th} anchor. $d_{user,anchor}$ can be the minimum value of the difference between the first data $\hat{c}_{user}$ and the K reference data $\hat{c}$ in the n^{th} anchor, or it can be the average value of the difference between the first data $\hat{c}_{user}$ and the K reference data $\hat{c}$ in the n^{th} anchor.

[0022] Alternatively, the difference between the first data and the n^{th} anchor can also be obtained by mutual information, Hilbert-Schmidt independence criterion (HSIC) metric, Kullback-Leibler (KL) scatter, graphical edit distance, Wasserstein distance, Jensen-Shannon divergence (JSD) distance, DNN-based algorithms, etc.

[0023] In the communication method provided by the present application, the BS or the UE can send the index of the M anchors with the smallest difference from the first data to the receiver, and the receiver indicates the BS or the UE to realize the proactive switching of the model or the mode.

[0024] In a possible implementation, the method further including: sending a third difference value, the third difference value being a difference value between the first anchor and the first data, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[0025] The optimal AI/ML model may not work when the difference between first data and the nearest anchor is greater than a threshold, so the UE can report the third difference value to the BS. According to the UE's report, BS can indicate the UE to switch to another model or fallback to non-AI mode. Similarly, the BS can report the third difference value to the UE. According to the BS's report, UE can indicate the BS to switch to another model or fallback to non-AI mode. In embodiments of the present application, the BS is allowed to provide additional assistance information to the UE for proactive switching.

[0026] In a possible implementation, the method further including: sending a first message or a second message, the first message being configured to indicate that the anchor with the smallest difference value from the first data has not changed during a time period, and the second message being configured to indicate that the anchor with the smallest difference value from the first data has changed during a time period.

[0027] If the nearest anchor index is not changed as compared to the previous value, the UE report "same as previous one" to BS, e.g. using 1 bit to indicate whether it is changed, value 1 means changed and value 0 means the same. If the nearest anchor index is changed, the UE can also report the nearest anchor index to the BS. Another reporting scheme is event-triggered reporting. The UE reports the nearest anchor index to the BS only when the nearest anchor index changes. The process of sending the first message and the second message from the BS to the UE is the same and will not be repeated in this application.

[0028] In the communication method provided by the present application, the air interface overhead is reduced by reporting the first message or the second message.

[0029] In a possible implementation, the sending at least one index of M anchor(s) differing least from first data includes: sending at least one index of M anchor(s) differing least from the first data when it is determined that the anchor with the smallest difference value from the first data has changed during the time period.

[0030] In the communication method provided in the present application, the index value(s) of the report M anchor(s) is an event-triggered report, which reduces the air interface overhead.

[0031] In a possible implementation, the method further including: receiving a third message; and switching to another

AI model or non-AI mode based on the third message.

[0032] When the UE sends the index(s) of the M anchor(s) with the smallest difference from the first data to the BS, the BS can send a third message to the UE to instruct the UE to switch to another AI model or non-AI mode. When the BS sends the index(s) of the M anchor(s) with the smallest difference from the first data to the UE, the UE can send a third message to the BS to instruct the BS to switch to another AI model or non-AI mode.

[0033] In the communication method provided by the present application, the BS or the UE can send the index of the M anchors with the smallest difference from the first data to the receiver, and the receiver indicates the BS or the UE to realize the proactive switching of the model or the mode.

[0034] In a possible implementation, the switching to another AI model or non-AI mode based on the third message includes: switching to a first AI model corresponding to a first anchor based on the third message, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[0035] In a possible application scenario, the BS sends the third message to the UE instructing the UE to perform the model switching. the UE receives the third message and switches the model to the first AI model corresponding to the anchor with the smallest difference from the first data.

[0036] In the communication method provided by the present application, the BS or UE may send a third message instructing the receiver to perform the model switching, enabling proactive model switching.

[0037] In a possible implementation, where the third message includes an index of a second AI model, and the switching to another AI model or non-AI mode based on the third message includes: switching to the second AI model based on the index of the second AI model.

[0038] In a possible application scenario, the BS may instruct the UE to switch to a specified AI model. The third message includes the index of the second AI model, and the UE receives the third message and switches the model to the second AI model based on the third message.

[0039] In the communication method provided by the present application, the BS or UE may send a third message instructing the receiver to perform the model switching, enabling proactive model switching.

[0040] In a possible implementation, the method further including: sending a fourth message, the fourth message being configured to indicate that a third difference value is greater than a predetermined threshold, the third difference value being a difference value between the first data and a first anchor, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[0041] In a possible application scenario, instead of sending the index of the anchor to the BS, the UE may send a fourth message to the BS to indicate that the fourth difference value is greater than the predetermined threshold, if the fourth difference value is greater than the predetermined threshold.

[0042] In the communication method provided by the present application, the UE or the BS may send a fourth message to indicate that the difference value between the first data and the nearest anchor is greater than a predetermined threshold to switch to the non-AI mode.

[0043] In a possible implementation, the fourth message includes an invalid anchor index.

[0044] Exemplarily, a valid anchor index = 0~N-1, and the index value of the anchor in the fourth message sent by the UE is N, indicating that none of the anchors has a difference value from the first data that is less than a predetermined threshold.

[0045] In a possible implementation, the fourth message includes information configured to indicate that a sender of the fourth message has switched to a non-AI mode.

[0046] Exemplarily, a specified bit of the fourth message can be used to indicate whether the UE has switched to the non-AI mode. If this specified bit is 1 it means that it has switched to a non-AI mode.

[0047] In a possible implementation, a value of M is predefined or configured.

[0048] In a possible implementation, one of the N anchors corresponds to one AI model and index.

[0049] The association between an anchor and an AI model can be determined implicitly or configured explicitly. Exemplarily, the association between the anchor and the AI/ML model is implicitly determined by making the anchor index have the same value as the model index, i.e. {model index k, anchor index k}. Exemplarily, the association between the anchor and the AI model is explicitly configured, i.e. {model index k, anchor index j}.

[0050] In the communication method provided by the present application, the BS or the UE can send the index of the M anchors with the smallest difference from the first data to the receiver, and the receiver indicates the BS or the UE to realize the proactive switching of the model or the mode.

[0051] In a possible implementation, the first data includes monitoring data or measured data of a user equipment or a network device.

[0052] In a possible implementation, the first data includes filtered measured data.

[0053] The first data can be raw measured data, or the measured quantities (measured data) are filtered by layer 3 filter. For the layer 3 filtering, the UE filters the measured quantities according to equation $F_n = (1 - a) \times F_{n-1} + a \times M_n$ for each measured quantity before it is used to evaluate the reporting criteria or the measurement report. For measurement object for 5th generation new radio（MeasObjectNR）, $a = \frac{1}{2^{(ki/4)}}$, where ki is the filterCoefficient for the corresponding measurement quantity of the ith QuantityConfigNR in quantityConfigNR-List, and i is indicated by QuantityConfigIndex. For other measurements, $a = \frac{1}{2^{(k/4)}}$, where k is the filterCoefficient for the corresponding measurement quantity received by the quantityConfig.

[0054] In a possible implementation, the index(es) of the M anchor(s) are sent through a physical uplink control channel (PUCCH) or a physical uplink shared channel (PUSCH).

[0055] In a possible implementation, the N anchors are configured by a radio resource control (RRC), a medium access control-control element (MAC-CE) or a downlink control information (DCI) signal.

[0056] In a possible implementation, the method is executed by a user equipment or a network device.

[0057] The above description describes the method provided by the present application in terms of the UE sending information to the BS, and the process of the BS sending information to the UE is similar to the above process, which will not be repeated in the present application.

[0058] According to a second aspect, an embodiment of the present application provides a communication method including: receiving at least one index of M anchor(s), the M anchor(s) being M anchor(s) among N anchor(s) having the smallest difference from first data, one of the M anchor(s) including one or more reference data, $N \geq 1$, $M \leq N$ and $M \geq 1$; and sending a third message, the third message being configured to indicate that the receiver switches the AI model or mode.

[0059] In a possible implementation, M first difference value(s) corresponding to the M anchor(s) are the smallest M of N first difference value(s), and the n^{th} first difference value among the N first difference value(s) is a difference value between the first data and the n^{th} anchor among the N anchor(s), and $1 \leq n \leq N$.

[0060] In a possible implementation, the first difference value corresponding to the n^{th} anchor is a minimum value or an average value of K second difference value(s) corresponding to the n^{th} anchor, and the n^{th} anchor includes K pieces of reference data, and the j^{th} second difference value among the K second difference value(s) is a difference value between the first data and the j^{th} reference data among the K pieces of reference data, $K \geq 1$, $1 \leq j \leq K$ and $1 \leq n \leq N$.

[0061] In a possible implementation, the communication method further includes: receiving a third difference value, the third difference value being a difference value between the first anchor and the first data, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[0062] In a possible implementation, the communication method further includes: receiving a first message or a second message, the first message being configured to indicate that the anchor with the smallest difference value from the first data has not changed during a time period, and the second message being configured to indicate that the anchor with the smallest difference value from the first data has changed during a time period.

[0063] In a possible implementation, the third message is configured to indicate a receiver to switch to a first AI model corresponding to a first anchor, and the first anchor is the anchor among the M anchors having the smallest difference from the first data.

[0064] In a possible implementation, the third message includes an index of a second AI model, and the third message is configured to indicate a receiver to switch to the second AI model.

[0065] In a possible implementation, the communication method further includes: receiving a fourth message, the fourth message being configured to indicate that a third difference value is greater than a predetermined threshold, the third difference value being a difference value between the first data and a first anchor, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[0066] In a possible implementation, the fourth message includes an invalid anchor index.

[0067] In a possible implementation, the fourth message includes information configured to indicate that a sender of the fourth message has switched to a non-AI mode.

[0068] In a possible implementation, a value of M is predefined or configured.

[0069] In a possible implementation, one of the N anchors corresponds to one AI model and index.

[0070] In a possible implementation, the first data includes monitoring data or measured data of a user equipment or a network device.

[0071] In a possible implementation, the first data includes filtered measured data.

[0072] In a possible implementation, the index(es) of the M anchor(s) are sent through a physical uplink control channel (PUCCH) or a physical uplink shared channel (PUSCH).

[0073] In a possible implementation, the N anchors are configured by a radio resource control (RRC), a medium access control-control element (MAC-CE) or a downlink control information (DCI) signal.

[0074] In a possible implementation, the method is executed by a user equipment or a network device.

[0075] For the beneficial effects of the second aspect, reference is made to the first aspect. Details are not described herein again.

[0076] According to a third aspect, this application provides a communication apparatus, including: an obtaining module configured to obtain N anchor(s), one of the N anchors including one or multiple pieces of reference data, $N \geq 1$; and a sending module configured to send at least one index of M anchor(s) differing least from first data, $M \leq N$ and $M \geq 1$.

[0077] In a possible implementation, M first difference value(s) corresponding to the M anchor(s) are the smallest M of N first difference value(s), and the n^{th} first difference value among the N first difference value(s) is a difference value between the first data and the n^{th} anchor among the N anchor(s), and $1 \leq n \leq N$.

[0078] In a possible implementation, the first difference value corresponding to the n^{th} anchor is a minimum value or an average value of K second difference value(s) corresponding to the n^{th} anchor, and the n^{th} anchor includes K pieces of reference data, and the j^{th} second difference value among the K second difference value(s) is a difference value between the first data and the j^{th} reference data among the K pieces of reference data, $K \geq 1$, $1 \leq j \leq K$ and $1 \leq n \leq N$.

[0079] In a possible implementation, the sending module is further configured to send a third difference value, the third difference value being a difference value between the first anchor and the first data, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[0080] In a possible implementation, the sending module is further configured to send a first message or a second message, the first message being configured to indicate that the anchor with the smallest difference value from the first data has not changed during a time period, and the second message being configured to indicate that the anchor with the smallest difference value from the first data has changed during a time period.

[0081] In a possible implementation, the sending module is further configured to send at least one index of M anchor(s) differing least from the first data when it is determined that the anchor with the smallest difference value from the first data has changed during the time period.

[0082] In a possible implementation, the obtaining module is further configured to receive a third message; and a processing module configured to switch to another AI model or non-AI mode based on the third message.

[0083] In a possible implementation, the processing module is further configured to switch to a first AI model corresponding to a first anchor based on the third message, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[0084] In a possible implementation, the third message includes an index of the second AI model, and the processing module is further configured to switch to the second AI model based on the index of the second AI model.

[0085] In a possible implementation, the sending module is further configured to send a fourth message, the fourth message being configured to indicate that a third difference value is greater than a predetermined threshold, the third difference value being a difference value between the first data and a first anchor, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[0086] In a possible implementation, the fourth message includes an invalid anchor index.

[0087] In a possible implementation, the fourth message includes information configured to indicate that a sender of the fourth message has switched to a non-AI mode.

[0088] In a possible implementation, a value of M is predefined or configured.

[0089] In a possible implementation, one of the N anchors corresponds to one AI model and index.

[0090] In a possible implementation, the first data includes monitoring data or measured data of a user equipment or a network device.

[0091] In a possible implementation, the first data includes filtered measured data.

[0092] In a possible implementation, the index(es) of the M anchor(s) are sent through a physical uplink control channel (PUCCH) or a physical uplink shared channel (PUSCH).

[0093] In a possible implementation, the N anchors are configured by a radio resource control (RRC), a medium access control-control element (MAC-CE) or a downlink control information (DCI) signal.

[0094] In a possible implementation, the apparatus is located on a user equipment or a network device.

[0095] According to a fourth aspect, this application provides a communication apparatus including: a receiving module configured to receive at least one index of M anchor(s), the M anchor(s) being M anchor(s) among N anchor(s) having the smallest difference from first data, one of the M anchor(s) including one or more reference data, $N \geq 1$, $M \leq N$ and $M \geq 1$; and a sending module configured to send a third message, the third message being configured to indicate that the receiver switches the AI model or mode.

[0096] In a possible implementation, M first difference value(s) corresponding to the M anchor(s) are the smallest M of N first difference value(s), and the n^{th} first difference value among the N first difference value(s) is a difference value between the first data and the n^{th} anchor among the N anchor(s), and $1 \leq n \leq N$.

[0097] In a possible implementation, the first difference value corresponding to the n^{th} anchor is a minimum value or an average value of K second difference value(s) corresponding to the n^{th} anchor, and the n^{th} anchor includes K pieces of reference data, and the j^{th} second difference value among the K second difference value(s) is a difference value between the first data and the j^{th} reference data among the K pieces of reference data, $K \geq 1$, $1 \leq j \leq K$ and $1 \leq n \leq N$.

[0098] In a possible implementation, the receiving module is further configured to receive a third difference value, the third difference value being a difference value between the first anchor and the first data, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[0099] In a possible implementation, the receiving module is further configured to receive a first message or a second message, the first message being configured to indicate that the anchor with the smallest difference value from the first data has not changed during a time period, and the second message being configured to indicate that the anchor with the smallest difference value from the first data has changed during a time period.

[00100] In a possible implementation, the third message is configured to indicate a receiver to switch to a first AI model corresponding to a first anchor, and the first anchor is the anchor among the M anchors having the smallest difference from the first data.

[00101] In a possible implementation, the third message includes an index of a second AI model, and the third message is configured to indicate a receiver to switch to the second AI model.

[00102] In a possible implementation, the receiving module is further configured to receive a fourth message, the fourth message being configured to indicate that a third difference value is greater than a predetermined threshold, the third difference value being a difference value between the first data and a first anchor, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[00103] In a possible implementation, the fourth message includes an invalid anchor index.

[00104] In a possible implementation, the fourth message includes information configured to indicate that a sender of the fourth message has switched to a non-AI mode.

[00105] In a possible implementation, a value of M is predefined or configured.

[00106] In a possible implementation, one of the N anchors corresponds to one AI model and index.

[00107] In a possible implementation, the first data includes monitoring data or measured data of a user equipment or a network device.

[00108] In a possible implementation, the first data includes filtered measured data.

[00109] In a possible implementation, the index(es) of the M anchor(s) are sent through a physical uplink control channel (PUCCH) or a physical uplink shared channel (PUSCH).

[00110] In a possible implementation, the N anchors are configured by a radio resource control (RRC), a medium access control-control element (MAC-CE) or a downlink control information (DCI) signal.

[00111] In a possible implementation, the apparatus is located on a user equipment or a network device.

[00112] According to a fifth aspect, a communication apparatus including a processor and a memory is provided. The processor is connected to the memory. The memory is configured to store instructions, and the processor is configured to execute the instructions. When the processor executes the instructions stored in the memory, the processor is enabled to perform the method in any possible implementation of the first aspect or the second aspect.

[00113] According to a sixth aspect, this application provides a communication system, which includes communication apparatus in any possible implementation of the third aspect, as well as communication apparatus in any possible implementation of the fourth aspect.

[00114] According to a seventh aspect, this application provides a computer readable storage medium, which includes instructions. When the instructions run on a processor, the processor is enabled to perform the method in any possible implementation of the first aspect or the second aspect.

[00115] According to an eighth aspect, this application provides a computer program product, which includes computer program code. When the computer program code runs on a computer, the computer is enabled to perform the method in any possible implementation of the first aspect or the second aspect.

[00116] It should be noted that all or a part of the above computer program code can be stored in on a first storage medium. The first storage medium can be packaged together with the processor or separately with the processor.

[00117] According to a ninth aspect, this application provides a chip system, which includes a memory and a processor. The memory is configured to store a computer program, and the processor is configured to invoke the computer program from the memory and run the computer program, so that an electronic device on which the chip system is disposed performs the method in any possible implementation of the first aspect or the second aspect.

DESCRIPTION OF DRAWINGS

[00118] FIG. 1 is a schematic diagram of a communication system according to an embodiment of the present application.

[00119] FIG. 2 is a schematic diagram of a communication system 100 according to an embodiment of the present application.

[00120] FIG. 3 is a schematic diagram of an ED 110 and a base station 170a, 170b and/or 170c according to an embodiment of the present application.

[00121] FIG. 4 is a schematic diagram of units or modules in a device according to an embodiment of the present application.

[00122] FIG. 5 is a schematic diagram of an AI-based communication device.

[00123] FIG. 6 is a schematic diagram of a device 500 receiving reference data samples from a device 600 according to an embodiment of the present application.

[00124] FIG. 7 is a schematic diagram of reference data samples including a plurality of groups according to an embodiment of the present application.

[00125] FIG. 8 is a schematic representation of a DNN-based approximation according to an embodiment of the present application.

[00126] FIG. 9 is a flowchart of a communication method according to an embodiment of the present application.

[00127] FIG. 10 is a flowchart of a communication method according to an embodiment of the present application.

[00128] FIG. 11 is a schematic diagram of projecting a high-dimensional signal to a low-dimensional signal according to an embodiment of the present application.

[00129] FIG. 12 is a flowchart of a communication method according to an embodiment of the present application.

[00130] FIG. 13 is a schematic diagram of a matrix U being determined according to an embodiment of the present application.

[00131] FIG. 14 is a schematic diagram of a first sampling matrix P_1 according to an embodiment of the present application.

[00132] FIG. 15 is a schematic diagram of a sampling matrix compression matrix U according to an embodiment of the present application.

[00133] FIG. 16 is a schematic diagram of a scoring distance on the low spectrum space according to an embodiment of the present application.

[00134] FIG. 17 is a flowchart of an embodiment of a communication method according to an embodiment of the present application.

[00135] FIG. 18 is a schematic diagram of the BS indicating the UE to perform a model switch according to an embodiment of the present application.

[00136] FIG. 19 is a schematic diagram of another BS indicating the UE to perform a model switch according to an embodiment of the present application.

[00137] FIG. 20 is a schematic diagram of another BS indicating the UE to perform a model switch according to an embodiment of the present application.

[00138] FIG. 21 is a schematic diagram of the UE indicating the BS to perform a model switch according to an embodiment of the present application.

[00139] FIG. 22 is a schematic block diagram of a communication apparatus according to an embodiment of the present application.

[00140] FIG. 23 is a schematic block diagram of another communication apparatus according to an embodiment of the present application.

[00141] FIG. 24 is a schematic block diagram of still another communication apparatus according to an embodiment of the present application.

DESCRIPTION OF EMBODIMENTS

[00142] The following describes the technical solutions in the present application with reference to the accompanying drawings.

[00143] The following describes the technical solutions in the present application with reference to the accompanying drawings. Obviously, the described embodiments are part of the embodiments of the present application, and not all of them. Based on the embodiments in the present application, all other embodiments obtained by a person of ordinary skill in the art without making creative labor shall fall within the scope of protection of the present application.

[00144] The present application will present aspects, embodiments, or features around systems that include multiple devices, components, modules, etc. It should be understood and appreciated that the individual systems may include additional devices, components, modules, etc., and/or may not include all of the devices, components, modules, etc, discussed in connection with the accompanying drawings. In addition, combinations of these options may be used.

[00145] In addition, in the embodiments of the present application, the word "exemplarily" and the phrase "as an example" are used to indicate, for example, illustration or description. Any embodiment or design solution described as "exemplarily" in this application should not be construed as being superior to or more advantageous than other embodiments or design solutions. Rather, the use of the word "example" is intended to present the concept in a specific manner.

