



(51) International Patent Classification:
G06Q 30/06 (2012.01)

(21) International Application Number:
PCT/CN2018/097205

(22) International Filing Date:
26 July 2018 (26.07.2018)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **BEIJING DIDI INFINITY TECHNOLOGY AND DEVELOPMENT CO., LTD.** [CN/CN]; Building 34, No. 8 Dongbeiwang West Road, Haidian District, Beijing 100193 (CN).

(72) Inventors: **LI, Haibo**; Room 201, 2/F, Building 3, No. 29 Dongbeiwang South Road, Haidian District, Beijing 100193 (CN). **QIU, Hui**; Room 201, 2/F, Building 3, No. 29 Dongbeiwang South Road, Haidian District, Beijing

100193 (CN). **QIN, Yuhao**; Room 201, 2/F, Building 3, No. 29 Dongbeiwang South Road, Haidian District, Beijing 100193 (CN). **JIANG, Menghua**; Room 201, 2/F, Building 3, No. 29 Dongbeiwang South Road, Haidian District, Beijing 100193 (CN).

(74) Agent: **METIS IP (CHENGDU) LLC**; (No. 846 South Tianfu Road), Tianfu Innovation Center, Chengdu, Sichuan 610213 (CN).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,

(54) Title: SYSTEMS AND METHODS FOR DISPATCHING SERVICE PROVIDERS

500

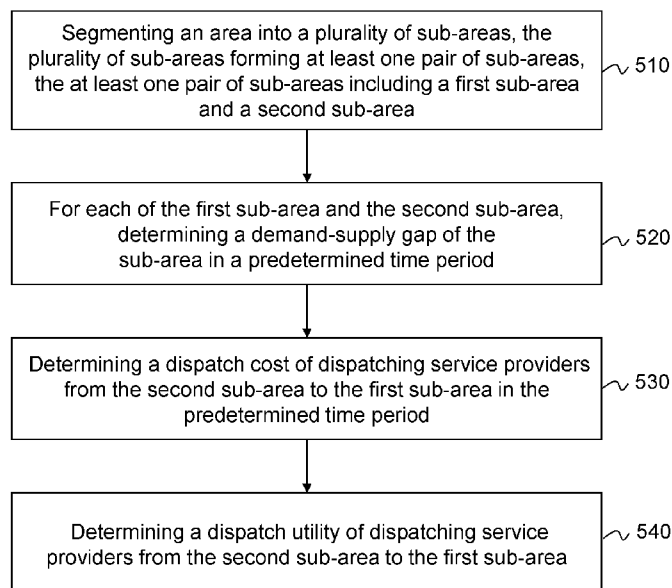


FIG. 5

(57) Abstract: A method for dispatching available service providers in an O2O service platform is provided. The method may include segmenting an area into a plurality of sub-areas. The sub-areas may form at least one pair of sub-areas including a first sub-area and a second sub-area. The method may also include determining, for each of the first and second sub-areas, a demand-supply gap of the sub-area in a predetermined time period. The method may further include determining a dispatch cost of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period. The method may further include determining a dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period.

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

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

SYSTEMS AND METHODS FOR DISPATCHING SERVICE PROVIDERS

TECHNICAL FIELD

[0001] The present disclosure generally relates to online to offline (O2O) service platforms, and in specifically, to systems and methods for dispatching available service providers in the O2O service platforms.

BACKGROUND

[0002] With the development of Internet technology, O2O services, such as online taxi hailing services and delivery services, play a more and more significant role in people's daily lives. In some scenarios, different areas may be different in their service demand and service supply. For example, in an area there may be a surplus service supply if the number of available service providers is greater than the number of service requesters, which may result in a waste of service resources in the area. In contrast, in another area there may be an insufficient service supply if the number of service requesters is larger than the number of available service providers, which may cause a long wait for the available services providers. Thus, it is desirable to provide effective systems and methods for provisioning available service providers between different areas, and thereby improving the resource utilization, the service efficiency, as well as the user experiences of service requesters.

SUMMARY

[0003] According to an aspect of the present disclosure, a system for dispatching available service providers in an O2O service platform is provided. The system may include at least one non-transitory computer-readable storage medium including a set of instructions and at least one processor in communication with the at least one non-transitory computer-readable storage medium. When executing the set of instructions,

the at least one processor may be configured to direct the system to segment an area into a plurality of sub-areas. The plurality of sub-areas may form at least one pair of sub-areas including a first sub-area and a second sub-area. The at least one processor may be further configured to direct the system to determine, for each of the first sub-area and the second sub-area, a demand-supply gap of the sub-area in a predetermined time period. The at least one processor may be further configured to direct the system to determine a dispatch cost of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period. The at least one processor may be further configured to direct the system to determine a dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period based on the demand-supply gap of the first sub-area in the predetermined time period, the demand-supply gap of the second sub-area in the predetermined time period, and the dispatch cost.

[0004] In some embodiments, the at least one pair of sub-areas may be a plurality of pairs of sub-areas. The at least one processor may be further configured to direct the system to determine, for each pair of the plurality of pairs of sub-areas, a dispatch utility of dispatching available service providers from the second sub-area of the pair to the first sub-area of the pair in the predetermined time period. The at least one processor may be further configured to direct the system to rank the dispatch utility of the plurality of pairs of sub-areas. The at least one processor may be further directed to determine a strategy for dispatching available service providers between the plurality of sub-areas based on the ranking result of the dispatch utility.

[0005] In some embodiments, to determine the demand-supply gap of the sub-area in the predetermine time period, the at least one processor may be further configured to direct the system to obtain a plurality of first sets of data related to the sub-area. Each of the plurality of first sets of data may correspond to a first historical time slice and at least

include a historical number of service orders initiated in the sub-area in the corresponding first historical time slice. The at least one processor may be further configured to direct the system to predict a number of service orders initiated in the sub-area in the predetermined time period by inputting the plurality of first sets of data related to the sub-area into a service demand prediction model. The at least one processor may be further configured to direct the system to obtain a plurality of second sets of data related to the sub-area. Each of the plurality of second sets of data may correspond to a second historical time slice and at least include a historical number of available service providers in the sub-area in the corresponding second historical time slice. The at least one processor may be further configured to direct the system to predict the number of available service providers in the sub-area in the predetermined time period by inputting the plurality of second sets of data related to the sub-area into a service supply prediction model. The at least one processor may be further configured to direct the system to determine the demand-supply gap of the sub-area in the predetermined time period based on the number of service orders initiated in the sub-area in the predetermined time period and the number of available service providers in the sub-area in the predetermined time period.

[0006] In some embodiments, to obtain the plurality of second sets of data related to the sub-area, the at least one processor may be further configured to direct the system to, for each of the plurality of second sets of data related to the sub-area, obtain a first number of available service providers entering the sub-area in the corresponding second historical time slice. The at least one processor may be further directed to obtain a second number of available service providers leaving the sub-area in the corresponding second historical time slice. The at least one processor may be further directed to determine the historical number of available service providers of the sub-area in the corresponding second historical time slice based on the first number and the second number.

[0007] In some embodiments, each of the plurality of first sets of data related to the sub-area may further include at least one of information related to peak-hours in the sub-area, location information related to the sub-area, road information related to the sub-area, traffic information of the sub-area in the corresponding first historical time slice, weather information related to the sub-area in the corresponding first historical time slice, news information related to the sub-area in the corresponding first historical time slice, policy information related to the sub-area in the corresponding first historical time slice, or date information related to the corresponding first historical time slice.

[0008] In some embodiments, each of the plurality of second sets of data related to the sub-area may further include at least one of information related to peak-hours in the sub-area, location information related to the sub-area, road information related to the sub-area, traffic information of the sub-area in the corresponding second historical time slice, weather information related to the sub-area in the corresponding second historical time slice, news information related to the sub-area in the corresponding second historical time slice, policy information related to the sub-area in the corresponding second historical time slice, date information related to the corresponding second historical time slice, a first number of available service providers entering the sub-area in the corresponding second historical time slice, or a second number of available service providers leaving the sub-area in the corresponding second historical time slice.

[0009] In some embodiments, the at least one processor may be further configured to direct the system to obtain a plurality of first sample sets of data related to a plurality of sample sub-areas. Each of the plurality of first sample sets of data may correspond to a sample historical time slice and at least include a historical number of service orders initiated in each of the plurality of sample sub-areas in the corresponding sample historical time slice. The at least one processor may be further configured to direct the system to obtain a preliminary model. The at least one processor may be further configured to

direct the system to generate the service demand prediction model by training the preliminary model using the plurality of first sample sets of data related to the plurality of sample sub-areas.

[0010] In some embodiments, the at least one processor may be further configured to direct the system to obtain a plurality of second sample sets of data related to a plurality of sample sub-areas. Each of the plurality of second sample sets of data may correspond to a sample historical time slice and at least include a historical number of available service providers in each of the plurality of sample sub-areas in the corresponding sample historical time slice. The at least one processor may be further configured to direct the system to obtain a preliminary model. The at least one processor may be further configured to direct the system to generate the service supply prediction model by training the preliminary model using the plurality of second sample sets of data related to the plurality of sample sub-areas.

[0011] In some embodiments, at least one of the service demand prediction model or the service supply prediction model may be a recurrent neural network (RNN) model.

[0012] In some embodiments, the plurality of first historical time slices and the plurality of second historical time slices may be within 1 hour to 72 hours before the predetermined time period.

[0013] In some embodiments, to determine the dispatch cost of dispatching available service providers from the second sub-area to the first sub-area, the at least one processor may be further configured to direct the system to determine a route from the second sub-area to the first sub-area. The at least one processor may be further configured to direct the system to obtain road condition information related to the route. The at least one processor may be further configured to direct the system to determine the dispatch cost of dispatching available service providers from the second sub-area to the first sub-area based on the road condition information related to the route.

