

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2023年6月1日 (01.06.2023)



(10) 国际公布号
WO 2023/093416 A1

- (51) 国际专利分类号:
H04L 41/16 (2022.01)
- (21) 国际申请号: PCT/CN2022/127034
- (22) 国际申请日: 2022年10月24日 (24.10.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202111407051.2 2021年11月24日 (24.11.2021) CN
- (71) 申请人: 大唐移动通信设备有限公司 (DATANG MOBILE COMMUNICATIONS EQUIPMENT CO., LTD.) [CN/CN]; 中国北京市海淀区上地东路5号院1号楼1层, Beijing 100085 (CN)。
- (72) 发明人: 刘莹莹 (LIU, Yingying); 中国北京市海淀区上地东路5号院1号楼1层, Beijing 100085 (CN)。 王文慧 (WANG, Wenhui); 中国北京市海淀区上地东路5号院1号楼1层, Beijing 100085 (CN)。 段小嫣 (DUAN, Xiaoyan); 中国北京市海淀区上地东路5号院1号楼1层, Beijing 100085 (CN)。
- (74) 代理人: 北京同立钧成知识产权代理有限公司 (LEADER PATENT & TRADEMARK FIRM); 中国北京市海淀区西直门北大街32号枫蓝国际A座8F-6, Beijing 100082 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): 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, 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) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (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,

(54) Title: METHOD AND APPARATUS FOR MODEL TRANSMISSION STATE ANALYSIS IN SUBSCRIPTION NETWORK, AND READABLE STORAGE MEDIUM

(54) 发明名称: 订阅网络中模型传输状态分析方法、装置及可读存储介质

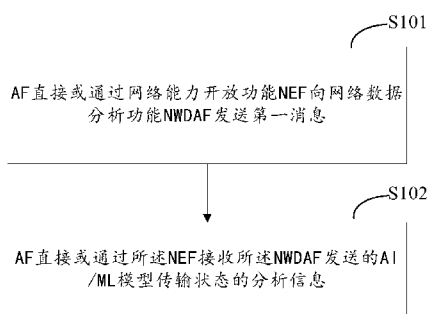


图3

- S101 An AF sends a first message to a network data analysis function (NWDAF) directly or by means of a network exposure function (NEF)
- S102 The AF receives, directly or by means of the NEF, analysis information of an AI/ML model transmission state sent by the NWDAF

(57) Abstract: The present invention provides a method and apparatus for model transmission state analysis in a subscription network and a readable storage medium. The method comprises: an application function (AF) sends a first message to a network data analysis function (NWDAF) directly or by means of a network exposure function (NEF); the AF receives, directly or by means of the NEF, analysis information of an AI/ML model transmission state sent by the NWDAF, the analysis information being determined by the NWDAF according to received data of the AI/ML model transmission state sent by other network function 5GC NF(s), and the analysis information being used for adjusting network strategy parameters and/or information of an application layer model. According to the present invention, the AI/ML model transmission state can be effectively analyzed, so that the network effectively adjusts the network transmission strategy on the basis of the AI/ML model transmission state, and a third party obtains analysis of the AI/ML model transmission state to adjust the application layer information.

CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN,
TD, TG)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：本公开提供一种订阅网络中模型传输状态分析方法、装置及可读存储介质，该方法包括：应用功能AF直接或通过网络能力开放功能NEF向网络数据分析功能NWDAF发送第一消息；AF直接或通过所述NEF接收所述NWDAF发送的AI/ML模型传输状态的分析信息，所述分析信息是所述NWDAF根据接收其他网络功能5GC NF(s)发送的AI/ML模型传输状态的数据确定的；其中，所述分析信息用于调整网络策略参数和/或应用层模型的信息。本公开能够有效地对AI/ML模型传输状态进行分析，进而使得网络基于AI/ML模型传输状态有效地调整网络传输策略以及第三方获得AI/ML模型传输状态的分析进行应用层信息的调整。

订阅网络中模型传输状态分析方法、装置及可读存储介质

本公开要求于 2021 年 11 月 24 日提交中国专利局、申请号为 202111407051.2、申请名称为“订阅网络中模型传输状态分析方法、装置及可读存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本公开中。

技术领域

本公开涉及通信技术领域，尤其涉及一种订阅网络中模型传输状态分析方法、装置及可读存储介质。

背景技术

近年来，由于人工智能的技术突破，人工智能的应用越来越广泛。对于移动终端，由于其具有严格的能耗、计算和内存成本限制，无法在终端运行重量级的人工智能（英文为：Artificial Intelligence，简称为：AI）/机器学习（英文为：Machine Learning，简称为：ML）模型（以下称为 AI/ML 模型），因此，目前采用的手段是将许多 AI/ML 模型的推理从移动终端传输到云或者其他终端，即需要把 AI/ML 模型传输给云或者其他终端。

此外，由于考虑到对传输数据的隐私保护以及缓解网络传输数据的压力等问题，对 AI/ML 模型的传输要求也越来越高。其中，5G 系统作为传输 AI/ML 模型的通道，为了提高 5G 网络的智能化能力，满足 SA1 #93e 通过的 TS 22.261 中 AI/ML 模型在 5G 系统中传输的要求，5G 系统需要支持将关于 AI-ML 会话的监测和状态信息暴露给第三方。

但是，现有技术中，无法有效地对 AI/ML 模型传输状态进行分析，进而使得网络无法基于 AI/ML 模型传输状态有效地调整网络传输策略以及第三方无法获得 AI/ML 模型传输状态的分析进行应用层信息的调整。

发明内容

本公开提供一种订阅网络中模型传输状态分析方法、装置及可读存储介

质，解决了无法有效地对 AI/ML 模型传输状态进行分析，进而使得网络无法基于 AI/ML 模型传输状态有效地调整网络传输策略和第三方无法获得 AI/ML 模型传输状态的分析进行应用层信息的调整的技术问题。

第一方面，本公开提供一种订阅网络中模型传输状态分析方法，所述方法应用于应用功能 AF，所述方法包括：

直接或通过网络能力开放功能 NEF 向网络数据分析功能 NWDAF 发送第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

直接或通过所述 NEF 接收所述 NWDAF 发送的 AI/ML 模型传输状态的分析信息，所述分析信息是所述 NWDAF 根据接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据确定的；

其中，所述分析信息用于调整网络策略参数和/或应用层模型的信息。

可选地，所述 AI/ML 模型传输状态的数据是所述 NWDAF 根据接收到的所述第一消息中请求的参数通过向所述 5GC NF (s) 发送第二消息得到的，所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据。

可选地，所述第一消息中请求的参数包括下述至少一项：网络数据分析标识、接收 AI/ML 模型的一个用户设备 UE 或一组 UE 的标识或满足分析条件的任意 UE、使用 AI/ML 模型的应用的标识、AI/ML 模型传输的区域、指示传输 AI/ML 模型服务质量流的协议数据单元 PDU 会话的网络切片、指示传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、AI/ML 模型传输的时间段、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于指示传输 AI/ML 模型的服务质量流的服务质量要求和/或用于指示传输 AI/ML 模型的特定的服务质量要求；

所述第二消息中包括下述至少一项：使用 AI/ML 模型的 UE 的当前位置、使用 AI/ML 模型的应用的标识、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、数据采集时间、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、

AI/ML 传输模型的大小、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、用于所述 AF 的服务流程；

所述分析信息包括下述至少一项：用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、使用 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、分析结果的有效时间、提供 AI/ML 模型传输的用户面功能 UPF、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称、AI/ML 传输模型的大小、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量；

其中，若 AI/ML 模型执行联邦学习，所述第一消息中请求的参数还包括：联邦学习群信息，所述联邦学习群信息包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识；

相应的，所述第二消息中还包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识；

相应的，所述分析信息还包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识。

本公开实施例中，通过向 NWDAF 发送订阅网络中 AI/ML 模型传输状态的分析信息的请求，接收 NWDAF 发送的根据采集到的 5GC NF (s) 数据确定的分析信息，根据分析信息来调整网络策略参数和/或应用层模型的信息，实现了有效地对 AI/ML 模型传输状态进行分析，进而使得网络能够基于 AI/ML 模型传输状态有效地调整网络传输策略以及使得第三方获得 AI/ML 模型传输状态的分析进行应用层信息的调整。

可选地，在接收到所述分析信息后，所述方法还包括：

根据所述分析信息，直接或通过所述 NEF 向策略控制功能 PCF 发送第一请求；

其中，所述第一请求用于请求更新用于 AI/ML 模型传输的网络策略参数；

5 所述网络策略参数用于优化 AI/ML 模型传输状态。

可选地，所述根据所述分析信息，直接或通过所述 NEF 向策略控制功能 PCF 发送第一请求，包括：

根据所述分析信息中传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及
10 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数中的至少一项，确定传输 AI/ML 模型的新的服务质量参数；所述新的服务质量参数包括下述至少一项：5G 服务质量标识符、反射式服务质量控制、
15 传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级；

根据所述分析信息中的传输 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、使用 AI/ML 模型的应用服务的 IP 地址信息、用于传输 AI/ML
20 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称，确定传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息；或者，若 AI/ML 模型执行联邦学习，根据所述分析信息中的指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识，确定传输 AI/ML
25 模型的 UE(s)和各个 AF 的区域信息及地址信息；

根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，确定数据网络接入标识 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息，所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息用于提供优化 AI/ML 模型传输状态的路径；

30 将所述新的服务质量参数、所述 DNAI 以及所述 DNAI 对应的 UE(s)和各

个 AF 的区域信息及地址信息作为所述第一请求中的参数直接或通过所述 NEF 发送给 PCF。

可选地，所述根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，确定数据网络接入标识 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息，包括：

根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，判断当前路由路径是否不佳；

若当前路由路径不佳，则根据传输 AI/ML 模型的 UE(s)和各个 AF 的地址信息以及 UE(s)的区域信息，确定传输 AI/ML 模型中传输双方的目的地址；

根据所述目的地址，确定最近的路径；

根据最近的路径，确定 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息。

可选地，所述第一请求，具体用于：

请求 PCF 根据所述新的服务质量参数调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级并指示所述 PCF 将调整后的第一更新结果直接或通过 NEF 反馈；其中，所述第一更新结果是 PCF 根据所述新的服务质量参数调整 PCC 规则的结果确定的；

相应的，所述方法还包括：

直接或通过 NEF 接收 PCF 发送的所述第一更新结果，所述第一更新结果包括所述第一请求被接受或所述第一请求被拒绝。

可选地，所述第一请求具体用于：

请求 PCF 确定会话管理功能网元 SMF 是否需要更新会话管理策略，若确定 SMF 需要更新会话管理策略，则确定 PCF 向 SMF 发送第二请求，所述第二请求中请求的参数包括下述至少一项：DNAI、流量导向政策标识符、流量路线信息；所述第二请求用于 SMF 根据新的会话管理策略确定选择的用户面功能 UPF 并提供相应的 DNAI、流量导向政策标识符、流量路线信息；

相应的，所述方法还包括：

直接或通过 NEF 接收 PCF 发送的第二更新结果，所述第二更新结果是 PCF 根据 SMF 发送的新的会话管理策略是否更新 UPF 路径确定的；

其中，所述第二更新结果包括所述第一请求被接受或被拒绝。

本公开实施例中，根据分析信息，确定新的服务质量参数，并将新的服务质量参数发送给 PCF，使得 PCF 依据新的服务质量参数相应调整 PCC 规则或通过 SMF 更新 SM 策略并提供 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息，通过接收 PCF 发送的用于表示第一请求被接受还是被拒绝的通知，能够实现基于 NWDAF 分析结果，请求调整网络策略从而优化 AI/ML 模型传输状态。

10 可选地，在接收到所述分析信息后，所述方法还包括：

根据所述分析信息，调整应用层模型的信息，所述应用层模型的信息包括下述至少一项：模型压缩、模型大小、模型传输时间段、模型编解码；所述应用层模型的信息用于更新服务质量参数；

15 根据调整后的应用层模型的信息，确定新的服务质量参数，所述新的服务质量参数包括：5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级；

直接或通过所述 NEF 向策略控制功能 PCF 发送第三请求；

20 其中，所述第三请求中请求的参数包括所述新的服务质量参数，所述第三请求用于请求更新服务质量参数。

可选地，所述第三请求，具体用于：

25 请求 PCF 根据所述新的服务质量参数调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、优先级；

相应的，所述方法还包括：

30 直接或通过 NEF 接收 PCF 发送的第三更新结果，第三更新结果是 PCF 基于调整 PCC 规则的结果确定的，所述第三更新结果包括所述第三请求被接受或所述第三请求被拒绝。

可选地，在所述调整应用层模型的信息之后，所述方法还包括：

将调整后的应用层模型的信息中的模型压缩、模型大小以及模型编解码直接发送给 PCF，所述调整后的应用层模型的信息用于支持 PCF 调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级；

相应的，所述方法还包括：

接收 PCF 发送的第四更新结果，第四更新结果是 PCF 根据调整后的应用层模型的信息调整 PCC 规则的结果确定的，所述第四更新结果包括所述第三请求被接受或所述第三请求被拒绝。

可选地，在所述调整应用层模型的信息之后，所述方法还包括：

将调整后的应用层模型的信息中的模型传输时间直接发送给 PCF，所述调整后的应用层模型的信息中的模型传输时间用于支持 PCF 调整 PCC 规则中的门状态参数；所述门状态参数用于支持 SMF 根据门状态中的传输开始时间和传输结束时间来更新会话管理策略；

相应的，所述方法还包括：

接收 PCF 发送的第五更新结果，第五更新结果是 PCF 通过接收 SMF 发送的新的会话管理策略的结果确定的，所述第五更新结果包括所述第三请求被接受或所述第三请求被拒绝。

本公开实施例中，根据分析信息，调整应用层模型的信息，依据调整后的应用层模型的信息，确定新的服务质量参数并将新的服务质量参数发送给 PCF，将使得 PCF 依据新的服务质量参数相应调整 PCC 规则；或者，直接将调整后的应用层模型的信息发送给 PCF，使得 PCF 依据模型压缩、模型大小、模型编解码调整 PCC 规则中上述的服务质量参数，或者，使得 PCF 依据模型传输时间调整 PCC 规则中的门状态参数，使得 SMF 根据门状态中的传输开始时间和传输结束时间来更新会话管理策略，通过接收 PCF 发送的用于表示第三请求被接受还是被拒绝的通知，能够实现基于 NWDAF 分析结果，调整应用层模型的信息，进而更新 QoS 要求，从而优化 AI/ML 模型传输状态。

第二方面，本公开提供一种订阅网络中模型传输状态分析方法，所述方

法应用于网络数据分析功能 NWDAF，所述方法包括：

直接或通过网络能力开放功能 NEF 接收应用功能 AF 发送的第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

5 根据所述第一消息中请求的参数，向 5G 核心网的其他网络功能 5GC NF (s) 发送第二消息，所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据；

接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据，并对 AI/ML 模型传输状态的数据进行分析，得到 AI/ML 模型传输状态的分析信息；

10 其中，所述分析信息用于通过 AF 调整网络策略参数和/或应用层模型的信息。

可选地，所述第一消息中请求的参数包括下述至少一项：网络数据分析标识、接收 AI/ML 模型的一个用户设备 UE 或一组 UE 的标识或满足分析条件的任意 UE、使用 AI/ML 模型的应用的标识、AI/ML 模型传输的区域、指示传输 AI/ML 模型服务质量流的协议数据单元 PDU 会话的网络切片、指示传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、AI/ML 模型传输的时间段、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于指示传输 AI/ML 模型的服务质量流的服务质量要求和/或用于指示传输 AI/ML 模型的特定的服务质量要求；

20 所述第二消息中包括下述至少一项：使用 AI/ML 模型的 UE 的当前位置、使用 AI/ML 模型的应用的标识、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、数据采集时间、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、用于所述 AF 的服务流程；

所述分析信息包括下述至少一项：用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、使用 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、分析结果的有效时间、提供 AI/ML 模型传输的用户面功能 UPF、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称、
5 AI/ML 传输模型的大小、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量；

其中，若 AI/ML 模型执行联邦学习，所述第一消息中请求的参数还包括：联邦学习群信息，所述联邦学习群信息包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加
15 联邦学习的应用标识；

相应的，所述第二消息中还包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识；

相应的，所述分析信息还包括下述至少一项：用于指示分析的联邦学
20 习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识。

第三方面，本公开提供一种订阅网络中模型传输状态分析装置，所述装置包括：存储器，收发机，处理器：

存储器，用于存储计算机程序；收发机，用于在所述处理器的控制下
25 收发数据；处理器，用于读取所述存储器中的计算机程序并执行以下操作：

直接或通过网络能力开放功能 NEF 向网络数据分析功能 NWDAF 发送第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

直接或通过所述 NEF 接收所述 NWDAF 发送的 AI/ML 模型传输状态的分析信息，所述分析信息是所述 NWDAF 根据接收 5G 核心网的其他网络功
30

能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据确定的;

其中, 所述分析信息用于调整网络策略参数和/或应用层模型的信息。

第四方面, 本公开提供一种订阅网络中模型传输状态分析装置, 所述装置包括存储器, 收发机, 处理器:

5 存储器, 用于存储计算机程序; 收发机, 用于在所述处理器的控制下收发数据; 处理器, 用于读取所述存储器中的计算机程序并执行以下操作:

直接或通过网络能力开放功能 NEF 接收应用功能 AF 发送的第一消息; 其中, 所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息;

10 根据所述第一消息中请求的参数, 向 5G 核心网的其他网络功能 5GC NF (s) 发送第二消息, 所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据;

接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据, 并对 AI/ML 模型传输状态的数据进行分析, 得到 AI/ML 模型传输状态的分析信息;

15 其中, 所述分析信息用于通过 AF 调整网络策略参数和/或应用层模型的信息。

第五方面, 本公开提供一种订阅网络中模型传输状态分析装置, 所述装置包括:

20 发送单元, 用于直接或通过网络能力开放功能 NEF 向网络数据分析功能 NWDAF 发送第一消息; 其中, 所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息;

分析单元, 用于直接或通过所述 NEF 接收所述 NWDAF 发送的 AI/ML 模型传输状态的分析信息, 所述分析信息是所述 NWDAF 根据接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据确定的;

25 其中, 所述分析信息用于调整网络策略参数和/或应用层模型的信息。

第六方面, 本公开提供一种订阅网络中模型传输状态分析装置, 所述装置包括:

接收单元, 用于直接或通过网络能力开放功能 NEF 接收应用功能 AF 发送的第一消息; 其中, 所述第一消息用于请求订阅网络中人工智能/机器学习

AI/ML 模型传输状态的分析信息;

发送单元, 用于根据所述第一消息中请求的参数, 向 5G 核心网的其他网络功能 5GC NF (s) 发送第二消息, 所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据;

5 分析单元, 用于接收 5G 核心网的其他网络功能 5GC NF(s) 发送的 AI/ML 模型传输状态的数据, 并对 AI/ML 模型传输状态的数据进行分析, 得到 AI/ML 模型传输状态的分析信息;

其中, 所述分析信息用于通过 AF 调整网络策略参数和/或应用层模型的信息。

10 第七方面, 本公开提供一种处理器可读存储介质, 所述处理器可读存储介质存储有计算机程序, 所述计算机程序用于使所述处理器执行第一方面或第二方面任一项所述的方法。

本公开提供一种订阅网络中模型传输状态分析方法、装置及可读存储介质, 直接或通过网络能力开放功能 NEF 向网络数据分析功能 NWDAF 发送第一消息; 其中, 所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息; 直接或通过所述 NEF 接收所述 NWDAF 发送的 AI/ML 模型传输状态的分析信息, 所述分析信息是所述 NWDAF 根据接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据确定的, 所述 AI/ML 模型传输状态的数据是所述 NWDAF 根据接收到的所述第一消息中请求的参数通过向所述 5GC NF (s) 发送第二消息得到的, 所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据; 其中, 所述分析信息用于调整网络策略参数和/或应用层模型的信息。通过向 NWDAF 发送订阅网络中 AI/ML 模型传输状态的分析信息的请求, 接收 NWDAF 发送的根据采集到的 5GC NF (s) 数据确定的分析信息, 根据分析信息来调整网络策略参数和/或应用层模型的信息, 实现了有效地对 AI/ML 模型传输状态进行分析, 进而使得网络能够基于 AI/ML 模型传输状态有效地调整网络传输策略以及使得第三方获得 AI/ML 模型传输状态的分析进行应用层信息的调整。

应当理解, 上述发明内容部分中所描述的内容并非旨在限定本发明的实施例的关键或重要特征, 亦非用于限制本发明的范围。本发明的其它特征将通过以下的描述变得容易理解。

附图说明

为了更清楚地说明本公开或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 为本公开实施例提供的订阅网络中模型传输状态分析方法的网络架构图；

图 2 为本公开实施例提供的支持网络数据分析的 5GC 的网络架构图；

10 图 3 为本公开实施例一提供的订阅网络中模型传输状态分析方法的流程示意图；

图 4 为本公开实施例一提供的订阅网络中模型传输状态分析方法的信令流程示意图；

15 图 5 为本公开实施例二提供的订阅网络中模型传输状态分析方法的信令流程示意图；

图 6 为本公开实施例三提供的订阅网络中模型传输状态分析方法的信令流程示意图；

图 7 为本公开实施例四提供的订阅网络中模型传输状态分析方法的流程示意图；

20 图 8 为本公开实施例提供的订阅网络中模型传输状态分析装置的结构示意图；

图 9 为本公开另一实施例提供的订阅网络中模型传输状态分析装置的结构示意图；

25 图 10 为本公开再一实施例提供的订阅网络中模型传输状态分析装置的结构示意图；

图 11 为本公开又一实施例提供的订阅网络中模型传输状态分析装置的结构示意图。

具体实施方式

30 本公开中术语“和/或”，描述关联对象的关联关系，表示可以存在三种

关系，例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。

下面将结合本公开实施例中的附图，对本公开实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本公开一部分实施例，并不是全部的实施例。基于本公开中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本公开保护的范围。

为了清楚理解本公开的技术方案，首先对现有技术的方案进行详细介绍。现有技术中，在 SA#93e 通过的 SA1 R18 需求中，需要 AI/ML 模型传输的场景至少存在以下几种情况：

10 场景 1、AI/ML 模型的分发和共享。由于任务或者环境等的变化，移动终端的内存等有限，不能预先机载所有的模型，因此需要移动终端通过 5G 系统从网络实时下载新的 AI/ML 模型。

15 场景 2、通过 5GS 的联邦学习算法。云服务器训练一个全局的模型时，需要聚合各个终端设备在本地训练的模型。每次训练迭代过程：一个终端设备从云服务器下载一个全局模型，用本地数据进行训练；终端将中间训练结果上报给云服务器；云服务器聚合来自所有终端的中间训练结果并更新全局模型，然后把全局模型再分发给终端；终端再执行下一次迭代。

20 场景 3、AI/ML 端点之间的 AI/ML 模型切分。一个 AI/ML 模型可以基于当前的任务或者环境切分成多个部分。趋势是将计算复杂，能耗大的部分由网络进行推理，需要隐私保护或者时延敏感的部分在终端推理。例如，终端下载/机载一个模型，先推理特定几层/部分，然后把中间结果发送给网络；网络再执行剩余层/部分，然后把推理结果反馈给终端。该场景在第一步或者中间传输部分模型，因此可能包含模型的传输。

25 因此，5G 系统作为传输 AI/ML 模型的通道，为了提高 5G 网络的智能化能力，满足 SA1 #93e 通过的 TS 22.261 中 AI/ML 模型在 5G 系统中传输的要求，5G 系统需要支持将关于 AI-ML 会话的监测和状态信息暴露给第三方，但是，目前，没有对 AI/ML 模型传输状态的分析，第三方也不能基于 AI/ML 模型传输状态有效地调整自身行为以及网络无法基于 AI/ML 模型传输状态有效地调整网络状态。

30 发明人进一步研究发现，要对 AI/ML 模型传输状态进行有效地分析，需

要应用功能（英文为：Application Function，简称为：AF）、网络能力开放功能（英文为：Network Exposure Function，简称为：NEF）、网络数据分析功能（英文为：Network data analytic function，简称为：NWDAF）以及各个网络功能（英文为：Network Function，简称为：NF）之间进行交互。如图 1 所示，AF 可以直接或通过 NEF 向 NWDAF 发送用于表示订阅网络中 AI/ML 模型传输状态分析的请求，NWDAF 通过向 5G 核心网（英文为：5G Core Network，简称为：5GC）中各个网络功能（英文为：Network Function，简称为：NF）（即 NF（s））采集数据来分析网络中 AI/ML 模型传输状态并反馈，能够实现对 AI/ML 模型传输状态有效地分析，进而使得网络能够基于 AI/ML 模型传输状态有效地调整网络状态和第三方能够获得 AI/ML 模型传输状态的分析进行自身行为数据的调整。

所以基于上述发明人的创造性研究，提出了本公开提出的订阅网络中模型传输状态分析方法，本公开中，直接或通过网络能力开放功能 NEF 向网络数据分析功能 NWDAF 发送第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；直接或通过所述 NEF 接收所述 NWDAF 发送的 AI/ML 模型传输状态的分析信息，所述分析信息是所述 NWDAF 根据接收 5G 核心网的其他网络功能（即 5GC NF（s））发送的 AI/ML 模型传输状态的数据确定的，所述 AI/ML 模型传输状态的数据是所述 NWDAF 根据接收到的所述第一消息中请求的参数通过向所述 5GC NF（s）发送第二消息得到的，所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据；其中，所述分析信息用于调整网络策略参数和/或应用层模型的信息。通过向 NWDAF 发送订阅网络中 AI/ML 模型传输状态的分析信息的请求，接收 NWDAF 发送的根据采集到的 5GC NF（s）数据确定的分析信息，根据分析信息来调整网络策略参数和/或应用层模型的信息，实现了有效地对 AI/ML 模型传输状态进行分析，进而使得网络能够基于 AI/ML 模型传输状态有效地调整网络传输策略以及使得第三方获得 AI/ML 模型传输状态的分析进行应用层信息的调整。

图 2 为本公开实施例提供的支持网络数据分析的 5GC 的网络架构图，如图 2 所示，本公开实施例中，NWDAF 是运营商管理的网络分析功能，NWDAF 能够向 5GC 网络功能、应用功能（英文为：Application Function，简称为：

AF) 和操作管理维护 (英文为: Operation Administration and Maintenance, 简称为: OAM) 提供数据分析服务。其中, 分析结果可以是历史统计信息或者预测信息。NWDAF 可以服务一个或多个网络切片。

其中, 在 5GC 中还包括其他多种功能。分别为用户平面功能(英文为: User Plane Function, 简称为: UPF)、会话管理功能(英文为: Session Management Function, 简称为: SMF)、接入和移动性管理功能 (英文为: Access and Mobility Management Function, 简称为: AMF)、统一数据库(英文为: Unified Data Repository, 简称为: UDR)、网络能力开放功能 (英文为: Network Exposure Function, 简称为: NEF)、AF、策略控制功能 (英文为: Policy Control Function, 简称为: PCF)及在线计费系统 (英文为: Online Charging System, 简称为: OCS)。其中, 这些其他功能均可统称为 NF。NWDAF 基于服务化接口与 5G 核心网中其他功能实体 5GC NF (s) 及 OAM 进行通信。

5GC 中可以有不同 NWDAF 实例提供不同类型的专用分析。为了让消费者 NF 能够发现合适的 NWDAF 实例来提供特定类型的分析, NWDAF 实例需在向网络数据库功能(英文为: Network Repository Function, 简称为: NRF) 注册时提供其支持 Analytic ID, Analytic ID 代表了分析类型 (或分析标识)。这样消费者 NF 可以在向 NRF 查询 NWDAF 实例时, 提供 Analytic ID 来指示需要何种类型的分析。5GC 网络功能和 OAM 决定如何使用网络数据分析功能 NWDAF 提供的数据分析来提高网络性能。

本公开实施例中, 在一种应用场景中, AF 请求 NWDAF 提供 AI/ML 模型传输状态分析, 分析结果 (或分析信息) 包含: 使用 AI/ML 模型的应用标识 (即 Application ID)、使用 AI/ML 模型的区域信息、传输 AI/ML 模型的时间段、所传输模型的大小、传输 AI/ML 模型的服务质量 (即 QoS) 相关信息、AI/ML 模型传输使用的网络切片、数据网络名称 (英文为: Data Network Name, 简称为: DNN) 信息; 如果有联邦学习, 还包括: 组标识 (即联邦学习 group ID)、参与联邦学习的 UE ID or UE group ID、提供模型或者参与联邦学习的应用服务器的地址信息。AF 根据 NWDAF 提供的数据分析, 请求调整 5GS 的网络策略或者调整应用层 AI/ML 模型参数来优化 AI/ML 模型传输状态, 5GC NF(s) 基于请求调整网络策略或者 AF 基于数据分析调整应用层 AI/ML 模型参数。

其中，AF 向 NWDAF 发送分析请求时，若 AF 在受信区，AF 可以直接向 NWDAF 发送请求；若 AF 在受信区，AF 可以通过 NEF 向 NWDAF 发送请求，即 AF 向 NEF 发送请求，再由 NEF 向 NWDAF 发送请求。

因此，直接或通过 NEF 向 NWDAF 发送订阅网络中 AI/ML 模型传输状态的分析信息的请求，接收 NWDAF 发送的根据采集到的 5GC NF (s) 数据确定的分析信息，根据分析信息来调整网络策略参数和/或应用层模型的信息，实现了有效地对 AI/ML 模型传输状态进行分析，进而使得网络能够基于 AI/ML 模型传输状态有效地调整网络传输策略以及使得第三方获得 AI/ML 模型传输状态的分析进行应用层信息的调整。

10 以下将参照附图来描述本公开的实施例。

图 3 为本公开实施例一提供的订阅网络中模型传输状态分析方法的流程示意图，如图 3 所示，本实施例提供的订阅网络中模型传输状态分析方法的执行主体为 AF，则本公开实施例提供的订阅网络中模型传输状态分析方法包括以下步骤：

15 步骤 101、AF 直接或通过网络能力开放功能 NEF 向网络数据分析功能 NWDAF 发送第一消息。

其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息。

本实施例中，第一消息中请求的参数包括下述至少一项：网络数据分析标识（即 Analytics ID）、接收 AI/ML 模型的一个用户设备 UE 或一组 UE 的标识或满足分析条件的任意 UE（即 Target of Analytics Reporting）、使用 AI/ML 模型的应用的标识（即 Application ID）、AI/ML 模型传输的区域（即 AoI (Area of Interest)）、指示传输 AI/ML 模型服务质量流的协议数据单元 PDU 会话的网络切片（即 S-NSSAI）、指示传输 AI/ML 模型服务质量流的 PDU 会话的数据网络（即 DNN）、AI/ML 模型传输的时间段（即 Model transmission duration）、AI/ML 模型传输的开始时间戳（即 Model transmission start）、AI/ML 模型传输的结束时间戳（即 Model transmission stop）、AI/ML 传输模型的大小（即 Model size）、用于指示传输 AI/ML 模型的服务质量流的服务质量要求（即 5QI (5G QoS Identifier)）和/或用于指示传输 AI/ML 模型的特定的服务质量要求（即 QoS Characteristics）。

20
25
30

其中，特定的服务质量要求，例如包传输时延、误包率等。参见下述表 1 所示的第一消息中请求的参数示例表。

表 1: 第一消息中请求的参数示例表

Parameter Name (即参数名)	Parameter Value (即参数值)	Parameter Description (即参数描述)
Analytics ID	定义新的 Analytics ID, 例如 AI/ML 模型传输状态 AI/ML Model Transfer Status, 或 AI/ML 模型传输性能 AI/ML Model Transfer Performance;	网络数据分析标识
Target of Analytics Reporting	UE ID, UE group ID, or “any UE”	接收 AI/ML 模型的一个 UE 或一组 UE 的标识, 或满足分析条件(例如 AoI 内)的任意 UE
Application ID	使用该 AI/ML 模型的应用的 Application ID	使用 AI/ML 模型的应用的标识
AoI (Area of Interest)		AI/ML 模型传输的区域
S-NSSAI		指示传输 AI/ML 模型 QoS flow 的 PDU Session 的网络切片
DNN		指示传输 AI/ML 模型 QoS flow 的 PDU Session 的数据网络
Model transmission duration		AI/ML 模型传输的时间段
Model transmission start		AI/ML 模型传输的开始时间戳
Model transmission stop		AI/ML 模型传输的结束时间戳
Model size		AI/ML 传输模型的大小
QoS requirements	包括以下参数的一种或多种:	
> 5QI (5G QoS Identifier)	传输该 AI/ML 模型的 QoS flow 的 5QI, 以及适用的附加参数(例如 GBR 5QI 的 GFBR(保证的流比特率)等)	用于指示传输 AI/ML 模型的 QoS flow 的 QoS 要求
> QoS Characteristics	QoS flow 的资源类型 Resource Type (GBR, non-GBR), PDB (Packet Delay Budget)和 PER (Packet Error Rate)值	用于指示传输 AI/ML 模型的特定的 QoS 要求, 例如包传输时延、误包率等

如果存在联邦学习，第一消息中请求的参数还可以包括：联邦学习群信息 (即 Federated Learning (FL) group information)；联邦学习群信息包括下述

至少一项：用于指示分析的联邦学习组的标识（即 Federated Learning (FL) group ID）、参加联邦学习的 UE 标识或 UE(s) 标识（即 Federated Learning (FL) UE ID or UE group ID）、参加联邦学习的应用标识（即 Federated Learning (FL) Application ID）。参见下述表 2 所示的第一消息中请求的参数示例表。

5

表 2: 第一消息中请求的参数示例表

Parameter Name	Parameter Value	Parameter Description
Analytics ID	定义新的 Analytics ID, 例如 AI/ML 模型传输状态 AI/ML Model Transfer Status, 或 AI/ML 模型传输性能 AI/ML Model Transfer Performance;	网络数据分析标识
Target of Analytics Reporting	UE ID, UE group ID, or “any UE”	接收 AI/ML 模型的一个 UE 或一组 UE 的标识, 或满足分析条件(例如 AoI 内)的任意 UE
Application ID	使用该 AI/ML 模型的应用的 Application ID	使用 AI/ML 模型的应用的标识
AoI (Area of Interest)		AI/ML 模型传输的区域
S-NSSAI		指示传输 AI/ML 模型 QoS flow 的 PDU Session 的网络切片
DNN		指示传输 AI/ML 模型 QoS flow 的 PDU Session 的数据网络
Model transmission duration		AI/ML 模型传输的时间段
Model transmission start		AI/ML 模型传输的开始时间戳
Model transmission stop		AI/ML 模型传输的结束时间戳
Model size		AI/ML 传输模型的大小
Federated Learning (FL) group information		如果有, 联邦学习群信息
>Federated Learning (FL) group ID		如果有, 联邦学习组标识, 用于指示分析的联邦学习组
>Federated Learning (FL) UE ID or UE group ID		如果有, 参加联邦学习的 UE(s)
>Federated Learning (FL) Application ID		如果有, 参加联邦学习的应用标识
QoS requirements	包括以下参数的一种或多种:	
> 5QI (5G QoS Identifier)	传输该 AI/ML 模型的 QoS	用于指示传输 AI/ML 模

	flow 的 5QI, 以及适用的附加参数(例如 GBR 5QI 的 GFBR(保证的流比特率)等)	型的 QoS flow 的 QoS 要求
> QoS Characteristics	QoS flow 的资源类型 Resource Type (GBR, non-GBR), PDB (Packet Delay Budget)和 PER (Packet Error Rate)值	用于指示传输 AI/ML 模型的特定的 QoS 要求, 例如包传输时延、误包率等