[00146] The phrases "in some possible embodiments", "in some possible application scenarios", etc., appearing in various places in this description, do not necessarily refer to the same embodiments, but rather mean "one or more, but not all, embodiments" unless otherwise specifically emphasized. Unless otherwise specifically emphasized, the terms "including", "comprising", "having", and variations thereof all mean "including but not limited to".

[00147] In the present application, "at least one" refers to one or more, and "multiple" refers to two or more. "and/or", describing the association of the associated objects, indicates that three relationships can exist. For example, A and/or B can mean A alone, both A and B, and B alone, where A and B can be singular or plural. The character "/" generally indicates that the preceding and following associated objects are in an "or" relationship.

[00148] The application scenarios described in the embodiments of the present application are intended to illustrate the technical solutions of the embodiments of the present application more clearly and do not constitute a limitation to the technical solutions provided by the embodiments of the present application. It is known to those of ordinary skill in the art that the technical solutions provided by the embodiments of the present application are equally applicable to similar technical problems as the system architecture evolves and new application scenarios emerge.

[00149] The technical solutions in embodiments of this application may be applied to various communications systems, such as a Global System for Mobile Communications (GSM), a Code Division Multiple Access (CDMA) system, a Wideband Code Division Multiple Access (WCDMA) system, a general packet radio service (GPRS) system, a Long Term Evolution (LTE) system, an LTE frequency division duplex (FDD) system, an LTE time division duplex (TDD) system, a Universal Mobile Telecommunications System (UMTS), a Worldwide Interoperability for Microwave Access (WiMAX) communications system, a wireless local area network (WLAN), a fifth generation (5G) wireless communications system, a new ratio (NR) wireless communications system, a sixth generation (6G) wireless communications system, or other evolving communications systems.

[00150] In order to better describe the solutions of embodiments in the present application, concepts and terms that may be involved in the present application will be described below.

[00151] (1) Data collection

[00152] Data is a very important component for artificial intelligence (AI)/machine learning (ML) techniques. Data collection is a process of collecting data by the network nodes, management entity, or UE for the purpose of AI/ML model training, data analytics and inference.

[00153] (2) AI/ML model training

[00154] AI/ML model training is a process to train an AI/ML model by learning the input/output relationship in a data driven manner and obtain the trained AI/ML Model for inference.

[00155] (3) AI/ML model inference

[00156] A process of using a trained AI/ML model to produce a set of outputs based on a set of inputs.

[00157] (4) AI/ML model validation

[00158] As a sub-process of training, validation is used to evaluate the quality of an AI/ML model using a dataset different from the one used for model training. Validation can help select model parameters that generalize beyond the dataset used for model training. The model parameter after training can be adjusted further by the validation process.

[00159] (5) AI/ML model testing

[00160] Similar to validation, testing is also a sub-process of training, and it is used to evaluate the performance of a final AI/ML model using a dataset different from the one used for model training and validation. Different from AI/ML model validation, testing does not assume subsequent tuning of the model.

[00161] (6) Online training

[00162] Online training means an AI/ML training process where the model being used for inference is typically continuously trained in (near) real-time with the arrival of new training samples.

[00163] (7) Offline training

[00164] Offline training is an AI/ML training process where the model is trained based on the collected dataset, and where the trained model is later used or delivered for inference.

[00165] (8) AI/ML model delivery/transfer

[00166] AI/ML model delivery/transfer is a generic term referring to the delivery of an AI/ML model from one entity to another entity in any manner. Delivery of an AI/ML model over the air interface includes either parameters of a model structure known at the receiving end or a new model with parameters. Delivery may contain a full model or a partial model.

[00167] (9) Life cycle management (LCM)

[00168] When the AI/ML model is trained and/or inferred at one device, it is necessary to monitor and manage the whole

AI/ML process to guarantee the performance gain obtained by AI/ML technologies. For example, due to the randomness of wireless channels and the mobility of UEs, the propagation environment of wireless signals changes frequently. Nevertheless, it is difficult for an AI/ML model to maintain optimal performance in all scenarios for all the time, and the performance may even deteriorate sharply in some scenarios. Therefore, the lifecycle management (LCM) of AI/ML models is essential for sustainable operation of AI/ML in the NR air-interface. Life cycle management covers the whole procedure of AI/ML technologies applied on one or more nodes. In specific, it includes at least one of the following sub-process: data collection, model training, model identification, model registration, model deployment, model configuration, model inference, model selection, model activation, deactivation, model switching, model fallback, model monitoring, model update, model transfer/delivery, and UE capability report. Model monitoring can be based on inference accuracy, including metrics related to intermediate key performance indicators (KPIs), and it can also be based on system performance, including metrics related to system performance KPIs, e.g., accuracy and relevance, overhead, complexity (computation and memory cost), latency (timeliness of monitoring result, from model failure to action) and power consumption. Moreover, data distribution may shift after deployment due to environmental changes, and thus the model based on input or output data distribution should also be considered.

[00169] (10) Supervised learning

[00170] The goal of supervised learning algorithms is to train a model that maps feature vectors (inputs) to labels (output), based on the training data which includes the example feature-label pairs. The supervised learning can analyze the training data and produce an inferred function, which can be used for mapping the inference data. Supervised learning can be further divided into two types: Classification and Regression. Classification is used when the output of the AI/ML model is categorical i.e., with two or more classes. Regression is used when the output of the AI/ML model is a real or continuous value.

[00171] (11) Unsupervised learning

[00172] In contrast to supervised learning where the AI/ML models learn to map the input to the target output, the unsupervised methods learn concise representations of the input data without the labelled data, which can be used for data exploration or to analyze or generate new data. One typical unsupervised learning is clustering which explores the hidden structure of input data and provides the classification results for the data.

[00173] (12) Reinforcement learning

[00174] Reinforcement learning is used to solve sequential decision-making problems. Reinforcement learning is a process of training the action of an intelligent agent from input (state) and a feedback signal (reward) in an environment. In reinforcement learning, an intelligent agent interacts with an environment by taking an action to maximize the cumulative reward. Whenever the intelligent agent takes one action, the current state in the environment may transfer to the new state, and the new state resulting from the action will bring the associated reward. Then the intelligent agent can take the next action based on the received reward and new state in the environment. During the training phase, the agent interacts with the environment to collect experience. The environments are often mimicked by the simulator since it is expensive to directly interact with the real system. In the inference phase, the agent can use the optimal decision-making rule learned from the training phase to achieve the maximal accumulated reward.

[00175] (13) Federated learning

[00176] Federated learning (FL) is a machine learning technique that is used to train an AI/ML model by a central node (e.g., server) and a plurality of decentralized edge nodes (e.g., UEs, next Generation NodeBs, "gNBs"). According to the wireless FL technique, a server may provide, to an edge node, a set of model parameters (e.g., weights, biases, gradients) that describe a global AI/ML model. The edge node may initialize a local AI/ML model with the received global AI/ML model parameters. The edge node may then train the local AI/ML model using local data samples to, thereby, produce a trained local AI/ML model. The edge node may then provide, to the serve, a set of AI/ML model parameters that describe the local AI/ML model. Upon receiving, from a plurality of edge nodes, a plurality of sets of AI/ML model parameters that describe respective local AI/ML models at the plurality of edge nodes, the server may aggregate the local AI/ML model parameters reported from the plurality of UEs and, based on such aggregation, update the global AI/ML model. A subsequent iteration progresses much like the first iteration. The server may transmit the aggregated global model to a plurality of edge nodes. The above procedure is performed multiple iterations until the global AI/ML model is considered to be finalized, e.g., the AI/ML model is converged or the training stopping conditions are satisfied. Notably, the wireless FL technique does not involve the exchange of local data samples. Indeed, the local data samples remain at respective edge nodes.

[00177] AI-based algorithms have been introduced into modern wireless communications to solve some wireless problems such as channel estimation, scheduling, channel state information (CSI) compression (from user equipment to base station), Multiple-in Multiple-Out (MIMO)'s beamforming, positioning, and so on. AI algorithm is a data-driven method that tunes some predefined architectures by a set of data samples called as training data set. The recent AI trains DNN (including CNN, RNN, transformer, etc.) architecture by setting the neurons with a SGD algorithm.

[00178] AI techniques (including ML techniques) in communication include AI-based communications in the physical layer and/or AI-based communications in the MAC layer. For the physical layer, the AI communication may aim to optimize component design and/or improve algorithm performance. For the MAC layer, the AI/ML based communication may aim to utilize the AI/ML capability for learning, prediction, and/or making a decision to solve a complicated optimization problem with possible better strategy and/or optimal solution, e.g. to optimize the functionality in the MAC layer, e.g. intelligent TRP management, intelligent beam management, intelligent channel resource allocation, intelligent power control, intelligent spectrum utilization, intelligent modulation and coding scheme (MCS), intelligent hybrid automatic repeat request (HARQ) strategy, intelligent transmit/receive (Tx/Rx) mode adaption, etc.

[00179] AI architecture may involve multiple nodes, where the multiple nodes may be organized in one of two modes, i.e., centralized and distributed, both of which may be deployed in an access network, a core network, or an edge computing system, or a third party network. A centralized training and computing architecture is restricted by possibly large communication overhead and strict user data privacy. A distributed training and computing architecture may include several frameworks, e.g., distributed machine learning and federated learning. In some embodiments, an AI architecture may include an intelligent controller which can perform as a single agent or a multi-agent, based on joint optimization or individual optimization. New protocols and signaling mechanisms are desired so that the corresponding interface link can be personalized with customized parameters to meet particular requirements while minimizing signaling overhead and maximizing the whole system spectrum efficiency by personalized AI technologies.

[00180] New protocols and signaling mechanisms are provided for operating within and switching between different modes of operation, including between AI and non-AI modes, and for measurement and feedback to accommodate the different possible measurements and information that may need to be fed back, depending upon the implementation.

[00181] It is now quite common for neural network models to become larger and deeper, which may easily require more computational resources than just one or two computers. Most neural network models would be trained on a powerful computation cloud. A user with a desired neural network architecture, raw training data set, and training goal may not have sufficient local computation resources to train their model locally. In order to access a powerful computation cloud, the user would have to transmit all the specifications of its neural network architecture, its training data set, and its training goal to the network cloud completely. It is mandated that the user must trust the cloud and grant the cloud full authorization to manipulate its intellectual property (neural network architecture, training data set, and training goal).

[00182] As data-driven method, AI-based algorithms inevitably suffer from low generalization: if a testing data sample were an outlier to the training data set, a neural network wouldn't make a good inference on the test data sample. Even if the AI model is trained on a large number of data sets, it may also not possess the necessary knowledge to perform effectively in other environments, especially in wireless communication where the channel information is changed rapidly.

[00183] In the present application, the AI model is exemplified by a DNN, i.e., a deep neural network or network. The specific AI model should not be construed as a limitation of the present application.

[00184] FIG. 1 is a schematic diagram of a communication system according to an embodiment of the present application.

[00185] Referring to FIG.1, as an illustrative example without limitation, a simplified schematic illustration of a communication system is provided. The communication system 100 includes a radio access network 120. The radio access network 120 may be a next generation (e.g. sixth generation (6G) or later) radio access network, or a legacy (e.g. 5G, 4G, 3G or 2G) radio access network. One or more communication electric devices (EDs) 110a-120j (generically referred to as 110) may be interconnected to one another or connected to one or more network nodes (170a, 170b, generically referred to as 170) in the radio access network 120. A core network 130 may be a part of the communication system and may be dependent or independent of the radio access technology used in the communication system 100. Also, the communication system 100 includes a public switched telephone network (PSTN) 140, the internet 150, and other networks 160.

[00186] FIG. 2 is a schematic diagram of a communication system 100 according to an embodiment of the present application.

[00187] FIG. 2 illustrates an example communication system 100. In general, the communication system 100 enables multiple wireless or wired elements to communicate data and other content. The purpose of the communication system 100 may be to provide content, such as voice, data, video, and/or text, via broadcast, multicast and unicast, etc. The communication system 100 may operate by sharing resources, such as carrier spectrum bandwidth, between its constituent elements. The communication system 100 may include a terrestrial communication system and/or a non-terrestrial communication system. The communication system 100 may provide a wide range of communication services and applications (such as earth monitoring, remote sensing, passive sensing and positioning, navigation and tracking, autonomous delivery and mobility, etc.). The communication system 100 may provide a high degree of availability and robustness through a joint operation of the terrestrial communication system and the non-terrestrial communication system. For example, integrating a non-terrestrial communication system (or components thereof) into a terrestrial communication system can result in what may be considered a heterogeneous network including multiple layers. Compared to conventional communication networks, the heterogeneous network may achieve better overall performance through efficient multi-link joint operation, more flexible functionality sharing, and faster physical layer link switching between terrestrial networks and non-terrestrial networks.

[00188] The terrestrial communication system and the non-terrestrial communication system can be regarded as sub-systems of the communication system. In the example shown, the communication system 100 includes electronic devices (EDs) 110a-110d (generically referred to as ED 110), radio access networks (RANs) 120a-120b, non-terrestrial communication network 120c, a core network 130, a public switched telephone network (PSTN) 140, the internet 150, and other networks 160. The RANs 120a-120b include respective base stations (BSs) 170a-170b, which may be generically referred to as terrestrial transmit and receive points (T-TRPs) 170a-170b. The non-terrestrial communication network 120c includes an access node 120c, which may be generically referred to as a non-terrestrial transmit and receive point (NT-TRP) 172.

[00189] Any ED 110 may be alternatively or additionally configured to interface, access, or communicate with any other T-TRP 170a-170b and NT-TRP 172, the internet 150, the core network 130, the PSTN 140, the other networks 160, or any combination of the preceding. In some examples, ED 110a may communicate an uplink and/or downlink transmission over an interface 190a with T-TRP 170a. In some examples, the EDs 110a, 110b and 110d may also communicate directly with one another via one or more sidelink air interfaces 190b. In some examples, ED 110d may communicate an uplink and/or downlink transmission over an interface 190c with NT-TRP 172.

[00190] The air interfaces 190a and 190b may use similar communication technology, such as any suitable radio access technology. For example, the communication system 100 may implement one or more channel access methods, such as code division multiple access (CDMA), time division multiple access (TDMA), frequency division multiple access (FDMA), orthogonal FDMA (OFDMA), or single-carrier FDMA (SC-FDMA) in the air interfaces 190a and 190b. The air interfaces 190a and 190b may utilize other higher dimension signal spaces, which may involve a combination of orthogonal and/or non-orthogonal dimensions. The air interface 190c can enable communication between the ED 110d and one or multiple NT-TRPs 172 via a wireless link or simply a link. For some examples, the link is a dedicated connection for unicast transmission, a connection for broadcast transmission, or a connection between a group of EDs and one or multiple NT-TRPs for multicast transmission.

[00191] The RANs 120a and 120b are in communication with the core network 130 to provide the EDs 110a 110b, and 110c with various services such as voice, data, and other services. The RANs 120a and 120b and/or the core network 130 may be in direct or indirect communication with one or more other RANs (not shown), which may or may not be directly served by core network 130, and may or may not employ the same radio access technology as RAN 120a, RAN 120b or both. The core network 130 may also serve as a gateway access between (i) the RANs 120a and 120b or EDs 110a 110b, and 110c or both, and (ii) other networks (such as the PSTN 140, the internet 150, and the other networks 160). In addition, some or all of the EDs 110a 110b, and 110c may include functionality for communicating with different wireless networks over different wireless links using different wireless technologies and/or protocols. Instead of wireless communication (or in addition thereto), the EDs 110a 110b, and 110c may communicate via wired communication channels to a service provider or switch (not shown), and to the internet 150. PSTN 140 may include circuit switched telephone networks for providing plain old telephone service (POTS). Internet 150 may include a network of computers and subnets (intranets) or both, and incorporate protocols, such as internet protocol (IP), transmission control protocol (TCP), and user datagram protocol (UDP). EDs 110a 110b, and 110c may be multimode devices capable of operation according to multiple radio access technologies, and incorporate multiple transceivers necessary to support such.

[00192] FIG. 3 is a schematic diagram of an ED 110 and a base station 170a, 170b and/or 170c according to an embodiment of the present application.

[00193] FIG. 3 illustrates another example of an ED 110 and a base station 170a, 170b and/or 170c. The ED 110 is used to connect persons, objects, machines, etc. The ED 110 may be widely used in various scenarios, for example, cellular communications, device-to-device (D2D), vehicle to everything (V2X), peer-to-peer (P2P), machine-to-machine (M2M), machine-type communications (MTC), internet of things (IoT), virtual reality (VR), augmented reality (AR), industrial control, self-driving, remote medical, smart grid, smart furniture, smart office, smart wearable, smart transportation, smart city, drones, robots, remote sensing, passive sensing, positioning, navigation and tracking, autonomous delivery and mobility, etc.

[00194] Each ED 110 represents any suitable end user device for wireless operation and may include such devices (or may be referred to) as a user equipment/device (UE), a wireless transmit/receive unit (WTRU), a mobile station, a fixed or mobile subscriber unit, a cellular telephone, a station (STA), a machine type communication (MTC) device, a personal digital assistant (PDA), a smartphone, a laptop, a computer, a tablet, a wireless sensor, a consumer electronics device, a smart book, a vehicle, a car, a truck, a bus, a train, or an IoT device, an industrial device, or apparatus (e.g. communication module, modem, or chip) in the forgoing devices, among other possibilities. Future generation EDs 110 may be referred to using other terms. The base station 170a and 170b is a T-TRP and will hereafter be referred to as T-TRP 170. Also shown in FIG.3, a NT-TRP will hereafter be referred to as NT-TRP 172. Each ED 110 connected to T-TRP 170 and/or NT-TRP 172 can be dynamically or semi-statically turned on (i.e., established, activated, or enabled), turned off (i.e., released, deactivated, or disabled) and/or configured in response to one or more of connection availability and connection necessity.

[00195] The ED 110 includes a transmitter 201 and a receiver 203 coupled to one or more antennas 204. Only one antenna 204 is illustrated. One, some, or all of the antennas may alternatively be panels. The transmitter 201 and the receiver 203 may be integrated, e.g. as a transceiver. The transceiver is configured to modulate data or other content for transmission by at least one antenna 204 or network interface controller (NIC). The transceiver is also configured to demodulate data or other content received by the at least one antenna 204. Each transceiver includes any suitable structure for generating signals for wireless or wired transmission and/or processing signals received wirelessly or by wire. Each antenna 204 includes any suitable structure for transmitting and/or receiving wireless or wired signals.

[00196] The ED 110 includes at least one memory 208. The memory 208 stores instructions and data used, generated, or collected by the ED 110. For example, the memory 208 can store software instructions or modules configured to implement some or all of the functionality and/or embodiments described herein and that are executed by the processing unit(s) 210. Each memory 208 includes any suitable volatile and/or non-volatile storage and retrieval device(s). Any suitable type of memory may be used, such as random access memory (RAM), read only memory (ROM), hard disk, optical disc, subscriber identity module (SIM) card, memory stick, secure digital (SD) memory card, on-processor cache, and the like.

[00197] The ED 110 may further include one or more input/output devices (not shown) or interfaces (such as a wired interface to the internet 150 in FIG. 1). The input/output devices permit interaction with a user or other devices in the network.

Each input/output device includes any suitable structure for providing information to or receiving information from a user, such as a speaker, microphone, keypad, keyboard, display, or touch screen, including network interface communications.

[00198] The ED 110 further includes a processor 210 for performing operations including those related to preparing a transmission for uplink transmission to the NT-TRP 172 and/or T-TRP 170, those related to processing downlink transmissions received from the NT-TRP 172 and/or T-TRP 170, and those related to processing sidelink transmission to and from another ED 110. Processing operations related to preparing a transmission for uplink transmission may include operations such as encoding, modulating, transmit beamforming, and generating symbols for transmission. Processing operations related to processing downlink transmissions may include operations such as receive beamforming, demodulating and decoding received symbols. Depending upon the embodiment, a downlink transmission may be received by the receiver 203, possibly using receive beamforming, and the processor 210 may extract signaling from the downlink transmission (e.g. by detecting and/or decoding the signaling). An example of signaling may be a reference signal transmitted by NT-TRP 172 and/or T-TRP 170. In some embodiments, the processor 276 implements the transmit beamforming and/or receive beamforming based on the indication of beam direction, e.g. beam angle information (BAI), received from T-TRP 170. In some embodiments, the processor 210 may perform operations relating to network access (e.g. initial access) and/or downlink synchronization, such as operations relating to detecting a synchronization sequence, decoding and obtaining the system information, etc. In some embodiments, the processor 210 may perform channel estimation, e.g. using a reference signal received from the NT-TRP 172 and/or T-TRP 170.

[00199] Although not illustrated, the processor 210 may form part of the transmitter 201 and/or receiver 203. Although not illustrated, the memory 208 may form part of the processor 210.

[00200] The processor 210, and the processing components of the transmitter 201 and receiver 203 may each be implemented by the same or different one or more processors that are configured to execute instructions stored in a memory (e.g. in memory 208). Alternatively, some or all of the processor 210, and the processing components of the transmitter 201 and receiver 203 may be implemented using dedicated circuitry, such as a programmed field-programmable gate array (FPGA), a graphical processing unit (GPU), or an application-specific integrated circuit (ASIC).