[0014] In some embodiments, the road condition information related to the route may include at least one of a distance of the route, an estimated travelling time along the route, or a traffic condition along the route.

[0015] In some embodiments, the determining the dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period may be performed according to an Equation as below:

$$Utility(1,2) = \frac{G[1] - G[2]}{1 + e^{\frac{eta(1,2)}{60}} + \theta}$$

wherein $G[1]$ refers to the demand-supply gap of the first sub-area in the predetermined time period, $G[2]$ refers to the demand-supply gap of the second sub-area in the predetermined time period, $eta(1,2)$ refers to the estimated travelling time along the route, θ refers to a penalty coefficient related to the traffic condition along the route, and $(1 + e^{\frac{eta(1,2)}{60}} + \theta)$ represents the dispatch cost of dispatching available service providers from the second sub-area to the first sub-area.

[0016] In some embodiments, the road condition related to the route may include the traffic condition along the route. To obtain the traffic condition along the route, the at least one processor may be further configured to direct the system to segment the route into a plurality of road segments. The at least one processor may be further configured to direct the system to determine, for each of the plurality of road segments, a travel speed of one or more service providers passing through the road segment. The at least one processor may be further configured to direct the system to determine the traffic condition of the route based on the travel speeds corresponding to the plurality of road segments.

[0017] In some embodiments, for each of the plurality of road segments, the determining a travel speed of one or more service providers passing through the road segment may be performed according to an Equation as below:

$$V_i = \frac{\sum_j^J \text{Length}(\text{Traj}_{j,i}) * W_j}{\sum_j^J \text{Time}(\text{Traj}_{j,i})}$$

where i refers to the i^{th} road segment of the plurality of road segments, V_i refers to the travel speed of one or more service providers passing through the i^{th} road segment, J refers to a number of service providers passing through the i^{th} road segment, j refers to the j^{th} service provider among the service providers passing through the i^{th} road segment, $\text{Traj}_{j,i}$ refers to a travelling trajectory of the j^{th} service provider on the i^{th} road segment, $\text{Length}(\text{Traj}_{j,i})$ refers to a length of the $\text{Traj}_{j,i}$, $\text{Time}(\text{Traj}_{j,i})$ refers to a travel time corresponding to the $\text{Traj}_{j,i}$, and W_j refers to a weight regarding the $\text{Traj}_{j,i}$.

[0018] According to another aspect of the present disclosure, a method for dispatching available service providers in an O2O service platform is provided. The method may include segmenting an area into a plurality of sub-areas. The plurality of sub-areas may form at least one pair of sub-areas including a first sub-area and a second sub-area. The method may also include determining, for each of the first sub-area and the second sub-area, a demand-supply gap of the sub-area in a predetermined time period. The method may further include determining a dispatch cost of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period. The method may further include determining a dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period based on the demand-supply gap of the first sub-area in the predetermined time period, the demand-supply gap of the second sub-area in the predetermined time period, and the dispatch cost.

[0019] According to still another aspect of the present disclosure, a non-transitory computer-readable storage medium embodying a computer program product is provided. The computer program product comprising instructions may be configured to cause a computing device to segment an area into a plurality of sub-areas. The plurality of sub-

areas may form at least one pair of sub-areas including a first sub-area and a second sub-area. The computer program product comprising instructions may also be configured to cause the computing device to determine, for each of the first sub-area and the second sub-area, a demand-supply gap of the sub-area in a predetermined time period. The computer program product comprising instructions may further be configured to cause the computing device determine a dispatch cost of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period. The computer program product comprising instructions may further be configured to cause the computing device determine a dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period based on the demand-supply gap of the first sub-area in the predetermined time period, the demand-supply gap of the second sub-area in the predetermined time period, and the dispatch cost.

[0020] Additional features will be set forth in part in the description which follows, and in part will become apparent to those skilled in the art upon examination of the following and the accompanying drawings or may be learned by production or operation of the examples. The features of the present disclosure may be realized and attained by practice or use of various aspects of the methodologies, instrumentalities and combinations set forth in the detailed examples discussed below.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] The present disclosure is further described in terms of exemplary embodiments. These exemplary embodiments are described in detail with reference to the drawings. These embodiments are non-limiting exemplary embodiments, in which like reference numerals represent similar structures throughout the several views of the drawings, and wherein:

[0022] FIG. 1 is a schematic diagram illustrating an exemplary O2O service system according to some embodiments of the present disclosure;

[0023] FIG. 2 is a schematic diagram illustrating exemplary hardware and software components of a computing device according to some embodiments of the present disclosure;

[0024] FIG. 3 is a schematic diagram illustrating exemplary hardware and/or software components of a mobile device on which a terminal may be implemented according to some embodiments of the present disclosure;

[0025] FIGs. 4A and 4B are block diagrams illustrating exemplary processing engines 112A and 112B according to some embodiments of the present disclosure;

[0026] FIG. 5 is a flowchart illustrating an exemplary process for determining a dispatch utility of dispatching available service providers from a second sub-area to a first sub-area according to some embodiments of the present disclosure;

[0027] FIG. 6 is a flowchart illustrating an exemplary process for determining a strategy for dispatching available service providers in an area according to some embodiments of the present disclosure;

[0028] FIG. 7 is a flowchart illustrating an exemplary process for determining a demand-supply gap of a sub-area in a predetermined time period according to some embodiments of the present disclosure;

[0029] FIG. 8 is a flowchart illustrating an exemplary process for generating a service demand prediction model according to some embodiments of the present disclosure;

[0030] FIG. 9 is a flowchart illustrating an exemplary process for generating a service supply prediction model according to some embodiments of the present disclosure; and

[0031] FIG. 10 is a flowchart illustrating an exemplary process for determining a dispatching cost of dispatching available service providers from a second sub-area to a first sub-area according to some embodiments of the present disclosure.

DETAILED DESCRIPTION

[0032] In the following detailed description, numerous specific details are set forth by way of examples in order to provide a thorough understanding of the relevant disclosure. However, it should be apparent to those skilled in the art that the present disclosure may be practiced without such details. In other instances, well-known methods, procedures, systems, components, and/or circuitry have been described at a relatively high-level, without detail, in order to avoid unnecessarily obscuring aspects of the present disclosure. Various modifications to the disclosed embodiments will be readily apparent to those skilled in the art, and the general principles defined herein may be applied to other embodiments and applications without departing from the spirit and scope of the present disclosure. Thus, the present disclosure is not limited to the embodiments shown, but to be accorded the widest scope consistent with the claims.

[0033] The terminology used herein is for the purpose of describing particular example embodiments only and is not intended to be limiting. As used herein, the singular forms “a,” “an,” and “the” may be intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprise,” “comprises,” and/or “comprising,” “include,” “includes,” and/or “including,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

[0034] It will be understood that the term “system,” “engine,” “unit,” “module,” and/or “block” used herein are one method to distinguish different components, elements, parts, section or assembly of different level in ascending order. However, the terms may be displaced by another expression if they achieve the same purpose.

[0035] Generally, the word “module,” “unit,” or “block,” as used herein, refers to logic embodied in hardware or firmware, or to a collection of software instructions. A module, a

unit, or a block described herein may be implemented as software and/or hardware and may be stored in any type of non-transitory computer-readable medium or other storage device. In some embodiments, a software module/unit/block may be compiled and linked into an executable program. It will be appreciated that software modules can be callable from other modules/units/blocks or from themselves, and/or may be invoked in response to detected events or interrupts. Software modules/units/blocks configured for execution on computing devices may be provided on a computer-readable medium, such as a compact disc, a digital video disc, a flash drive, a magnetic disc, or any other tangible medium, or as a digital download (and can be originally stored in a compressed or installable format that needs installation, decompression, or decryption prior to execution). Such software code may be stored, partially or fully, on a storage device of the executing computing device, for execution by the computing device. Software instructions may be embedded in a firmware, such as an erasable programmable read-only memory (EPROM). It will be further appreciated that hardware modules/units/blocks may be included in connected logic components, such as gates and flip-flops, and/or can be included of programmable units, such as programmable gate arrays or processors. The modules/units/blocks or computing device functionality described herein may be implemented as software modules/units/blocks, but may be represented in hardware or firmware. In general, the modules/units/blocks described herein refer to logical modules/units/blocks that may be combined with other modules/units/blocks or divided into sub-modules/sub-units/sub-blocks despite their physical organization or storage. The description may be applicable to a system, an engine, or a portion thereof.

[0036] It will be understood that when a unit, engine, module or block is referred to as being “on,” “connected to,” or “coupled to,” another unit, engine, module, or block, it may be directly on, connected or coupled to, or communicate with the other unit, engine, module, or block, or an intervening unit, engine, module, or block may be present, unless

the context clearly indicates otherwise. As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0037] These and other features, and characteristics of the present disclosure, as well as the methods of operation and functions of the related elements of structure and the combination of parts and economies of manufacture, may become more apparent upon consideration of the following description with reference to the accompanying drawings, all of which form a part of this disclosure. It is to be expressly understood, however, that the drawings are for the purpose of illustration and description only and are not intended to limit the scope of the present disclosure. It is understood that the drawings are not to scale.

[0038] The flowcharts used in the present disclosure illustrate operations that systems implement according to some embodiments in the present disclosure. It is to be expressly understood, the operations of the flowchart may be implemented not in order. Conversely, the operations may be implemented in inverted order, or simultaneously. Moreover, one or more other operations may be added to the flowcharts. One or more operations may be removed from the flowcharts.