本实施例中, 如果 AF 是不受信的 (即 AF 不在受信区), AF 向 NEF 发送 AI/ML 模型传输状态订阅请求, 比如 Nnef_AnalyticsExposure_Subscribe (即分析开放订阅) or Nnef_AnalyticsExposure_Fetch (即分析开放获取) 请求。NEF 向 NWDAF 发送第一消息, 该第一消息可以为 AI/ML 模型开放传输状态

5 订阅 Nnwdaf_AnalyticsSubscription_Subscribe (即分析订阅订阅) or Nnwdaf_AnalyticsInfo_Request (即分析信息) 请求, 该请求中可以携带如表所示参数, 请求订阅网络中 AI/ML 模型传输状态的分析信息。如果 AF 是受信的 (即 AF 在受信区), AF 直接向 NWDAF 发送第一消息。

其中, 该 AI/ML 模型传输状态订阅请求中可以携带如表 1 或表 2 所示参

10 数, 请求订阅网络中 AI/ML 模型传输状态的分析信息。

步骤 102、AF 直接或通过所述 NEF 接收所述 NWDAF 发送的 AI/ML 模型传输状态的分析信息。

其中, 所述分析信息是所述 NWDAF 根据接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据确定的。

15 可选地, 所述 AI/ML 模型传输状态的数据是所述 NWDAF 根据接收到的所述第一消息中请求的参数通过向所述 5GC NF (s) 发送第二消息得到的, 所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据。

本实施例中, 第二消息中包括下述至少一项: 使用 AI/ML 模型的 UE 的当前位置 (即 UE location)、使用 AI/ML 模型的应用的标识 (即

20 Application ID, 可以是服务器的标识, 也可以是 AF 的标识)、传输 AI/ML 模型的服务质量流标识符 (即 QFI)、传输 AI/ML 模型的上行链路方向比特率 (即 bit rate for UL direction) 及传输 AI/ML 模型的下行链路方向比特率 (即 bit rate for DL direction)、AI/ML 模型的上行链路方向分组延迟 (即 Packet delay for UL direction) 及 AI/ML 模型的下行链路方向分组延迟 (即

25 Packet delay for the DL direction)、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量 (QoS Sustainability)、AI/ML 模型的分组传输数量 (packet transmission)、AI/ML 模型的分组重传数量 (即 packet

- retransmission)、数据采集时间(即 Timestamp)、AI/ML 模型传输的时长(即 AI/ML 模型传输的时间段)、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、用于所述 AF 的服务流程(即 IP filter information)。
- 5 参见下述表 3 所示的第二消息示例表。

表 3: 第二消息示例表

Information (即数据)	Source (即来源)	Description (即数据描述)
UE location	AMF	使用 AI/ML 模型的 UE 的当前位置
Application ID	SMF, AF	使用 AI/ML 模型的应用的标识
QoS requirements		
>QFI	SMF	传输 AI/ML 模型的 QoS flow identifier
>QoS flow Bit Rate	UPF	传输 AI/ML 模型的 bit rate for UL direction; and 传输 AI/ML 模型的 bit rate for DL direction.
>QoS flow Packet Delay	UPF	AI/ML 模型的 Packet delay for UL direction; and AI/ML 模型的 Packet delay for the DL direction.
>QoS Sustainability	OAM TS 28.554 [10]	QoS Flows 在 AI/ML 模型传输的时间段内异常释放的数量
>Packet transmission	UPF	AI/ML 模型的 packet transmission 数量
>Packet retransmission	UPF	AI/ML 模型的 packet retransmission 数量
Timestamp	UPF, AF	数据采集时间
Model transmission duration	UPF, AF	AI/ML 模型传输的时长
Model transmission start	UPF, AF	AI/ML 模型传输的开始时间戳
Model transmission stop	UPF, AF	AI/ML 模型传输的结束时间戳
Model size	UPF, AF	AI/ML 传输模型的大小
S-NSSAI	SMF	用于传输 AI/ML 模型 QoS flow 的 PDU Session 的网络切片
DNN	SMF	用于传输 AI/ML 模型 QoS flow 的 PDU Session 的 DNN
IP filter information	AF	明确用于该应用的 service flow

- 10 如果存在联邦学习, 所述第二消息中还包括下述至少一项: 指示用于指定分析的联邦学习组的标识(即 Federated Learning (FL) group ID)、参加联邦学习的 UE 或 UE(s)(即 Federated Learning (FL) UE ID or UE group ID)、参加联邦学习的应用标识(即 Federated Learning (FL) Application ID)。参见下述表 4 所示的第二消息示例表。

表 4: 第二消息示例表

Information (即数据)	Source (即来源)	Description (即数据描述)
UE location	AMF	使用 AI/ML 模型的 UE 的当前位置
Application ID	SMF, AF	使用 AI/ML 模型的应用的标识
Federated Learning (FL) group ID	SMF, AF	如果有, 指示分析的联邦学习组
Federated Learning (FL) UE ID or UE group ID	SMF, AF	如果有, 参加联邦学习的 UE(s)
Federated Learning (FL) Application ID	SMF, AF	如果有, 参加联邦学习的应用标识
QoS requirements		
>QFI	SMF	传输 AI/ML 模型的 QoS flow identifier
>QoS flow Bit Rate	UPF	传输 AI/ML 模型的 bit rate for UL direction; and 传输 AI/ML 模型的 bit rate for DL direction.
>QoS flow Packet Delay	UPF	AI/ML 模型的 Packet delay for UL direction; and AI/ML 模型的 Packet delay for the DL direction.
>QoS Sustainability	OAM TS 28.554 [10]	QoS Flows 在 AI/ML 模型传输的时间段内异常释放的数量
>Packet transmission	UPF	AI/ML 模型的 packet transmission 数量
>Packet retransmission	UPF	AI/ML 模型的 packet retransmission 数量
Timestamp	UPF, AF	数据采集时间
Model transmission duration	UPF, AF	AI/ML 模型传输的时长
Model transmission start	UPF, AF	AI/ML 模型传输的开始时间戳
Model transmission stop	UPF, AF	AI/ML 模型传输的结束时间戳
Model size	UPF, AF	AI/ML 传输模型的大小
S-NSSAI	SMF	用于传输 AI/ML 模型 QoS flow 的 PDU Session 的网络切片
DNN	SMF	用于传输 AI/ML 模型 QoS flow 的 PDU Session 的 DNN
IP filter information	AF	明确用于该应用的 service flow

本实施例中, 如果 AF 是不受信的, 接收 NEF 发送的 AI/ML 模型传输状态的分析信息, 该分析信息是 NWDAF 向 NEF 发送的。如果 AF 是受信的, 直接接收 NWDAF 发送的 AI/ML 模型传输状态的分析信息。

- 5 其中, 所述分析信息包括下述至少一项: 用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、使用 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、分析结果的有效时间 (即 Validity period)、提供 AI/ML 模型传输的用户面功能 UPF (即 UPF Info)、用于传输 AI/ML 模型服务质量流的

PDU 会话的数据网络名称、AI/ML 传输模型的大小、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、服务质量要求（即 QoS requirements）；服务质量要求包括：传输 AI/ML 模型的服务质量流标识符（即 QFI）、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量。参见下述表 5 所示的分析信息示例表。

10

表 5: 分析信息示例表

Information	Description
AI/ML 模型传输状态的分析信息列表	
> S-NSSAI	用于传输 AI/ML 模型 QoS flow 的 PDU Session 的网络切片
> Application ID	使用 AI/ML 模型的应用的标识
> Validity area	使用 AI/ML 模型的区域信息
> Validity period	分析结果的有效时间
> UPF Info	提供 AI/ML 模型传输的 UPF
> DNN	用于传输 AI/ML 模型 QoS flow 的 PDU Session 的 DNN
> Model size	AI/ML 传输模型的大小
> Model transmission duration	AI/ML 模型传输的时长
> Model transmission start	AI/ML 模型传输的开始时间戳
> Model transmission stop	AI/ML 模型传输的结束时间戳
> QoS requirements	QoS 要求
>> QFI	传输 AI/ML 模型的 QoS flow identifier
>> QoS flow Bit Rate	传输 AI/ML 模型的 bit rate for UL direction; and 传输 AI/ML 模型的 bit rate for DL direction.
>> QoS flow Packet Delay	AI/ML 模型的 Packet delay for UL direction; and AI/ML 模型的 Packet delay for the DL direction.

>>QoS Sustainability	QoS Flows 在 AI/ML 模型传输的时间段内异常释放的数量
>>QoS Sustainability Reporting Threshold(s)	达到 QoS Flows 在 AI/ML 模型传输时间段内异常释放的报告阈值的次数
>>Packet transmission	AI/ML 模型的 packet transmission 数量
>>Packet retransmission	AI/ML 模型的 packet retransmission 数量

如果存在联邦学习，所述分析信息还包括下述至少一项：用于指定分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s) 标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识（即 Application Server Instance Address）。参见下述表 6 所示的分析信息示例表。

5

表 6: 分析信息示例表

Information	Description
AI/ML 模型传输状态的分析信息列表	
> S-NSSAI	用于传输 AI/ML 模型 QoS flow 的 PDU Session 的网络切片
> Application ID	使用 AI/ML 模型的应用的标识
> Validity area	使用 AI/ML 模型的区域信息
> Validity period	分析结果的有效时间
> UPF Info	提供 AI/ML 模型传输的 UPF
> DNN	用于传输 AI/ML 模型 QoS flow 的 PDU Session 的 DNN
> Model size	AI/ML 传输模型的大小
> Model transmission duration	AI/ML 模型传输的时长
> Model transmission start	AI/ML 模型传输的开始时间戳
> Model transmission stop	AI/ML 模型传输的结束时间戳
> Federated Learning (FL) group ID	如果有，指示联邦学习的 group ID
>> List of FL UE(s)	如果有，参与联邦学习的 UE ID or UE group ID
>> Application Server Instance Address	如果有，指示提供模型或者参与联邦学习的 Application Server Instance(s) (IP address of the Application Server) or FQDN of Application Server.

>QoS requirements	QoS 要求
>>QFI	传输 AI/ML 模型的 QoS flow identifier
>>QoS flow Bit Rate	传输 AI/ML 模型的 bit rate for UL direction; and 传输 AI/ML 模型的 bit rate for DL direction.
>>QoS flow Packet Delay	AI/ML 模型的 Packet delay for UL direction; and AI/ML 模型的 Packet delay for the DL direction.
>>QoS Sustainability	QoS Flows 在 AI/ML 模型传输的时间段内异常释放的数量
>>QoS Sustainability Reporting Threshold(s)	达到 QoS Flows 在 AI/ML 模型传输时间段内异常释放的报告阈值的次数
>>Packet transmission	AI/ML 模型的 packet transmission 数量
>>Packet retransmission	AI/ML 模型的 packet retransmission 数量

本实施例中，通过向 NWDAF 发送订阅网络中 AI/ML 模型传输状态的分析信息的请求，接收 NWDAF 发送的根据采集到的 5GC NF (s) 数据确定的分析信息，根据分析信息来调整网络策略参数和/或应用层模型的信息，实现了有效地对 AI/ML 模型传输状态进行分析，进而使得网络能够基于 AI/ML 模型传输状态有效地调整网络传输策略以及使得第三方获得 AI/ML 模型传输状态的分析进行应用层信息的调整。

示例性地，参见图 4 所示，图 4 为本公开实施例一提供的订阅网络中模型传输状态分析方法的信令流程示意图，图 4 为订阅网络中模型传输状态分析方法中 AF 与 NWDAF、NEF 和 NF 间的信令交互图。本实施例提供的订阅网络中模型传输状态分析方法包括以下步骤（即实施例一对应的信令交互过程：AF 请求 NWDAF 提供 AI/ML 模型传输状态分析）：（其中，步骤 4011 至 4016 是 AF 在不受信任区域，步骤 4021 至 4024 是 AF 在受信任区域。）

步骤 4011、如果 AF 是不受信的，AF 向 NEF 发送 AI/ML 模型传输状态订阅 Nnef_AnalyticsExposure_Subscribe（即分析开放订阅） or Nnef_AnalyticsExposure_Fetch（即分析开放获取）请求。

本实施例中，该请求中可以携带如表 1 或表 2 所示参数，请求订阅网络中 AI/ML 模型传输状态的分析信息。

步骤 4012、NEF 向 NWDAF 发送 AI/ML 模型开放传输状态订阅 Nnwdaf_AnalyticsSubscription_Subscribe（即分析订阅订阅） or

Nnwdaf_AnalyticsInfo_Request (即分析信息) 请求。

其中, AI/ML 模型开放传输状态订阅 Nnwdaf_AnalyticsSubscription_Subscribe or Nnwdaf_AnalyticsInfo_Request 请求可以作为第一消息。

5 本实施例中, 该请求中可以携带如表 1 或表 2 所示参数, 请求订阅网络中 AI/ML 模型传输状态的分析信息。

步骤 4013、NWDAF 调用 Nnf_EventExposure_Subscribe (即事件开放订阅) 向 5GC NF(s) 采集数据。

10 其中, 采集数据如表 3 或表 4 所示, 用于分析网络中 AI/ML 模型传输状态。NWDAF 向所述 5GC NF (s) 发送第二消息的方式可以是 NWDAF 调用 Nnf_EventExposure_Subscribe。

步骤 4014、5GC NF(s) 调用 Nnf_EventExposure_Notify (即事件开放通知) 向 NWDAF 反馈所需数据。

15 步骤 4015、NWDAF 调用 Nnwdaf_AnalyticsSubscription_Notify (即分析订阅通知) or Nnwdaf_AnalyticsInfo_Request response (即分析信息请求响应) 向 NEF 发送 AI/ML 模型传输状态的分析信息。

步骤 4016、NEF 调用 Nnef_AnalyticsExposure_Notify (即分析开放通知) or Nnef_AnalyticsExposure_Fetch response (即分析开放获取响应) 向 AF 发送 AI/ML 模型传输状态的分析信息。

20 其中, 分析信息如表 5 或表 6 所示。

步骤 4021、如果 AF 是可信的, AF 直接向 NWDAF 发送 AI/ML 模型传输状态订阅请求, 执行操作如步骤 4012 描述。此外, 消费者还可以是 PCF、SMF。

步骤 4022、执行操作如步骤 4013 描述。即步骤 4013。

25 步骤 4023、执行操作如步骤 4014 描述。即步骤 4014。

步骤 4024、NWDAF 直接向 AF 发送 AI/ML 模型传输状态的分析信息, 执行操作如步骤 4016 描述。

实施例二, 在接收到所述分析信息后, 所述方法还包括:

30 根据所述分析信息, 直接或通过所述 NEF 向策略控制功能 PCF 发送第一请求。

其中,所述第一请求用于请求更新用于 AI/ML 模型传输的网络策略参数;所述网络策略参数用于优化 AI/ML 模型传输状态。

具体地, AF 基于 NWDAF 分析结果, AF 请求调整网络策略从而优化 AI/ML 模型传输状态。具体地,如果 AF 在受信区, AF 直接向 PCF 发送第一请求,请求 PCF 更新用于 AI/ML 模型传输的网络策略参数。如果 AF 不在受信区, AF 通过 NEF 向 PCF 发送第一请求,请求 PCF 更新用于 AI/ML 模型传输的网络策略参数。

可选地,根据所述分析信息,直接或通过所述 NEF 向策略控制功能 PCF 发送第一请求,可以通过以下步骤实现:

10 步骤 a1、根据所述分析信息中传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数中的至少一项,确定传输 AI/ML 模型的新的服务质量参数;所述新的服务质量参数包括下述至少一项: 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级。

20 步骤 a2、根据所述分析信息中的传输 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、使用 AI/ML 模型的应用服务的 IP 地址信息、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称,确定传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息;或者,若 AI/ML 模型执行联邦学习,根据所述分析信息中的指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识,确定传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息。

30 步骤 a3、根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息,确定数据网络接入标识 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息,所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址

信息用于提供优化 AI/ML 模型传输状态的路径。

步骤 a4、将所述新的服务质量参数、所述 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息作为所述第一请求中的参数直接或通过所述 NEF 发送给 PCF。

- 5 具体地，AF 依据从 NWDAF 获取的关于传输 AI/ML 模型分析信息，比如：传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数，确定传输 AI/ML 模型的新的 QoS 参数，并向 PCF 提供新的 QoS 参数。

- AF 依据从 NWDAF 获取的关于传输 AI/ML 模型分析信息，比如：传输 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、使用 AI/ML 模型的应用服务的 IP 地址信息、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称，或者，AF 依据从 NWDAF 获取的关于传输 AI/ML 模型分析信息，比如：指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识，确定传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，进而确定数据网络接入标识 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息。AF 将携带有新的 QoS 参数、DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息的第一请求直接或通过所述 NEF 发送给 PCF，请求 PCF 依据第一请求中携带的请求的参数，来更新用于 AI/ML 模型传输的相关策略参数从而优化 AI/ML 模型传输状态。

- 25 可选地，步骤 a3 可以通过以下步骤实现：

步骤 a31、根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，判断当前路由路径是否不佳；

- 步骤 a32、若当前路由路径不佳，则根据传输 AI/ML 模型的 UE(s)和各个 AF 的地址信息以及 UE(s)的区域信息，确定传输 AI/ML 模型中传输双方的目的地址；
- 30

步骤 a33、根据所述目的地址，确定最近的路径；

步骤 a34、根据最近的路径，确定 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息。

具体地，判断当前路由路径不佳，选择能够提供更优的服务体验或性能
5 的 DNAI，向 PCF 提供传输 AI/ML 模型的 DNAI 以及对应的 UE(s)和 AF(s)的区域信息、地址信息。