[00201] The T-TRP 170 may be known by other names in some implementations, such as a base station, a base transceiver station (BTS), a radio base station, a network node, a network device, a device on the network side, a transmit/receive node, a Node B, an evolved NodeB (eNodeB or eNB), a Home eNodeB, a next Generation NodeB (gNB), a transmission point (TP), a site controller, an access point (AP), or a wireless router, a relay station, a remote radio head, a terrestrial node, a terrestrial network device, or a terrestrial base station, base band unit (BBU), remote radio unit (RRU), active antenna unit (AAU), remote radio head (RRH), central unit (CU), distribute unit (DU), positioning node, among other possibilities. The T-TRP 170 may be macro BSs, pico BSs, relay nodes, donor nodes, or the like, or combinations thereof. The T-TRP 170 may refer to the forging devices or apparatus (e.g. communication module, modem, or chip) in the forgoing devices.

[00202] In some embodiments, the parts of the T-TRP 170 may be distributed. For example, some of the modules of the T-TRP 170 may be located remote from the equipment housing the antennas of the T-TRP 170, and may be coupled to the equipment housing the antennas over a communication link (not shown) sometimes known as front haul, such as common public radio interface (CPRI). Therefore, in some embodiments, the term T-TRP 170 may also refer to modules on the network side that perform processing operations, such as determining the location of the ED 110, resource allocation (scheduling), message generation, and encoding/decoding, and that are not necessarily part of the equipment housing the antennas of the T-TRP 170. The modules may also be coupled to other T-TRPs. In some embodiments, the T-TRP 170 may actually be a plurality of T-TRPs that are operating together to serve the ED 110, e.g. through coordinated multipoint transmissions.

[00203] The T-TRP 170 includes at least one transmitter 252 and at least one receiver 254 coupled to one or more antennas 256. Only one antenna 256 is illustrated. One, some, or all of the antennas may alternatively be panels. The transmitter 252 and the receiver 254 may be integrated as a transceiver. The T-TRP 170 further includes a processor 260 for performing operations including those related to: preparing a transmission for downlink transmission to the ED 110, processing an uplink transmission received from the ED 110, preparing a transmission for backhaul transmission to NT-TRP 172, and processing a transmission received over backhaul from the NT-TRP 172. Processing operations related to preparing a transmission for downlink or backhaul transmission may include operations such as encoding, modulating, precoding (e.g. MIMO precoding), transmit beamforming, and generating symbols for transmission. Processing operations related to processing received transmissions in the uplink or over backhaul may include operations such as receive beamforming, and demodulating and decoding received symbols. The processor 260 may also perform operations relating to network access (e.g. initial access) and/or downlink synchronization, such as generating the content of synchronization signal blocks (SSBs), generating the system information, etc. In some embodiments, the processor 260 also generates the indication of beam direction, e.g. BAI, which may be scheduled for transmission by scheduler 253. The processor 260 performs other network-side processing operations described herein, such as determining the location of the ED 110, determining where to deploy NT-TRP 172, etc. In some embodiments, the processor 260 may generate signaling, e.g. to configure one or more parameters of the ED 110 and/or one or more parameters of the NT-TRP 172. Any signaling generated by the processor 260 is sent by the transmitter 252. Note that “signaling”, as used herein, may alternatively be called control signaling. Dynamic signaling may be transmitted in a control channel, e.g. a physical downlink control channel (PDCCH), and static or semi-static higher layer signaling may be included in a packet transmitted in a data channel, e.g. in a physical downlink shared channel (PDSCH).

[00204] A scheduler 253 may be coupled to the processor 260. The scheduler 253 may be included within or operated separately from the T-TRP 170, which may schedule uplink, downlink, and/or backhaul transmissions, including issuing scheduling grants and/or configuring scheduling-free ("configured grant") resources. The T-TRP 170 further includes a memory 258 for storing information and data. The memory 258 stores instructions and data used, generated, or collected by the T-TRP 170. For example, the memory 258 can store software instructions or modules configured to implement some or all of the functionality and/or embodiments described herein and that are executed by the processor 260.

[00205] Although not illustrated, the processor 260 may form part of the transmitter 252 and/or receiver 254. Also, although not illustrated, the processor 260 may implement the scheduler 253. Although not illustrated, the memory 258 may form part of the processor 260.

[00206] The processor 260, the scheduler 253, and the processing components of the transmitter 252 and receiver 254 may each be implemented by the same or different one or more processors that are configured to execute instructions stored in a memory, e.g. in memory 258. Alternatively, some or all of the processor 260, the scheduler 253, and the processing components of the transmitter 252 and receiver 254 may be implemented using dedicated circuitry, such as a FPGA, a GPU, or an ASIC.

[00207] Although the NT-TRP 172 is illustrated as a drone only as an example, the NT-TRP 172 may be implemented in any suitable non-terrestrial form. Also, the NT-TRP 172 may be known by other names in some implementations, such as a non-terrestrial node, a non-terrestrial network device, or a non-terrestrial base station. The NT-TRP 172 includes a transmitter 272 and a receiver 274 coupled to one or more antennas 280. Only one antenna 280 is illustrated. One, some, or all of the antennas may alternatively be panels. The transmitter 272 and the receiver 274 may be integrated as a transceiver. The NT-TRP 172 further includes a processor 276 for performing operations including those related to: preparing a transmission for downlink transmission to the ED 110, processing an uplink transmission received from the ED 110, preparing a transmission for backhaul transmission to T-TRP 170, and processing a transmission received over backhaul from the T-TRP 170. Processing operations related to preparing a transmission for downlink or backhaul transmission may include operations such as encoding, modulating, precoding (e.g. MIMO precoding), transmit beamforming, and generating symbols for transmission. Processing operations related to processing received transmissions in the uplink or over backhaul may include operations such as receive beamforming, and demodulating and decoding received symbols. In some embodiments, the processor 276 implements the transmit beamforming and/or receive beamforming based on beam direction information (e.g. BAI) received from T-TRP 170. In some embodiments, the processor 276 may generate signaling, e.g. to configure one or more parameters of the ED 110. In some embodiments, the NT-TRP 172 implements physical layer processing, but does not implement higher layer functions such as functions at the medium access control (MAC) or radio link control (RLC) layer. As this is only an example, more generally, the NT-TRP 172 may implement higher layer functions in addition to physical layer processing.

[00208] The NT-TRP 172 further includes a memory 278 for storing information and data. Although not illustrated, the processor 276 may form part of the transmitter 272 and/or receiver 274. Although not illustrated, the memory 278 may form part of the processor 276.

[00209] The processor 276 and the processing components of the transmitter 272 and receiver 274 may each be implemented by the same or different one or more processors that are configured to execute instructions stored in a memory, e.g. in memory 278. Alternatively, some or all of the processor 276 and the processing components of the transmitter 272 and receiver 274 may be implemented using dedicated circuitry, such as a programmed FPGA, a GPU, or an ASIC. In some embodiments, the NT-TRP 172 may actually be a plurality of NT-TRPs that are operating together to serve the ED 110, e.g. through coordinated multipoint transmissions.

[00210] The T-TRP 170, the NT-TRP 172, and/or the ED 110 may include other components, but these have been omitted for the sake of clarity.

[00211] FIG. 4 is a schematic diagram of units or modules in a device according to an embodiment of the present application.

[00212] One or more steps of the embodiment methods provided may be performed by corresponding units or modules, according to FIG. 4. FIG. 4 illustrates units or modules in a device, such as in ED 110, T-TRP 170, or NT-TRP 172. For example, a signal may be transmitted by a transmitting unit or a transmitting module. For example, a signal may be transmitted by a transmitting unit or a transmitting module. A signal may be received by a receiving unit or a receiving module. A signal may be processed by a processing unit or a processing module. Other steps may be performed by an artificial intelligence (AI) or machine learning (ML) module. The respective units or modules may be implemented using hardware, one or more components or devices that execute software, or a combination thereof. For instance, one or more of the units or modules may be an integrated circuit, such as a programmed FPGA, a GPU, or an ASIC. It will be appreciated that where the modules are implemented using software for execution by a processor for example, they may be retrieved by a processor, in whole or part as needed, individually or together for processing, in single or multiple instances, and that the modules themselves may include instructions for further deployment and instantiation.

[00213] Additional details regarding the EDs 110, T-TRP 170, and NT-TRP 172 are known to those of skill in the art. As such, these details are omitted here.

[00214] FIG. 5 is a schematic diagram of an AI-based communication device.

[00215] A wireless system includes a plurality of connected devices. A device 500 is either base station (BS) or user equipment (UE). The device 500 may have three systems: sensing system 510, communication system 520, and/or AI system 530. The sensing system 510 senses and collects signals and data, the communication system 520 transmits and receives signals and data, and the AI system 530 trains and infers the AI implementations. An exemplary AI implementation is based on two cycles of deep learning, a training cycle and an inference cycle. In some possible application scenarios, the training cycle can also be referred to as the learning cycle and the inference cycle can also be referred to as the reasoning cycle.

[00216] Deep learning consists of two cycles: training (or learning) and inference (or reasoning). In a training cycle, the coefficients of neurons are learned from training data to fulfill a specific training goal or target. In the inference or reasoning cycle, an input data sample is fed into a trained neural network that would output a prediction.

[00217] During a training cycle, the AI system 530 of the device 500 may train the DNN or DNNs where the sensing system 510 of the device 500 may generate signals and/or data. The communication system 520 of the device 500 may receive the signals or data from another device or other devices. During and/or after the AI system 530 finishes training, the communication of the device may transmit the training results to another device or other devices.

[00218] During an inference cycle, the AI system 530 of a device 500 may perform one inference or a series of inferences with one DNN or DNNs to fulfill one task or tasks, where the sensing system 510 of the device 500 may generate signals and/or data, the communication system 520 of the device 500 may receive signals or data from another device or other devices. After the AI system 530 of the device 500 finishes inferencing, the communication system 520 of the device 500 may transmit the inferencing results to another device or other devices.

[00219] The AI implementations may either switch between the two cycles or stay in the two cycles simultaneously. For example, the AI system 530 of the device 500 may train the second DNN but still performs inference on the first DNN.

[00220] During the training cycle, the AI system 530 of the device 500 can work in single-user mode. In this mode, the AI system 530 trains the DNN or DNN(s) with the data provided by the sensing system 510 of the device 500. Examples of the data include local sensing data and local channel data. Local sensing data includes RGB data, light detection and ranging (LiDAR) data, temperature data, air pressure data, electric outrage data, etc. Local channel data includes channel state information (CSI), received signal strength indicator (RSSI), latency data, etc.

[00221] Alternatively, the AI system 530 of the device 500 may work in a cooperative mode. In this mode, the AI system 530 trains the DNN or DNN(s) with the data that the communication system 520 of the device 500 receives. Example data includes sensing data, channel data, neuron data and latent output data. Sensing data includes RGB data, LiDAR data, temperature data, air pressure data, electric outrage data, etc. Channel data includes CSI, RSSI, delay data, etc. Neuron data includes a number of neurons or a number of gradients. Latent output data includes several latent outputs.

[00222] FIG. 6 is a schematic diagram of a device 500 receiving reference data samples from a device 600 according to an embodiment of the present application. The AI system 530 of the device 500 in cooperative mode may use data such as: accumulating the sensing data that the communication system 520 of the device 500 received into one training data set;

accumulating the channel data that the communication system 520 of the device 500 received into one training data set; setting local neurons by the neurons that the communication system 520 of the device 500 received, which is a typical federated learning scheme; inputting the latent outputs that the communication system 520 of the device 500 received to its DNN(s).

[00223] Alternatively, the AI system 530 of the device 500 in a cooperative mode may use the data that the communication system 520 of the device 500 received together with its local ones, such as: mixing the local sensing data that the sensing system 510 of the device 500 provided with the sensing data that the communication system 520 of the device 500 received into one training data set; mixing the local channel data that the sensing system 510 of the device 500 provided with the channel data that the communication system 520 of the device 500 received into one training data set; averaging the local neurons that the AI system 530 of the device 500 possessed with the neurons that the communication system 520 of the device 500 received, which is a typical federated learning scheme; averaging the local latent outputs that the AI system 530 of the device 500 possessed and inputting them to its DNN(s).

[00224] FIG. 7 is a schematic diagram of reference data samples consisting of a plurality of groups according to an embodiment of the present application. During the training cycle, the communication system 520 of the device 500 may receive some reference data samples in both single-user or cooperative mode. Some devices transmit the reference data samples in broadcast, multicast, or unicast channels. The other devices transmits an indicator or indicators about which layer or layers to which the reference data samples are related, where, for example, there are three groups of the reference data samples: the first group of the reference data samples is indicated to be related to the input layer to the DNN, the second group of the reference data samples is indicated to be related to one latent layer output of the DNN, and the third group of the reference data samples is indicated to be related to the layer output from the DNN.

[00225] The AI system 530 of the device 500 may measure the distances between its local data samples and reference data samples group by group. The AI system 530 of the device 500 may randomly, non-randomly, uniformly, or non-uniformly sample its local layer inputs, local latent layer outputs, and/or layer outputs. Then the AI system 530 of the device 500 measures the distance between the local samples and the reference samples that the communication system 520 of the device 500 received. If the average distances of all the groups are consistently below a predefined threshold or thresholds, the AI system 530 of the device 500 may tell that the current training procedure works as expected, otherwise the AI system 530 may tell it is abnormal.

[00226] In a case where a device has no AI system but has sensing and communication systems, the sensing system of the device may be still able to measure the distances between its local data sample(s) and the reference data sample(s) related to the layer input to the DNN. If the average distance on the layer input is below a predefined threshold, the sensing system of the device may consider that the sensing device is catching "good" data, otherwise bad data. The communication system of the device may transmit only good data to other devices and may not transmit bad data to other devices, or the communication system of the device may label the sensing data with the distance before transmitting them to other devices.

[00227] The UE can report information about its data to the BS, which then determines whether that data differs significantly from the training data. If the difference is too large, the BS can switch the operating mode from AI to non-AI mode, or to another AI model. However, UE's direct reporting of raw data may be considered an invasion of user privacy. It is inefficient or against privacy policy to transmit raw data cross the air. Therefore, how to transmit data state information securely and efficiently is an urgent technical problem to be solved.

[00228] To protect raw data and save bandwidth, a group of the reference data samples are encoded or compressed to a lower dimensional space than their original space. The encoder or compressor can be linear or non-linear. A linear encoder can be realized with some standard basis such as Fourier Basis, DCT, wavelets, or a linear encoder can be with some customized basis. These bases may consist of a unitary matrix (orthonormal). A non-linear encoder can be realized with some DNNs. FIG. 8 is a schematic representation of a DNN-based approximation according to an embodiment of the present application.

[00229] Unlike the traditional compression schemes built for reliable reconstruction, the encoder deliberately avoids a reliable reconstruction but preserves as much topological distances as possible, when the data is compressed into a lower dimensional space. That is, the relative distance between two data samples in their original signal space may be well preserved after being encoded into a low-dimensional space.

[00230] FIG. 9 is a flowchart of a communication method according to an embodiment of the present application.

[00231] 710, sending a first coefficient.

[00232] The first coefficient is determined based on first data and a reference basis, and a dimension of the first coefficient is less than a dimension of the first data.

[00233] The first data includes monitoring data or measured data of the user equipment or the network device. Further, the first data is the monitoring data or measured data related to the AI models. The network device in this embodiment can be a BS.

[00234] If the first data is the data sent by the UE to the BS over the uplink, the data is the monitoring or measured data of the UE. If the first data is the data sent by the BS to the UE over the downlink, the data is the monitoring or measured data of the BS.

[00235] One or multiple reference bases are predefined or configured. The reference basis is one of the predefined or configured multiple reference bases. For example, the reference basis can be configured by the BS for the UE. The reference basis can be an orthogonal basis, and any two columns of the reference basis are perfectly orthogonal to each other. One typical orthogonal basis is the DFT basis.

[00236] 720, performing communication based on the first coefficient.

[00237] FIG. 10 is a flowchart of a communication method according to an embodiment of the present application. To protect raw data and save bandwidth, a group of the reference data samples are encoded or compressed to a lower dimensional space than their original space. The encoder or compressor can be linear or non-linear. A linear encoder can be realized with some standard basis such as Fourier basis, discrete cosine transform (DCT), wavelets. Alternatively, a linear encoder can be with some customized basis, and these bases may consist of a unitary matrix (orthonormal). A non-linear encoder can be realized with some DNNs.

[00238] In the embodiments given below, the UE projects a high-dimensional signal into a low-dimensional one (coefficients $\hat{c}$) by a transformation (orthonormal basis U). Reporting coefficients instead of raw data is efficient and conducive to privacy protection.

[00239] 810, one or multiple reference bases are configured or predefined.

[00240] Coefficients of reference basis indicator (CRBI) are used to indicate coefficients with respect to a reference basis (e.g. orthogonal basis). Let $\{u_1, u_2, \cdots, u_r\}$ be an orthonormal set of vectors in the subspace Rn. This set forms a basis U for the subspace Rn. An element represented by basis U in the subspace Rn can be written as a finite weighted linear combination of elements of the basis. The coefficients of this weighed linear combination are referred to as components or coordinates ($\hat{c}$) of the vector with respect to the basis U.

[00241] FIG. 11 is a schematic diagram of projecting a high-dimensional signal to a low-dimensional signal according to an embodiment of the present application. For example, $\hat{H} = U\hat{c}$, where $\hat{H}$ is the original space of n × 1, U is the orthogonal basis of n × r, and $\hat{c}$ is the spectrum subspace of r × 1. n is an integer greater than 1, and r<<n. $\hat{H}$ is the data to be reported by UE, e.g. sensing data, measured data, AI/machine learning (ML) data, channel data, environment data, etc. U is a reference basis as well as an orthogonal basis, and any two columns of U are perfectly orthogonal to each other. Embodiments of the present application can use columns as a basis, which can easily be applied to a basis matrix whose rows are the basis, simply U^H. One typical orthogonal basis is the discrete fourier transform (DFT) basis. $\hat{c}$ is the CRBI, which is a reference coefficient.

[00242] $\hat{H}$ is denoted as an n-by-1 reference sample and U is n-by-r matrix. $\hat{H}$ can be represented by a weighted linear combination of each columns of U: $\hat{H} = U\hat{c}$, where $\hat{c}$ is r-by-1 spectrum coefficients or weights. In the case of r<<n, $\hat{c}$ is an equivalent low-dimensional space signal (vector) of $\hat{H}$. The matrix U is unitary s.t. $U^H U = I$ and $\hat{c} = U^H\hat{H}$. Then, the matrix U^H is the encoder or compressor that compresses a high-dimensional (n-by-1) reference sample $\hat{H}$ into a low-dimensional (r-by-1) $\hat{c}$.

[00243] In one possible implementation scenario, multiple reference bases (U_A, U_B, U_C, …) are configured or predefined. The BS configures which reference basis to use, e.g., U_X. The UE reports CRBI based on U_X. According to the formula

$\hat{H} = U\hat{c}$, the UE knows U and $\hat{H}$, so the coefficients $\hat{c}$ can be calculated.

[00244] In one possible implementation scenario, one reference matrix U is configured or predefined, and one or multiple pruning bases are indicated or predefined as the reference basis. The reference matrix Y is a matrix of size M rows and N columns. A pruning basis for the reference basis is a K-column of Y, such as the first K-column of Y, where K is configured and K≤N. Optionally, it can be specified which K columns of Y are selected as the pruning basis.

[00245] 820, UE determines its coefficients of the reference basis.

[00246] A reference basis (U) is configured or predefined. The BS can configure one or more reference signals, and the UE can obtain raw data $\hat{H}$ by measuring the reference signal(s). Optionally, the reference signal(s) may also not be configured, and the UE can acquire the raw data $\hat{H}$ by sensing it. The UE determines its CRBI by $\hat{c} = U^H\hat{H}$. U is a unitary matrix that satisfies the conjugate transpose of the matrix equal to the inverse of the matrix, i.e., $U^H U = I$, and I is the unit matrix.

[00247] The UE may obtain one or multiple reporting data from a single time slot. Based on an observation interval in time (or unrestricted), the UE shall derive CRBI values reported in uplink slot. Exemplarily, the UE reports CRBI values in uplink time slot n. The UE may obtain the corresponding one or multiple CRBI values by measuring the data in the configured time window n-5 to n-1. The UE can choose to report the multiple CRBI values or report the average/maximum/minimum of the multiple CRBI values.

[00248] 830, UE reports CRBI or an index of the CRBI.

[00249] Exemplarily, the UE obtains P reporting data from the time window of n-5 to n-1, and P CRBI values corresponding to the P reporting data can be obtained by $\hat{c} = U^H\hat{H}$ .The UE can choose to report the average, maximum, or minimum of the P CRBI values. The reporting data includes monitoring data or measured data of the UE.

[00250] The UE can report the CRBI directly, or report the index corresponding to the CRBI. The BS can configure a physical uplink control channel (PUCCH) or a physical uplink shared channel (PUSCH) for the UE to report the CRBI. The CRBI reporting supports periodic, aperiodic, and semi-persistent.