[0039] Embodiments of the present disclosure may be applied to different transportation systems including but not limited to land transportation, sea transportation, air transportation, space transportation, or the like, or any combination thereof. A vehicle of the transportation systems may include a rickshaw, travel tool, taxi, chauffeured car, hitch, bus, rail transportation (e.g., a train, a bullet train, high-speed rail, and subway), ship, airplane, spaceship, hot-air balloon, driverless vehicle, or the like, or any combination thereof. The transportation system may also include any transportation system that applies management and/or distribution, for example, a system for sending and/or receiving an express.

[0040] The application scenarios of different embodiments of the present disclosure may include but not limited to one or more webpages, browser plugins and/or extensions, client terminals, custom systems, intracompany analysis systems, artificial intelligence robots, or the like, or any combination thereof. It should be understood that application scenarios of the system and method disclosed herein are only some examples or embodiments.

Those having ordinary skills in the art, without further creative efforts, may apply these drawings to other application scenarios. For example, other similar server.

[0041] The term “passenger,” “requester,” “requestor,” “service requester,” “service requestor” and “customer” in the present disclosure are used interchangeably to refer to an individual, an entity or a tool that may request or order a service. Also, the term “driver,” “provider,” “service provider,” and “supplier” in the present disclosure are used interchangeably to refer to an individual, an entity or a tool that may provide a service or facilitate the providing of the service. The term “user” in the present disclosure may refer to an individual, an entity or a tool that may request a service, order a service, provide a service, or facilitate the providing of the service. For example, the user may be a requester, a passenger, a driver, an operator, or the like, or any combination thereof. In the present disclosure, “requester” and “requester terminal” may be used interchangeably, and “provider” and “provider terminal” may be used interchangeably.

[0042] The term “request,” “service,” “service request,” and “order” in the present disclosure are used interchangeably to refer to a request that may be initiated by a passenger, a requester, a service requester, a customer, a driver, a provider, a service provider, a supplier, or the like, or any combination thereof. The service request may be accepted by any one of a passenger, a requester, a service requester, a customer, a driver, a provider, a service provider, or a supplier. The service request may be chargeable or free.

[0043] The present disclosure provides systems and methods for dispatching available service providers an area in an O2O service platform. The systems and methods may segment the area into a plurality of sub-areas. The sub-areas may form at least one pair of sub-areas including a first sub-area and a second sub-area. The systems and methods may determine, for each of the first sub-area and the second sub-area, a demand-supply gap of the sub-area in a predetermined time period. The systems and methods may also determine a dispatch cost of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period. The dispatch cost may be related to a route between the first and second sub-areas. For example, the dispatch cost may increase with the distance of the route, the estimated travelling time along the route, or the congestion of the route, or the like, or any combination thereof. Based on the demand-supply gaps of the first and second sub-areas and the dispatch cost, the systems and methods may determine a dispatch utility of dispatching available service providers from the second sub-area to the first sub-area. The dispatch utility, which takes both the dispatch cost and the supply/demand conditions of the first and second sub-areas into consideration, may serve as a basis to determine a dispatch strategy of available service providers among the sub-areas of the area. As such, the resource utilization, the service efficiency, and user experience in the area may be improved.

[0044] FIG. 1 is a block diagram illustrating an exemplary O2O service system 100 according to some embodiments of the present disclosure. For example, the O2O service system 100 may be an online transportation service platform for transportation services. The O2O service system 100 may include a server 110, a network 120, a requester terminal 130, a provider terminal 140, a vehicle 150, a storage device 160, and a navigation system 170.

[0045] The O2O service system 100 may provide a plurality of services. Exemplary service may include a taxi-hailing service, a chauffeur service, an express car service, a carpool service, a bus service, a driver hire service, and a shuttle service. In some embodiments, the O2O service may be any online service, such as booking a meal, shopping, or the like, or any combination thereof.

[0046] In some embodiments, the server 110 may be a single server or a server group. The server group may be centralized, or distributed (e.g., the server 110 may be a distributed system). In some embodiments, the server 110 may be local or remote. For example, the server 110 may access information and/or data stored in the requester terminal 130, the provider terminal 140, and/or the storage device 160 via the network 120. As another example, the server 110 may be directly connected to the requester terminal 130, the provider terminal 140, and/or the storage device 160 to access stored information and/or data. In some embodiments, the server 110 may be implemented on a cloud platform. Merely by way of example, the cloud platform may include a private cloud, a public cloud, a hybrid cloud, a community cloud, a distributed cloud, an inter-cloud, a multi-cloud, or the like, or any combination thereof. In some embodiments, the server 110 may be implemented on a computing device 200 having one or more components illustrated in FIG. 2 in the present disclosure.

[0047] In some embodiments, the server 110 may include a processing engine 112. The processing engine 112 may process information and/or data related to the O2O service system 100 to perform one or more functions described in the present disclosure. For example, the processing engine 112 may analyze information related to a plurality of sub-areas of an area to determine a dispatch strategy for dispatching available service providers among the plurality of sub-areas. In some embodiments, the processing engine 112 may include one or more processing engines (e.g., single-core processing engine(s) or multi-core processor(s)). Merely by way of example, the processing engine 112 may

include a central processing unit (CPU), an application-specific integrated circuit (ASIC), an application-specific instruction-set processor (ASIP), a graphics processing unit (GPU), a physics processing unit (PPU), a digital signal processor (DSP), a field-programmable gate array (FPGA), a programmable logic device (PLD), a controller, a microcontroller unit, a reduced instruction-set computer (RISC), a microprocessor, or the like, or any combination thereof.

[0048] The network 120 may facilitate exchange of information and/or data. In some embodiments, one or more components of the O2O service system 100 (e.g., the server 110, the requester terminal 130, the provider terminal 140, the vehicle 150, the storage device 160, and the navigation system 170) may transmit information and/or data to other component(s) of the O2O service system 100 via the network 120. For example, the server 110 may receive a service request from the requester terminal 130 via the network 120. In some embodiments, the network 120 may be any type of wired or wireless network, or combination thereof. Merely by way of example, the network 120 may include a cable network, a wireline network, an optical fiber network, a telecommunications network, an intranet, an Internet, a local area network (LAN), a wide area network (WAN), a wireless local area network (WLAN), a metropolitan area network (MAN), a wide area network (WAN), a public telephone switched network (PSTN), a Bluetooth network, a ZigBee network, a near field communication (NFC) network, or the like, or any combination thereof. In some embodiments, the network 120 may include one or more network access points. For example, the network 120 may include wired or wireless network access points such as base stations and/or internet exchange points 120-1, 120-2, through which one or more components of the O2O service system 100 may be connected to the network 120 to exchange data and/or information.

[0049] In some embodiments, a passenger may be an owner of the requester terminal 130. In some embodiments, the owner of the requester terminal 130 may be someone

other than the passenger. For example, an owner A of the requester terminal 130 may use the requester terminal 130 to transmit a service request for a passenger B or receive a service confirmation and/or information or instructions from the server 110. In some embodiments, a service provider may be a user of the provider terminal 140. In some embodiments, the user of the provider terminal 140 may be someone other than the service provider. For example, a user C of the provider terminal 140 may use the provider terminal 140 to receive a service request for a service provider D, and/or information or instructions from the server 110. In some embodiments, "passenger" and "passenger terminal" may be used interchangeably, and "service provider" and "provider terminal" may be used interchangeably. In some embodiments, the provider terminal may be associated with one or more service providers (e.g., a night-shift service provider, or a day-shift service provider).

[0050] In some embodiments, the requester terminal 130 may include a mobile device 130-1, a tablet computer 130-2, a laptop computer 130-3, a built-in device in a vehicle 130-4, a wearable device 130-5, or the like, or any combination thereof. In some embodiments, the mobile device 130-1 may include a smart home device, a smart mobile device, a virtual reality device, an augmented reality device, or the like, or any combination thereof. In some embodiments, the smart home device may include a smart lighting device, a control device of an intelligent electrical apparatus, a smart monitoring device, a smart television, a smart video camera, an interphone, or the like, or any combination thereof. In some embodiments, the smart mobile device may include a smartphone, a personal digital assistant (PDA), a gaming device, a navigation device, a point of sale (POS) device, or the like, or any combination thereof. In some embodiments, the virtual reality device and/or the augmented reality device may include a virtual reality helmet, virtual reality glasses, a virtual reality patch, an augmented reality helmet, augmented reality glasses, an augmented reality patch, or the like, or any combination thereof. For

example, the virtual reality device and/or the augmented reality device may include Google™ Glasses, an Oculus Rift, a HoloLens, a Gear VR, etc. In some embodiments, the built-in device in the vehicle 130-4 may include an onboard computer, an onboard television, etc. In some embodiments, the requester terminal 130 may be a device with positioning technology for locating the position of the passenger and/or the requester terminal 130. In some embodiments, the wearable device 130-5 may include a smart bracelet, a smart footgear, smart glasses, a smart helmet, a smart watch, smart clothing, a smart backpack, a smart accessory, or the like, or any combination thereof.