可选地，所述第一请求，具体用于：

请求 PCF 根据所述新的服务质量参数调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、
10 传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级并指示所述 PCF 将调整后的第一更新结果直接或通过 NEF 反馈；其中，所述第一更新结果是 PCF 根据所述新的服务质量参数调整 PCC 规则确定的；

相应的，所述方法还包括：

15 直接或通过 NEF 接收 PCF 发送的所述第一更新结果，所述第一更新结果包括所述第一请求被接受或所述第一请求被拒绝。

具体地，AF 请求 PCF 依据提供的新的 QoS 参数，相应调整 PCC 规则中的请求 PCF 根据所述新的服务质量参数调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、
20 传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级，并直接或通过 NEF 通知第一更新结果，即向 AF 通知这个请求是被接受还是被拒绝。可选地，所述第一请求具体用于：

请求 PCF 确定会话管理功能网元 SMF 是否需要更新会话管理策略，若
25 确定 SMF 需要更新会话管理策略，则确定 PCF 向 SMF 发送第二请求，所述第二请求中请求的参数包括下述至少一项：DNAI、流量导向政策标识符、流量路线信息；所述第二请求用于用于 SMF 根据新的会话管理策略确定选择的用户面功能 UPF 并提供相应的 DNAI、流量导向政策标识符、流量路线信息；

相应的，所述方法还包括：

30 直接或通过 NEF 接收 PCF 发送的第二更新结果，所述第二更新结果是

PCF 根据 SMF 发送的新的会话管理策略是否更新 UPF 路径确定的；

其中，所述第二更新结果包括所述第一请求被接受或被拒绝。

具体地，AF 请求 PCF 确定会话管理功能网元 SMF 是否需要更新会话管理策略，如果 PCF 确定 SMF 需要更新策略信息，PCF 将会向 SMF 发起携带有 DNAI、流量导向政策标识符、流量路线信息等请求参数的第二请求，SMF 根据新的会话管理策略确定选择的用户面功能 UPF 并提供相应的 DNAI、流量导向政策标识符、流量路线信息来更新会话管理策略，即更新 SM 策略。并直接或通过 NEF 通知第二更新结果，即向 AF 通知这个请求是被接受还是被拒绝。

10 示例性地，参见图 5 所示，图 5 为本公开实施例二提供的订阅网络中模型传输状态分析方法的信令流程示意图，图 5 为订阅网络中模型传输状态分析方法中 AF 与 NEF 和 PCF 间的信令交互图。本实施例提供的订阅网络中模型传输状态分析方法包括以下步骤（即实施例二对应的信令交互过程：基于 NWDAF 分析结果，AF 请求调整网络策略从而优化 AI/ML 模型传输状态）：
15 （其中，步骤 5011 至 5014 是 AF 在不受信任区域，步骤 5021 至 5022 是 AF 在受信任区域。）

步骤 5010、AF 从 NWDAF 签约并检索到 AI/ML 模型传输状态分析。

本实施例中，AF 通过 NEF 或者直接向 NWDAF 订阅并获得 AI/ML 模型传输状态分析，如上述步骤 4011 至 4024 所述。

20 步骤 5011、AF 向 NEF 发送 Nnef_AFsessionWithQoS_Update（即基于服务质量的 AF 会话更新）请求。

本实施例中，如果 AF 不受信，对于已经建立和有 QoS 要求的 AF session（即 AF 会话），AF 可以向 NEF 发送 Nnef_AFsessionWithQoS_Update（即基于服务质量的 AF 会话更新）请求，来更新用于 AI/ML 模型传输的相关策略
25 参数从而优化 AI/ML 模型传输状态。

具体的，AF 依据从 NWDAF 获取的关于传输 AI/ML 模型分析信息：QoS flow Bit Rate（即服务质量流比特率，如：传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率）、QoS flow Packet Delay（即服务质量流分组延迟，如：AI/ML 模型的上行链路方向分组延迟及 AI/ML 模

型的下行链路方向分组延迟)、QoS Sustainability、Packet transmission、Packet retransmission、QoS Sustainability、QoS Sustainability Reporting Threshold(s), 确定传输 AI/ML 模型的新的 QoS 参数,并向 PCF 提供新的 QoS 参数: 5G QoS Identifier (5QI, 即 5G 服务质量标识符)、Reflective QoS Control (即反射式服务质量控制)、UL-maximum bitrate (即传输 AI/ML 模型的上行链路方向最大比特率)、DL-maximum bitrate (即传输 AI/ML 模型的下行链路方向最大比特率)、UL-guaranteed bitrate(即传输 AI/ML 模型的上行链路方向最低比特率)、DL-guaranteed bitrate(传输 AI/ML 模型的下行链路方向最低比特率)、Priority Level(即优先级),即随着预测结果的增减而增减使 5GS 满足模型传输的 QoS 要求。

此外, AF 依据从 NWDAF 获取的关于传输 AI/ML 模型分析信息: 传输 AI/ML 模型的 AF ID, 使用 AI/ML 模型的 UE 区域信息, 使用 AI/ML 模型的应用服务的 IP 地址信息, AI/ML 模型传输使用的网络切片、DNN 信息, 确定传输 AI/ML 模型的 UE(s)和 AF(s)的区域信息、地址信息。或者, 分析信息中也可能有关于联邦学习 group 信息的信息: Federated Learning (FL) group ID、Federated Learning (FL) UE ID or UE group ID、Federated Learning (FL) 的 Application ID, 确定传输 AI/ML 模型的 UE(s)和 AF(s)的区域信息、地址信息, 判断当前路由路径不佳, 选择能够提供更优的服务体验或性能的 DNAI , 向 PCF 提供传输 AI/ML 模型的 DNAI 以及对应的 UE(s)和 AF(s)的区域信息、地址信息。

其中,判断当前路由路径不佳的过程可以为: AF 基于 AF 的 IP 地址信息, UE(s)的区域信息 (它可对应 AMF、SMF、UPF 等的 ID, UPF 的 N6 接口连接 DN), 确定当前的路由路径; 例如, 有些 UE/服务器如果下次传输前加入/离开 (联邦群组), 则有些路径则不佳, 或者跳数过多, 路径就不佳。

选择能够提供更优的服务体验或性能的 DNAI 的过程可以为: 基于收到的 IP 地址信息, UE(s)的区域信息, 可以确定传输双方的目的地址, 就能找出更优的 (最近的) 路由路径, 对应出 DNAI。

步骤 5012、NEF 向 PCF 发送 Npcf_PolicyAuthorization_Update request(即策略授权更新请求)。

本实施例中，NEF 将上述信息通过 Npcf_PolicyAuthorization_Update request（即策略授权更新请求）发送给 PCF，来更新用于 AI/ML 模型传输的相关策略参数从而优化 AI/ML 模型传输状态。

其中，Npcf_PolicyAuthorization_Update request 可以作为第二请求。

5 步骤 5013、PCF 向 NEF 通知结果（即 PCF 向 NEF 发送 Npcf_PolicyAuthorization_Update response（策略授权更新响应））。

本实施例中，PCF 依据 NEF 提供的信息，具体的，PCF 依据提供的新的 QoS 参数，相应调整 PCC 规则中的：5G QoS Identifier (5QI)、Reflective QoS Control、UL-maximum bitrate、DL-maximum bitrate、UL-guaranteed bitrate、
10 DL-guaranteed bitrate, Priority Level, 并发送 Npcf_PolicyAuthorization_Update response（策略授权更新响应）向 NEF 通知结果。

如果 PCF 确定 SMF 需要更新策略信息，PCF 将会向 SMF 发起 Npcf_SMPolicyControl_UpdateNotify request（即会话策略控制更新通知请求）（DNAI, Per DNAI: Traffic steering policy identifier, Per DNAI: N6 traffic
15 routing information), 更新 SM 策略, SMF 依据此策略选择 UPF, 并提供 DNAI, Per DNAI: Traffic steering policy identifier(即流量导向政策标识符), Per DNAI: N6 traffic routing information（流量路线信息）。其中，Npcf_SMPolicyControl_UpdateNotify request 可以作为第二请求。

步骤 5014、NEF 向 AF 发送 Nnef_AFsessionWithQoS_Update response。

20 本实施例中，NEF 发送 Nnef_AFsessionWithQoS_Update response（即基于服务质量的 AF 会话更新响应），向 AF 通知这个请求是被接受还是被拒绝。

步骤 5021、AF 直接向 PCF 发送 Npcf_PolicyAuthorization_Update request。

本实施例中，如果 AF 是受信的，AF 直接向 PCF 发送 Npcf_PolicyAuthorization_Update request 来更新用于 AI/ML 模型传输的相关
25 策略参数从而优化 AI/ML 模型传输状态，执行操作如步骤 5012 描述。

步骤 5022、PCF 直接向 AF 通知结果（即 PCF 直接向 AF 发送 Npcf_PolicyAuthorization_Update response（即策略授权更新响应））。

本实施例中，首先，PCF 依据 AF 提供的信息，执行操作如步骤 5012 描

述。PCF 直接向 AF 通知这个请求是被接受还是被拒绝。

实施例三，在接收到所述分析信息后，所述方法还可以通过以下步骤实现：

5 步骤 b1、根据所述分析信息，调整应用层模型的信息，所述应用层模型的信息包括下述至少一项：模型压缩、模型大小、模型传输时间段、模型编解码；所述应用层模型的信息用于更新服务质量参数；

10 步骤 b2、根据调整后的应用层模型的信息，确定所述新的服务质量参数，所述新的服务质量参数包括：5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级；

步骤 b3、直接或通过所述 NEF 向策略控制功能 PCF 发送第三请求；

其中，所述第三请求中请求的参数包括所述新的服务质量参数，所述第三请求用于请求更新服务质量参数。

15 具体地，AF 基于 NWDAF 提供的 QoS 要求信息，调整应用层模型的信息来更新服务质量参数，比如：调整模型压缩、模型大小、模型传输时间段、模型编解码等。根据调整后的应用层模型的信息，确定所述新的服务质量参数，直接或通过 NEF，PCF 发送携带有新的服务质量参数的第三请求。

可选地，所述第三请求，具体用于：

20 请求 PCF 根据所述新的服务质量参数调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、优先级；

相应的，所述方法还包括：

25 直接或通过 NEF 接收 PCF 发送的第三更新结果，第三更新结果是 PCF 基于调整 PCC 规则的结果确定的，所述第三更新结果包括所述第三请求被接受或所述第三请求被拒绝。

30 具体地，AF 请求 PCF 依据提供的新的 QoS 参数，相应调整 PCC（即策略和付费控制）规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最

大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、优先级，PCF 调整 PCC 规则并直接或通过 NEF 向 AF 通知这个请求是被接受还是被拒绝。

5 可选地，在所述调整应用层模型的信息之后，所述方法还可以通过以下步骤实现：

将调整后的应用层模型的信息中的模型压缩、模型大小以及模型编解码直接发送给 PCF，所述调整后的应用层模型的信息用于支持 PCF 调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级；

相应的，所述方法还包括：

接收 PCF 发送的第四更新结果，第四更新结果是 PCF 根据调整后的应用层模型的信息调整 PCC 规则的结果确定的，所述第四更新结果包括所述第三请求被接受或所述第三请求被拒绝。

具体地，PCF 依据模型压缩、模型大小、模型编解码调整 PCC 规则中的上述 QoS 参数。比如：5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级。PCF 调整 PCC 规则并直接或通过 NEF 向 AF 通知这个请求是被接受还是被拒绝。

可选地，在所述调整应用层模型的信息之后，所述方法还包括：

将调整后的应用层模型的信息中的模型传输时间直接发送给 PCF，所述调整后的应用层模型的信息中的模型传输时间用于支持 PCF 调整 PCC 规则中的门状态参数；所述门状态参数用于支持 SMF 根据门状态中的传输开始时间和传输结束时间来更新会话管理策略；

相应的，所述方法还包括：

接收 PCF 发送的第五更新结果，第五更新结果是 PCF 通过接收 SMF 发送的新的会话管理策略的结果确定的，所述第五更新结果包括所述第三请求被接受或所述第三请求被拒绝。

具体地，PCF 依据模型传输时间调整 PCC 规则中的 Gate status（即门状态参数），更新 SM 策略，SMF 依据此影响流的传输开始和结束时间，反馈给 PCF，并 PCF 接或通过 NEF 向 AF 通知这个请求是被接受还是被拒绝。

5 示例性地，参见图 6 所示，图 6 为本公开实施例三提供的订阅网络中模型传输状态分析方法的信令流程示意图，图 6 为订阅网络中模型传输状态分析方法中 AF 与 NEF 和 PCF 间的信令交互图。本实施例提供的订阅网络中模型传输状态分析方法包括以下步骤（即实施例三对应的信令交互过程：AF 基于 NWDAF 提供的分析信息，调整应用层模型的信息，如，调整模型压缩、模型大小、模型传输时间段、模型编解码等，进而更新 QoS 要求，类似实施
10 例二的步骤流程）：（其中，步骤 6012 至 6014 是 AF 在不受信任区域，步骤 6021 至 6022 是 AF 在受信任区域。）

步骤 6010、AF 从 NWDAF 签约并检索到 AI/ML 模型传输状态分析。

具体地 AF 通过 NEF 或者直接向 NWDAF 订阅并获得 AI/ML 模型传输状态分析，如上述步骤 4011 至 4024 所述。

15 步骤 6011、AF 基于分析信息，调整应用层行为。

本实施例中，AF 基于 NWDAF 提供的 QoS 要求信息，调整应用层模型的信息，如，调整模型压缩、模型大小、模型传输时间段、模型编解码等。

步骤 6012、AF 向 NEF 发送 Nnef_AFsessionWithQoS_Update 请求。

步骤 6013、NEF 向 PCF 发送 Npcf_PolicyAuthorization_Update request。

20 其中，Npcf_PolicyAuthorization_Update request 可以作为第三请求。

步骤 6014、PCF 向 NEF 通知结果（即 PCF 向 NEF 发送 Npcf_PolicyAuthorization_Update response。

步骤 6015、NEF 向 AF 发送 Nnef_AFsessionWithQoS_Update response。

步骤 6021、AF 直接向 PCF 发送 Npcf_PolicyAuthorization_Update request。

25 步骤 6022、PCF 直接向 AF 通知结果（即 PCF 直接向 AF 发送 Npcf_PolicyAuthorization_Update response（即策略授权更新响应））。

具体地，AF 请求 session 进行 QoS 更新。具体的，AF 基于调整的模型压缩、模型大小、模型传输时间段、模型编解码等，确定传输 AI/ML 模型的

新的 QoS 参数，并向 PCF 提供新的 QoS 参数：5G QoS Identifier (5QI)、Reflective QoS Control、UL-maximum bitrate、DL-maximum bitrate、UL-guaranteed bitrate、DL-guaranteed bitrate，Priority Level。或者，AF 将模型压缩、模型大小、模型传输时间段、模型编解码调整信息等直接发送给 PCF。

- 5 步骤 6013、PCF 依据提供的新的 QoS 参数，相应调整 PCC 规则中的：5G QoS Identifier (5QI)、Reflective QoS Control、UL-maximum bitrate、DL-maximum bitrate、UL-guaranteed bitrate、DL-guaranteed bitrate，Priority Level；或者 PCF 依据模型压缩、模型大小、模型编解码调整 PCC 规则中的上述 QoS 参数；或者 PCF 依据模型传输时间调整 PCC 规则中的 Gate status，
10 更新 SM 策略，SMF 依据此影响流的传输开始和结束时间；最后，通过实施例二中步骤 5014 或 5022 的方式向 AF 通知这个请求是被接受还是被拒绝。

本公开，NWDAF 接收 AF 发送的 AI/ML 模型传输状态分析请求，以及请求中包含的参数；NWDAF 从 5GC NF(s)采集的输入数据，用于分析网络中 AI/ML 模型传输状态；NWDAF 进行分析并向 AF 发送 AI/ML 模型传输状态
15 分析信息；AF 基于 AI/ML 模型传输状态分析信息请求 AF session 的 QoS 更新，并调整相应的策略控制功能 PCF、会话管理功能网元 SMF 等网元的策略；AF 基于 AI/ML 模型传输状态分析信息调整应用层模型信息的相关参数，进而调整 QoS，从而优化 AI/ML 模型传输状态。使得第三方可以获得 AI/ML 模型传输状态，还可以基于模型传输的分析结果，网络调整自身行为来满足
20 AI/ML 模型传输要求，并且第三方还可以基于模型传输的分析结果，调整模型应用层的行为来实现 AI/ML 模型的高效传输，保证 AI/ML 模型传输的业务体验和业务性能。

图 7 为本公开实施例四提供的订阅网络中模型传输状态分析方法的流程示意图，如图 7 所示，本实施例提供的订阅网络中模型传输状态分析方法的
25 执行主体为 NWDAF，则本公开实施例提供的订阅网络中模型传输状态分析方法包括以下步骤：

步骤 701、NWDAF 直接或通过网络能力开放功能 NEF 接收应用功能 AF 发送的第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

- 30 步骤 702、NWDAF 根据所述第一消息中请求的参数，向 5G 核心网的其

他网络功能 5GC NF (s) 发送第二消息, 所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据。

步骤 703、NWDAF 接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据, 并对 AI/ML 模型传输状态的数据进行分析, 得到 AI/ML 模型传输状态的分析信息。

其中, 所述分析信息用于通过 AF 调整网络策略参数和/或应用层模型的信息。

可选地, 所述第一消息中请求的参数包括下述至少一项: 网络数据分析标识、接收 AI/ML 模型的一个用户设备 UE 或一组 UE 的标识或满足分析条件的任意 UE、使用 AI/ML 模型的应用的标识、AI/ML 模型传输的区域、指示传输 AI/ML 模型服务质量流的协议数据单元 PDU 会话的网络切片、指示传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、AI/ML 模型传输的时间段、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于指示传输 AI/ML 模型的服务质量流的服务质量要求和/或用于指示传输 AI/ML 模型的特定的服务质量要求;