[00251] In some possible application scenarios, the UE reports the index corresponding to the CRBI. In this scenario, one or multiple CRBI tables are predefined or configured. A reference basis can be associated with one CRBI table or with multiple CRBI tables. When a reference basis is associated with multiple CRBI tables, the BS indicates which CRBI table to use.

[00252] CRBI index of the CRBI table is reported by the UE. As shown in Table 1, 4 bits are used to indicate the CRBI index. Although the CRBI values in Table 1 are all denoted by the same $\{c_0, c_1, \ldots, c_r\}$, each CRBI index corresponds to a different CRBI value. In some possible implementations, the value of r in $\{c_0, c_1, \ldots, c_r\}$ is different in different rows of a CRBI table, e.g., some are $\{c_0, c_1, \ldots, c_5\}$ and some are $\{c_0, c_1, \ldots, c_6\}$.

Table 1

CRBI index	CRBI
0	$\{c_0, c_1, \ldots, c_r\}$
1	$\{c_0, c_1, \ldots, c_r\}$
…	…
15	$\{c_0, c_1, \ldots, c_r\}$

[00253] In some possible implementations, one CRBI index may correspond to a CRBI range, and Table 1 should not be construed as a limitation of this application.

[00254] In the communication method provided in this embodiment, the UE can report its data information to the BS with minimum air interface overhead, and then the BS determines whether the data is significantly different from the training data, improving the efficiency of data reporting and protecting the privacy of the data.

[00255] FIG. 12 is a flowchart of a communication method according to an embodiment of the present application. In this embodiment, a differential CRBI index reporting can be used.

[00256] 910, determining a reference CRBI index.

[00257] The reference CRBI index can be indicated by a BS, or it can be configured or predefined.

[00258] 920, reporting an offset level to the BS.

[00259] The UE reports the offset level to the BS. According to the offset level and reference CRBI index, the BS knows the current data CRBI index. Exemplarily, the differential CRBI can be obtained by equation (1).

[00260] offset level = current data CRBI index – reference CRBI index (1)

[00261] In the communication method provided in this embodiment, the UE can report its data information to the BS with minimum air interface overhead, and then the BS determines whether the data is significantly different from the training data, improving the efficiency of data reporting and protecting the privacy of the data.

[00262] In addition, the communication method provided in this application can also be applied to downlink (DL) transmission where the BS indicates the CRBI or CRBI index to the UE for indicating the data information at the BS side. Specific implementations can refer to the descriptions in FIG. 9 to FIG. 12 and will not be repeated in this application.

[00263] FIG. 13 is a schematic diagram of a matrix U being determined according to an embodiment of the present application.

[00264] Each column of the matrix U can be a standard basis such as Fourier basis, DCT basis, wavelet basis, and the like. Or the r columns of the matrix U can be built on the distribution of the group of the reference samples x. An example procedure to calculate the matrix U on the distribution of $x_1, x_2, \ldots$ may be as follows:

[00265] Accumulating a sufficient amount (M) n-by-1 samples $x_1, x_2, \ldots x_M$, and M<<n; Juxtaposing them into a n-

by-M matrix $\mathbb{x} = [x_1 \quad x_2 \quad \cdots x_M]$ and the order of data samples doesn't matter; Applying a rank-reduced singular value decomposition (SVD) on $\mathbb{x}$: $\mathbb{x} = \boldsymbol{U\Sigma V}^H$, where U is n-by-r unitary (orthonormal) matrix representing a commonality among all the M reference samples x.

[00266] Because a group of the reference data samples corresponds to one layer output, each group of the reference data samples has its own matrix U. The first group $\mathbb{x}_1 = [x_{1,1} \quad x_{1,2} \quad \cdots x_{1,M_1}]$ has the matrix U_1 and compressed versions $\mathbb{c}_1 = [c_{1,1} \quad c_{1,2} \quad \cdots c_{1,M_1}]$, and the second group $\mathbb{x}_2 = [x_{2,1} \quad x_{2,2} \quad \cdots x_{2,M_2}]$ has the matrix U_2 compressed versions $[c_{2,1} \quad c_{2,2} \quad \cdots c_{2,M_2}]$.

[00267] The communication system of the device receives the first matrix U_1 and the first group of reference samples (compressed) $\mathbb{c}_1$, and the second matrix U_2 and the second group of reference samples (compressed) $\mathbb{c}_2$.

[00268] FIG. 14 is a schematic diagram of a first sampling matrix P_1 according to an embodiment of the present application.

[00269] The first matrix U_1 is n_1-by-r_1 and the second matrix U_2 is n_2-by-r_2. If n_1 and/or n_2 are very big numbers, the first sampling matrix P_1 can be applied to the first matrix U_1, and the second sampling matrix P_2 can be applied to the second matrix U_2. The first sampling matrix P_1 is m_1-by-n_1 ($m_1 \ll n_1$), and each row of which has only one "1" to indicate the position of $x_{1,i}$ to be sampled. The second sampling matrix P_2 is m_2-by-n_2 ($m_2 \ll n_2$), and each row of which has only one "1" to indicate the position of $x_{2,i}$ to be sampled. The first sampling matrix P_1 can "compress" the first matrix U_1 (n_1-by-r_1) into a m_1-by-r_1 θ_1 as $\theta_1 = P_1 U_1$. Because θ_1 is much smaller than U_1 (because $m_1 \ll n_1$), θ_1 can be a better alternative to U_1. The second sampling matrix P_2 can "compress" the second matrix U_2 (n_2-by-r_2) into an m_2-by-r_2 θ_2 as $\theta_2 = P_2 U_2$. Because θ_2 is much smaller than U_2 (because $m_2 \ll n_2$), θ_2 can be a better alternative to U_2.

[00270] FIG. 15 is a schematic diagram of a sampling matrix compression matrix U according to an embodiment of the present application.

[00271] In one possible implementation, the communication system of the device receives the first compact matrix θ_1, the first sampling matrix P_1, and the first group of reference samples (compressed) $\mathbb{c}_1$. The communication system of the device receives the second compact matrix θ_2, the second sampling matrix P_2, and the second group of reference samples (compressed) $\mathbb{c}_2$.

[00272] Alternatively, the communication system of the device receives the left inverse of the first compact matrix $\boldsymbol{\theta}_1^{\,+}$, the first sampling matrix P_1, and the first group of reference samples (compressed) $\mathbb{c}_1$. The communication system of the device receives the inverse of the second compact matrix $\boldsymbol{\theta}_2^{\,+}$, the second sampling matrix P_2, and the second group of reference samples (compressed) $\mathbb{c}_2$.

[00273] FIG. 16 is a schematic diagram of a scoring distance on the low spectrum space according to an embodiment of the present application.

[00274] The communication system of the device may receive the first scoring function $d_1(c_{1,i}, c_{1,j})$ that measures the distance between two samples, $c_{1,i}$ and $c_{1,j}$ of the first group. The communication system of the device may receive the second scoring function $d_2(c_{2,i}, c_{2,j})$ that measures the distance between two samples, $c_{2,i}$ and $c_{2,j}$ of the second group. The first scoring function d_1 and the second scoring function d_2 may be the same or different. The first scoring function $d_1(,)$ and the second scoring function $d_2(,)$ may be dot product, inner product, Euclidean distance, and so on. Or the first scoring function $d_1(,)$ and the second scoring function $d_2(,)$ may be DNN-based.

[00275] Alternatively, the communication system of the device may receive the first scoring function $d_1(\mathbb{c}_1, \mathbb{c}_1')$ that measures the distance between two distributions, $\mathbb{c}_1$ and $\mathbb{c}_1'$ of the first group. The communication system of the device may receive the second scoring function $d_2(\mathbb{c}_2, \mathbb{c}_2')$ that measures the distance between two distributions, $\mathbb{c}_2$ and $\mathbb{c}_2'$ of the second group. The first scoring function d_1 and the second scoring function d_2 may be the same or different. The first scoring function $d_1(,)$ and the second scoring function $d_2(,)$ may be mutual information, Hilbert-Schmidt independence criterion (HSIC) metric, KL divergence, graph edit distance, Wasserstein distance, Jensen-Shannon divergence (JSD) distance, and so on. Or the first scoring function $d_1(,)$ and the second scoring function $d_2(,)$ may be DNN-based.

[00276] After the communication system 520 of the device 500 receives the groups of the reference data samples, the AI system 530 of the device 500 can measure the distances between its local data samples and reference data samples.

[00277] The AI system 530 of the device 500 samples the local data on the layer indicated by the indicator with the first group of the reference data samples, where the AI system 530 of the device 500 may sample the m_1 positions indicated by the first sampling matrix $\boldsymbol{P}_1$ into a m_1-by-1 local sample $\widehat{x_{1,1}}$, then the AI system 530 may calculate the low-dimensional space $\widehat{c_{1,1}} = \boldsymbol{\theta}_1{}^+\widehat{x_{1,1}}$. The AI system 530 may sample each data of an epoch batch or randomly sample K_1 data of an epoch batch: $\hat{\mathbb{c}}_1 = [\widehat{c_{1,1}} \quad \widehat{c_{1,2}} \quad \cdots\widehat{c_{1,K_1}}]$.

[00278] The AI system 530 samples the local data as indicated by the indicator with the second group of the reference data samples, where the AI system 530 may sample the m_2 positions indicated by the second sampling matrix $\boldsymbol{P}_2$ into a m_2-by-1 local sample $\widehat{x_{2,1}}$. Then the AI system 530 may calculate the low-dimensional space $\widehat{c_{2,1}} = \boldsymbol{\theta}_2{}^+\widehat{x_{2,1}}$. The AI system 530 may sample each data of an epoch batch or randomly sample K_2 data of an epoch batch: $\hat{\mathbb{c}}_2 = [\widehat{c_{2,1}} \quad \widehat{c_{2,2}} \quad \cdots\widehat{c_{2,K_2}}]$.

[00279] The AI system 530 can obtain the left inverse of the compact matrix $\boldsymbol{\theta}_1{}^+$ and $\boldsymbol{\theta}_2{}^+$ in several ways. For example, the communication system 520 receives the inverse of the compact matrix. Alternatively, the communication system 520 receives the compact matrix $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, and then left inverses the first compact matrix $\boldsymbol{\theta}_1$ into $\boldsymbol{\theta}_1{}^+$ and the second compact matrix $\boldsymbol{\theta}_2$ into $\boldsymbol{\theta}_2{}^{-1}$. Alternatively, the communication system 520 receives the first matrix $\boldsymbol{U}_1$ and the second matrix $\boldsymbol{U}_2$ and the first sampling matrix $\boldsymbol{P}_1$ and the second sampling matrix $\boldsymbol{P}_2$. The AI system 530 calculates the left inverse of the first compact matrix $\boldsymbol{\theta}_1{}^{+} = (\boldsymbol{P}_1\boldsymbol{U}_1)^{+}$ and the left inverse of the second compact matrix $\boldsymbol{\theta}_2{}^{+} = (\boldsymbol{P}_2\boldsymbol{U}_2)^{+}$. Alternatively, the communication system 520 receives the first matrix $\boldsymbol{U}_1$ and the second matrix $\boldsymbol{U}_2$. The AI system 530 generates the first sampling matrix $\boldsymbol{P}_1$ locally and the second sampling matrix $\boldsymbol{P}_2$ locally. The AI system 530 calculates the left inverse of the first compact matrix $\boldsymbol{\theta}_1{}^{+} = (\boldsymbol{P}_1\boldsymbol{U}_1)^{+}$ and the left inverse of the second compact matrix $\boldsymbol{\theta}_2{}^{+} = (\boldsymbol{P}_2\boldsymbol{U}_2)^{+}$. Alternatively, the communication system receives the first matrix $\boldsymbol{U}_1$ and the second matrix $\boldsymbol{U}_2$; the AI system doesn't sample the local data, mathematically the first sampling matrix $\boldsymbol{P}_1$ being an identity matrix $\boldsymbol{I}$ and the second sampling matrix $\boldsymbol{P}_2$ being an identity matrix $\boldsymbol{I}$; the AI system calculates the left inverse of the first compact matrix $\boldsymbol{\theta}_1{}^{+} = (\boldsymbol{P}_1\boldsymbol{U}_1)^{+} = (\boldsymbol{U}_1)^{+}$ and the left inverse of the second compact matrix $\boldsymbol{\theta}_2{}^{+} = (\boldsymbol{P}_2\boldsymbol{U}_2)^{+} = (\boldsymbol{U}_2)^{+}$; if the first matrix $\boldsymbol{U}_1$ is unitary, $\boldsymbol{\theta}_1{}^{+} = (\boldsymbol{U}_1)^{+} = (\boldsymbol{U}_1)^{H}$; if the second matrix $\boldsymbol{U}_2$ is unitary, $\boldsymbol{\theta}_2{}^{+} = (\boldsymbol{U}_2)^{+} = (\boldsymbol{U}_2)^{H}$.

[00280] The AI system 530 of the device 500 may measure the distance between the local data samples $\hat{\mathbb{c}}_1 = [\widehat{c_{1,1}} \quad \widehat{c_{1,2}} \quad \cdots \widehat{c_{1,K_1}}]$ and $\mathbb{c}_1 = [c_{1,1} \quad c_{1,2} \quad \cdots c_{1,M_1}]$ for the first group. The AI system 530 of the device 530 may measure the distance between the local data samples $\hat{\mathbb{c}}_2 = [\widehat{c_{2,1}} \quad \widehat{c_{2,2}} \quad \cdots \widehat{c_{2,K_2}}]$ and $\mathbb{c}_2 = [c_{2,1} \quad c_{2,2} \quad \cdots c_{2,M_2}]$ for the second group, where the measuring method is based on the scoring functions $d_1(,)$ and $d_2(,)$ received by the communication system 520 of the device 500.

[00281] If the scoring functions $d_1(,)$ and $d_2(,)$ measure the distance between the two samples, an example is $\delta_1 = \frac{\sum_{k=1}^{k=K_1} \underbrace{min}_{j=1,2,\ldots,M_1} \left(d_1(\widehat{c_{1,k}}, c_{1,j})\right)}{K_1}$ the average minimum distance for the first group and $\delta_2 = \frac{\sum_{k=1}^{k=K_2} \underbrace{min}_{j=1,2,\ldots,M_2} \left(d_2(\widehat{c_{2,k}}, c_{2,j})\right)}{K_2}$ the average minimum distance for the second group. If the scoring functions $d_1(,)$ and $d_2(,)$ measure the distance between the two distributions, an example is $\delta_1 = d_1(\hat{\mathbb{c}}_1, \mathbb{c}_1)$ for the first group and $\delta_2 = d_2(\hat{\mathbb{c}}_2, \mathbb{c}_2)$ for the second group. Optionally, the AI system 530 may calculate the higher order such as root-mean-square (RMS), standard deviation of δ_1 and δ_2.

[00282] FIG. 17 is a flowchart of an embodiment of a communication method according to an embodiment of the present application.

[00283] 1710, obtaining N anchor(s).

[00284] Each of the N anchors includes one or more reference data, and N≥1.

[00285] In one possible implementation scenario, if the BS assists the UE in model switching, N anchor(s) can be configured by the BS to the UE. In one possible implementation scenario, if the UE assists the BS in model switching, N anchor(s) can be reported by the UE to the BS. Configuration signal may be radio resource control (RRC), medium access control-control element (MAC-CE), or downlink control information (DCI), and may be broadcast, multicast, or unicast.

[00286] 1720, sending at least one index of M anchor(s) differing least from first data.

[00287] The first data includes monitoring data or measured data of the user equipment or the network device. Further, the first data is the monitoring data or measured data related to the AI models. The network device in this embodiment may be a base station (BS). If the first data is the data sent by the UE to the BS over the uplink, the data is the monitoring or measured data of the UE. If the first data is the data sent by the BS to the UE over the downlink, the data is the monitoring or measured data of the BS.

[00288] The M anchor(s) are the M anchor(s) among the N anchor(s) with the smallest difference from the first data. Each of the N anchor(s) corresponds to a configured AI model. M and N are integers greater than or equal to 1, and M≤N.

[00289] The UE or BS reports the nearest anchor index can be periodical, semi-persistent, or aperiodic. The reporting can be performed on a physical uplink control channel (PUCCH) or a physical uplink shared channel (PUSCH).

[00290] FIG. 18 is a schematic diagram of the BS indicating the UE to perform a model switch according to an embodiment of the present application. In the communication field, a dual-sided model usually includes a model of the BS and a model of the UE, such as a machine learning model that combines the encoder of the BS and the decoder of the UE. In this model, the BS encoder is used to encode the raw data into encoded data while the UE decoder is used to decode the received encoded data into raw data. For example, the encoded data may be raw data processed into another format of data (e.g., compressed raw data). This dual-sided model can be trained jointly to optimize the encoder and decoder, thus improving the performance and efficiency of the communication system. Multiple decoder model candidates are configured for a UE, and the optimal decoder may depend on the location (surrounding environment) of the UE due to the limited AI/ML capability to support large AI/ML models at the UE side.

[00291] 1810, BS configures multiple candidate models and corresponding data anchors.

[00292] BS configures multiple candidate AI/ML models to UEs, each model is configured with a model index. The configuration signal may be radio resource control (RRC), medium access control-control element (MAC-CE), or downlink control information (DCI), and may be broadcast, multicast, or unicast.

[00293] The BS configures an associated data anchor for each candidate AI/ML model. An anchor is a set of reference data, e.g. reference coefficients ($\hat{c}$). The reference data is reference coefficients of reference basis, and reference basis (U) is configured or pre-defined. A reference data ($\hat{c}$) can be a vector, e.g. a one-dimensional array, where the size of the vector is r, r is pre-defined or configured. The size of the set is K (c_j, j=1,2,...K), where K is pre-defined or configured. Each of the N anchors includes K reference coefficients.

[00294] The association between a data anchor and a candidate AI/ML model can be determined implicitly or configured explicitly. Exemplarily, the association between the data anchor and the candidate AI/ML model is implicitly determined by making the anchor index have the same value as the model index, i.e. {model index k, anchor index k}. Exemplarily, the association between the data anchor and the candidate AI/ML model is explicitly configured, and the BS configures the data anchor j is associated to candidate model k, i.e. {model index k, anchor index j}.

[00295] 1820, UE reports the nearest anchor.

[00296] The reference data (e.g. coefficient ($\hat{c}_{user}$)) is reference coefficients of reference basis (orthonormal basis U). During the information interaction between the UE and the BS, the UE can project the high-dimensional signal into the low-dimensional signal (coefficients $\hat{c}_{user}$) through a transformation (orthonormal basis U). The transformation equation is $\hat{H} = U\hat{c}_{user}$, where $\hat{H}$ is the reporting data, U is the reference basis and $\hat{c}_{user}$ is the reference coefficient. One column of U is one of the basis, meaning that any two columns of U are perfectly orthogonal to each other. The reporting data includes monitoring data or measured data of the UE.

[00297] A reference basis (U) is configured or pre-defined. For example, the BS can configure a reference signal about the reference basis. Optionally, this reference signal may also be sensed by the UE. UE determines its coefficients of reference basis indicator (CRBI) by $\hat{c}_{user} = U^H\hat{H}$. U is a unitary matrix that satisfies the conjugate transpose of the matrix equal to the inverse of the matrix, i.e., $U^H U = I$, I is the unit matrix.

[00298] UE calculates the difference between its data (e.g. coefficient ($\hat{c}_{user}$)) and the reference data c_j (j=1,2,...K) in the anchor, where the calculation method or function can be indicated by BS or pre-defined. For example, the difference can be calculated by either of the equations (2), (3), and (4). $d_{user,j}$ is the difference between the reporting data and the reference data $\hat{c}_j$ in the anchor. $\hat{c}_{user}$ is the reporting data of the UE. $\hat{c}_j$ is the j[th] reference data in the anchor. <> represents the inner product. $\| \|$ represents the modulus length of a vector. f represents other custom functions. $1 \le j \le K$ and $1 \le i \le r$.

[00299] $$d_{user,j} = \sum_i^r \left\| \hat{c}_{user}(i) - \hat{c}_j(i) \right\|^2 \quad (2)$$

[00300] $$d_{user,j} = \left\langle \hat{c}_{user}, \hat{c}_j \right\rangle \quad (3)$$

[00301] $$d_{user,j} = f\left(\hat{c}_{user}, \hat{c}_j\right) \quad (4)$$

[00302] Equations (2)-(4) are only examples, and the UE calculates the difference between its data (e.g. coefficient ($\hat{c}_{user}$)) and the reference data c_j (j=1,2,...K) in the anchor can also be calculated by a dot product, Euclidean distance, or DNN-based algorithm, etc., and the above examples should not be construed as a limitation of the present application.

[00303] An anchor is a set of reference data and the UE calculates the difference between its data and the anchor according to a method that can be indicated by the BS or predefined, such as equation (5) or (6). $d_{user,anchor}$ is the difference between the reporting data and the anchor. $d_{user,anchor}$ can be the minimum value of the difference between the reporting data $\hat{c}_{user}$ and the K reference data $\hat{c}$ in the anchor, or it can be the average value of the difference between the reporting data

$\hat{c}_{user}$ and the K reference data $\hat{c}$ in the anchor.