[0051] The provider terminal 140 may include a plurality of provider terminals 140-1, 140-2, ..., 140-n. In some embodiments, the provider terminal 140 may be similar to, or the same device as the requester terminal 130. In some embodiments, the provider terminal 140 may be customized to be able to implement the O2O service system 100. In some embodiments, the provider terminal 140 may be a device with positioning technology for locating the service provider, the provider terminal 140, and/or a vehicle 150 associated with the provider terminal 140. In some embodiments, the requester terminal 130 and/or the provider terminal 140 may communicate with another positioning device to determine the position of the passenger, the requester terminal 130, the service provider, and/or the provider terminal 140. In some embodiments, the requester terminal 130 and/or the provider terminal 140 may periodically transmit the positioning information to the server 110. In some embodiments, the provider terminal 140 may also periodically transmit the availability status to the server 110. The availability status may indicate whether a vehicle 150 associated with the provider terminal 140 is available to carry a passenger. For example, the requester terminal 130 and/or the provider terminal 140 may transmit the positioning information and the availability status to the server 110 every thirty minutes. As another example, the requester terminal 130 and/or the provider terminal 140 may

transmit the positioning information and the availability status to the server 110 each time the user logs into the mobile application associated with the O2O service system 100.

[0052] In some embodiments, the provider terminal 140 may correspond to one or more vehicles 150. The vehicles 150 may carry the passenger and travel to the destination. The vehicles 150 may include a plurality of vehicles 150-1, 150-2, ..., 150-n. One vehicle may correspond to one type of services (e.g., a taxi-hailing service, a chauffeur service, an express car service, a carpool service, a bus service, a driver hire service, or a shuttle service).

[0053] The storage device 160 may store data and/or instructions. In some embodiments, the storage device 160 may store data obtained from the requester terminal 130 and/or the provider terminal 140. In some embodiments, the storage device 160 may store data and/or instructions that the server 110 may execute or use to perform exemplary methods described in the present disclosure. In some embodiments, storage device 160 may include a mass storage device, a removable storage device, a volatile read-and-write memory, a read-only memory (ROM), or the like, or any combination thereof. Exemplary mass storage devices may include a magnetic disk, an optical disk, a solid-state drive, etc. Exemplary removable storage devices may include a flash drive, a floppy disk, an optical disk, a memory card, a zip disk, a magnetic tape, etc. Exemplary volatile read-and-write memory may include a random-access memory (RAM). Exemplary RAM may include a dynamic RAM (DRAM), a double data rate synchronous dynamic RAM (DDR SDRAM), a static RAM (SRAM), a thyristor RAM (T-RAM), and a zero-capacitor RAM (Z-RAM), etc. Exemplary ROM may include a mask ROM (MROM), a programmable ROM (PROM), an erasable programmable ROM (EPROM), an electrically-erasable programmable ROM (EEPROM), a compact disk ROM (CD-ROM), and a digital versatile disk ROM, etc. In some embodiments, the storage device 160 may be implemented on a cloud platform. Merely by way of example, the cloud platform may

include a private cloud, a public cloud, a hybrid cloud, a community cloud, a distributed cloud, an inter-cloud, a multi-cloud, or the like, or any combination thereof.

[0054] In some embodiments, the storage device 160 may be connected to the network 120 to communicate with one or more components of the O2O service system 100 (e.g., the server 110, the requester terminal 130, or the provider terminal 140). One or more components of the O2O service system 100 may access the data or instructions stored in the storage device 160 via the network 120. In some embodiments, the storage device 160 may be directly connected to or communicate with one or more components of the O2O service system 100 (e.g., the server 110, the requester terminal 130, the provider terminal 140). In some embodiments, the storage device 160 may be part of the server 110.

[0055] The navigation system 170 may determine information associated with an object, for example, one or more of the requester terminal 130, the provider terminal 140, the vehicle 150, etc. In some embodiments, the navigation system 170 may be a global positioning system (GPS), a global navigation satellite system (GLONASS), a compass navigation system (COMPASS), a BeiDou navigation satellite system, a Galileo positioning system, a quasi-zenith satellite system (QZSS), etc. The information may include a location, an elevation, a velocity, or an acceleration of the object, or a current time. The navigation system 170 may include one or more satellites, for example, a satellite 170-1, a satellite 170-2, and a satellite 170-3. The satellites 170-1 through 170-3 may determine the information mentioned above independently or jointly. The satellite navigation system 170 may transmit the information mentioned above to the network 120, the requester terminal 130, the provider terminal 140, or the vehicle 150 via wireless connections.

[0056] In some embodiments, one or more components of the O2O service system 100 (e.g., the server 110, the requester terminal 130, the provider terminal 140) may have permissions to access the storage device 160. In some embodiments, one or more

components of the O2O service system 100 may read and/or modify information related to the passenger, service provider, and/or the public when one or more conditions are met. For example, the server 110 may read and/or modify one or more passengers' information after a service is completed. As another example, the server 110 may read and/or modify one or more service providers' information after a service is completed.

[0057] One of ordinary skill in the art would understand that when an element (or component) of the O2O service system 100 performs, the element may perform through electrical signals and/or electromagnetic signals. For example, when a requester terminal 130 transmits out a service request to the server 110, a processor of the requester terminal 130 may generate an electrical signal encoding the request. The processor of the requester terminal 130 may then transmit the electrical signal to an output port. If the requester terminal 130 communicates with the server 110 via a wired network, the output port may be physically connected to a cable, which further may transmit the electrical signal to an input port of the server 110. If the requester terminal 130 communicates with the server 110 via a wireless network, the output port of the requester terminal 130 may be one or more antennas, which convert the electrical signal to electromagnetic signal. Similarly, a provider terminal 130 may receive an instruction and/or service request from the server 110 via electrical signal or electromagnetic signals. Within an electronic device, such as the requester terminal 130, the provider terminal 140, and/or the server 110, when a processor thereof processes an instruction, transmits out an instruction, and/or performs an action, the instruction and/or action is conducted via electrical signals. For example, when the processor retrieves or saves data from a storage medium, it may transmit out electrical signals to a read/write device of the storage medium, which may read or write structured data in the storage medium. The structured data may be transmitted to the processor in the form of electrical signals via a bus of the electronic device. Here, an

electrical signal may refer to one electrical signal, a series of electrical signals, and/or a plurality of discrete electrical signals.

[0058] FIG. 2 illustrates a schematic diagram of an exemplary computing device 200 according to some embodiments of the present disclosure. The computing device 200 may be a computer, such as the server 110 in FIG. 1 and/or a computer with specific functions, configured to implement any particular system according to some embodiments of the present disclosure. The computing device 200 may be configured to implement any component that performs one or more functions disclosed in the present disclosure. For example, the server 110 (e.g., the processing engine 112) may be implemented in hardware devices, software programs, firmware, or any combination thereof of a computer like the computing device 200. For brevity, FIG. 2 depicts only one computing device. In some embodiments, the functions of the computing device may be implemented by a group of similar platforms in a distributed mode to disperse the processing load of the system.

[0059] The computing device 200 may include a communication terminal 250 that may connect with a network that may implement the data communication. The computing device 200 may also include a processor 220 that is configured to execute instructions and includes one or more processors. The schematic computer platform may include an internal communication bus 210, different types of program storage units and data storage units (e.g., a hard disk 270, a read-only memory (ROM) 230, a random-access memory (RAM) 240), various data files applicable to computer processing and/or communication, and some program instructions executed possibly by the processor 220. The computing device 200 may also include an I/O device 260 that may support the input and output of data flows between the computing device 200 and other components. Moreover, the computing device 200 may receive programs and data via the communication network.

[0060] FIG. 3 is a schematic diagram illustrating exemplary hardware and/or software components of an exemplary mobile device on which a terminal (e.g., a requester terminal 130 and/or a provider terminal 140) may be implemented according to some embodiments of the present disclosure. As illustrated in FIG. 3, the mobile device 300 may include, a communication platform 310, a display 320, a graphics processing unit (GPU) 330, a central processing unit (CPU) 340, an I/O 350, a memory 360, a mobile operating system (OS) 370, application (s), and a storage 390. In some embodiments, any other suitable component, including but not limited to a system bus or a controller (not shown), may also be included in the mobile device 300.

[0061] In some embodiments, the mobile operating system 370 (e.g., iOS™, Android™, Windows Phone™, etc.) and one or more applications 380 may be loaded into the memory 360 from the storage 390 in order to be executed by the CPU 340. The applications 380 may include a browser or any other suitable mobile apps for receiving and rendering information relating to image processing or other information from the O2O service system 100. User interactions with the information stream may be achieved via the I/O 350 and provided to the storage device 160, the server 110 and/or other components of the O2O service system 100. In some embodiments, the mobile device 300 may be an exemplary embodiment corresponding to the requester terminal 130 or the provider terminal 140.

[0062] To implement various modules, units, and their functionalities described in the present disclosure, computer hardware platforms may be used as the hardware platform(s) for one or more of the elements described herein. A computer with user interface elements may be used to implement a personal computer (PC) or any other type of work station or terminal device. A computer may also act as a system if appropriately programmed.

[0063] FIGs. 4A and 4B are block diagrams illustrating exemplary processing engines 112A and 112B according to some embodiments of the present disclosure. In some

embodiments, the processing engines 112A and 112B may be embodiments of the processing engine 112 as described in connection with FIG. 1.

[0064] In some embodiments, the processing engine 112A may be configured to analyze the service demand and/or supply in a plurality of sub-areas of an area, and optionally determine a dispatch strategy of available service providers among the sub-areas. The processing engine 112B may be configured to generate a service demand prediction model and/or a service supply prediction model. In some embodiments, the processing engines 112A and 112B may respectively be implemented on a computing device 200 (e.g., the processor 220) illustrated in FIG. 2 or a CPU 340 as illustrated in FIG. 3. Merely by way of example, the processing engine 112A may be implemented on a CPU 340 of a mobile device and the processing engine 112B may be implemented on a computing device 200. Alternatively, the processing engines 112A and 112B may be implemented on the same computing device 200 or the same CPU 340.

[0065] As shown in FIG. 4A, the processing engine 112A may include a segmentation module 401 and a determination module 402.

