C L A I M S

What is claimed is:

1. A method for Artificial Intelligence (AI) computing information exchanging, performed by a first network element, the method comprising:

receiving, from a second network element, a first message comprising AI computing information of the second network element, wherein the first network element deploys a first machine learning model for performing load prediction, and wherein the AI computing information comprises at least one of: an input to a second machine learning model deployed on the second network element for load prediction of the second network element; or model configuration information of the second machine learning model, wherein the input to the second machine learning model comprises at least one of:

current load information of a current cell associated with and served by the first network element;

current load information of a neighbor cell of the current cell;

historical load information of the current cell; or

historical load information of the neighbor cell; and

adjusting the first machine learning model according to the AI computing information, such that a load prediction result of the first machine learning model is consistent with a load prediction result of the second machine learning model.

2. The method of claim 1, wherein the model configuration information of the second machine learning model comprises at least one of:

a machine learning model type;

an AI algorithm information;

information of a hardware platform on which the machine learning model is run; or

training configuration information of the machine learning model.

3. The method of claim 2, wherein the machine learning model type comprises at least one of:

a time level model type;

a spatial level model type;

a historical level model type; or

a resemble model type.

4. The method of claim 2, wherein the AI algorithm information indicates at least one of:

an Autoregressive Integrated Moving Average (ARIMA) algorithm;

a Prophet model algorithm;

a Random Forest algorithm;

a Long Short Term Memory (LSTM) algorithm; or

an Ensemble Learning algorithm.

5. The method of claim 2, wherein the hardware platform information comprises at least one of:

a graphic processing unit (GPU) information;

a field programmable Gate Array (FPGA) information;

an application specific integrated circuit (ASIC) information; or

a system on a chip (SoC) information.

6. The method of claim 2, wherein the training configuration information comprises at least one of:

a gradient configuration;

a weight configuration;

a derivation configuration; or

a size of training parameters of the machine learning model.

7. The method of claim 1, wherein the first message comprises one of:

a Next-Generation Radio Access Network (NG-RAN) Node Configuration Update message;

a Next generation nodeB Distribution Unit (gNB-DU) Configuration Update message;

a Next generation nodeB Central Unit (gNB-CU) Configuration Update message;

an EN-DC Configuration Update message; or

an AI Configuration Update message.

8. The method of claim 7, further comprising:

transmitting, to the second network element, a response message to the first message.

9. The method of claim 8, wherein the response message comprises one of:

a NG-RAN Node Configuration Update Acknowledge message;

an AI Configuration Update Acknowledge message;

a gNB-DU Configuration Update Acknowledge message;

a gNB-CU Configuration Update Acknowledge message; or

an EN-DC Configuration Update Acknowledge message.

10. The method of claim 1, wherein the first message comprises a one-way AI Configuration Transfer message.

11. The method of claim 1, wherein the first network element or the second network element comprises a base station.

12. The method of claim 11, wherein the base station comprises at least one of:

a New Generation NodeB (gNB);

an Evolved NodeB (eNB); or

a NodeB.

13. The method of claim 1, wherein the first network element comprises a Distribution Unit (DU)

of a base station and the second network element comprises a Central Unit (CU) of the base station.

14. The method of claim 1, wherein the first network element comprises a Central Unit (CU) of a base station and the second network element comprises a Distribution Unit (DU) of the base station.

15. The method of claim 1, wherein the first network element is configured to be a secondary node and the second network element is configured to be a master node, and wherein the first network element and the second network element form a dual connectivity configuration.

16. The method of claim 1, further comprising causing the second network element to adapt a machine learning model running on the second network element according to the AI computing information message.