[00304] $$d_{user,anchor} = \min_{j=1,...,K}(d_{user,j}) \quad (5)$$

[00305] $$d_{user,anchor} = Avg_{j=1,...,K}(d_{user,j}) \quad (6)$$

[00306] Alternatively, the difference between the reporting data and the anchor can also be obtained by mutual information, Hilbert-Schmidt independence criterion (HSIC) metric, Kullback-Leibler (KL) scatter, graphical edit distance, Wasserstein distance, Jensen-Shannon divergence (JSD) distance, DNN-based algorithms, etc.

[00307] UE calculates the difference (d_j) between its data and the anchors j (j=1, 2, ..., N), where N is the number of anchors. UE can report to the BS the index of the anchor with the smallest difference from the UE's data, or it can report to the BS M indexes of the anchors with the smallest difference from the UE's data, e.g. 1st, 2nd, Mth smallest. M can be configured by the BS.

[00308] The UE's data can be raw measured data, or the measured quantities (measured data) are filtered by layer 3 filter. For the layer 3 filtering, the UE filters the measured quantities according to equation (7) for each measured quantity before it is used to evaluate the reporting criteria or the measurement report. M_n is the latest received measurement result from the physical layer. F_n is the updated filtered measurement result that is used for evaluation of reporting criteria or for measurement reporting. F_{n-1} is the old filtered measurement result, where F_0 is set to M_1 when the first measurement result from the physical layer is received. For measurement object for 5th generation new radio（MeasObjectNR）, $a = \frac{1}{2^{(ki/4)}}$, where ki is the filterCoefficient for the corresponding measurement quantity of the ith QuantityConfigNR in quantityConfigNR-List, and i is indicated by QuantityConfigIndex. For other measurements, $a = \frac{1}{2^{(k/4)}}$, where k is the filterCoefficient for the corresponding measurement quantity received by the quantityConfig.

[00309] $$F_n = (1 - a) \times F_{n-1} + a \times M_n \quad (7)$$

[00310] The QuantityConfig-List is a data structure for configuring multiple measurement parameters of a device. It includes multiple QuantityConfigs, and each QuantityConfig describes a set of measurement parameters and a measurement reporting method, which is used to guide the device in performing and reporting measurements. QuantityConfigIndex is an index value used to identify the configuration of the device measurement parameter. The QuantityConfig is used to describe the parameters that the device needs to measure and the way the measurements are reported. QuantityConfig includes Filter Coefficient, which is used to describe the way the device smoothes the measurement results.

[00311] UE reports the nearest anchor index can be periodical, semi-persistent, or aperiodic. The reporting can be performed on a physical uplink control channel (PUCCH) or a physical uplink shared channel (PUSCH).

[00312] 1830, BS assists UE in model switching.

[00313] According to the UE's report, BS can assist the UE in model switching. When there are multiple AI/ML models at the UE side, the BS assists the UE in model switching or switching to non-AI mode. When there is only one AI model at the UE side, the BS assists the UE in switching between AI and non-AI modes.

[00314] FIG. 19 is a schematic diagram of another BS indicating the UE to perform a model switch according to an embodiment of the present application. When UE's measured data is moving, and moving toward anchor index 2, the BS indicates the UE to switch its model from model index 1 to model index 2.

[00315] Another reporting scheme can be differential reporting. If the nearest anchor index is not changed as compared to the previous value, the UE report "same as previous one" to BS, e.g. using 1 bit to indicate whether it is changed, value 1 means changed and value 0 means the same. If the nearest anchor index is changed, the UE can also report the nearest anchor index to the BS.

[00316] Another reporting scheme is event-triggered reporting. The UE reports the nearest anchor index to the BS only when the nearest anchor index changes.

[00317] Embodiments of the present application provide a method that enables proactive model switching by data anchors.

[00318] In addition to reporting the index of the nearest data anchor, the UE can also report the difference with the nearest data anchor. The optimal AI/ML model may not work when the difference between its data and the nearest anchor is greater than a threshold, so the UE reports this information to the BS. According to the UE's report, BS can indicate the UE to switch to another model or fallback to non-AI mode. FIG. 20 is a schematic diagram of another BS indicating the UE to perform a model switch according to an embodiment of the present application. The BS indicates the UE to switch to non-AI mode when the UE's measured data is moving and moving farther away from all anchors.

[00319] Embodiments of the present application allow the UE to provide additional assistance information to the BS for proactive switching.

[00320] FIG. 21 is a schematic diagram of the UE indicating the BS to perform a model switch according to an embodiment of the present application.

[00321] 2110, UE reports one or multiple data anchors to BS.

[00322] UE reports one or multiple data anchors to BS, each data anchor is associated with an index. A reporting signal may be radio resource control (RRC), medium access control-control element (MAC-CE), or downlink control information (DCI), and may be broadcast, multicast or unicast.

[00323] The UE configures an associated data anchor for each candidate AI/ML model. An anchor is a set of reference data, e.g. reference coefficients ($\hat{c}$). The reference data is reference coefficients of reference basis, and reference basis (U) is configured or pre-defined. A reference data ($\hat{c}$) can be a vector, e.g. a one-dimensional array, where the size of the vector is r, r is pre-defined or configured. The size of the set is K (c_j, j=1,2,...K), where K is pre-defined or configured. Each of the N anchors includes K reference coefficients.

[00324] The association between a data anchor and a candidate AI/ML model can be determined implicitly or configured explicitly. Exemplarily, the association between the data anchor and the candidate AI/ML model is implicitly determined by making the anchor index have the same value as the model index, i.e. {model index k, anchor index k}. Exemplarily, the association between the data anchor and the candidate AI/ML model is explicitly configured, and the UE configures the data anchor j is associated to candidate model k, i.e. {model index k, anchor index j}.

[00325] 2120, BS indicates the nearest anchor index to the UE.

[00326] The reference data (e.g. coefficient ($\hat{c}_{BS}$)) is reference coefficients of reference basis (orthonormal basis U). During the information interaction between the UE and the BS, the BS can project the high-dimensional signal into the low-dimensional signal (coefficients $\hat{c}_{BS}$) through a transformation (orthonormal basis U). The transformation equation is $\hat{B} = U\hat{c}_{BS}$, where $\hat{B}$ is the BS's measured data, U is the reference basis and $\hat{c}_{BS}$ is the reference coefficient. The measured data at BS could be obtained by sensing measurement or uplink (UL) channel measurement by sounding reference signal (SRS). One column of U is one of the basis, meaning that any two columns of U are perfectly orthogonal to each other.

[00327] A reference basis (U) is configured or pre-defined. For example, the UE can configure a reference signal about the reference basis. Optionally, this reference signal may also be sensed by the BS. BS determines its coefficients of reference basis indicator (CRBI) by $\hat{c}_{BS} = U^H\hat{B}$. U is a unitary matrix that satisfies the conjugate transpose of the matrix equal to the inverse of the matrix, i.e., $U^H U = I$, I is the unit matrix.

[00328] BS calculates the difference between its data (e.g. coefficient ($\hat{c}_{BS}$)) and the reference data c_j (j=1,2,...K) in the anchor, where the calculation method or function can be indicated by UE or pre-defined. For example, the difference can be calculated by either of the equations (8), (9), and (10). $d_{BS,j}$ is the difference between the BS's data and the reference data $\hat{c}_j$ in the anchor. $\hat{c}_{BS}$ is the BS's data. $\hat{c}_j$ is the j[th] reference data in the anchor. <> represents the inner product. $\| \|$ represents the modulus length of a vector. f represents other custom functions. $1 \le j \le K$ and $1 \le i \le r$.

[00329] $$d_{BS,j} = \sum_i^r \left\| \hat{c}_{BS}(i) - \hat{c}_j(i) \right\|^2 \quad (8)$$

[00330] $$d_{BS,j} = \left\langle \hat{c}_{BS}, \hat{c}_j \right\rangle \quad (9)$$

[00331] $$d_{BS,j} = f\left(\hat{c}_{BS}, \hat{c}_j\right) \quad (10)$$

[00332] It should be understood that equations (8)-(10) are only examples, and the BS calculates the difference between its data (e.g. coefficient ($\hat{c}_{BS}$)) and the reference data c_j (j=1,2,...K) in the anchor can also be calculated by a dot product, Euclidean distance, or DNN-based algorithm, etc., and the above examples should not be construed as a limitation of the present application.

[00333] An anchor is a set of reference data and the BS calculates the difference between its data and the anchor according to a method that can be indicated by the UE or predefined, such as equation (11) or (12). $d_{BS,anchor}$ is the difference between the BS's data and the anchor. $d_{BS,anchor}$ can be the minimum value of the difference between the BS's data $\hat{c}_{BS}$ and the K reference data $\hat{c}$ in the anchor, or it can be the average value of the difference between the BS's data $\hat{c}_{BS}$ and the K reference data $\hat{c}$ in the anchor.

[00334] $$d_{BS,anchor} = \min_{j=1,\ldots,K}(d_{BS,j}) \quad (11)$$

[00335] $$d_{BS,anchor} = \mathrm{Avg}_{j=1,\ldots,K}(d_{BS,j}) \quad (12)$$

[00336] Alternatively, the difference between the BS's data and the anchor can also be obtained by mutual information, Hilbert-Schmidt independence criterion (HSIC) metric, Kullback-Leibler (KL) scatter, graphical edit distance, Wasserstein distance, Jensen-Shannon divergence (JSD) distance, DNN-based algorithms, etc.

[00337] BS calculates the difference (d_j) between its data and the anchors j (j=1, 2, ..., N), where N is the number of anchors. BS can report to the UE the index of the anchor with the smallest difference from the BS's data, or it can report to the UE M indexes of the anchors with the smallest difference from the BS's data, e.g. 1st, 2nd, Mth smallest. M can be configured by the UE.

[00338] The BS's data can be raw measured data, or the measured quantities (measured data) are filtered by layer 3 filter. For the layer 3 filtering, the BS filters the measured quantities according to equation (13) for each measured quantity before it is used to evaluate the reporting criteria or the measurement report. M_n is the latest received measurement result from the physical layer. F_n is the updated filtered measurement result that is used for evaluation of reporting criteria or for measurement reporting. F_{n-1} is the old filtered measurement result, where F_0 is set to M_1 when the first measurement result from the physical layer is received. For measurement object for 5th generation new radio（MeasObjectNR）, $a = \frac{1}{2^{(ki/4)}}$, where ki is the filterCoefficient for the corresponding measurement quantity of the ith QuantityConfigNR in quantityConfigNR-List, and i is indicated by QuantityConfigIndex. For other measurements, $a = \frac{1}{2^{(k/4)}}$, where k is the filterCoefficient for the corresponding measurement quantity received by the quantityConfig.

[00339] $F_n = (1 - a) \times F_{n-1} + a \times M_n$ (13)

[00340] BS reports the nearest anchor index can be periodical, semi-persistent, or aperiodic. The reporting can be performed on a PUCCH or a PUSCH.

[00341] 2130, UE assists BS in model switching.

[00342] According to the BS's report, UE can assist the BS in model switching. When there are multiple AI/ML models at the BS side, the UE assists the BS in model switching or switching to non-AI mode. When there is only one AI model at the UE side, the UE assists the BS in switching between AI and non-AI modes.

[00343] The implementation of the UE assisting the BS in model switching is similar to the implementation of the BS assisting the UE in model switching. A specific description can be found in the description of FIGS. 3 to 5, which will not be repeated in this application.

[00344] Embodiments of the present application provide a method that enables proactive model switching by data anchors.

[00345] In addition to reporting the index of the nearest data anchor, the BS can also report the difference with the nearest data anchor. The optimal AI/ML model may not work when the difference between its data and the nearest anchor is greater than a threshold, so the BS reports this information to the UE. According to the BS's report, UE can indicate the BS to switch to another model or fallback to non-AI mode. In embodiments of the present application, the BS is allowed to provide additional assistance information to the UE for proactive switching.

[00346] FIG. 22 is a schematic block diagram of a communication apparatus 2200 according to an embodiment of this application. The communication apparatus 2200 includes: an obtaining module 2210 configured to obtain N anchor(s), one of the N anchors comprising one or multiple pieces of reference data, $N \geq 1$; and a sending module 2220 configured to send at least one index of M anchor(s) differing least from first data, $M \leq N$ and $M \geq 1$.

[00347] In a possible implementation, M first difference value(s) corresponding to the M anchor(s) are the smallest M of N first difference value(s), and the n^{th} first difference value among the N first difference value(s) is a difference value between the first data and the n^{th} anchor among the N anchor(s), and $1 \leq n \leq N$.

[00348] In a possible implementation, the first difference value corresponding to the n^{th} anchor is a minimum value or an average value of K second difference value(s) corresponding to the n^{th} anchor, and the n^{th} anchor includes K pieces of reference data, and the j^{th} second difference value among the K second difference value(s) is a difference value between the first data and the j^{th} reference data among the K pieces of reference data, $K \geq 1$, $1 \leq j \leq K$ and $1 \leq n \leq N$.

[00349] In a possible implementation, the sending module is further configured to send a third difference value, the third difference value being a difference value between the first anchor and the first data, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[00350] In a possible implementation, the sending module is further configured to send a first message or a second message, the first message being configured to indicate that the anchor with the smallest difference value from the first data has not changed during a time period, and the second message being configured to indicate that the anchor with the smallest difference value from the first data has changed during a time period.

[00351] In a possible implementation, the sending module is further configured to send at least one index of M anchor(s) differing least from the first data when it is determined that the anchor with the smallest difference value from the first data has changed during the time period.

[00352] In a possible implementation, the obtaining module is further configured to receive a third message; and a processing module 2230 configured to switch to another AI model or non-AI mode based on the third message.

[00353] In a possible implementation, the processing module is further configured to switch to a first AI model corresponding to a first anchor based on the third message, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[00354] In a possible implementation, the third message includes an index of the second AI model, and the processing module is further configured to switch to the second AI model based on the index of the second AI model.

[00355] In a possible implementation, the sending module is further configured to send a fourth message, the fourth message being configured to indicate that a third difference value is greater than a predetermined threshold, the third difference value being a difference value between the first data and a first anchor, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[00356] In a possible implementation, the fourth message includes an invalid anchor index.

[00357] In a possible implementation, the fourth message includes information configured to indicate that a sender of the fourth message has switched to a non-AI mode.

[00358] In a possible implementation, a value of M is predefined or configured.

[00359] In a possible implementation, one of the N anchors corresponds to one AI model and index.

[00360] In a possible implementation, the first data includes monitoring data or measured data of a user equipment or a network device.

[00361] In a possible implementation, the first data includes filtered measured data.

[00362] In a possible implementation, the index(es) of the M anchor(s) are sent through a physical uplink control channel (PUCCH) or a physical uplink shared channel (PUSCH).

[00363] In a possible implementation, the N anchors are configured by a radio resource control (RRC), a medium access control-control element (MAC-CE) or a downlink control information (DCI) signal.

[00364] In a possible implementation, the apparatus is located on a user equipment or a network device.

[00365] FIG. 23 is a schematic block diagram of a communication apparatus 2300 according to an embodiment of this application. The communication apparatus 2300 includes: a receiving module 2310 configured to receive at least one index of M anchor(s), the M anchor(s) being M anchor(s) among N anchor(s) having the smallest difference from first data, one of the M anchor(s) including one or more reference data, $N\geq1$, $M\leq N$ and $M\geq1$; and a sending module 2320 configured to send a third message, the third message being configured to indicate that the receiver switches the AI model or mode.

[00366] In a possible implementation, M first difference value(s) corresponding to the M anchor(s) are the smallest M of N first difference value(s), and the n^{th} first difference value among the N first difference value(s) is a difference value between the first data and the n^{th} anchor among the N anchor(s), and $1\leq n\leq N$.

[00367] In a possible implementation, the first difference value corresponding to the n^{th} anchor is a minimum value or an average value of K second difference value(s) corresponding to the n^{th} anchor, and the n^{th} anchor includes K pieces of reference data, and the j^{th} second difference value among the K second difference value(s) is a difference value between the first data and the j^{th} reference data among the K pieces of reference data, $K\geq1$, $1\leq j\leq K$ and $1\leq n\leq N$.

[00368] In a possible implementation, the receiving module is further configured to receive a third difference value, the third difference value being a difference value between the first anchor and the first data, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[00369] In a possible implementation, the receiving module is further configured to receive a first message or a second message, the first message being configured to indicate that the anchor with the smallest difference value from the first data has not changed during a time period, and the second message being configured to indicate that the anchor with the smallest difference value from the first data has changed during a time period.

[00370] In a possible implementation, the third message is configured to indicate a receiver to switch to a first AI model corresponding to a first anchor, and the first anchor is the anchor among the M anchors having the smallest difference from the first data.

[00371] In a possible implementation, the third message includes an index of a second AI model, and the third message is configured to indicate a receiver to switch to the second AI model.

[00372] In a possible implementation, the receiving module is further configured to receive a fourth message, the fourth message being configured to indicate that a third difference value is greater than a predetermined threshold, the third difference value being a difference value between the first data and a first anchor, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

[00373] In a possible implementation, the fourth message includes an invalid anchor index.

[00374] In a possible implementation, the fourth message includes information configured to indicate that a sender of the fourth message has switched to a non-AI mode.

[00375] In a possible implementation, a value of M is predefined or configured.

[00376] In a possible implementation, one of the N anchors corresponds to one AI model and index.

[00377] In a possible implementation, the first data includes monitoring data or measured data of a user equipment or a network device.

[00378] In a possible implementation, the first data includes filtered measured data.

[00379] In a possible implementation, the index(es) of the M anchor(s) are sent through a physical uplink control channel (PUCCH) or a physical uplink shared channel (PUSCH).

[00380] In a possible implementation, the N anchors are configured by a radio resource control (RRC), a medium access control-control element (MAC-CE) or a downlink control information (DCI) signal.

[00381] In a possible implementation, the apparatus is located on a user equipment or a network device.

[00382] As shown in FIG. 24, a communication apparatus 2400 may include a processor 2410 and a transceiver 2420. Optionally, the communication apparatus 2400 may further include a memory 2430. The memory 2430 may be configured to store indication information, or may be configured to store code, instructions, and the like that is to be executed by the processor 2410.

[00383] The memory 2430 may include a random memory, a flash memory, a read-only memory, a programmable read-only memory, a non-volatile memory, a register, or the like. The processor 2410 may be a central processing unit (CPU).

[00384] For other functions and operations of the communication apparatus 2400, refer to processes of the method embodiments from FIG. 5 to FIG. 21, which are not described again herein to avoid repetition.

[00385] An embodiment of the present application further provides a communication system. The communication system includes communication apparatus 2200 and communication apparatus 2300, or the communication system includes communication apparatus 2400.

[00386] An embodiment of the present application further provides a computer storage medium, and the computer storage medium may store a program instruction for performing the steps in the foregoing methods.

[00387] Optionally, the storage medium may be specifically the memory 2430.

[00388] An embodiment of the present application further provides a computer program product. The computer program product includes computer program code. When the computer program code runs on a computer, the computer is enabled to perform the steps in the foregoing methods.

[00389] Optionally, all or a part of computer program code can be stored in on a first storage medium. The first storage medium can be packaged together with the processor or separately with the processor.

[00390] An embodiment of the present application further provides a chip system, where the chip system includes an input/output interface, at least one processor, at least one memory, and a bus. The at least one memory is configured to store instructions, and the at least one processor is configured to invoke the instructions of the at least one memory to perform operations in the methods in the foregoing embodiments.

[00391] A person of ordinary skill in the art may understand that all or some of the processes of the methods in the embodiments may be implemented by a computer program instructing related hardware. The program may be stored in a computer-readable storage medium. When the program runs, the processes of the methods in the embodiments are performed. The foregoing storage medium may include: a magnetic disk, an optical disc, a read-only memory (ROM), or a random-access memory (RAM).

[00392] In the several embodiments provided in the present application, it should be understood that the disclosed system, apparatus, and method may be implemented in other manners. For example, the described apparatus embodiment is merely exemplary. For example, the unit division is merely logical function division and may be other division in actual implementation. For example, a plurality of units or components may be combined or integrated into another system, or some features may be ignored or not performed. In addition, the displayed or discussed mutual couplings or direct couplings or communication connections may be implemented by using some interfaces. The indirect couplings or communication connections between the apparatuses or units may be implemented in electronic, mechanical, or other forms.

[00393] The units described as separate parts may or may not be physically separate, and parts displayed as units may or may not be physical units, may be located in one position, or may be distributed on a plurality of network units. Some or all of the units may be selected according to actual needs to achieve the objectives of the solutions of the embodiments.

[00394] In addition, functional units in the embodiments of the present invention may be integrated into one processing unit, or each of the units may exist alone physically, or two or more units are integrated into one unit.

[00395] The foregoing are merely exemplary embodiments of the present invention. A person skilled in the art may make various modifications and variations to the present invention without departing from and scope of the present invention.