[0066] The segmentation module 401 may be configured to segment an area into a plurality of sub-areas. The plurality of sub-areas may form at least one pair of sub-areas, the at least one pair of sub-areas including a first sub-area and a second sub-area. In some embodiments, the segmentation module 401 may segment the area according to one or more parameters, such as a size of area, a density of population, a division of administrative area, a density of office buildings, a density of residential buildings, or the like, or any combination thereof. In some embodiments, the segmentation module 401 may segment the area into a plurality of unit areas, and then generate the sub-areas by clustering the unit areas. More descriptions regarding the segmentation of the area may be found elsewhere in the present disclosure. See, e.g., operation 510 in FIG. 5 and the relevant descriptions thereof.

[0067] The determination module 402 may be configured to determine, for each of the first sub-area and the second sub-area, a demand-supply gap of the sub-area in a predetermined time. In some embodiments, the determination module 402 may predict a service demand and a service supply for the sub-area based on a service demand prediction model and a service supply prediction model, respectively. Further, the determination module 402 may determine the demand-supply gap based on the predicted service demand and the predicted service demand. More descriptions regarding the determination of the demand-supply gap may be found elsewhere in the present disclosure. See, e.g., operation 520 in FIG. 5 and the relevant descriptions thereof.

[0068] The determination module 402 may also be configured to determine a dispatch cost of dispatching service providers in the second sub-area to the first sub-area in the predetermined time period. In some embodiments, the determination module 402 may determine the dispatch cost based on a route from the second sub-area to the first sub-area. More descriptions regarding the determination of the dispatch cost may be found elsewhere in the present disclosure. See, e.g., operation 530 and process 1000 and the relevant descriptions thereof.

[0069] The determination module 402 may further be configured to determine a dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period. The dispatch utility may be used to measure the worth (or value) of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period. More descriptions regarding the determination of the dispatch utility may be found elsewhere in the present disclosure. See, e.g., operation 540 in FIG. 5 and the relevant descriptions thereof.

[0070] Additionally or alternatively, the determination module 402 may be configured to determine a strategy for dispatching available service providers among the plurality of sub-areas in the area. More descriptions regarding the determination of the dispatch strategy

may be found elsewhere in the present disclosure. See, e.g., FIG. 8 and the relevant descriptions thereof.

[0071] As shown in FIG. 4B, the processing engine 112B may include an obtaining module 403 and a training module 404.

[0072] The obtaining module 403 may be configured obtain information used to generate the service demand prediction model and/or the service supply prediction model. In some embodiments, the obtaining module 403 may obtain a plurality of first sets of data and a plurality of second sets of data related to the sub-area. Each first set of data may correspond to a first historical time slice and at least include a historical number of service orders initiated in the sub-area in the corresponding first historical time slice. Each second set of data may correspond to a second historical time slice and at least include a historical number of available service providers in the sub-area in the corresponding second historical time slice. Details regarding the obtaining of the information used to generate the service demand prediction model and/or the service supply prediction model may be found elsewhere in the present disclosure. See, e.g., operations 810 and 910 and the relevant descriptions thereof.

[0073] The training module 404 may be configured to train a model. For example, the training module 404 may generate the service demand prediction model by training a preliminary model using the plurality of first sets of data. As another example, the training module 405 may generate the service supply prediction model by training a preliminary model using the plurality of second sets of data. Details regarding the generation of the service demand prediction model and the service supply prediction model may be found elsewhere in the present disclosure. See, e.g., operations 830 and 930 and the relevant descriptions thereof.

[0074] It should be noted that the above description of the processing engines 112A and 112B is provided for the purposes of illustration, and is not intended to limit the scope of

the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teachings of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure. In some embodiments, any module mentioned above may be implemented in two or more separate units. For example, the functions of determination module 402 may be implemented in two separate units, one of which is configured to determine the demand-supply gap of the first and the second sub-area, and the other is configured to determine the dispatch cost of dispatching available service providers from the second sub-area to the first sub-area. In some embodiments, the processing engine 112A and/or the processing engine 112B may further include one or more additional modules (e.g., a storage module). Additionally or alternatively, one or more modules mentioned above may be omitted. In some embodiments, the processing engines 112A and 112B may be integrated as one processing engine.

[0075] FIG. 5 is a flowchart illustrating an exemplary process for determining a dispatch utility of dispatching available service providers from a second sub-area to a first sub-area according to some embodiments of the present disclosure. At least a portion of process 500 may be implemented on the computing device 200 as illustrated in FIG. 2 or the mobile device 300 as illustrated in FIG. 3. In some embodiments, one or more operations of process 500 may be implemented in the O2O service system 100 as illustrated in FIG. 1. In some embodiments, one or more operations in the process 500 may be stored in a storage device (e.g., the storage device 160, the ROM 230, the RAM 240, the storage 390) as a form of instructions, and invoked and/or executed by the server 110 (e.g., the processing engine 112 in the server 110, or the processor 220 of the computing device 200) or the CPU 340 of the mobile device 300. In some embodiments, the instructions may be transmitted in a form of electronic current or electrical signals.

[0076] In 510, the processing engine 112A (e.g., the segmentation module 401) may segment an area into a plurality of sub-areas. In some embodiments, the area segmented by the segmentation module 401 may be an administrative area (e.g. country, province, city or district), a geographic area (e.g. Qinghai-Tibet Plateau, North China Plain, Sichuan Basin), an economic area (e.g. Yangtze River Delta economic area, Pearl River Delta economic area, Beijing-Tianjin economic area), or the like, or any combination thereof.

[0077] In some embodiments, the segmentation module 401 may segment the area according to one or more parameters, such as a size of area, a density of population, a division of administrative area, a density of office buildings, a density of residential buildings, or the like, or any combination thereof. In some embodiments, the segmentation module 401 may evenly divide the area into the sub-areas according to the parameter(s). For example, the segmentation module 401 may segment the area into sub-areas with the same size in term of the parameter(s). As another example, the segmentation module 401 may segment the area into sub-areas each of which has similar density of office buildings or population. As yet another example, the segmentation module 401 may segment the area into sub-areas having the same shape, such as a triangle, a rectangle, a diamond, a hexagon, or the like, or any combination thereof. In some embodiments, the segmentation module 401 may segment the area into a plurality of hexagonal sub-areas each of which has a predetermined length of sides (e.g. 100 meters, 200 meters, 300 meters, or 400 meters). In some embodiments, there may be no gap between two adjacent sub-areas.

[0078] In some embodiments, the segmentation module 401 may segment the area into a plurality of unit areas, and then generate the sub-areas by clustering the unit areas. The unit areas may have the same size and the same shape with each other. The clustering of the unit areas may be performed in any suitable manner. In some embodiments, the

clustering of the unit areas may be performed based on the location information of the unit areas and the number of historical service orders initiated (or completed) in each of the unit areas. For example, the segmentation module 401 may cluster a plurality of adjacent unit areas into a sub-area if the total number of historical service orders initiated (or completed) in the adjacent unit areas during a preset historical time period is smaller than a threshold.

[0079] The plurality of sub-areas may form one or more pairs of sub-areas. Each pair of sub-areas may include a first sub-area and a second sub-area. In some embodiments, a pair of a first sub-area and a second sub-area may be regarded as being different from a pair of the second sub-area and the first sub-area, considering that the dispatch utility for dispatching available service providers from the first sub-area to the second sub-area may be different from the dispatch utility for dispatching available service providers from the second sub-area to the first sub-area. Alternatively, the pair of the first sub-area and the second sub-area may be regarded as being the same as the pair of the second sub-area and the first sub-area. For illustration purposes, the present invention assumes that the pair of the first sub-area and the second sub-area is different from the pair of the second sub-area and the first sub-area, and describes the determination of the dispatch utility of for dispatching available service providers from the second sub-area to the first sub-area as an example. It should be noted that it is not intended to be limiting, with minor modifications known by a person skilled in the art, the embodiments of the present disclosure can be applied to a scenario that the pair of the first sub-area and the second sub-area is regarded as being the same as the pair of the second sub-area and the first sub-area.

[0080] In some embodiments, for each of the sub-areas in the area, the segmentation module 401 may pair the sub-area with each of the rest sub-area(s) to form a pair of sub-areas. For example, an area S may be segmented into sub-areas A, B, and C. The

sub-areas *A* to *C* may form six pairs of sub-areas, i.e., a pair of sub-areas *A* and *B*, a pair of sub-areas *B* and *A*, a pair of sub-areas *A* and *C*, a pair of sub-areas *C* and *A*, a pair of sub-areas *B* and *C*, and a pair of sub-areas *C* and *B*. In some embodiments, the segmentation module 401 may only pair a first sub-area and a second sub-area if they satisfy a certain condition. For example, the segmentation module 401 may pair the first and second sub-areas if the distance between them is less than a threshold distance. As another example, the segmentation module 401 may pair the first and second sub-area according to the supply and demand condition in the two sub-areas. Merely by way of example, only if the first sub-area has an insufficient service supply and the second sub-area has a surplus service supply, the segmentation module 401 may pair the first sub-area and the second sub-area. In some embodiments, a sub-area may have an insufficient service supply if the service demand in the sub-area is greater than the service supply in the sub-area. A sub-area may have a surplus service supply if the service supply in the sub-area is greater than the service demand in the sub-area.

[0081] In 520, for each of the first sub-area and the second sub-area, the processing engine 112A (e.g., the determination module 402) may determine a demand-supply gap of the sub-area in a predetermined time period.