所述第二消息中包括下述至少一项: 使用 AI/ML 模型的 UE 的当前位置、使用 AI/ML 模型的应用的标识、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、数据采集时间、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、用于所述 AF 的服务流程;

所述分析信息包括下述至少一项: 用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、使用 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、分析结果的有效时间、提供 AI/ML 模型传输的用户面功能 UPF、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称、AI/ML 传输模型的大小、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML

模型传输的结束时间戳、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量；

其中，若 AI/ML 模型执行联邦学习，所述第一消息中请求的参数还包括：联邦学习群信息，所述联邦学习群信息包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s) 标识、参加联邦学习的应用标识；

相应的，所述第二消息中还包括下述至少一项：用于指定分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s) 标识、参加联邦学习的应用标识；

相应的，所述分析信息还包括下述至少一项：用于指定分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s) 标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识。

本实施例中，通过接收 AF 发送的订阅网络中 AI/ML 模型传输状态的分析信息的请求，根据所述第一消息中请求的参数，向 5G 核心网的其他网络功能 5GC NF (s) 采集数据，接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据，并对 AI/ML 模型传输状态的数据进行分析，得到 AI/ML 模型传输状态的分析信息，实现了有效地对 AI/ML 模型传输状态进行分析，并使得 AF 根据分析信息来调整网络策略参数和/或应用层模型的信息，进而使得网络能够基于 AI/ML 模型传输状态有效地调整网络传输策略以及使得第三方获得 AI/ML 模型传输状态的分析进行应用程序信息的调整。

在此需要说明的是，本公开实施例提供的订阅网络中模型传输状态分析方法，能够实现图 4 所示方法实施例所实现的所有方法步骤，且能够达到相同的技术效果，在此不再对本实施例中与方法实施例相同的部分及有益效果进行具体赘述。

图 8 为本公开实施例提供的订阅网络中模型传输状态分析装置的结构示意图，如图 8 所示，本实施例提供的订阅网络中模型传输状态分析装置应用

于 AF。则本实施例提供的订阅网络中模型传输状态分析装置包括：收发机 800，用于在处理器 810 的控制下接收和发送数据。

其中，在图 8 中，总线架构可以包括任意数量的互联的总线和桥，具体由处理器 810 代表的一个或多个处理器和存储器 820 代表的存储器的各种电路链接在一起。总线架构还可以将诸如外围设备、稳压器和功率管理电路等之类的各种其他电路链接在一起，这些都是本领域所公知的，因此，本文不再对其进行进一步描述。总线接口提供接口。收发机 800 可以是多个元件，即包括发送机和接收机，提供用于在传输介质上与各种其他装置通信的单元，这些传输介质包括无线信道、有线信道、光缆等传输介质。处理器 810 负责管理总线架构和通常的处理，存储器 820 可以存储处理器 810 在执行操作时所使用的数据。

处理器 810 可以是中央处理器(CPU)、专用集成电路(Application Specific Integrated Circuit, ASIC)、现场可编程门阵列(Field - Programmable Gate Array, FPGA)或复杂可编程逻辑器件(Complex Programmable Logic Device, CPLD)，处理器也可以采用多核架构。

本实施例中，存储器 820，用于存储计算机程序；收发机 800，用于在处理器 810 的控制下收发数据；处理器 810，用于读取存储器中的计算机程序并执行以下操作：

直接或通过网络能力开放功能 NEF 向网络数据分析功能 NWDAF 发送第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

直接或通过所述 NEF 接收所述 NWDAF 发送的 AI/ML 模型传输状态的分析信息，所述分析信息是所述 NWDAF 根据接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据确定的，所述 AI/ML 模型传输状态的数据是所述 NWDAF 根据接收到的所述第一消息中请求的参数通过向所述 5GC NF (s) 发送第二消息得到的，所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据；

其中，所述分析信息用于调整网络策略参数和/或应用层模型的信息。

可选地，所述第一消息中请求的参数包括下述至少一项：网络数据分析

标识、接收 AI/ML 模型的一个用户设备 UE 或一组 UE 的标识或满足分析条件的任意 UE、使用 AI/ML 模型的应用的标识、AI/ML 模型传输的区域、指示传输 AI/ML 模型服务质量流的协议数据单元 PDU 会话的网络切片、指示传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、AI/ML 模型传输的时间段、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于指示传输 AI/ML 模型的服务质量流的服务质量要求和/或用于指示传输 AI/ML 模型的特定的服务质量要求；

所述第二消息中包括下述至少一项：使用 AI/ML 模型的 UE 的当前位置、使用 AI/ML 模型的应用的标识、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、数据采集时间、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、用于所述 AF 的服务流程；

所述分析信息包括下述至少一项：用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、使用 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、分析结果的有效时间、提供 AI/ML 模型传输的用户面功能 UPF、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称、AI/ML 传输模型的大小、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量；

其中，若 AI/ML 模型执行联邦学习，所述第一消息中请求的参数还包括：联邦学习群信息，所述联邦学习群信息包括下述至少一项：用于指指示分析

的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识;

相应的, 所述第二消息中还包括下述至少一项: 用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识;

相应的, 所述分析信息还包括下述至少一项: 用于指定分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识。

可选地, 处理器 810, 还用于:

在接收到所述分析信息后, 根据所述分析信息, 直接或通过所述 NEF 向策略控制功能 PCF 发送第一请求;

其中, 所述第一请求用于请求更新用于 AI/ML 模型传输的网络策略参数; 所述网络策略参数用于优化 AI/ML 模型传输状态。

可选地, 处理器 810, 用于根据所述分析信息, 直接或通过所述 NEF 向策略控制功能 PCF 发送第一请求时, 具体包括:

根据所述分析信息中传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数中的至少一项, 确定传输 AI/ML 模型的新的服务质量参数; 所述新的服务质量参数包括下述至少一项: 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级;

根据所述分析信息中的传输 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、使用 AI/ML 模型的应用服务的 IP 地址信息、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称, 确定传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息; 或者, 若 AI/ML 模型执行联邦学习, 根据所述分析信息

中的指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识，确定传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息；

根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，确定
5 数据网络接入标识 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息，所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息用于提供优化 AI/ML 模型传输状态的路径；

将所述新的服务质量参数、所述 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息作为所述第一请求中的参数直接或通过所述
10 NEF 发送给 PCF。

可选地，处理器 810，用于根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，确定数据网络接入标识 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息时，具体包括：

根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，判断
15 当前路由路径是否不佳；

若当前路由路径不佳，则根据传输 AI/ML 模型的 UE(s)和各个 AF 的地址信息以及 UE(s)的区域信息，确定传输 AI/ML 模型中传输双方的目的地址；

根据所述目的地址，确定最近的路径；

根据最近的路径，确定 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的
20 区域信息及地址信息。

可选地，所述第一请求，具体用于：

请求 PCF 根据所述新的服务质量参数调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向
25 向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级并指示所述 PCF 将调整后的第一更新结果直接或通过 NEF 反馈；其中，所述第一更新结果是 PCF 根据所述新的服务质量参数调整 PCC 规则确定的；

相应的，处理器 810，还具体：

直接或通过 NEF 接收 PCF 发送的所述第一更新结果，所述第一更新结果
30 包括所述第一请求被接受或所述第一请求被拒绝。

可选地，所述第一请求具体用于：

请求 PCF 确定会话管理功能网元 SMF 是否需要更新会话管理策略，若确定 SMF 需要更新会话管理策略，则确定 PCF 向 SMF 发送第二请求，所述第二请求中请求的参数包括下述至少一项：DNAI、流量导向政策标识符、流量路线信息；所述第二请求用于 SMF 根据新的会话管理策略确定选择的用户面功能 UPF 并提供相应的 DNAI、流量导向政策标识符、流量路线信息；

相应的，处理器 810，还具体：

直接或通过 NEF 接收 PCF 发送的第二更新结果，所述第二更新结果是 PCF 根据 SMF 发送的新的会话管理策略是否更新 UPF 路径确定的；

10 其中，所述第二更新结果包括所述第一请求被接受或被拒绝。

可选地，处理器 810，还用于：

在接收到所述分析信息后，根据所述分析信息，调整应用层模型的信息，所述应用层模型的信息包括下述至少一项：模型压缩、模型大小、模型传输时间段、模型编解码；所述应用层模型的信息用于更新服务质量参数；

15 根据调整后的应用层模型的信息，确定所述新的服务质量参数，所述新的服务质量参数包括：5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级；

20 直接或通过所述 NEF 向策略控制功能 PCF 发送第三请求；

其中，所述第三请求中请求的参数包括所述新的服务质量参数，所述第三请求用于请求更新服务质量参数。

可选地，所述第三请求，具体用于：

25 请求 PCF 根据所述新的服务质量参数调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、优先级；

相应的，处理器 810，还用于：

30 直接或通过 NEF 接收 PCF 发送的第三更新结果，第三更新结果是 PCF 基于调整 PCC 规则的结果确定的，所述第三更新结果包括所述第三请求被接

受或所述第三请求被拒绝。

可选地，处理器 810，还用于：

在所述调整应用层模型的信息之后，将调整后的应用层模型的信息中的模型压缩、模型大小以及模型编解码直接发送给 PCF，所述调整后的应用层模型的信息用于支持 PCF 调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级；

相应的，处理器 810，还用于：

接收 PCF 发送的第四更新结果，第四更新结果是 PCF 根据调整后的应用层模型的信息调整 PCC 规则的结果确定的，所述第四更新结果包括所述第三请求被接受或所述第三请求被拒绝。

可选地，处理器 810，还用于：

在所述调整应用层模型的信息之后，将调整后的应用层模型的信息中的模型传输时间直接发送给 PCF，所述调整后的应用层模型的信息中的模型传输时间用于支持 PCF 调整 PCC 规则中的门状态参数；所述门状态参数用于支持 SMF 根据门状态中的传输开始时间和传输结束时间来更新会话管理策略；

相应的，处理器 810，还用于：

接收 PCF 发送的第五更新结果，第五更新结果是 PCF 通过接收 SMF 发送的新的会话管理策略的结果确定的，所述第五更新结果包括所述第三请求被接受或所述第三请求被拒绝。

在此需要说明的是，本公开提供的订阅网络中模型传输状态分析装置，能够实现图 3-图 6 所示方法实施例所实现的所有方法步骤，且能够达到相同的技术效果，在此不再对本实施例中与方法实施例相同的部分及有益效果进行具体赘述。

图 9 为本公开另一实施例提供的订阅网络中模型传输状态分析装置的结构示意图，如图 9 所示，本实施例提供的订阅网络中模型传输状态分析装置应用于 AF，则本实施例提供的订阅网络中模型传输状态分析装置 900 包括：

发送单元 901，用于直接或通过网络能力开放功能 NEF 向网络数据分析功能 NWDAF 发送第一消息；其中，所述第一消息用于请求订阅网络中人工

智能/机器学习 AI/ML 模型传输状态的分析信息;

分析单元 902, 用于直接或通过所述 NEF 接收所述 NWDAF 发送的 AI/ML 模型传输状态的分析信息, 所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据;

5 其中, 所述分析信息用于调整网络策略参数和/或应用层模型的信息。

可选地, 所述第一消息中请求的参数包括下述至少一项: 网络数据分析标识、接收 AI/ML 模型的一个用户设备 UE 或一组 UE 的标识或满足分析条件的任意 UE、使用 AI/ML 模型的应用的标识、AI/ML 模型传输的区域、指示传输 AI/ML 模型服务质量流的协议数据单元 PDU 会话的网络切片、指示
10 传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、AI/ML 模型传输的时间段、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于指示传输 AI/ML 模型的服务质量流的服务质量要求和/或用于指示传输 AI/ML 模型的特定的服务质量要求;

所述第二消息中包括下述至少一项: 使用 AI/ML 模型的 UE 的当前位置、
15 使用 AI/ML 模型的应用的标识、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、数据采集时间、AI/ML 模型传输
20 的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、用于所述 AF 的服务流程;

所述分析信息包括下述至少一项: 用于传输 AI/ML 模型服务质量流的
25 PDU 会话的网络切片、使用 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、分析结果的有效时间、提供 AI/ML 模型传输的用户面功能 UPF、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称、AI/ML 传输模型的大小、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、服务质量要求、传输 AI/ML 模型的服务质量流标识
30 符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方

向比特率、AI/ML模型的上行链路方向分组延迟及AI/ML模型的下行链路方向分组延迟、服务质量流在AI/ML模型传输的时间段内异常释放的数量、达到服务质量流在AI/ML模型传输的时间段内异常释放的报告阈值的次数、AI/ML模型的分组传输数量、AI/ML模型的分组重传数量；

5 其中,若AI/ML模型执行联邦学习,所述第一消息中请求的参数还包括:联邦学习群信息,所述联邦学习群信息包括下述至少一项:用于指示分析的联邦学习组的标识、参加联邦学习的UE标识或UE(s)标识、参加联邦学习的应用标识;

10 相应的,所述第二消息中还包括下述至少一项:用于指定分析的联邦学习组的标识、参加联邦学习的UE标识或UE(s)标识、参加联邦学习的应用标识;

相应的,所述分析信息还包括下述至少一项:用于指定分析的联邦学习组的标识、参加联邦学习的UE标识或UE(s)标识、指示提供AI/ML模型或者参与联邦学习的各个应用标识。

15 可选地,发送单元,还用于:

在接收到所述分析信息后,根据所述分析信息,直接或通过所述NEF向策略控制功能PCF发送第一请求;

其中,所述第一请求用于请求更新用于AI/ML模型传输的网络策略参数;所述网络策略参数用于优化AI/ML模型传输状态。

20 可选地,发送单元,具体用于:根据所述分析信息中传输AI/ML模型的上行链路方向比特率及传输AI/ML模型的下行链路方向比特率、AI/ML模型的上行链路方向分组延迟及AI/ML模型的下行链路方向分组延迟、服务质量流在AI/ML模型传输的时间段内异常释放的数量、AI/ML模型的分组传输数量、AI/ML模型的分组重传数量、达到服务质量流在AI/ML模型传输的时间段内异常释放的报告阈值的次数中的至少一项,确定传输AI/ML模型的新的服务质量参数;所述新的服务质量参数包括下述至少一项:5G服务质量标识符、反射式服务质量控制、传输AI/ML模型的上行链路方向最大比特率、传输AI/ML模型的下行链路方向最大比特率、传输AI/ML模型的上行链路方向最低比特率、传输AI/ML模型的下行链路方向最低比特率、服务质量流的优
25
30 先级;

根据所述分析信息中的传输 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、使用 AI/ML 模型的应用服务的 IP 地址信息、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称，确定传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息；或者，若 AI/ML 模型执行联邦学习，根据所述分析信息中的指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识，确定传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息；

根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，确定数据网络接入标识 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息，所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息用于提供优化 AI/ML 模型传输状态的路径；

将所述新的服务质量参数、所述 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息作为所述第一请求中的参数直接或通过所述 NEF 发送给 PCF。

可选地，发送单元，还具体用于：

根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，判断当前路由路径是否不佳；

若当前路由路径不佳，则根据传输 AI/ML 模型的 UE(s)和各个 AF 的地址信息以及 UE(s)的区域信息，确定传输 AI/ML 模型中传输双方的目的地址；

根据所述目的地址，确定最近的路径；

根据最近的路径，确定 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息。

可选地，所述第一请求，具体用于：

请求 PCF 根据所述新的服务质量参数调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级并指示所述 PCF 将调整后的第一更新结果直接或通过 NEF 反馈；其中，所述第一更新结果是 PCF 根据所述新的服务质量参数调整 PCC 规则确定的；

相应的，发送单元，还用于：

直接或通过 NEF 接收 PCF 发送的所述第一更新结果，所述第一更新结果包括所述第一请求被接受或所述第一请求被拒绝。

可选地，所述第一请求具体用于：

- 5 请求 PCF 确定会话管理功能网元 SMF 是否需要更新会话管理策略，若确定 SMF 需要更新会话管理策略，则确定 PCF 向 SMF 发送第二请求，所述第二请求中请求的参数包括下述至少一项：DNAI、流量导向政策标识符、流量路线信息；所述第二请求用于 SMF 根据新的会话管理策略确定选择的用户面功能 UPF 并提供相应的 DNAI、流量导向政策标识符、流量路线信息；

- 10 相应的，接收单元，还用于：

直接或通过 NEF 接收 PCF 发送的第二更新结果，所述第二更新结果是 PCF 根据 SMF 发送的新的会话管理策略是否更新 UPF 路径确定的；

其中，所述第二更新结果包括所述第一请求被接受或被拒绝。

可选地，该装置还包括：确定单元；确定单元，用于：

- 15 在接收到所述分析信息后，根据所述分析信息，调整应用层模型的信息，所述应用层模型的信息包括下述至少一项：模型压缩、模型大小、模型传输时间段、模型编解码；所述应用层模型的信息用于更新服务质量参数；

根据调整后的应用层模型的信息，确定所述新的服务质量参数，所述新的服务质量参数包括：5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级；

直接或通过所述 NEF 向策略控制功能 PCF 发送第三请求；

- 25 其中，所述第三请求中请求的参数包括所述新的服务质量参数，所述第三请求用于请求更新服务质量参数。

可选地，所述第三请求，具体用于：

- 30 请求 PCF 根据所述新的服务质量参数调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、优先级；

相应的，接收单元，还用于：

直接或通过 NEF 接收 PCF 发送的第三更新结果，第三更新结果是 PCF 基于调整 PCC 规则的结果确定的，所述第三更新结果包括所述第三请求被接受或所述第三请求被拒绝。

5 可选地，发送单元，还用于：

在所述调整应用层模型的信息之后，将调整后的应用层模型的信息中的模型压缩、模型大小以及模型编解码直接发送给 PCF，所述调整后的应用层模型的信息用于支持 PCF 调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级；