CLAIMS

What is claimed is:

1. A communication method, comprising:

obtaining N anchor(s), one of the N anchors comprising one or multiple pieces of reference data, $N \geq 1$; and

sending at least one index of M anchor(s) differing least from first data, $M \leq N$ and $M \geq 1$.

2. The method according to claim 1, wherein M first difference value(s) corresponding to the M anchor(s) are the smallest M of N first difference value(s), and the n^{th} first difference value among the N first difference value(s) is a difference value between the first data and the n^{th} anchor among the N anchor(s), and $1 \leq n \leq N$.

3. The method according to claim 2, wherein the first difference value corresponding to the n^{th} anchor is a minimum value or an average value of K second difference value(s) corresponding to the n^{th} anchor, and the n^{th} anchor comprises K pieces of reference data, and the j^{th} second difference value among the K second difference value(s) is a difference value between the first data and the j^{th} reference data among the K pieces of reference data, $K \geq 1$, $1 \leq j \leq K$ and $1 \leq n \leq N$.

4. The method according to any one of claims 1 to 3, further comprising:

sending a third difference value, the third difference value being a difference value between the first anchor and the first data, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

5. The method according to any one of claims 1 to 4, further comprising:

sending a first message or a second message, the first message being configured to indicate that the anchor with the smallest difference value from the first data has not changed during a time period, and the second message being configured to indicate that the anchor with the smallest difference value from the first data has changed during a time period.

6. The method according to any one of claims 1 to 5, wherein the sending at least one index of M anchor(s) differing least from first data comprises:

sending at least one index of M anchor(s) differing least from the first data when it is determined that the anchor with the smallest difference value from the first data has changed during the time period.

7. The method according to any one of claims 1 to 6, further comprising:

receiving a third message; and

switching to another AI model or non-AI mode based on the third message.

8. The method according to claim 7, wherein the switching to another AI model or non-AI mode based on the third message comprises:

switching to a first AI model corresponding to a first anchor based on the third message, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

9. The method according to claim 7, wherein the third message comprises an index of a second AI model, and

the switching to another AI model or non-AI mode based on the third message comprises:

switching to the second AI model based on the index of the second AI model.

10. The method according to any one of claims 1 to 9, further comprising:

sending a fourth message, the fourth message being configured to indicate that a third difference value is greater than a predetermined threshold, the third difference value being a difference value between the first data and a first anchor, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

11. The method according to claim 10, wherein the fourth message comprises an invalid anchor index.

12. The method according to claim 10, wherein the fourth message comprises information configured to indicate that a sender of the fourth message has switched to a non-AI mode.

13.The method according to any one of claims 1 to 12, wherein a value of M is predefined or configured.

14.The method according to any one of claims 1 to 13, wherein one of the N anchors corresponds to one AI model and index.

15.The method according to any one of claims 1 to 14, wherein the first data comprises monitoring data or measured data of a user equipment or a network device.

16. The method according to claim 15, wherein the first data comprises filtered measured data.

17. The method according to any one of claims 1 to 16, wherein the index(es) of the M anchor(s) are sent through a physical uplink control channel (PUCCH) or a physical uplink shared channel (PUSCH).

18. The method according to any one of claims 1 to 17, wherein the N anchors are configured by a radio resource control (RRC), a medium access control-control element (MAC-CE) or a downlink control information (DCI) signal.

19. The method according to any one of claims 1 to 18, wherein the method is executed by a user equipment or a network device.

20. A communication apparatus, comprising:

an obtaining module configured to obtain N anchor(s), one of the N anchors comprising one or multiple pieces of reference data, $N \geq 1$; and

a sending module configured to send at least one index of M anchor(s) differing least from first data, $M \leq N$ and $M \geq 1$.

21. The communication apparatus according to claim 20, wherein M first difference value(s) corresponding to the M anchor(s) are the smallest M of N first difference value(s), and the n^{th} first difference value among the N first difference value(s)

is a difference value between the first data and the n^{th} anchor among the N anchor(s), and $1 \le n \le N$.

22. The communication apparatus according to claim 21, wherein the first difference value corresponding to the n^{th} anchor is a minimum value or an average value of K second difference value(s) corresponding to the n^{th} anchor, and the n^{th} anchor comprises K pieces of reference data, and the j^{th} second difference value among the K second difference value(s) is a difference value between the first data and the j^{th} reference data among the K pieces of reference data, $K \ge 1$, $1 \le j \le K$ and $1 \le n \le N$.

23. The communication apparatus according to any one of claims 20 to 22, wherein the sending module is further configured to send a third difference value, the third difference value being a difference value between the first anchor and the first data, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

24. The communication apparatus according to any one of claims 20 to 23, wherein the sending module is further configured to send a first message or a second message, the first message being configured to indicate that the anchor with the smallest difference value from the first data has not changed during a time period, and the second message being configured to indicate that the anchor with the smallest difference value from the first data has changed during a time period.

25. The communication apparatus according to any one of claims 20 to 24, wherein the sending module is further configured to send at least one index of M anchor(s) differing least from the first data when it is determined that the anchor with the smallest difference value from the first data has changed during the time period.

26. The communication apparatus according to any one of claims 20 to 25, wherein the obtaining module is further configured to receive a third message; and

a processing module configured to switch to another AI model or non-AI mode based on the third message.

27. The communication apparatus according to claim 26, wherein the processing module is further configured to switch to a first AI model corresponding to a first anchor based on the third message, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

28. The communication apparatus according to claim 26, wherein the third message comprises an index of the second AI model, and the processing module is further configured to switch to the second AI model based on the index of the second AI model.

29. The communication apparatus according to any one of claims 20 to 28, wherein the sending module is further configured to send a fourth message, the fourth message being configured to indicate that a third difference value is greater than a predetermined threshold, the third difference value being a difference value between the first data and a first anchor, the first anchor being an anchor with the smallest difference value from the first data among the N anchor(s).

30. The communication apparatus according to claim 29, wherein the fourth message comprises an invalid anchor index.

31. The communication apparatus according to claim 29, wherein the fourth message comprises information configured

to indicate that a sender of the fourth message has switched to a non-AI mode.

32. The communication apparatus according to any one of claims 20 to 31, wherein a value of M is predefined or configured.

33. The communication apparatus according to any one of claims 20 to 32, wherein one of the N anchors corresponding to one AI model and index.

34. The communication apparatus according to any one of claims 20 to 33, wherein the first data comprises monitoring data or measured data of a user equipment or a network device.

35. The communication apparatus according to claim 34, wherein the first data comprises filtered measured data.

36. The communication apparatus according to any one of claims 20 to 35, wherein the index(es) of the M anchor(s) are sent through a physical uplink control channel (PUCCH) or a physical uplink shared channel (PUSCH).

37. The communication apparatus according to any one of claims 20 to 36, wherein the N anchors are configured by a radio resource control (RRC), a medium access control-control element (MAC-CE) or a downlink control information (DCI) signal.

38. The communication apparatus according to any one of claims 20 to 37, wherein the apparatus is located on a user equipment or a network device.

39. A communication apparatus, comprising a processor and a memory, and the processor is connected to the memory; wherein the memory is configured to store instructions, and the processor is configured to execute the instructions; and when the processor executes the instructions stored in the memory, the processor is enabled to perform the method according to any one of claims 1 to 19.

40. A computer-readable storage medium, wherein the computer-readable storage medium stores instructions, and when the instructions run on a processor, the processor is enabled to perform the method according to any one of claims 1 to 19.

41. A computer program product, comprising computer program code, and when the computer program code runs on a computer, the computer is enabled to perform the method according to any one of claims 1 to 19.

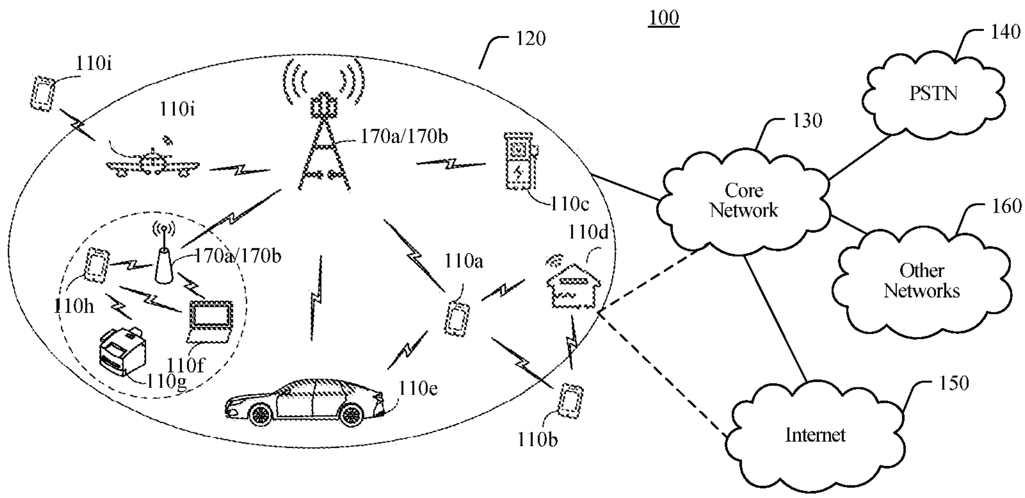
100
120
140
PSTN
110i
110i
130
170a/170b
Core
Network
110c
160
110d
170a/170b
110a
Other
Networks
110h
110f
110g
110e
150
110b
Internet

FIG. 1

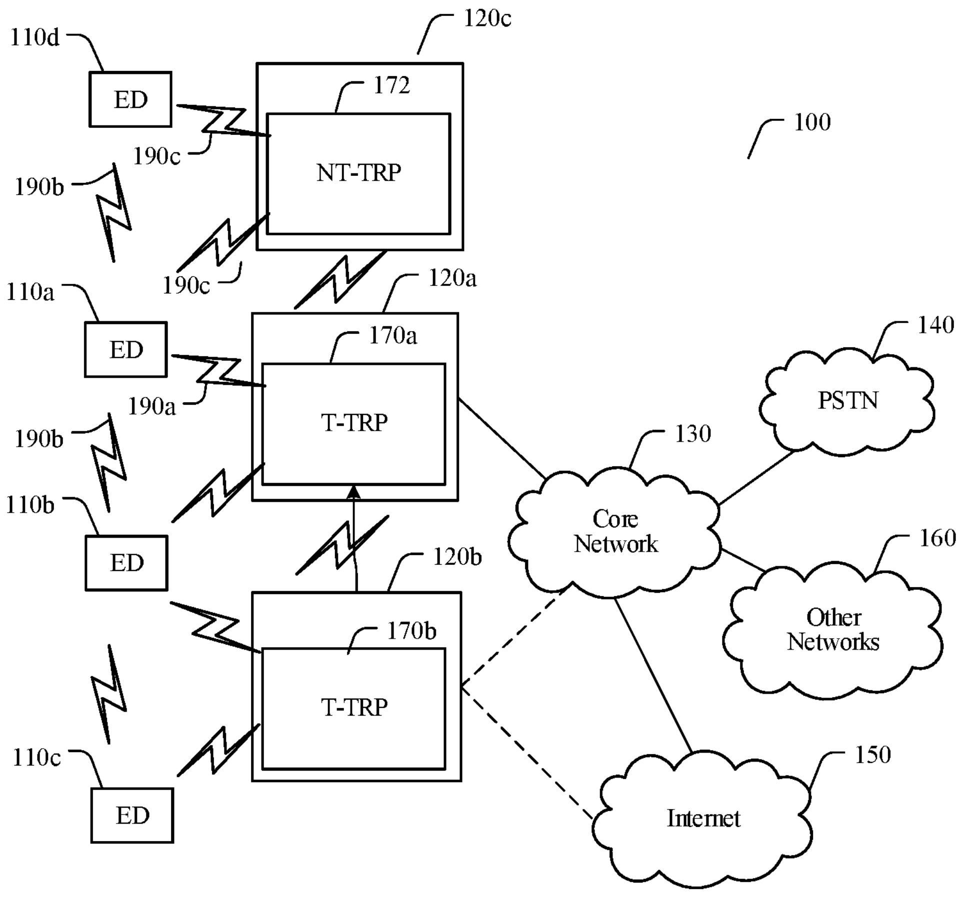
120c
110d
ED
172
190c
NT-TRP
100
190b
190c
120a
110a
ED
170a
140
PSTN
190a
T-TRP
190b
130
110b
Core
Network
ED
120b
160
170b
Other
Networks
T-TRP
110c
150
ED
Internet

FIG. 2

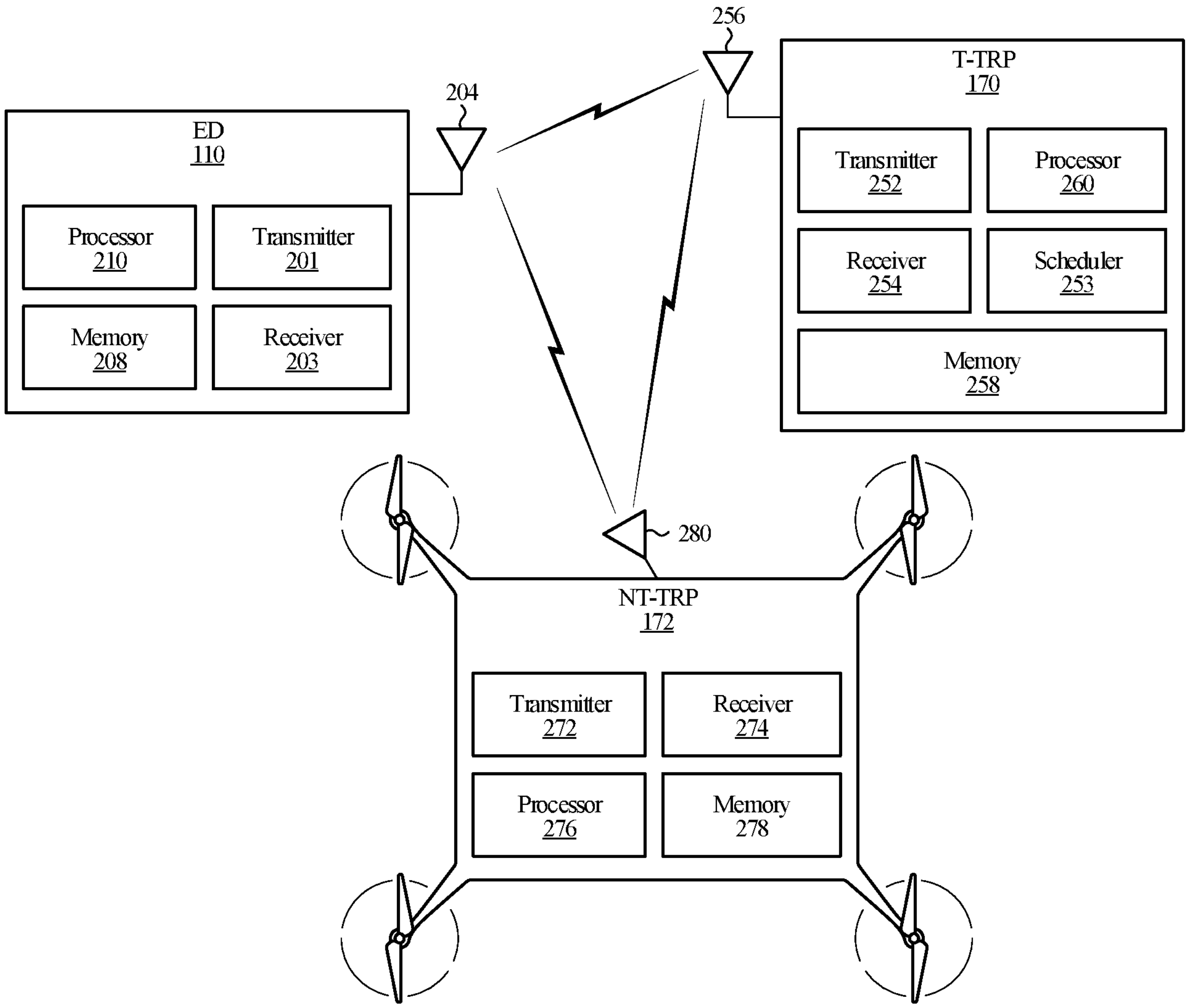
256
T-TRP
170
204
ED
110
Transmitter
252
Processor
260
Processor
210
Transmitter
201
Receiver
254
Scheduler
253
Memory
208
Receiver
203
Memory
258
280
NT-TRP
172
Transmitter
272
Receiver
274
Processor
276
Memory
278

FIG.3

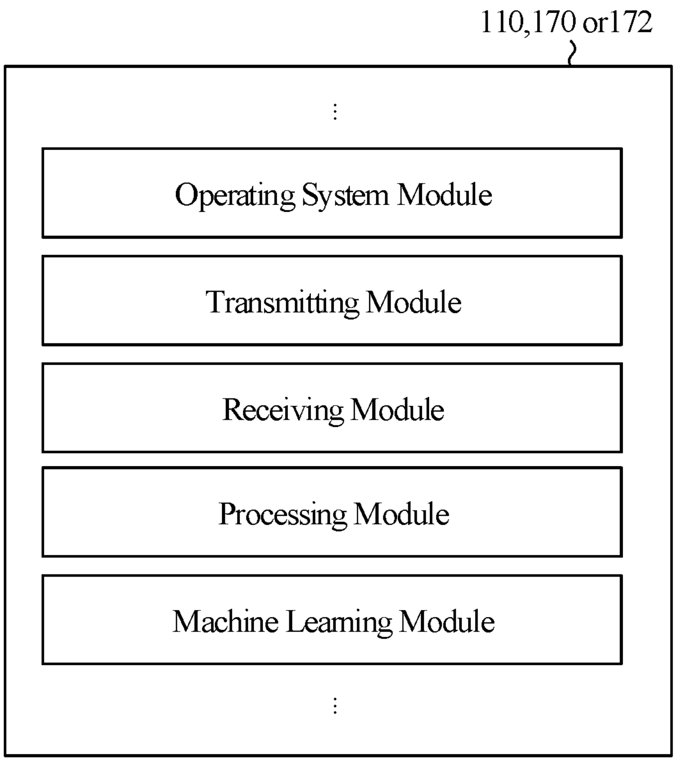
110,170 or172
⋮
Operating System Module
Transmitting Module
Receiving Module
Processing Module
Machine Learning Module
⋮

FIG.4

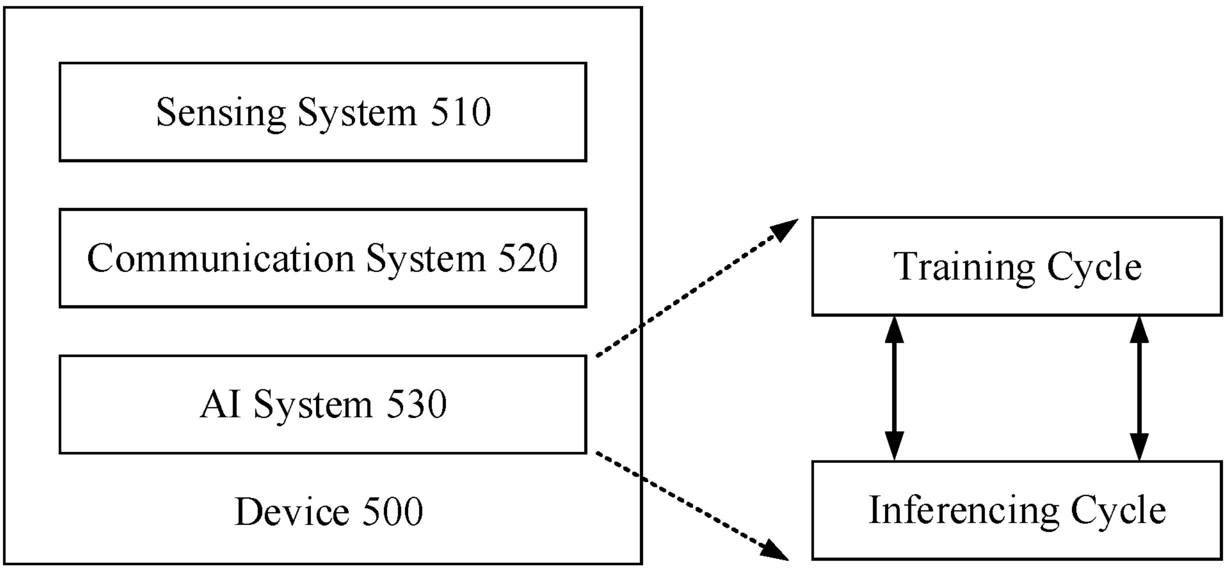
Sensing System 510
Communication System 520
AI System 530
Device 500
Training Cycle
Inferencing Cycle