[0082] As used herein, the demand-supply gap of the sub-area may be used interchangeably with the supply-demand gap to refer to a difference between the service demand and the service supply in the sub-area. In some embodiments, the service demand of the sub-area may be measured by, such as, a number of service orders initiated in the sub-area, a number of requester terminals 130 in the sub-area, a number of requester terminals 130 on which an application for an O2O service is launched in the sub-area, or the like, or any combination thereof. The service supply of the sub-area may be measured by, such as, a number of provider terminals 140 in the sub-area, a number of provider terminals 140 on which an application for an O2O service is launched in the sub-

area, a number of available service providers in the sub-area, or the like, or any combination thereof. The demand-supply gap may be measured by, such as a numerical difference of the service demand and the service supply, a ratio of the service demand to the service supply, a ratio of the service supply to the service demand, or the like.

[0083] An available service provider may refer to a service provider is able or about to be able to provide an O2O service. For example, the available service provider may be a service provider who is waiting for an O2O service order. As another example, the available service provider may be a service provider in the process of providing service, but will finish the last service order within a certain period (e.g., 0.5 minutes, 1 minute). As yet another example, the available service provider may be a service provider in the process of providing service, but can accept an O2O service order at the same time. For some kinds of O2O services, such as but not limited to a carpooling service, a delivery service, and a take-way service, a service provider may provide services for a plurality of service requesters simultaneously.

[0084] In some embodiments, the service demand and the service supply of the sub-area may be measured by the number of service orders initiated in the sub-area and the number of available service providers in the sub-area, respectively. The demand-supply gap of the sub-area may be a numerical difference between the number of service orders initiated in the sub-area and the number of available service providers in the sub-area, which may be determined by subtracting the number of service orders initiated in the sub-area from the number of available service providers of the sub-area or subtracting the number of available service providers of the sub-area from the number of service orders initiated in the sub-area. Alternatively, the demand-supply gap of the sub-area may be a ratio of the number of service orders initiated in the sub-area to the number of available service providers of the sub-area, or a ratio of the number of available service providers of the sub-area to the number of service orders initiated in the sub-area.

[0085] In some embodiments, the predetermined time period may be any time point or any time period with respect to the present moment. For example, the predetermined time period may be the current time point or a very short time period (e.g., the duration being shorter than a threshold) that can be regarded as the current time period. As another example, the predetermined time period may be, for example, 30 minutes, 45 minutes, 60 minutes, or any time period after the present moment. In some embodiments, the duration of the predetermined time period may be the same or different depend on the time of the present moment. For example, the processing engine 112A may set a shorter predetermined time period for a moment in peak hours (e.g., 8:00 a.m. to 9:00 a.m.) than a moment in normal hours (e.g., 10: 00 a.m. to 11:00 a.m.).

[0086] In some embodiments, to determine the demand-supply gap of the sub-area in the predetermine period, the determination module 402 may determine the number of service orders initiated in the sub-area in the predetermined time period and the number of available service providers in the sub-area in the predetermined time, respectively. The determination module 402 may then determine the demand-supply gap based on the number of service orders initiated in the sub-area in the predetermined time and the number of available service providers in the sub-area in the predetermined time. Details regarding the determination of the demand-supply gap may be found elsewhere in the present disclosure (e.g., FIG. 7 and the relevant descriptions thereof).

[0087] In 530, the processing engine 112A (e.g., the determination module 402) may determine a dispatch cost of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period (also referred to as the dispatch cost for brevity).

[0088] The dispatch cost may be configured to measure the cost (e.g., a time cost and/or resource cost) for dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period. A higher dispatch cost may indicate that

it is more difficult and costly to dispatch available service providers from the second sub-area to the first sub-area. In some embodiments, the dispatch cost may be determined based on a route from second sub-area to the first sub-area. Merely by way of example, the dispatch cost may increase with the distance along the route and/or the estimated travelling time along the route. As another example, the dispatch cost may increase if the traffic condition along the route is congested. In some embodiments, the dispatch cost may be determined based on road condition information related to the route according to one or more operations in process 1000.

[0089] In 540, the processing engine 112A (e.g., the determination module 402) may determine a dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period based on the demand-supply gap of the first sub-area in the predetermined time period, the demand-supply gap of the second sub-area in the predetermined time period, and the dispatch cost. For brevity, the demand-supply gap of the first sub-area in the predetermined time period may be referred to as a first demand-supply gap, the demand-supply gap of the second sub-area in the predetermined time period may be referred to as a second demand-supply gap, and the dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period may be referred to as the dispatch utility.

[0090] The dispatch utility may be used to measure the worth (or value) of dispatching available service providers from the second sub-area to the first sub-area. In some embodiments, a positive dispatch utility may indicate that it is worthy (or valuable) to dispatch available service providers from the second sub-area to the first sub-area, and the worth (or value) may increase with the value of the positive dispatch utility. On the other hand, a negative dispatch utility may indicate that it is not worthy (or valuable) to dispatch available service providers from the second sub-area to the first sub-area.

[0091] In some embodiments, the dispatch utility may be associated with at least one of the first demand-supply gap, the second demand-supply gap, or the dispatch cost. For example, if the first demand-supply gap shows that the first sub-area has a surplus supply in the predetermined time period, there may be no need to dispatch available service providers to the first sub-area in the predetermined time period and the dispatch utility may be negative. As another example, if the first demand-supply gap shows that the first sub-area has an insufficient supply, the dispatch utility may also depend on the second demand-supply gap of the second sub-area. Merely by way of example, the dispatch utility may be positive if the second sub-area has a surplus supply, and the dispatch utility may be negative if the second sub-area also has an insufficient supply. On the other hand, when other conditions (e.g., the first and the second demand-supply gap) are certain, the dispatch utility may decrease with the increase of the dispatch cost.

[0092] In some embodiments, the determination module 402 may determine the dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period according to Equation (1) as below:

$$Utility(1,2) = \frac{G[1]-G[2]}{Cost} \quad (1)$$

where $G[1]$ refers to the first demand-supply gap of the first sub-area in the predetermined time period, $G[2]$ refers to the second demand-supply gap of the second sub-area in the predetermined time period, and $Cost$ refers to the dispatch cost of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period.

[0093] According to Equation (1), the dispatch utility may have a positive relation with the difference between $G[1]$ and $G[2]$. In some embodiments, both $G[1]$ and $G[2]$ may be determined by subtracting the service supply from the service demand in the corresponding sub-area. If $G[1]$ is greater than $G[2]$, the dispatch utility may have a positive value, showing that there is a greater service need in the first sub-area than the

second sub-area and it is worthy to dispatch available service providers from the second sub-area to the first sub-area. The dispatch utility has a negative relation with the dispatch cost. The increase of the dispatch cost reduces the value of the dispatch utility when the denominator is fixed.

[0094] It should be noted that the Equation (1) is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. With minor modifications known by a person skilled in the art, the dispatch utility may be determined according to an equation other than Equation (1) illustrated above. In some embodiments, one or more parameters of Equation (1) may be omitted and/or one or more additional parameters may be added into Equation (1). For example, a parameter referring to a threshold demand-supply gap may be added into the Equation (1), and the Equation (1) may be modified into a piecewise function. In the piecewise function, $Utility(1, 2)$ may be equal to negative infinity if $G[1]$ is smaller than the threshold demand-supply gap, and $Utility(1, 2)$ may be determined according to the original Equation (1) if $G[1]$ is greater than or equal to the threshold demand-supply gap.

[0095] It should be noted that the above description regarding the process 500 is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teachings of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure.

[0096] In some embodiments, one or more operations may be omitted and/or one or more additional operations may be added. For example, operation 530 may be omitted and the dispatch utility in operation 540 may be determined on the first demand-supply gap and the second demand-supply gap. As another example, one or more operation in process 600 may be added into the process 500 to determine a dispatch strategy in the area. As yet another example, before performing operations 530, the processing engine 112A (e.g.,

the determination module 402) may analyze the first demand-supply gap and/or the second demand-supply gap to determine whether there is a need to determine the dispatch utility. Merely by way of example, the determination module 402 may determine there is no need to determine the dispatch utility if the first sub-area has a surplus supply or the second sub-area has an insufficient supply. Upon a determination that there is no need to determine the dispatch utility, operations 530 and 540 may be omitted.

[0097] FIG. 6 is a flowchart illustrating an exemplary process for determining a strategy for dispatching available service providers in an area according to some embodiments of the present disclosure. At least a portion of process 600 may be implemented on the computing device 200 as illustrated in FIG. 2 or the mobile device 300 as illustrated in FIG. 3. In some embodiments, one or more operations of process 600 may be implemented in the O2O service system 100 as illustrated in FIG. 1. In some embodiments, one or more operations in the process 600 may be stored in a storage device (e.g., the storage device 160, the ROM 230, the RAM 240, the storage 390) as a form of instructions, and invoked and/or executed by the server 110 (e.g., the processing engine 112 in the server 110, or the processor 220 of the computing device 200) or the CPU 340 of the mobile device 300. In some embodiments, the instructions may be transmitted in a form of electronic current or electrical signals.

[0098] In 610, the processing engine 112A (e.g., the determination module 402) may determine, for each pair of a plurality of pairs of sub-areas, a dispatch utility of dispatching available service providers from the second sub-area of the pair to the first sub-area of the pair.

[0099] As describe in connection with operation 510, in some embodiments, a plurality of sub-areas in an area may form a plurality of pairs of sub-areas. Each pair may include a first sub-area and a second sub-area. For each pair of sub-areas, the determination module 402 may determine a dispatch utility of dispatching available service providers

from the second sub-area of the pair to the first sub-area of the pair in the predetermined time period by performing operations 520 to 540 as described in connection with FIG. 5.