相应的，接收单元，还用于：

接收 PCF 发送的第四更新结果，第四更新结果是 PCF 根据调整后的应用层模型的信息调整 PCC 规则的结果确定的，所述第四更新结果包括所述第三请求被接受或所述第三请求被拒绝。

15 可选地，发送单元，还用于：

在所述调整应用层模型的信息之后，将调整后的应用层模型的信息中的模型传输时间直接发送给 PCF，所述调整后的应用层模型的信息中的模型传输时间用于支持 PCF 调整 PCC 规则中的门状态参数；所述门状态参数用于支持 SMF 根据门状态中的传输开始时间和传输结束时间来更新会话管理策略；

相应的，接收单元，还用于：

接收 PCF 发送的第五更新结果，第五更新结果是 PCF 通过接收 SMF 发送的新的会话管理策略的结果确定的，所述第五更新结果包括所述第三请求被接受或所述第三请求被拒绝。

25 在此需要说明的是，本公开提供的订阅网络中模型传输状态分析装置，能够实现图 3-图 6 方法实施例所实现的所有方法步骤，且能够达到相同的技术效果，在此不再对本实施例中与方法实施例相同的部分及有益效果进行具体赘述。

图 10 为本公开再一实施例提供的订阅网络中模型传输状态分析装置的结构示意图，如图 10 所示，本实施例提供的订阅网络中模型传输状态分析装

置应用于 NWDAF。则本实施例提供的订阅网络中模型传输状态分析装置包括：收发机 1000，用于在处理器 1010 的控制下接收和发送数据。

其中，在图 10 中，总线架构可以包括任意数量的互联的总线和桥，具体由处理器 1010 代表的一个或多个处理器和存储器 1020 代表的存储器的各种电路链接在一起。总线架构还可以将诸如外围设备、稳压器和功率管理电路等之类的各种其他电路链接在一起，这些都是本领域所公知的，因此，本文不再对其进行进一步描述。总线接口提供接口。收发机 1000 可以是多个元件，即包括发送机和接收机，提供用于在传输介质上与各种其他装置通信的单元，这些传输介质包括无线信道、有线信道、光缆等传输介质。处理器 1010 负责管理总线架构和通常的处理，存储器 1020 可以存储处理器 1010 在执行操作时所使用的数据。

处理器 1010 可以是中央处理器（CPU）、专用集成电路（Application Specific Integrated Circuit, ASIC）、现场可编程门阵列（Field-Programmable Gate Array, FPGA）或复杂可编程逻辑器件（Complex Programmable Logic Device, CPLD），处理器也可以采用多核架构。

本实施例中，存储器 1020，用于存储计算机程序；收发机 1000，用于在控制器的控制下收发数据；处理器 1010，用于读取存储器中的计算机程序并执行以下操作：

直接或通过网络能力开放功能 NEF 接收应用功能 AF 发送的第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

根据所述第一消息中请求的参数，向 5G 核心网的其他网络功能 5GC NF (s) 发送第二消息，所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据；

接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据，并对 AI/ML 模型传输状态的数据进行分析，得到 AI/ML 模型传输状态的分析信息；

其中，所述分析信息用于通过 AF 调整网络策略参数和/或应用层模型的信息。

可选地，所述第一消息中请求的参数包括下述至少一项：网络数据分析

标识、接收 AI/ML 模型的一个用户设备 UE 或一组 UE 的标识或满足分析条件的任意 UE、使用 AI/ML 模型的应用的标识、AI/ML 模型传输的区域、指示传输 AI/ML 模型服务质量流的协议数据单元 PDU 会话的网络切片、指示传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、AI/ML 模型传输的时间段、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于指示传输 AI/ML 模型的服务质量流的服务质量要求和/或用于指示传输 AI/ML 模型的特定的服务质量要求；

所述第二消息中包括下述至少一项：使用 AI/ML 模型的 UE 的当前位置、使用 AI/ML 模型的应用的标识、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、数据采集时间、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、用于所述 AF 的服务流程；

所述分析信息包括下述至少一项：用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、使用 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、分析结果的有效时间、提供 AI/ML 模型传输的用户面功能 UPF、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称、AI/ML 传输模型的大小、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、服务质量要求、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量；

其中，若 AI/ML 模型执行联邦学习，所述第一消息中请求的参数还包括：联邦学习群信息，所述联邦学习群信息包括下述至少一项：用于指示分析的

联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识；

相应的，所述第二消息中还包括下述至少一项：用于指定分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识；

相应的，所述分析信息还包括下述至少一项：用于指定分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识。

在此需要说明的是，本公开提供的订阅网络中模型传输状态分析装置，能够实现图 4、图 7 所示方法实施例所实现的所有方法步骤，且能够达到相同的技术效果，在此不再对本实施例中与方法实施例相同的部分及有益效果进行具体赘述。

图 11 为本公开又一实施例提供的订阅网络中模型传输状态分析装置的结构示意图，如图 11 所示，本公开实施例提供的订阅网络中模型传输状态分析装置应用于 NWDAF，则本实施例提供的订阅网络中模型传输状态分析装置 1100 包括：

接收单元 1101，用于直接或通过网络能力开放功能 NEF 接收应用功能 AF 发送的第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

发送单元 1102，用于根据所述第一消息中请求的参数，向 5G 核心网的其他网络功能 5GC NF (s) 发送第二消息，所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据；

分析单元 1103，还用于接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据，并对 AI/ML 模型传输状态的数据进行分析，得到 AI/ML 模型传输状态的分析信息；

其中，所述分析信息用于通过 AF 调整网络策略参数和/或应用层模型的信息。

可选地，所述第一消息中请求的参数包括下述至少一项：网络数据分析标识、接收 AI/ML 模型的一个用户设备 UE 或一组 UE 的标识或满足分析条件的任意 UE、使用 AI/ML 模型的应用的标识、AI/ML 模型传输的区域、指

示传输 AI/ML 模型服务质量流的协议数据单元 PDU 会话的网络切片、指示传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、AI/ML 模型传输的时间段、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于指示传输 AI/ML 模型的服务质量流的服务质量要求和/或用于指示传输 AI/ML 模型的特定的服务质量要求；

所述第二消息中包括下述至少一项：使用 AI/ML 模型的 UE 的当前位置、使用 AI/ML 模型的应用的标识、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、数据采集时间、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、用于所述 AF 的服务流程；

所述分析信息包括下述至少一项：用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、使用 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、分析结果的有效时间、提供 AI/ML 模型传输的用户面功能 UPF、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称、AI/ML 传输模型的大小、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量；

其中，若 AI/ML 模型执行联邦学习，所述第一消息中请求的参数还包括：联邦学习群信息，所述联邦学习群信息包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s) 标识、参加联邦学习的应用标识；

相应的，所述第二消息中还包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识；

5 相应的，所述分析信息还包括下述至少一项：用于指定分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识。

10 在此需要说明的是，本公开提供的订阅网络中模型传输状态分析装置，能够实现图 4、图 7 方法实施例所实现的所有方法步骤，且能够达到相同的技术效果，在此不再对本实施例中与方法实施例相同的部分及有益效果进行具体赘述。

需要说明的是，本公开实施例对单元的划分是示意性的，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式。另外，在本公开各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

15 集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个处理器可读取存储介质中。基于这样的理解，本公开的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备）或处理器（processor）执行本公开各个实施例方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（Read-Only Memory，ROM）、随机存取存储器（Random Access Memory，RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

25 本公开实施例还提供一种处理器可读取存储介质。处理器可读取存储介质存储有计算机程序，计算机程序用于使处理器执行上述任一种方法实施例。

其中，处理器可读取存储介质可以是处理器能够存取的任何可用介质或数据存储设备，包括但不限于磁性存储器（例如软盘、硬盘、磁带、磁光盘（MO）等）、光学存储器（例如 CD、DVD、BD、HVD 等）、以及半导体存储器（例如 ROM、EPROM、EEPROM、非易失性存储器（NAND FLASH）、固态硬盘

盘（SSD））等。

本领域内的技术人员应明白，本公开的实施例可提供为方法、系统、或计算机程序产品。因此，本公开可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本公开可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、5 光学存储器等）上实施的计算机程序产品的形式。

本公开是参照根据本公开实施例的方法、设备（系统）、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机可执行指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的10 流程和/或方框的结合。可提供这些计算机可执行指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

15 这些处理器可执行指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的处理器可读存储器中，使得存储在该处理器可读存储器中的指令产生包括指令装置的制品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

20 这些处理器可执行指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

25 显然，本领域的技术人员可以对本公开进行各种改动和变型而不脱离本公开的精神和范围。这样，倘若本公开的这些修改和变型属于本公开权利要求及其等同技术的范围之内，则本公开也意图包含这些改动和变型在内。

权 利 要 求 书

1. 一种订阅网络中模型传输状态分析方法，其特征在于，所述方法应用于应用功能 AF，所述方法包括：

直接或通过网络能力开放功能 NEF 向网络数据分析功能 NWDAF 发送第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

直接或通过所述 NEF 接收所述 NWDAF 发送的 AI/ML 模型传输状态的分析信息，所述分析信息是所述 NWDAF 根据接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据确定的；

10 其中，所述分析信息用于调整网络策略参数和/或应用层模型的信息。

2. 根据权利要求 1 所述的方法，其特征在于，所述 AI/ML 模型传输状态的数据是所述 NWDAF 根据接收到的所述第一消息中请求的参数通过向所述 5GC NF (s) 发送第二消息得到的，所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据。

15 3. 根据权利要求 2 所述的方法，其特征在于，所述第一消息中请求的参数包括下述至少一项：网络数据分析标识、接收 AI/ML 模型的一个用户设备 UE 或一组 UE 的标识或满足分析条件的任意 UE、使用 AI/ML 模型的应用的标识、AI/ML 模型传输的区域、指示传输 AI/ML 模型服务质量流的协议数据单元 PDU 会话的网络切片、指示传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、AI/ML 模型传输的时间段、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于指示传输 AI/ML 模型的服务质量流的服务质量要求和/或用于指示传输 AI/ML 模型的特定的服务质量要求；

25 所述第二消息中包括下述至少一项：使用 AI/ML 模型的 UE 的当前位置、使用 AI/ML 模型的应用的标识、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、数据采集时间、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、

AI/ML 传输模型的大小、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、用于所述 AF 的服务流程;

所述分析信息包括下述至少一项: 用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、使用 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、分析结果的有效时间、提供 AI/ML 模型传输的用户面功能 UPF、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称、AI/ML 传输模型的大小、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量;

其中,若 AI/ML 模型执行联邦学习,所述第一消息中请求的参数还包括: 联邦学习群信息, 所述联邦学习群信息包括下述至少一项: 用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识;

相应的, 所述第二消息中还包括下述至少一项: 用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识;

相应的, 所述分析信息还包括下述至少一项: 用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识。

4. 根据权利要求 1-3 任一项所述的方法, 其特征在于, 在接收到所述分析信息后, 所述方法还包括:

根据所述分析信息, 直接或通过所述 NEF 向策略控制功能 PCF 发送第一请求;

其中, 所述第一请求用于请求更新用于 AI/ML 模型传输的网络策略参数; 所述网络策略参数用于优化 AI/ML 模型传输状态。

5. 根据权利要求 4 所述的方法，其特征在于，所述根据所述分析信息，直接或通过所述 NEF 向策略控制功能 PCF 发送第一请求，包括：

根据所述分析信息中传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数中的至少一项，确定传输 AI/ML 模型的新的服务质量参数；所述新的服务质量参数包括下述至少一项：5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级；

根据所述分析信息中的传输 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、使用 AI/ML 模型的应用服务的 IP 地址信息、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称，确定传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息；或者，若 AI/ML 模型执行联邦学习，根据所述分析信息中的指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识，确定传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息；

根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，确定数据网络接入标识 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息，所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息用于提供优化 AI/ML 模型传输状态的路径；

将所述新的服务质量参数、所述 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息作为所述第一请求中的参数直接或通过所述 NEF 发送给 PCF。

6. 根据权利要求 5 所述的方法，其特征在于，所述根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，确定数据网络接入标识 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息，包括：

根据传输 AI/ML 模型的 UE(s)和各个 AF 的区域信息及地址信息，判断当前路由路径是否不佳；

若当前路由路径不佳，则根据传输 AI/ML 模型的 UE(s)和各个 AF 的地址信息以及 UE(s)的区域信息，确定传输 AI/ML 模型中传输双方的目的地址；

5 根据所述目的地址，确定最近的路径；

根据最近的路径，确定 DNAI 以及所述 DNAI 对应的 UE(s)和各个 AF 的区域信息及地址信息。

7. 根据权利要求 5 所述的方法，其特征在于，所述第一请求，具体用于：

10 请求 PCF 根据所述新的服务质量参数调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级并指示所述 PCF 将调整后的第一更新结果直接或通过 NEF 反馈；其中，所述第一更新结果是 PCF 根据所述新的服务质量参数调整 PCC 规则的结果确定的；

15 相应的，所述方法还包括：

直接或通过 NEF 接收 PCF 发送的所述第一更新结果，所述第一更新结果包括所述第一请求被接受或所述第一请求被拒绝。

8. 根据权利要求 5 所述的方法，其特征在于，所述第一请求具体用于：

20 请求 PCF 确定会话管理功能网元 SMF 是否需要更新会话管理策略，若确定 SMF 需要更新会话管理策略，则确定 PCF 向 SMF 发送第二请求，所述第二请求中请求的参数包括下述至少一项：DNAI、流量导向政策标识符、流量路线信息；所述第二请求用于 SMF 根据新的会话管理策略确定选择的用户面功能 UPF 并提供相应的 DNAI、流量导向政策标识符、流量路线信息；

25 相应的，所述方法还包括：

直接或通过 NEF 接收 PCF 发送的第二更新结果，所述第二更新结果是 PCF 根据 SMF 发送的新的会话管理策略是否更新 UPF 路径确定的；

其中，所述第二更新结果包括所述第一请求被接受或被拒绝。

9. 根据权利要求 1-3 任一项所述的方法，其特征在于，在接收到所述分析信息后，所述方法还包括：

根据所述分析信息，调整应用层模型的信息，所述应用层模型的信息包括下述至少一项：模型压缩、模型大小、模型传输时间段、模型编解码；所述应用层模型的信息用于更新服务质量参数；

根据调整后的应用层模型的信息，确定新的服务质量参数，所述新的服务质量参数包括：5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级；

直接或通过所述 NEF 向策略控制功能 PCF 发送第三请求；

10 其中，所述第三请求中请求的参数包括所述新的服务质量参数，所述第三请求用于请求更新服务质量参数。

10. 根据权利要求 9 所述的方法，其特征在于，所述第三请求，具体用于：

15 请求 PCF 根据所述新的服务质量参数调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、优先级；

相应的，所述方法还包括：

20 直接或通过 NEF 接收 PCF 发送的第三更新结果，第三更新结果是 PCF 基于调整 PCC 规则的结果确定的，所述第三更新结果包括所述第三请求被接受或所述第三请求被拒绝。

11. 根据权利要求 9 所述的方法，其特征在于，在所述调整应用层模型的信息之后，所述方法还包括：

25 将调整后的应用层模型的信息中的模型压缩、模型大小以及模型编解码直接发送给 PCF，所述调整后的应用层模型的信息用于支持 PCF 调整 PCC 规则中的 5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链路方向最低比特率、服务质量流的优先级；

30 相应的，所述方法还包括：

接收 PCF 发送的第四更新结果，第四更新结果是 PCF 根据调整后的应用层模型的信息调整 PCC 规则的结果确定的，所述第四更新结果包括所述第三请求被接受或所述第三请求被拒绝。

12. 根据权利要求 9 所述的方法，其特征在于，在所述调整应用层模型的信息之后，所述方法还包括：

将调整后的应用层模型的信息中的模型传输时间直接发送给 PCF，所述调整后的应用层模型的信息中的模型传输时间用于支持 PCF 调整 PCC 规则中的门状态参数；所述门状态参数用于支持 SMF 根据门状态中的传输开始时间和传输结束时间来更新会话管理策略；

10 相应的，所述方法还包括：

接收 PCF 发送的第五更新结果，第五更新结果是 PCF 通过接收 SMF 发送的新的会话管理策略的结果确定的，所述第五更新结果包括所述第三请求被接受或所述第三请求被拒绝。

13. 一种订阅网络中模型传输状态分析方法，其特征在于，所述方法应用于网络数据分析功能 NWDAF，所述方法包括：

直接或通过网络能力开放功能 NEF 接收应用功能 AF 发送的第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

20 根据所述第一消息中请求的参数，向 5G 核心网的其他网络功能 5GC NF (s) 发送第二消息，所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据；

接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据，并对 AI/ML 模型传输状态的数据进行分析，得到 AI/ML 模型传输状态的分析信息；

25 其中，所述分析信息用于通过 AF 调整网络策略参数和/或应用层模型的信息。

14. 根据权利要求 13 所述的方法，其特征在于，所述第一消息中请求的参数包括下述至少一项：网络数据分析标识、接收 AI/ML 模型的一个用户设备 UE 或一组 UE 的标识或满足分析条件的任意 UE、使用 AI/ML 模型的应用的标识、AI/ML 模型传输的区域、指示传输 AI/ML 模型服务质量流的协议数

据单元 PDU 会话的网络切片、指示传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、AI/ML 模型传输的时间段、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于指示传输 AI/ML 模型的服务质量流的服务质量要求和/或用于指示传输 AI/ML 模型的特定的

5 服务质量要求;

所述第二消息中包括下述至少一项: 使用 AI/ML 模型的 UE 的当前位置、使用 AI/ML 模型的应用的标识、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、数据采集时间、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的

10 PDU 会话的数据网络、用于所述 AF 的服务流程;

所述分析信息包括下述至少一项: 用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、使用 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、分析结果的有效时间、提供 AI/ML 模型传输的用户面功能 UPF、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称、