FIG. 5

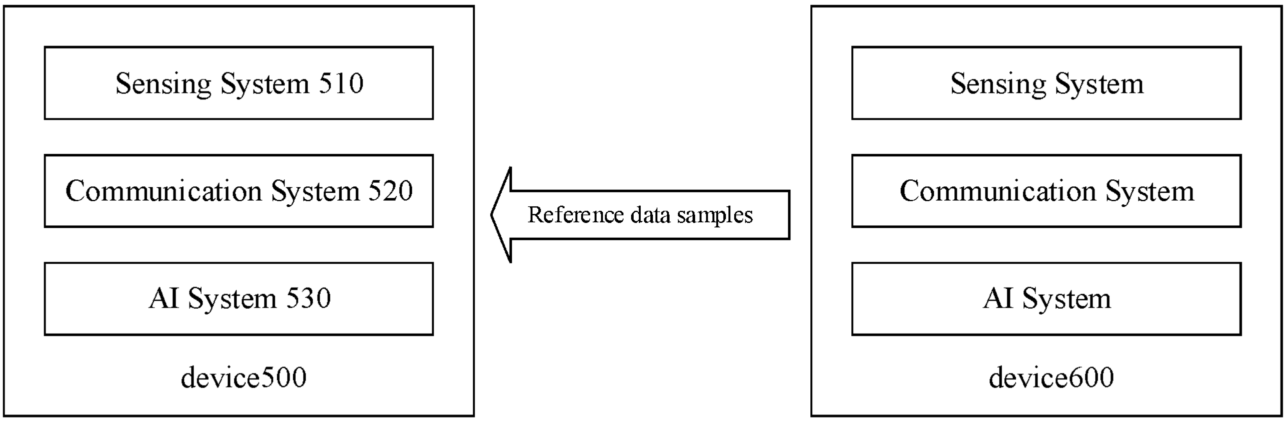
Sensing System 510
Communication System 520
AI System 530
device500
Reference data samples
Sensing System
Communication System
AI System
device600

FIG.6

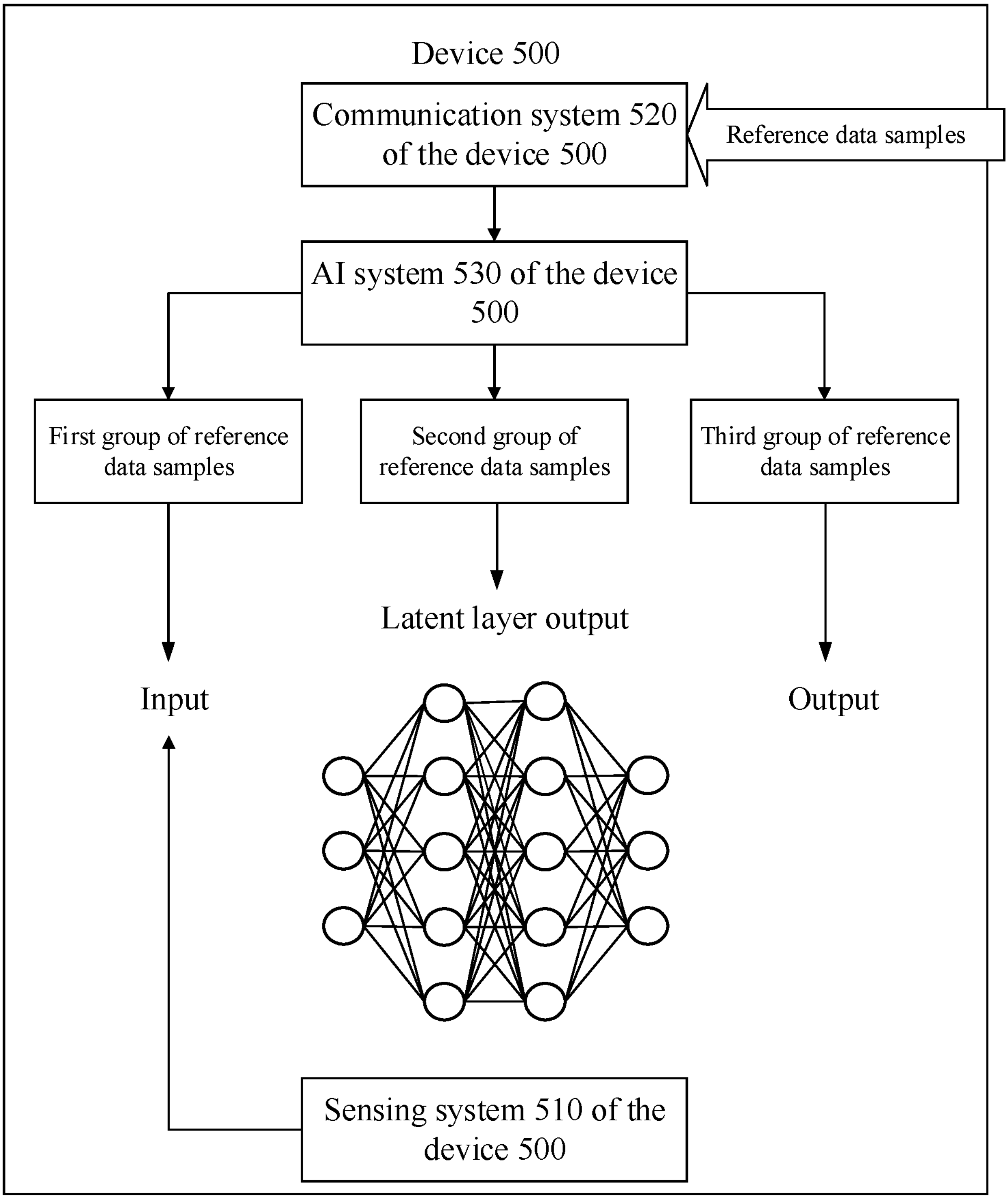
Device 500
Communication system 520 of the device 500
Reference data samples
AI system 530 of the device 500
First group of reference data samples
Second group of reference data samples
Third group of reference data samples
Latent layer output
Input
Output
Sensing system 510 of the device 500

FIG.7

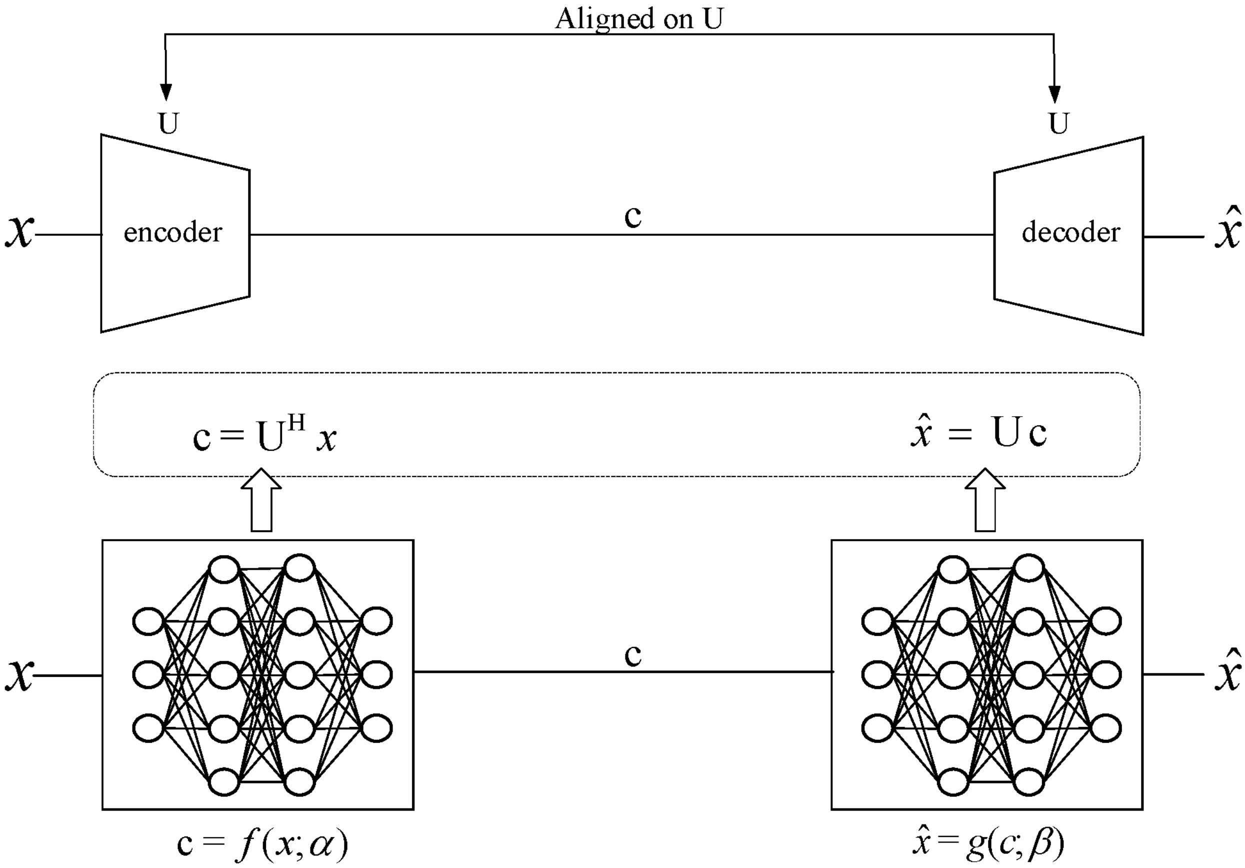
Aligned on U
U
U
x
encoder
c
decoder
$\hat{x}$
$c = U^H x$
$\hat{x} = Uc$
x
c
$\hat{x}$
$c = f(x;\alpha)$
$\hat{x} = g(c;\beta)$

FIG.8

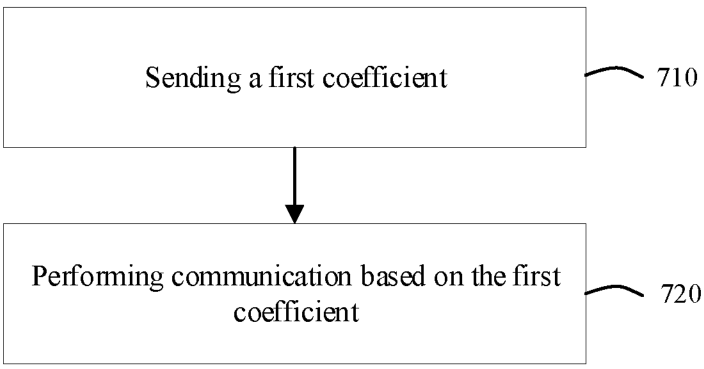
Sending a first coefficient
710
Performing communication based on the first coefficient
720

FIG.9

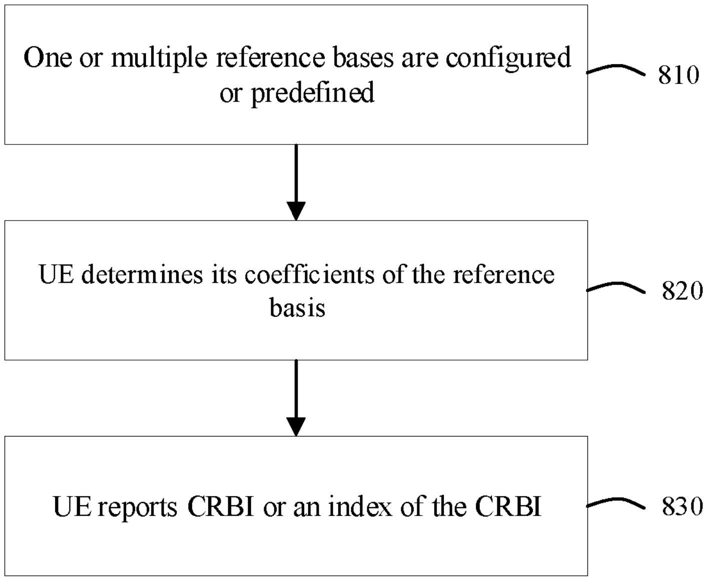
One or multiple reference bases are configured or predefined
810
UE determines its coefficients of the reference basis
820
UE reports CRBI or an index of the CRBI
830

FIG. 10

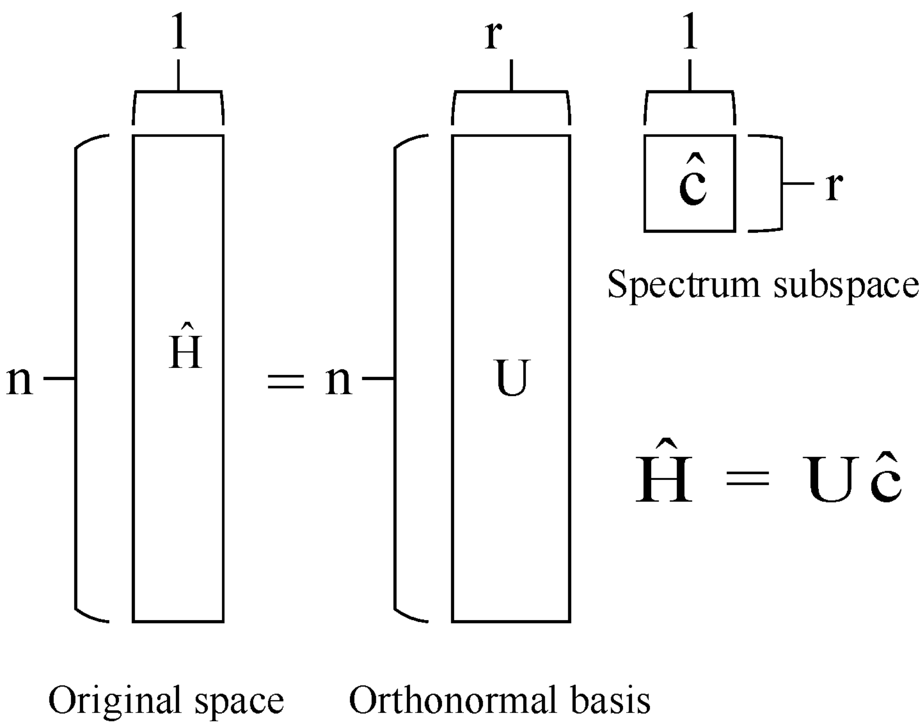
1
r
1
Ĥ
n
= n
U
ĉ
r
Spectrum subspace
Ĥ = U ĉ
Original space
Orthonormal basis

FIG. 11

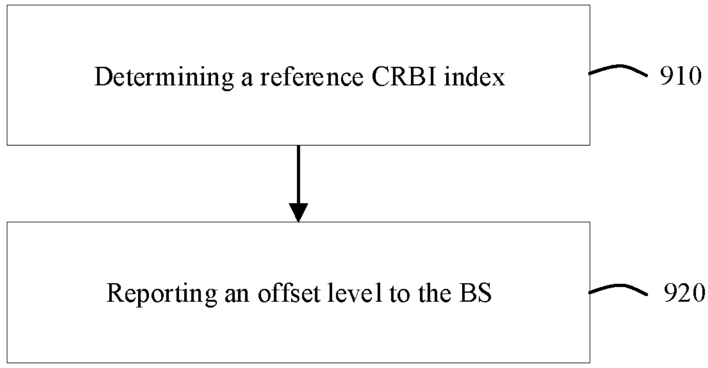
Determining a reference CRBI index
910
Reporting an offset level to the BS
920

FIG.12

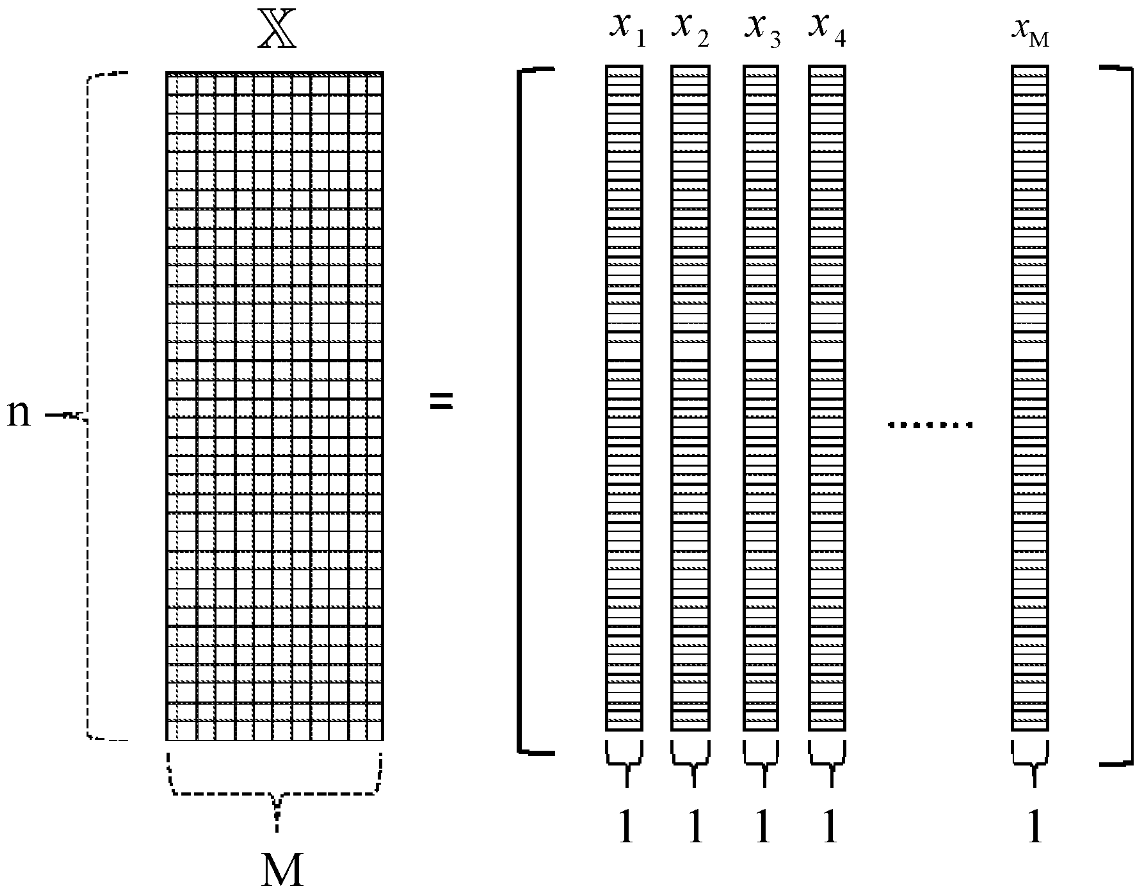
$\mathbb{X}$
n
M
=
x_1
x_2
x_3
x_4
x_M
1
1
1
1
1

FIG.13

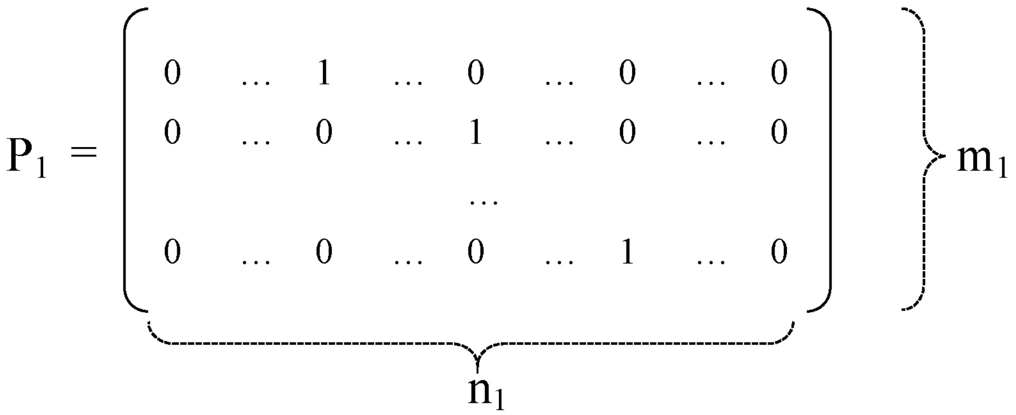
P1 =
0 … 1 … 0 … 0 … 0
0 … 0 … 1 … 0 … 0
…
0 … 0 … 0 … 1 … 0
m1
n1