[00100] In 620, the processing engine 112A (e.g., the determination module 402) may rank the dispatch utility corresponding to each pair of sub-areas.

[00101] As described in connection with operation 540, the dispatch utility of a pair may measure the worth (or value) of dispatching available service providers from the second sub-area to the first sub-area of the pair. If a pair of A and B has a higher dispatch utility than a pair of C and D, it can be speculated that it is more worthy to dispatch available service providers from B to A than dispatching available service providers from D to C. In some embodiments, the determination module 402 may rank the dispatch utilities of the pairs of sub-areas in, for example, a descending order.

[00102] In 630, the processing engine 112A (e.g., the determination module 402) may determine a strategy for dispatching available service providers between the plurality of sub-areas based on the ranking result.

[00103] In some embodiments, the determination module 402 may select one or more candidate pairs among the pairs of sub-areas according to the ranking result. For example, the candidate pair(s) may include top N (e.g., 1, 3, 5, 10%, 30%, or 50%) pair(s) having top N dispatch utilities in the ranking result. As another example, the candidate pair(s) may include one or more pairs having a dispatch utility greater than a threshold. The determination module 402 may then determine a dispatch strategy for each of the candidate pair(s).

[00104] In some embodiments, the dispatch strategy for a candidate pair may be used to dispatch available service providers from the second sub-area of the candidate pair to the first sub-area of the candidate pair. For example, the determination module 402 may transmit a message to available service providers in the second sub-area of the candidate pair to suggest them to go to the first sub-area of the candidate pair. As another

example, the determination module 402 may allocate a service order in the first sub-area of the candidate pair to an available service provider in the second sub-area of the candidate pair if the available service provider is located within a predetermined distance to the start location of the service order. As yet another example, the determination module 402 may increase the service price in the first sub-area of the candidate pair to attract the available service providers in the second sub-area of the candidate pair to go to the first sub-area.

[00105] Additionally or alternatively, the dispatch strategy may be used to dispatch other types of service resources (e.g., service requesters and/or service orders) between the first and second sub-areas of the candidate pair. Merely by way of example, the determination module 402 may dispatch service requesters from the first sub-area of the candidate pair to the second sub-area of the candidate pair. For example, the determination module 402 may transmit a message to service requesters in the first sub-area of the candidate pair to suggest them to go to the second sub-area of the candidate pair because the second sub-area has more service providers. As another example, in order to attract the service requester in the first sub-area to go to the second sub-area, the determination module 402 may transmit a coupon that can be used in the second sub-area to the service requesters in the first sub-area.

[00106] It should be noted that the above description of the process 600 is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teachings of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure. In some embodiments, the process 600 may be performed on a plurality of pairs of sub-areas that satisfy one or more conditions. Merely by way of example, the distance of the

first sub-area and the second sub-area of each pair may have to be smaller than a threshold.

[00107] FIG. 7 is a flowchart illustrating an exemplary process for determining a demand-supply gap of a sub-area in a predetermined time period according to some embodiments of the present disclosure. At least a portion of process 700 may be implemented on the computing device 200 as illustrated in FIG. 2 or the mobile device 300 as illustrated in FIG. 3. In some embodiments, one or more operations of process 700 may be implemented in the O2O service system 100 as illustrated in FIG. 1. In some embodiments, one or more operations in the process 700 may be stored in a storage device (e.g., the storage device 160, the ROM 230, the RAM 240, the storage 390) as a form of instructions, and invoked and/or executed by the server 110 (e.g., the processing engine 112 in the server 110, or the processor 220 of the computing device 200) or the CPU 340 of the mobile device 300. In some embodiments, the instructions may be transmitted in a form of electronic current or electrical signals.

[00108] In 710, the processing engine 112A (e.g., the determination module 402) may obtain a plurality of first sets of data related to the sub-area. Each of the plurality of first sets of data may correspond to a first historical time slice and include information related to the sub-area.

[00109] In some embodiments, each of the first sets of data may include at least a historical number of service orders initiated in the sub-area in the corresponding first historical time slice. The historical number of service orders initiated in the sub-area in the first historical time slice may represent the service demand in the sub-area in the first historical time slice. In some embodiments, the size of the sub-area may affect the service demand in the sub-area. For example, a sub-area having a larger size may have a greater historical number of service orders initiated in the sub-area than a sub-area having a smaller size. In order to eliminate or reduce the effect of the area size, the

determination module 402 may normalize the historical number of service orders initiated in the sub-area in each first historical time slice according to the size of the sub-area. For example, the determination module 402 may normalize of the historical number of service orders according to Equation (2) as below:

$$H[i].T[j].O_n = H[i].T[j].O * H[i].T[j].Weight \quad (2)$$

where $H[i].T[j]$ refers to the i^{th} sub-area in the j^{th} first historical time slice, $H[i].T[j].O$ refers to the historical number of service orders initiated in the i^{th} sub-area in the j^{th} first historical time slice, $H[i].T[j].O_n$ refers to the normalized historical number of service orders initiated the i^{th} sub-area in the j^{th} first historical time slice, and $H[i].T[j].Weight$ refers to a normalization coefficient related to the size of the sub-area. In some embodiments, $H[i].T[j].Weight$ may be determined according to Equation (3) as below:

$$H[i].T[j].Weight = \begin{cases} 1, & \text{if } H[i].T[j].AREA < Threshold \\ 2^{-\left(\frac{H[i].T[j].AREA}{Threshold} - 1\right)}, & \text{if } H[i].T[j].AREA \geq Threshold \end{cases} \quad (3)$$

where $H[i].T[j].AREA$ refers to the size of the i^{th} sub-area, and $Threshold$ refers to a threshold of area size. When the size of the i^{th} sub-area is smaller than the $Threshold$, the $H[i].T[j].Weight$ may be equal to 1. When the size of the i^{th} sub-area is equal or larger than the $Threshold$, the $H[i].T[j].Weight$ may decrease with the increase of the size of the i^{th} sub-area.

[00110] In some embodiments, each of the first sets of data related to the sub-area may further include other information related to the sub-area, such as information related to peak-hours in the sub-area, location information related to the sub-area, road information related to the sub-area, traffic information related to the sub-area in the corresponding first historical time slice, weather information related to the sub-area in the corresponding first historical time slice, news information related to the sub-area in the corresponding first historical time slice, policy information related to the sub-area in the corresponding first

historical time slice, date information related to the corresponding first historical time slice, or the like, or any combination thereof.

[00111] The information related to peak-hours in the sub-area may include a determination as to whether the sub-area has a morning peak and/or an evening peak, the time of the morning peak and/or the evening peak in the sub-area, or the like, or any combination thereof. The location information related to the sub-area may include a position (e.g., the coordinate) of the sub-area, a position of a location (e.g. a center point) in the sub-area, a position of a side of the sub-area, a position of a road in the sub-area, or the like, or any combination thereof. The road information related to the sub-area may include a road type, a position, and/or a speed limit of the road(s) in the sub-area. The traffic information may include a traffic volume, a traffic congestion condition, a number of traffic accidents and their locations, a vehicle speed (e.g., an average speed, an instantaneous speed) information of the sub-area in the corresponding first historical time slice, or the like, or any combination thereof. The weather information may include an index of air quality, a temperature, visibility, humidity, a pressure, a wind speed, an index of PM 2.5, an amount of precipitation, a type of precipitation (e.g., snow, rain), a percentage likelihood of precipitation of the sub-area in the corresponding first historical time slice, or the like, or any combination thereof. The policy information may include laws and rules related to traffic, to vehicle management (e.g., only vehicles with certain plate numbers (e.g., even or odd) can be driven in certain areas), to speed limits in the sub-area in the corresponding first historical time slice, or the like, or any combination thereof. The news information may include information related to one or more events (e.g., a concert, an exhibition, a competition, a market promotion) hold in sub-area in the corresponding first historical time slice. The date information may include a characteristic of the corresponding first historical time slice, for example, being a weekday, a weekend, a holiday, and/or a festival.

[00112] The first historical time slices may be a series of time periods before the predetermined time period. Each of the first historical time slices may have the same duration. In some embodiments, the plurality of first historical time slices may be within a certain time period before the predetermined time period. The certain time period may be any suitable value, for example, 1 hour, 12 hours, 24 hours, 72 hours, 100 hours, or the like. In some embodiments, the first historical time slices may be within 1 hour to 72 hours before the predetermined time period. In some embodiments, a day may be divided uniformly into a certain number of time slices, for example, 48 time slices each of which lasts for 30 minutes, 24 time slices each of which lasts for 1 hour, or the like. The first historical time slices may be a plurality of time slices before the time slice corresponding to the predetermined time period. For example, if the predetermined time period is 4:30 p.m. to 5:00 p.m., the first historical time slices may be 48 time slices before 4:30 p.m. each of which has a duration of 30 minutes.

[00113] In 720, the processing engine 112A (e.g., the determination module 402) may predict a number of service orders initiated in the sub-area in the predetermined time period by inputting the first sets of data related to the sub-area into a service demand prediction model.

[00114] The service demand prediction model may be used to determine the number of service orders initiated in the sub-area in the predetermined time period based on the input. In some embodiments, the service demand prediction model may be a recurrent neural network (RNN) model, for example, a Long Short-Term Memory (LSTM) model. In some embodiments, the service demand prediction model may be obtained from one or more components of the O2O service system 100 (e.g., the storage device 150, the ROM 230, and/or the RAM 240), or an external source via the network 120. Additionally or alternatively, the service demand prediction model may be generated by the processing engine 112B by training a preliminary model using sample data. Details regarding the

generation of the service demand prediction model may be found elsewhere in the present disclosure (e.g., FIG. 8 and the relevant descriptions thereof).