17. A first network element comprising one or more processors and one or more memories, wherein the one or more processors are configured to read computer code from the one or more memories to:

receive, from a second network element, a first message comprising AI computing information of the second network element, wherein the first network element deploys a first machine learning model for performing load prediction, and wherein the AI computing information comprises at least one of: an input to a second machine learning model deployed on the second network element for load prediction of the second network element; or model configuration information of the second

machine learning model;

transmit, to the second network element, a response message to the first message;

wherein the input to the second machine learning model comprises at least one of:

current load information of a current cell associated with and served by the first network element;

current load information of a neighbor cell of the current cell;

historical load information of the current cell; or

historical load information of the neighbor cell;

wherein the model configuration information of the machine learning model comprises at least one of:

a machine learning model type;

an AI algorithm information;

information of a hardware platform on which the machine learning model is run; or

training configuration information of the machine learning model;

wherein the first message comprises one of:

a Next-Generation Radio Access Network (NG-RAN) Node Configuration Update message;

a Next generation nodeB Distribution Unit (gNB-DU) Configuration Update message;

a Next generation nodeB Central Unit (gNB-CU) Configuration Update message;

an EN-DC Configuration Update message; or

an AI Configuration Update message; and

wherein the response message comprises one of:

a NG-RAN Node Configuration Update Acknowledge message;

an AI Configuration Update Acknowledge message;

a gNB-DU Configuration Update Acknowledge message;

a gNB-CU Configuration Update Acknowledge message; or

an EN-DC Configuration Update Acknowledge message and

adjust the first machine learning model according to the AI computing information, such that a load prediction result of the first machine learning model is consistent with a load prediction result of the second machine learning model.

18. The first network element of claim 17, wherein:

the AI algorithm information indicates at least one of:

an ARIMA algorithm;

a Prophet model algorithm;

a Random Forest algorithm;

a LSTM algorithm; or

an Ensemble Learning algorithm;

the hardware platform information comprises at least one of:

a GPU information;

an FPGA information;

an ASIC information; or

a SoC information; and

the training configuration information comprises at least one of:

a gradient configuration;

a weight configuration;

a derivation configuration; or

a size of training parameters of the machine learning model.

19. A non-transitory storage medium for storing computer readable instructions, the computer readable instructions, when executed by a processor in a first network element, causing the processor to:

receive, from a second network element, a first message comprising AI computing information of the second network element, wherein the first network element deploys a first machine learning model for performing load prediction, and wherein the AI computing information comprises at least one of: an input to the second machine learning model deployed on the second network element for load prediction of the second network element; or model configuration information of the second machine learning model;

transmit, to the second network element, a response message to the first message from the second network element;

wherein the input to the second machine learning model comprises at least one of:

current load information of a current cell associated with and served by the first network element;

current load information of a neighbor cell of the current cell;

historical load information of the current cell; or

historical load information of the neighbor cell;

wherein the model configuration information of the machine learning model comprises at least one of:

a machine learning model type;

an AI algorithm information;

information of a hardware platform on which the machine learning model is run; or

training configuration information of the machine learning model;

wherein the first message comprises one of:

a Next-Generation Radio Access Network (NG-RAN) Node Configuration Update message;

a Next generation nodeB Distribution Unit (gNB-DU) Configuration Update message;

a Next generation nodeB Central Unit (gNB-CU) Configuration Update message;

an EN-DC Configuration Update message; or

an AI Configuration Update message; and

wherein the response message comprises one of:

a NG-RAN Node Configuration Update Acknowledge message;

an AI Configuration Update Acknowledge message;

a gNB-DU Configuration Update Acknowledge message;

a gNB-CU Configuration Update Acknowledge message; or

an EN-DC Configuration Update Acknowledge message; and

adjust the first machine learning model according to the AI computing information, such that a load prediction result of the first machine learning model is consistent with a load prediction result of the second machine learning model.

20. The non-transitory storage medium of claim 19, wherein:

the AI algorithm information indicates at least one of:

an ARIMA algorithm;

a Prophet model algorithm;

a Random Forest algorithm;

a LSTM algorithm; or

an Ensemble Learning algorithm;

the hardware platform information comprises at least one of:

a GPU information;

an FPGA information;

an ASIC information; or

a SoC information; and

the training configuration information comprises at least one of:

a gradient configuration;

a weight configuration;

a derivation configuration; or

a size of training parameters of the machine learning model.