20 AI/ML 传输模型的大小、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常

25 释放的数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量;

其中, 若 AI/ML 模型执行联邦学习, 所述第一消息中请求的参数还包括: 联邦学习群信息, 所述联邦学习群信息包括下述至少一项: 用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s) 标识、参加

30 联邦学习的应用标识;

相应的，所述第二消息中还包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识；

相应的，所述分析信息还包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识。

15. 一种订阅网络中模型传输状态分析装置，其特征在于，所述装置包括存储器，收发机，处理器：

存储器，用于存储计算机程序；收发机，用于在所述处理器的控制下收发数据；处理器，用于读取所述存储器中的计算机程序并执行以下操作：

直接或通过网络能力开放功能 NEF 向网络数据分析功能 NWDAF 发送第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

直接或通过所述 NEF 接收所述 NWDAF 发送的 AI/ML 模型传输状态的分析信息，所述分析信息是所述 NWDAF 根据接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据确定的；

其中，所述分析信息用于调整网络策略参数和/或应用层模型的信息。

16. 一种订阅网络中模型传输状态分析装置，其特征在于，所述装置包括存储器，收发机，处理器：

存储器，用于存储计算机程序；收发机，用于在所述处理器的控制下收发数据；处理器，用于读取所述存储器中的计算机程序并执行以下操作：

直接或通过网络能力开放功能 NEF 接收应用功能 AF 发送的第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

25 根据所述第一消息中请求的参数，向 5G 核心网的其他网络功能 5GC NF (s) 发送第二消息，所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据；

接收 5G 核心网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据，并对 AI/ML 模型传输状态的数据进行分析，得到 AI/ML 模型传输状态的分析信息；

其中，所述分析信息用于通过 AF 调整网络策略参数和/或应用层模型的信息。

17. 一种订阅网络中模型传输状态分析装置，其特征在于，所述装置包括：

5 发送单元，用于直接或通过网络能力开放功能 NEF 向网络数据分析功能 NWDAF 发送第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

分析单元，用于直接或通过所述 NEF 接收所述 NWDAF 发送的 AI/ML 模型传输状态的分析信息，所述分析信息是所述 NWDAF 根据接收 5G 核心
10 网的其他网络功能 5GC NF (s) 发送的 AI/ML 模型传输状态的数据确定的；
其中，所述分析信息用于调整网络策略参数和/或应用层模型的信息。

18. 根据权利要求 17 所述的装置，其特征在于，所述 AI/ML 模型传输状态的数据是所述 NWDAF 根据接收到的所述第一消息中请求的参数通过向所述 5GC NF (s) 发送第二消息得到的，所述第二消息用于采集用于分析网络
15 中 AI/ML 模型传输状态的数据。

19. 根据权利要求 17 或 18 所述的装置，其特征在于，发送单元，还用于：

在接收到所述分析信息后，根据所述分析信息，直接或通过所述 NEF 向策略控制功能 PCF 发送第一请求；

20 其中，所述第一请求用于请求更新用于 AI/ML 模型传输的网络策略参数；所述网络策略参数用于优化 AI/ML 模型传输状态。

20. 根据权利要求 17 或 18 所述的装置，其特征在于，所述装置还包括：确定单元；确定单元，用于：

在接收到所述分析信息后，根据所述分析信息，调整应用层模型的信息，
25 所述应用层模型的信息包括下述至少一项：模型压缩、模型大小、模型传输时间段、模型编解码；所述应用层模型的信息用于更新服务质量参数；

根据调整后的应用层模型的信息，确定所述新的服务质量参数，所述新的服务质量参数包括：5G 服务质量标识符、反射式服务质量控制、传输 AI/ML 模型的上行链路方向最大比特率、传输 AI/ML 模型的下行链路方向最大比特
30 率、传输 AI/ML 模型的上行链路方向最低比特率、传输 AI/ML 模型的下行链

路方向最低比特率、服务质量流的优先级；

直接或通过所述 NEF 向策略控制功能 PCF 发送第三请求；

其中，所述第三请求中请求的参数包括所述新的服务质量参数，所述第三请求用于请求更新服务质量参数。

5 21. 一种订阅网络中模型传输状态分析装置，其特征在于，所述装置包括：

接收单元，用于直接或通过网络能力开放功能 NEF 接收应用功能 AF 发送的第一消息；其中，所述第一消息用于请求订阅网络中人工智能/机器学习 AI/ML 模型传输状态的分析信息；

10 发送单元，用于根据所述第一消息中请求的参数，向 5G 核心网的其他网络功能 5GC NF (s) 发送第二消息，所述第二消息用于采集用于分析网络中 AI/ML 模型传输状态的数据；

分析单元，用于接收 5G 核心网的其他网络功能 5GC NF(s) 发送的 AI/ML 模型传输状态的数据，并对 AI/ML 模型传输状态的数据进行分析，得到 AI/ML 模型传输状态的分析信息；

15 其中，所述分析信息用于通过 AF 调整网络策略参数和/或应用层模型的信息。

22. 根据权利要求 21 所述的装置，其特征在于，所述第一消息中请求的参数包括下述至少一项：网络数据分析标识、接收 AI/ML 模型的一个用户设备 UE 或一组 UE 的标识或满足分析条件的任意 UE、使用 AI/ML 模型的应用的标识、AI/ML 模型传输的区域、指示传输 AI/ML 模型服务质量流的协议数据单元 PDU 会话的网络切片、指示传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、AI/ML 模型传输的时间段、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于指示传输 AI/ML 模型的服务质量流的服务质量要求和/或用于指示传输 AI/ML 模型的特定的服务质量要求；

25 所述第二消息中包括下述至少一项：使用 AI/ML 模型的 UE 的当前位置、使用 AI/ML 模型的应用的标识、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下

行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量、数据采集时间、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、AI/ML 传输模型的大小、用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络、用于所述 AF 的服务流程；

所述分析信息包括下述至少一项：用于传输 AI/ML 模型服务质量流的 PDU 会话的网络切片、使用 AI/ML 模型的应用的标识、使用 AI/ML 模型的区域信息、分析结果的有效时间、提供 AI/ML 模型传输的用户面功能 UPF、用于传输 AI/ML 模型服务质量流的 PDU 会话的数据网络名称、AI/ML 传输模型的大小、AI/ML 模型传输的时长、AI/ML 模型传输的开始时间戳、AI/ML 模型传输的结束时间戳、传输 AI/ML 模型的服务质量流标识符、传输 AI/ML 模型的上行链路方向比特率及传输 AI/ML 模型的下行链路方向比特率、AI/ML 模型的上行链路方向分组延迟及 AI/ML 模型的下行链路方向分组延迟、服务质量流在 AI/ML 模型传输的时间段内异常释放的数量、达到服务质量流在 AI/ML 模型传输的时间段内异常释放的报告阈值的次数、AI/ML 模型的分组传输数量、AI/ML 模型的分组重传数量；

其中，若 AI/ML 模型执行联邦学习，所述第一消息中请求的参数还包括：联邦学习群信息，所述联邦学习群信息包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识；

相应的，所述第二消息中还包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、参加联邦学习的应用标识；

相应的，所述分析信息还包括下述至少一项：用于指示分析的联邦学习组的标识、参加联邦学习的 UE 标识或 UE(s)标识、指示提供 AI/ML 模型或者参与联邦学习的各个应用标识。