FIG.14

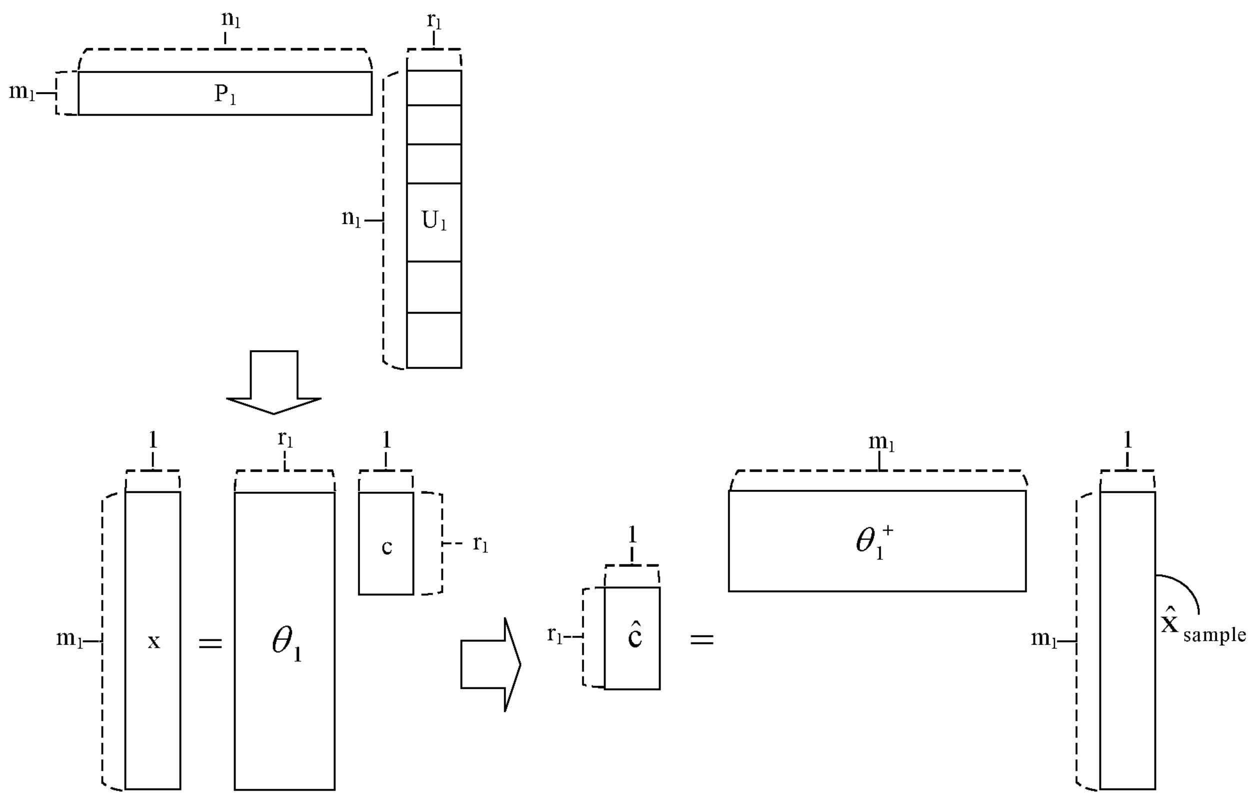
n1
r1
m1
P1
n1
U1
1
r1
1
c
r1
m1
x
=
θ1
1
r1
ĉ
=
m1
θ1+
1
m1
x̂ sample

FIG.15

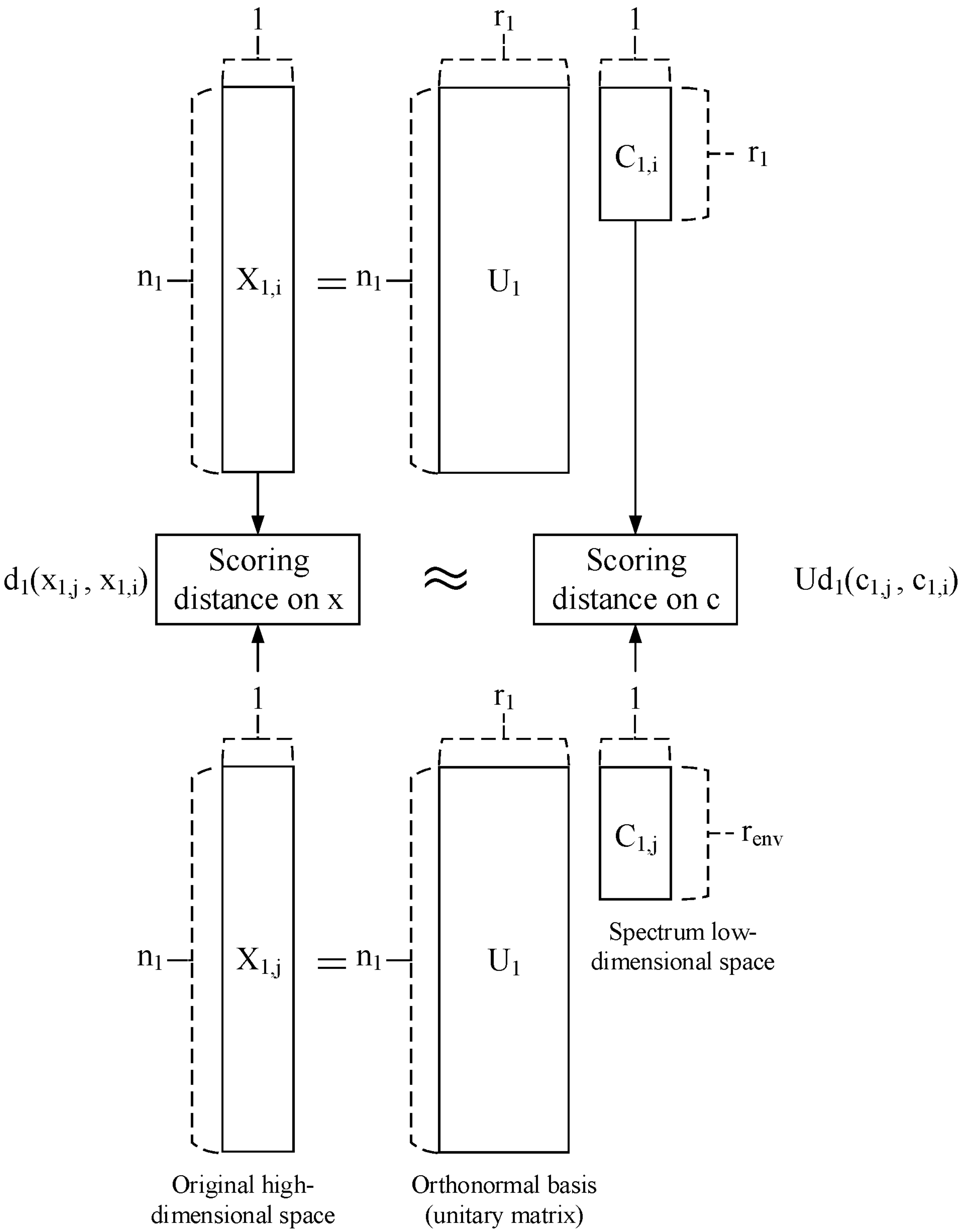
1
r_1
1
$C_{1,i}$
r_1
n_1
$X_{1,i}$
$= n_1$
U_1
$d_1(x_{1,j}, x_{1,i})$
Scoring distance on x
≈
Scoring distance on c
$Ud_1(c_{1,j}, c_{1,i})$
1
r_1
1
$C_{1,j}$
r_{env}
Spectrum low-dimensional space
n_1
$X_{1,j}$
$= n_1$
U_1
Original high-dimensional space
Orthonormal basis (unitary matrix)

FIG.16

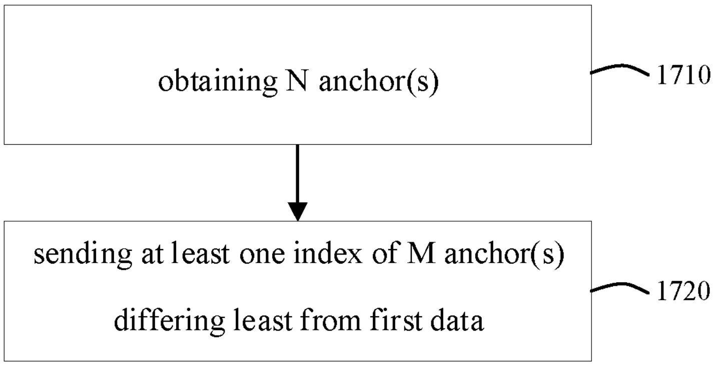
obtaining N anchor(s)
1710
sending at least one index of M anchor(s) differing least from first data
1720

FIG. 17

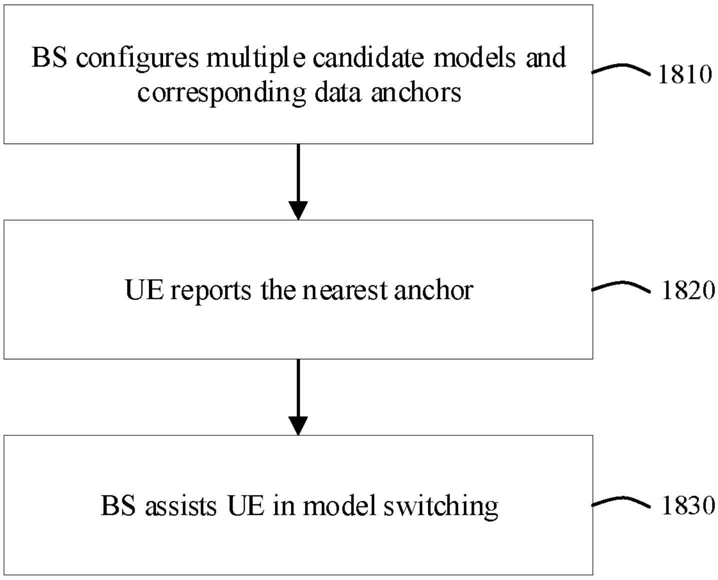
BS configures multiple candidate models and corresponding data anchors
1810
UE reports the nearest anchor
1820
BS assists UE in model switching
1830

FIG. 18

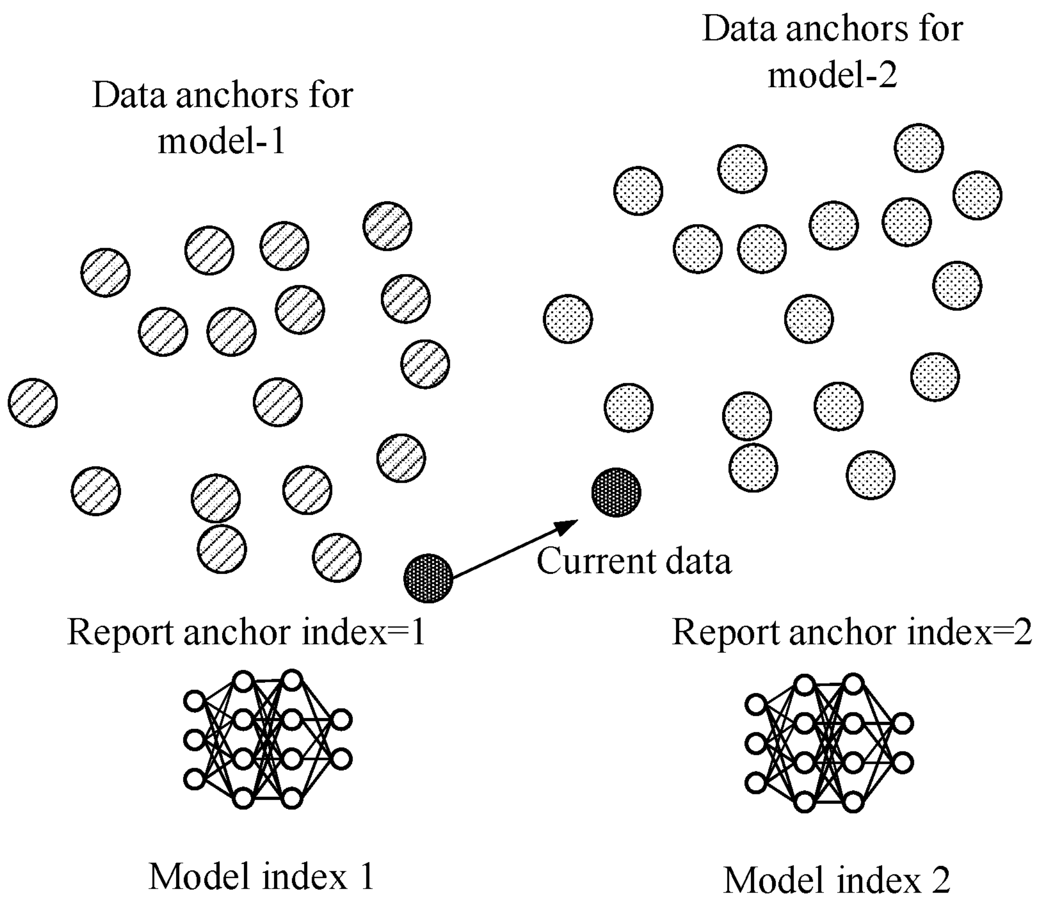
Data anchors for model-1
Data anchors for model-2
Current data
Report anchor index=1
Report anchor index=2
Model index 1
Model index 2

FIG. 19

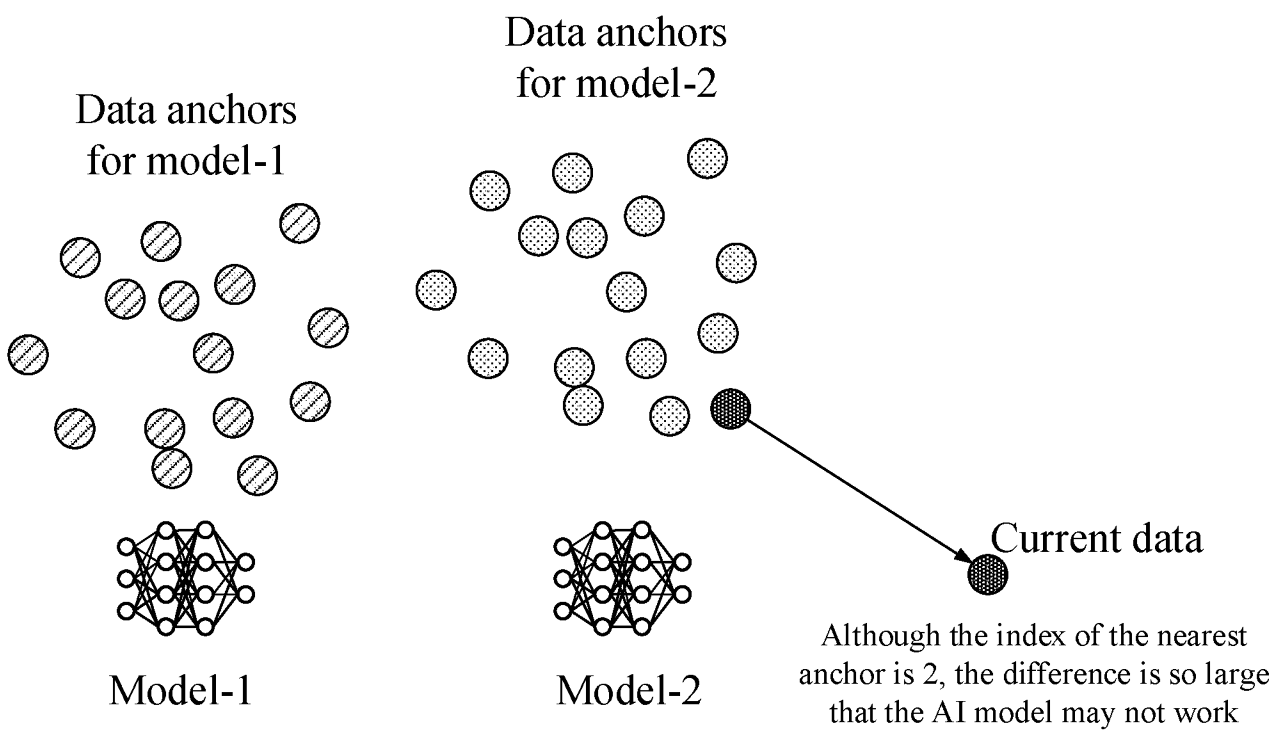
Data anchors for model-1
Data anchors for model-2
Current data
Model-1
Model-2
Although the index of the nearest anchor is 2, the difference is so large that the AI model may not work

FIG. 20

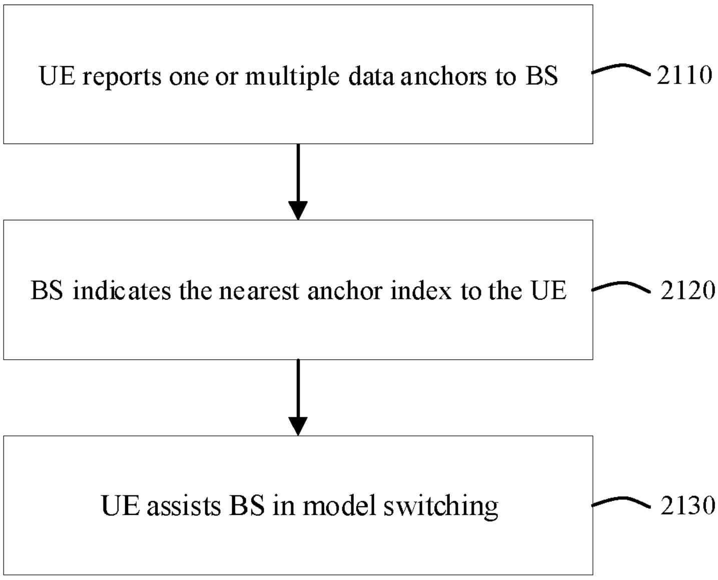

FIG. 21

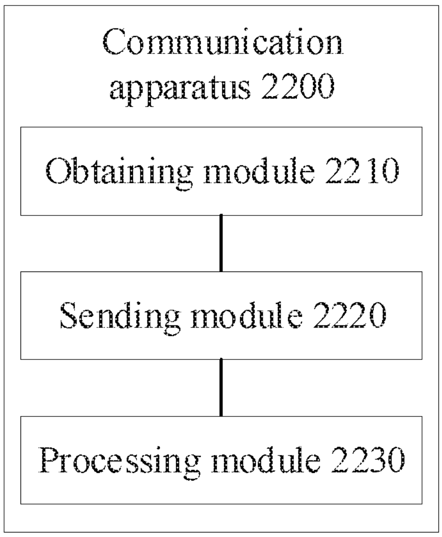

FIG. 22

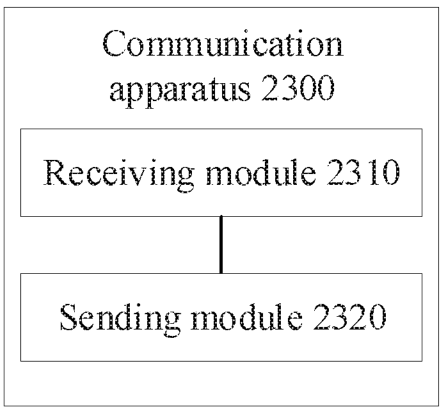

FIG. 23

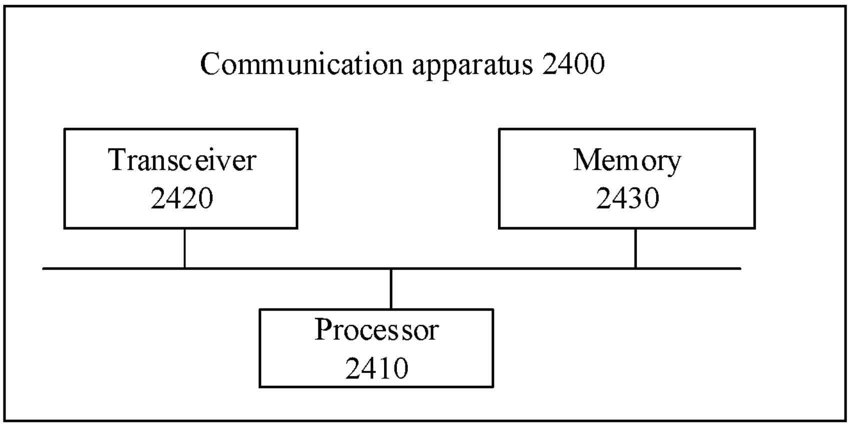
Communication apparatus 2400
Transceiver
2420
Memory
2430
Processor
2410

FIG. 24

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2023/124985

A. CLASSIFICATION OF SUBJECT MATTER

H04W 24/02(2009.01)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)

IPC: H04W; H04L; H04B

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)

CNTXT;CNKI;VEN;ENTXTC;3GPP: Artificial intelligence, Machine Learning, AI/ML, mode, model, switch+, anchor, index

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2023010302 A1 (QUALCOMM INC. et.al) 09 February 2023 (2023-02-09) description paragraphs 88-98	1-41
A	CN 115250502 A (INTEL CORPORATION) 28 October 2022 (2022-10-28) the whole document	1-41
A	US 2022322195 A1 (ERICSSON TELEFON AB L M) 06 October 2022 (2022-10-06) the whole document	1-41
A	WO 2022240089 A1 (LG ELECTRONICS INC.) 17 November 2022 (2022-11-17) the whole document	1-41
A	WO 2023019380 A1 (QUALCOMM INC. et.al) 23 February 2023 (2023-02-23) the whole document	1-41
A	Fraunhofer HHI. "[FS_AI4Media] Scenario for transmission of AI/ML model data" *3GPP TSG SA4 123-e Meeting S4-230565,* 12 April 2023 (2023-04-12), the whole document	1-41

☐ 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
- "D" document cited by the applicant in the international application
- "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
16 January 2024	**20 January 2024**

Name and mailing address of the ISA/CN	Authorized officer
CHINA NATIONAL INTELLECTUAL PROPERTY ADMINISTRATION **6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088, China**	**CAI,GuoLi** Telephone No. **(+86) 010-62411422**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2023/124985

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
WO	2023010302	A1	09 February 2023	None			
CN	115250502	A	28 October 2022	None			
US	2022322195	A1	06 October 2022	EP	3987841	A1	27 April 2022
				WO	2020254859	A1	24 December 2020
WO	2022240089	A1	17 November 2022	KR	20230160875	A	24 November 2023
WO	2023019380	A1	23 February 2023	None			