23. 一种处理器可读存储介质，其特征在于，所述处理器可读存储介质存储有计算机程序，所述计算机程序用于使所述处理器执行权利要求 1 至 14 任一项所述的方法。

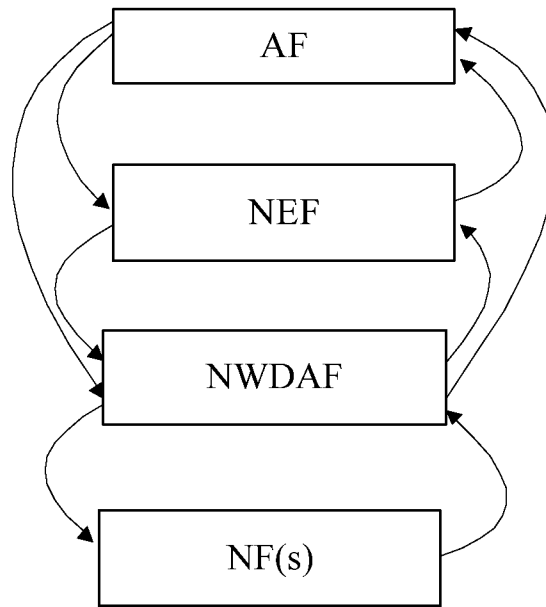


图 1

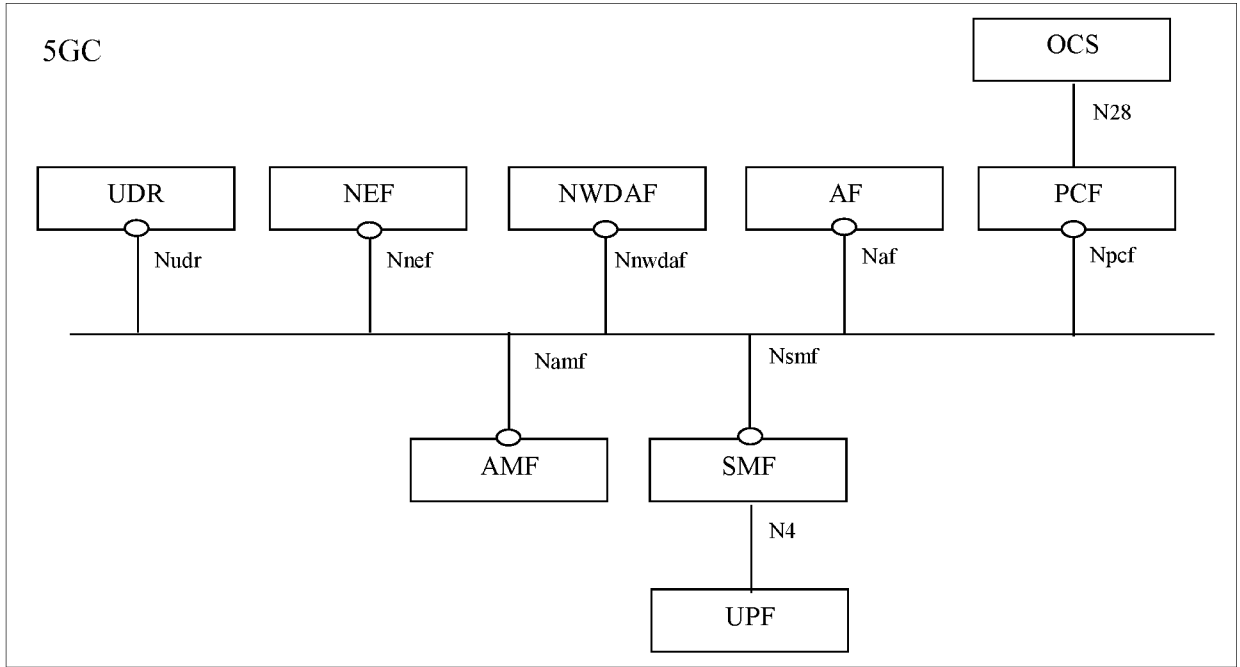


图 2

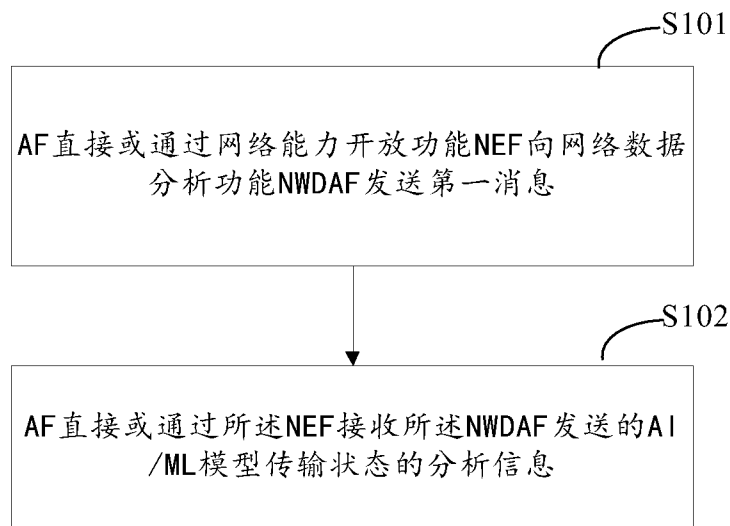


图 3

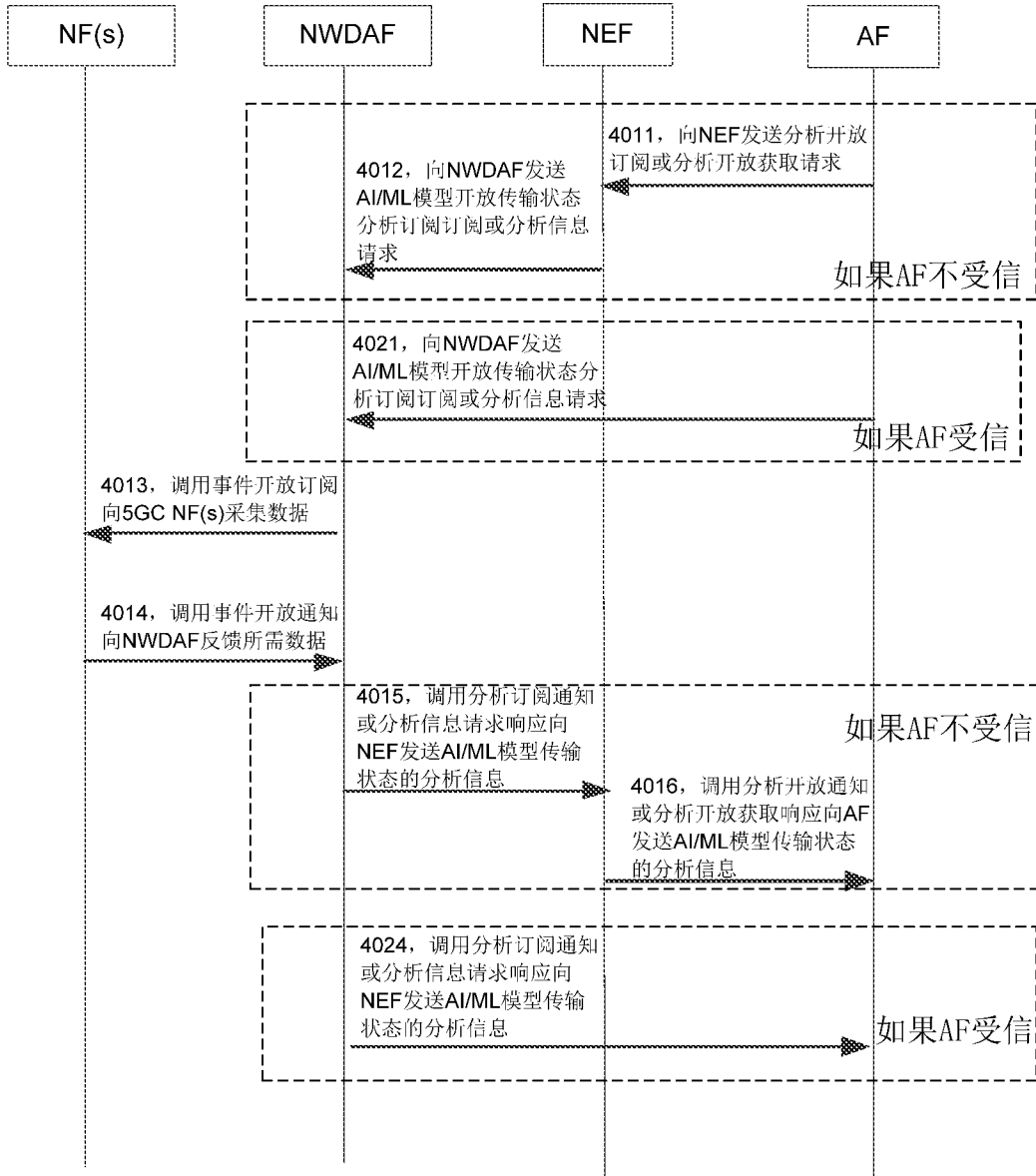


图 4

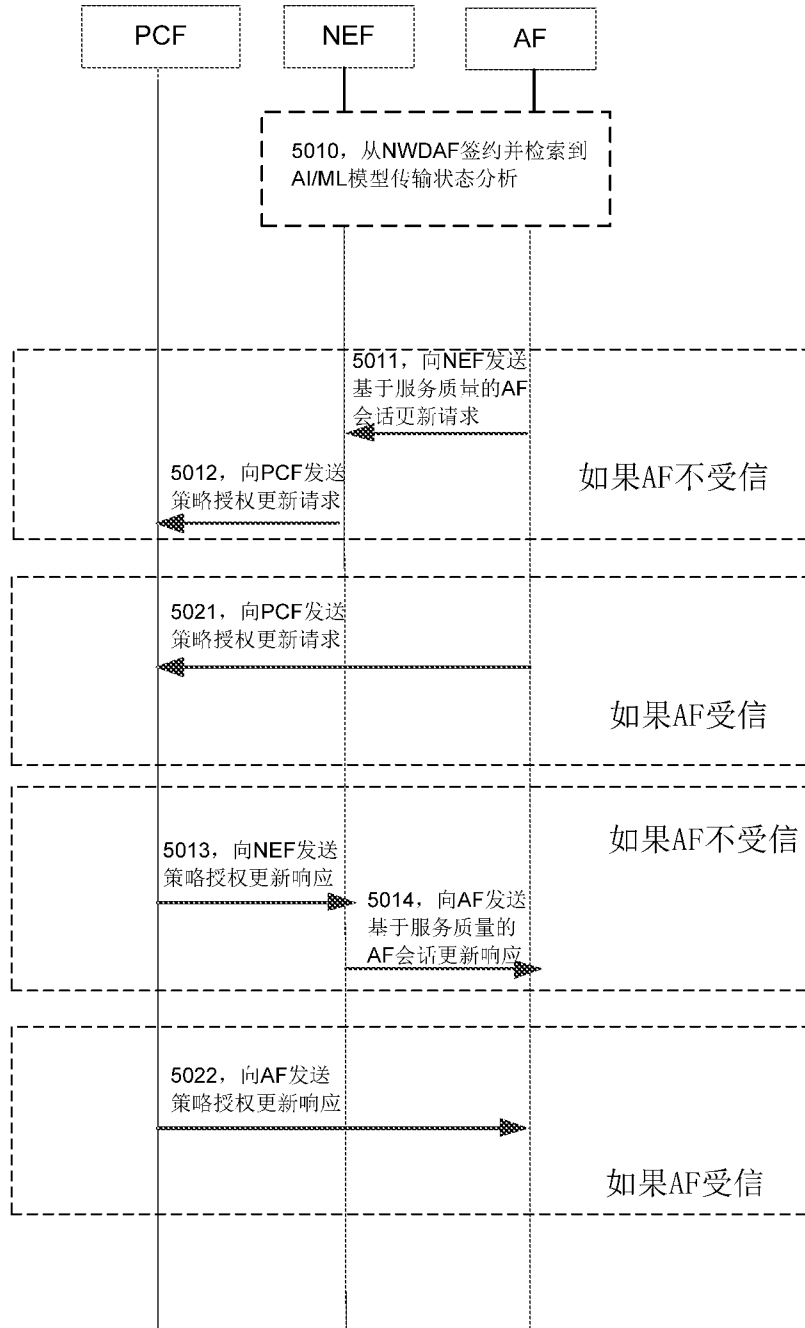


图 5

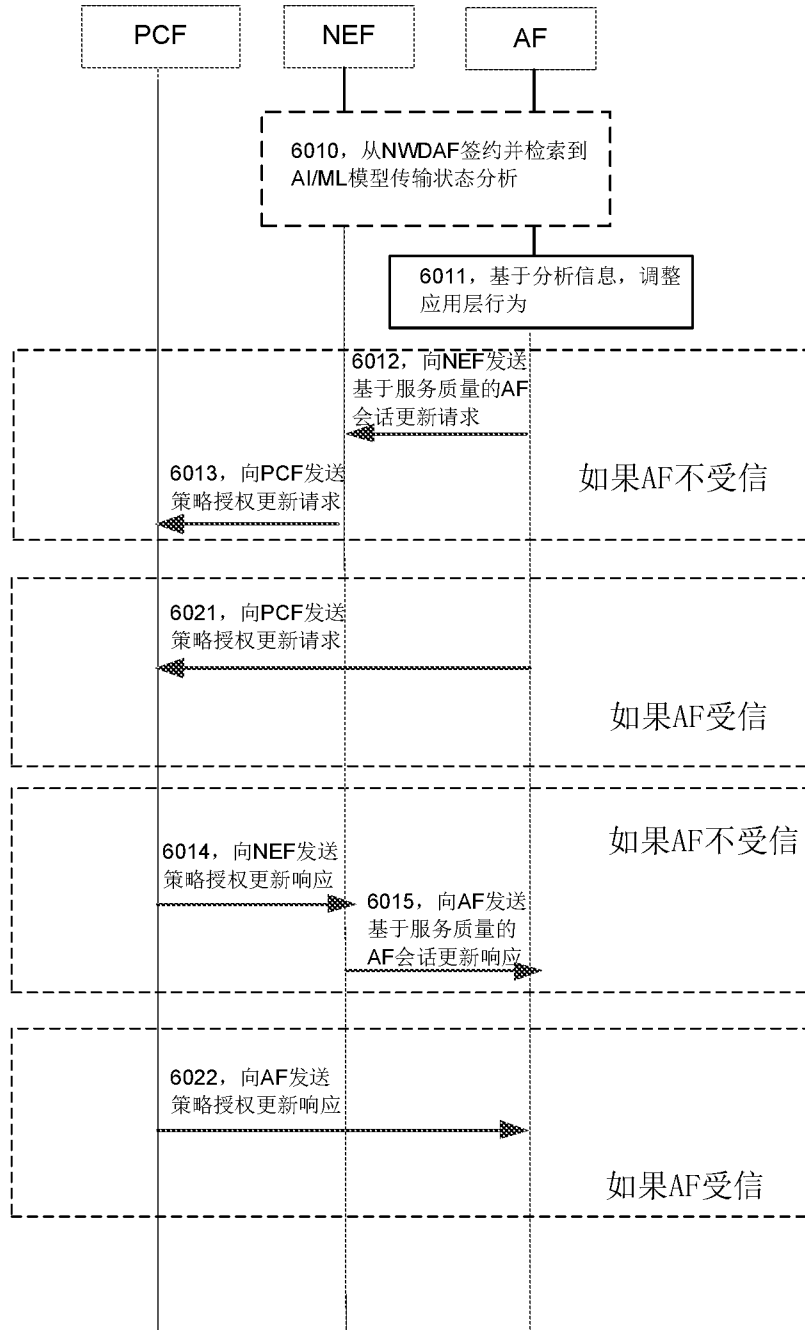


图 6

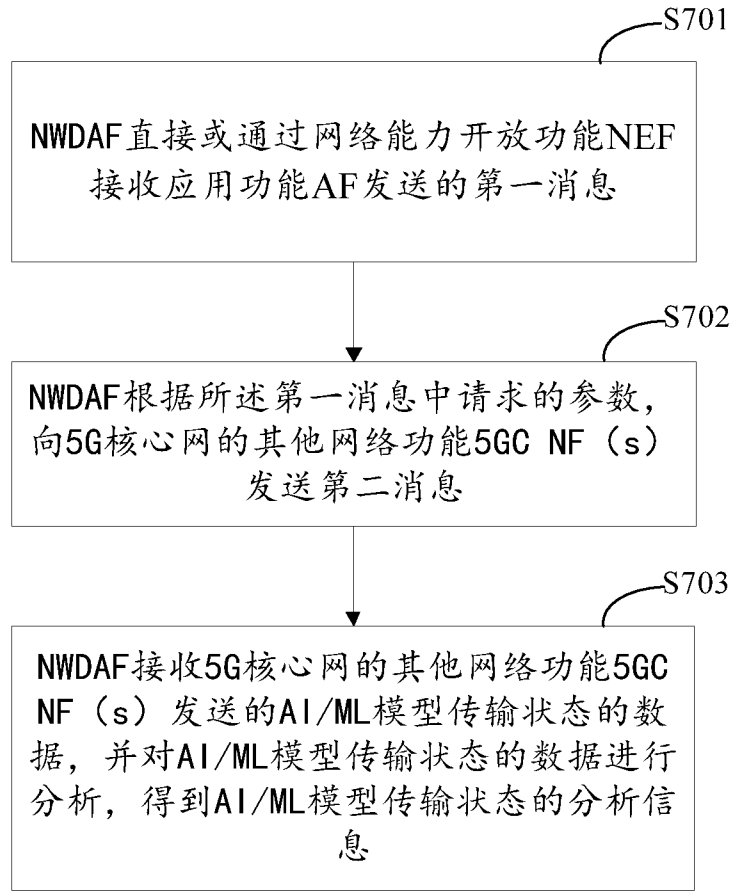


图 7

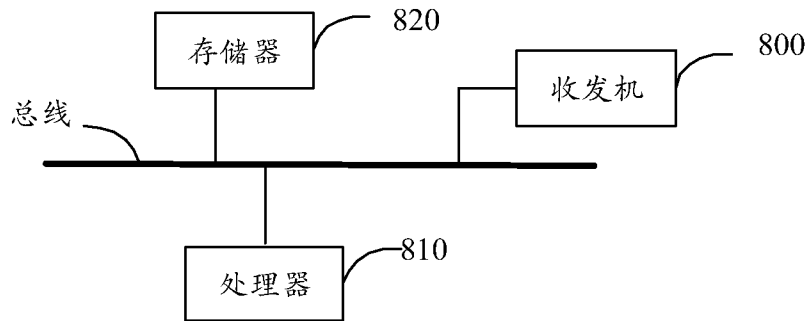


图 8

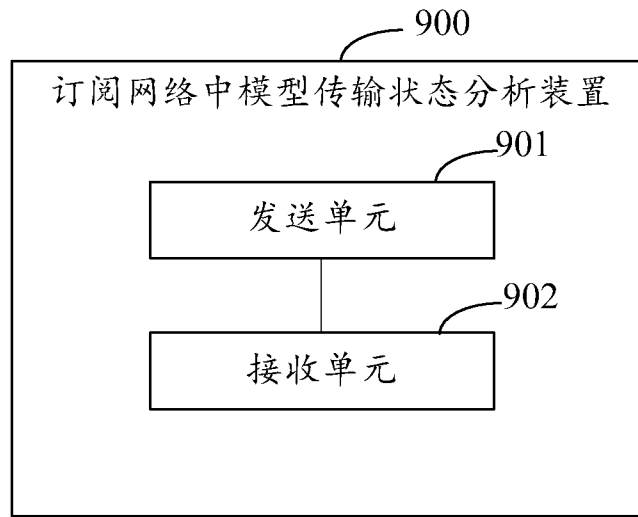


图 9

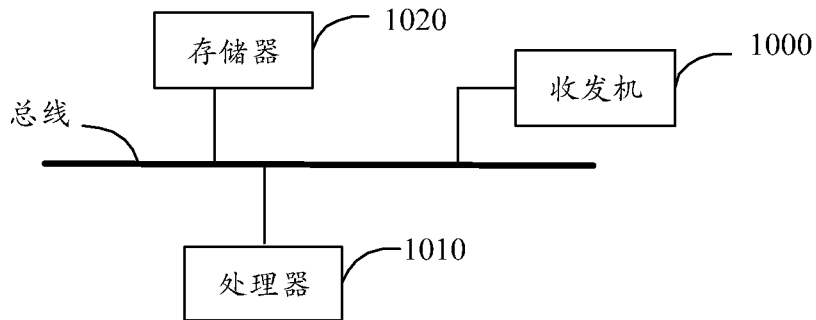


图 10

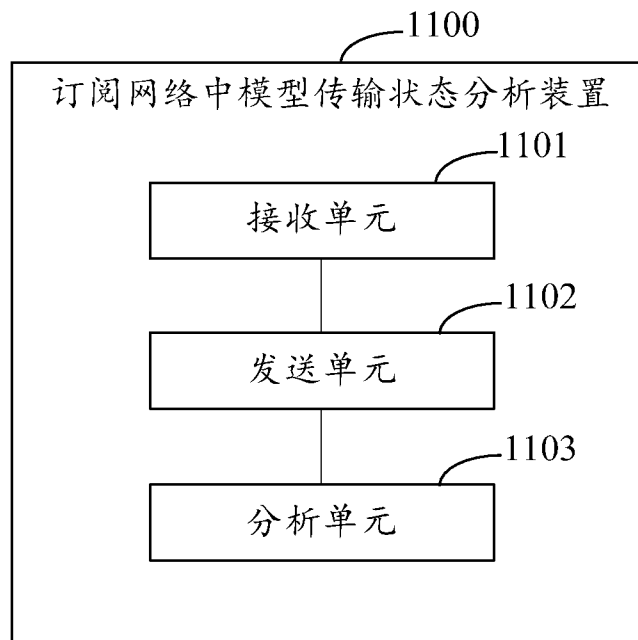


图 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/127034

A. CLASSIFICATION OF SUBJECT MATTER		
H04L 41/16(2022.01)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)		
H04L		
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; ENTXTC; ENTXT; 3GPP; VEN; CNABS; CNKI; 3GPP: 传输, 状态, 机器学习, 人工智能, 模型, 订阅, 5G, AF, NEF, NWDAF, statu+, monitor+, qos, model, Transmission, AI, ML, intelligence, learning, subscrib+		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2021099367 A1 (SAMSUNG ELECTRONICS CO., LTD.) 01 April 2021 (2021-04-01) entire document	1-23
A	US 2021014141 A1 (VERIZON PATENT AND LICENSING INC.) 14 January 2021 (2021-01-14) entire document	1-23
A	WO 2021032495 A1 (TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)) 25 February 2021 (2021-02-25) entire document	1-23
A	US 2021021494 A1 (INTEL CORP.) 21 January 2021 (2021-01-21) entire document	1-23
A	WO 2021141291 A1 (SAMSUNG ELECTRONICS CO., LTD.) 15 July 2021 (2021-07-15) entire document	1-23
A	""Study on traffic characteristics and performance requirements for AI/ML model transfer in 5GS" 3GPP TR 22.874 V1.0.0 (2021-03), 18 March 2021 (2021-03-18), entire document	1-23
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
04 January 2023		18 January 2023
Name and mailing address of the ISA/CN		Authorized officer
China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China		
Facsimile No. (86-10)62019451		Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/127034

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2021099367	A1	01 April 2021	EP	3997843	A1	18 May 2022
				KR	20210037416	A	06 April 2021
				US	2022224612	A1	14 July 2022
				WO	2021060840	A1	01 April 2021

US	2021014141	A1	14 January 2021	None			

WO	2021032495	A1	25 February 2021	US	2022294606	A1	15 September 2022
				EP	4014433	A1	22 June 2022

US	2021021494	A1	21 January 2021	None			

WO	2021141291	A1	15 July 2021	KR	20210088303	A	14 July 2021
				EP	4087201	A1	09 November 2022

国际检索报告

国际申请号

PCT/CN2022/127034

<p>A. 主题的分类</p> <p>H04L 41/16 (2022.01) i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																							
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>H04L</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CNXTX;ENTXTC;ENTXT;3GPP;VEN;CNABS;CNKI;3GPP: 传输, 状态, 机器学习, 人工智能, 模型, 订阅, 5G, AF, NEF, NWDAF, statu+, monitor+, qos, model, Transmission, AI, ML, intelligence, learning, subscrib+</p>																							
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>US 2021099367 A1 (SAMSUNG ELECTRONICS CO LTD) 2021年4月1日 (2021 - 04 - 01) 全文</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>US 2021014141 A1 (VERIZON PATENT & LICENSING INC) 2021年1月14日 (2021 - 01 - 14) 全文</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>WO 2021032495 A1 (ERICSSON TELEFON AB L M) 2021年2月25日 (2021 - 02 - 25) 全文</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>US 2021021494 A1 (INTEL CORP) 2021年1月21日 (2021 - 01 - 21) 全文</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>WO 2021141291 A1 (SAMSUNG ELECTRONICS CO LTD) 2021年7月15日 (2021 - 07 - 15) 全文</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>""Study on traffic characteristics and performance requirements for AI/ML model transfer in 5G"" 3GPP TR 22.874 V1.0.0 (2021-03), 2021年3月18日 (2021 - 03 - 18), 全文</td> <td>1-23</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	US 2021099367 A1 (SAMSUNG ELECTRONICS CO LTD) 2021年4月1日 (2021 - 04 - 01) 全文	1-23	A	US 2021014141 A1 (VERIZON PATENT & LICENSING INC) 2021年1月14日 (2021 - 01 - 14) 全文	1-23	A	WO 2021032495 A1 (ERICSSON TELEFON AB L M) 2021年2月25日 (2021 - 02 - 25) 全文	1-23	A	US 2021021494 A1 (INTEL CORP) 2021年1月21日 (2021 - 01 - 21) 全文	1-23	A	WO 2021141291 A1 (SAMSUNG ELECTRONICS CO LTD) 2021年7月15日 (2021 - 07 - 15) 全文	1-23	A	""Study on traffic characteristics and performance requirements for AI/ML model transfer in 5G"" 3GPP TR 22.874 V1.0.0 (2021-03), 2021年3月18日 (2021 - 03 - 18), 全文	1-23
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
A	US 2021099367 A1 (SAMSUNG ELECTRONICS CO LTD) 2021年4月1日 (2021 - 04 - 01) 全文	1-23																					
A	US 2021014141 A1 (VERIZON PATENT & LICENSING INC) 2021年1月14日 (2021 - 01 - 14) 全文	1-23																					
A	WO 2021032495 A1 (ERICSSON TELEFON AB L M) 2021年2月25日 (2021 - 02 - 25) 全文	1-23																					
A	US 2021021494 A1 (INTEL CORP) 2021年1月21日 (2021 - 01 - 21) 全文	1-23																					
A	WO 2021141291 A1 (SAMSUNG ELECTRONICS CO LTD) 2021年7月15日 (2021 - 07 - 15) 全文	1-23																					
A	""Study on traffic characteristics and performance requirements for AI/ML model transfer in 5G"" 3GPP TR 22.874 V1.0.0 (2021-03), 2021年3月18日 (2021 - 03 - 18), 全文	1-23																					
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																							
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件</p> <p>“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</p> <p>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</p> <p>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</p> <p>“&” 同族专利的文件</p>																							
<p>国际检索实际完成的日期</p> <p>2023年1月4日</p>		<p>国际检索报告邮寄日期</p> <p>2023年1月18日</p>																					
<p>ISA/CN的名称和邮寄地址</p> <p>中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>		<p>授权官员</p> <p>贺利良</p> <p>电话号码 86-(010)-62412304</p>																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/127034

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
US	2021099367	A1	2021年4月1日	EP	3997843	A1	2022年5月18日
				KR	20210037416	A	2021年4月6日
				US	2022224612	A1	2022年7月14日
				WO	2021060840	A1	2021年4月1日

US	2021014141	A1	2021年1月14日	无			

WO	2021032495	A1	2021年2月25日	US	2022294606	A1	2022年9月15日
				EP	4014433	A1	2022年6月22日

US	2021021494	A1	2021年1月21日	无			

WO	2021141291	A1	2021年7月15日	KR	20210088303	A	2021年7月14日
				EP	4087201	A1	2022年11月9日