[00115] In 730, the processing engine 112A may (e.g., the determination module 402) obtain a plurality of second sets of data related to the sub-area. Each of the plurality of second sets of data may correspond to a second historical time slice and include information related to the sub-area.

[00116] In some embodiments, each of the second sets of data may include at least a historical number of available service providers in the sub-area in the corresponding second historical time slice. The historical number of available service providers in the sub-area in the second historical time slice may represent the service supply in the sub-area in the second historical time slice. In some embodiments, the historical number of available service providers in the sub-area in a second historical time slice may be a difference between a first number of available service providers entering the sub-area in the second historical time slice and a second number of available service providers leaving the sub-area in the second historical time slice. Normally, an available service provider may not stay in a sub-area for a long period (e.g., 20 minutes or 30 minutes). The service supply of the sub-area in the second historical time slice may be determined without considering the available service providers who always stayed in the sub-area in the second historical time slice.

[00117] In some embodiments, the determination module 402 may obtain or determine the first number of available service providers and the second number of available service providers, respectively. The determination module 402 may further determine the historical number of available service providers in the sub-area in the second historical time slice based on the second number and the first number of available service providers. Optionally, the determination module 402 may normalize the historical number of available service providers in each second set of data to eliminate or reduce the effect of the size of

the sub-area. The normalization of the historical number of available service providers may be substantially similar to that of the historical number of service orders as described in connection with operation 710, and the descriptions thereof are not repeated here.

[00118] In certain embodiments, the sub-area may be a polygon having a plurality of sides, such as a hexagon. The first number of available service providers may be an accumulated number of the number of available service providers entering each side of the sub-area in the second historical time slice. The second number of available service providers may be an accumulated number of the number of available service providers leaving each side of the sub-area in the second historical time slice. The historical number of available service providers in the sub-area in the second historical time slice may be determined according to Equation (4) as below:

$$H[i].T[j].A = \sum_K (H[i].T[j].E[k].in - H[i].T[j].E[k].out) \quad (4)$$

where $H[i].T[j].A$ refers to the historical number of available service providers in the i^{th} sub-area in the j^{th} second historical time slice, $E[k]$ refers to the k^{th} side of the sub-area, K refers to the total number of sides of the sub-area, $H[i].T[j].E[k].in$ refers to the number of available service providers entering the k^{th} side of the i^{th} sub-area in the j^{th} second historical time slice, and $H[i].T[j].E[k].out$ refers to the number of available service providers leaving the k^{th} side of the i^{th} sub-area in the j^{th} second historical time slice. In some embodiments, the determination module 402 may further normalize the $H[i].T[j].A$ by multiplying Equation (4) with a normalization coefficient related to the size of the sub-area (e.g., a normalization coefficient determined according to Equation (3)).

[00119] In some embodiments, each of the second sets of data related to the sub-area may further include other information related to the sub-area, such as information related to peak-hours in the sub-area, location information related to the sub-area, road information related to the sub-area, traffic information of the sub-area in the corresponding

second historical time slice, weather information related to the sub-area in the corresponding second historical time slice, news information related to the sub-area in the corresponding second historical time slice, policy information related to the sub-area in the corresponding second historical time slice, date information related to the corresponding second historical time slice, a first number of available service providers entering the sub-area in the corresponding second historical time slice, or a second number of available service providers leaving the sub-area in the corresponding second historical time slice.

[00120] The information related to peak-hours, the location information, the road information, the traffic information, the weather information, the news information, the policy information, and the date information related to the sub-area may be substantially similar to that described in connection with operation 710, and the descriptions thereof are not repeated here. In some embodiments, the sub-area may be a polygon including a plurality of sides. The second set of data corresponding to a second historical time slice may include a first number of available service providers entering each side of the sub-area in the second historical time slice and/or a second number of available service providers leaving each side of the sub-area in the second historical time slice.

[00121] The second historical time slices may be a series of time periods before the predetermined time period. Each of the second historical time slices may have the same duration. In some embodiments, the second historical time slices may be the same as the first historical time slices. For example, both the first historical time slices and the second historical time slices may be 48 time slices before the predetermined time period, each of which lasts for 30 minutes. Alternatively, the second historical time slices may be different from the first historical time slices. For example, the duration of each second historical time slice may be slightly different from the duration of each first historical time slice. As used herein, if the difference between the durations of each first historical time slice and each second historical time slice is smaller than a threshold (e.g., 1 minute, 3

minutes), the durations of the first and second historical time slices may be regarded as being slightly different from each other. As another example, the numbers of the first historical time slices and the second historical time slice may be different. Merely by way of example, the first historical time slices may be 48 time slices before the predetermined time period, while the second historical time slices may be 50 time slices before the predetermined time period.

[00122] In 740, the processing engine 112A (e.g., the determination module 402) may predict a number of available service providers in the sub-area in the predetermined time period by inputting the second sets of data related to the sub-area into a service supply prediction model.

[00123] The service supply prediction model may be used to determine the number of available service providers in the sub-area in the predetermined time period based on the input. In some embodiments, the service supply prediction model may be an RNN model, for example, an LSTM model. In some embodiments, the service supply prediction model may be obtained from one or more components of the O2O service system 100 (e.g., the storage device 150, the ROM 230, and/or the Ram 240), or an external source via the network 120. Additionally or alternatively, the service supply prediction model may be generated by the processing engine 112B by training a preliminary model using sample data. Details regarding the generation of the service supply prediction model may be found elsewhere in the present disclosure (e.g., FIG. 9 and the relevant descriptions thereof).

[00124] In 750, the processing engine 112A (e.g., the determination module 402) may determine the demand-supply gap of the sub-area in the predetermined time period based on the number of service orders initiated in the sub-area in the predetermined time period and the number of available service providers in the sub-area in the predetermined time period. For example, the demand-supply gap may be a numerical difference between the

number of service orders initiated in the sub-area in the predetermined time period and the number of available service providers in the sub-area in the predetermined time period.

[00125] It should be noted that the above description of the process 700 is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teachings of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure. In some embodiments, one or more operations in the process 700 may be omitted and/or one or more additional operations may be added to the process 700. In some embodiments, the operations of the process 700 may be performed in any order. For example, operations 710 and 730 may be performed simultaneously or operation 730 may be performed before operation 710. In some embodiments, the determination module 402 may determine historical number available service providers in the sub-area in each second set of data by taking the available service providers stayed in the sub-area in the corresponding second historical time slice into consideration.

[00126] FIG. 8 is a flowchart illustrating an exemplary process for generating a service demand prediction model according to some embodiments of the present disclosure. At least a portion of process 800 may be implemented on the computing device 200 as illustrated in FIG. 2 or the mobile device 300 as illustrated in FIG. 3. In some embodiments, one or more operations of process 800 may be implemented in the O2O service system 100 as illustrated in FIG. 1. In some embodiments, one or more operations in the process 800 may be stored in a storage device (e.g., the storage device 160, the ROM 230, the RAM 240, the storage 390) as a form of instructions, and invoked and/or executed by the server 110 (e.g., the processing engine 112B in the server 110, or the processor 220 of the computing device 200) or the CPU 340 of the mobile device 300. In some embodiments, the service demand prediction model may be used in operation

720 to predict the number of service orders initiated in the sub-area in the predetermined time period.

[00127] In 810, the processing engine 112B (e.g., the obtaining module 403) may obtain a plurality of first sample set of data related to a plurality of sample sub-areas. In some embodiments, the sample sub-areas may include one or more of the sub-areas determined in operation 510. Additionally or alternatively, the sample sub-areas may include one or more sub-areas segmented by the processing engine 112A from another area.

[00128] Each of the first sample sets of data may correspond to a sample historical time slice and include information related to the sample sub-areas. In some embodiments, each first sample set of data may at least include a historical number of service orders initiated in each sample sub-area in the corresponding sample historical time slice. The historical number of service orders initiated in a sample sub-area may be substantially similar to the historical number of service orders initiated in a sub-area as described in connection with operation 710, and the descriptions thereof are not repeated.

[00129] In some embodiments, each of the first sample sets of data related to the sample sub-areas may include other information related to each sample sub-area, such as information related to peak-hours in the sample sub-area, location information related to the sample sub-area, road information related to the sample sub-area, traffic information related to the sample sub-area in the corresponding sample historical time slice, weather information related to the sample sub-area in the corresponding sample historical time slice, news information related to the sample sub-area in the corresponding sample historical time slice, policy information related to the sample sub-area in the corresponding sample historical time slice, date information related to the corresponding sample historical time slice, or the like, or any combination thereof. The information related to a sample sub-area included each first sample set of data may be substantially similar to the

information related to a sub-area included in a first set of data as described in connection with operation 710, and the descriptions thereof are not repeated.

[00130] The sample historical time slices may be a series of time periods before the present moment. Each of the sample historical time slices may have the same duration (e.g., 15 minutes, 30 minutes, 45 minutes, or 60 minutes). The duration of the sample historical time slices may be a default parameter of the O2O service system 100 or be adjusted according to different situations. In some embodiments, the sample historical time slices may be within a certain time period before the present moment. The certain time period may be, for example, one month, two months, three months, five months, or any other suitable value. In some embodiments, a day may be divided uniformly into a certain number of time slices, for example, 48 time slices each of which lasts for 30 minutes, 24 time slices each of which lasts for 1 hour, or the like. The sample historical time slices may be a plurality of time slices before the time slice corresponding to the present moment. For example, the sample historical time slices may be 2,880 time slices within two months before the present moment, each of which has a duration of 30 minutes.

[00131] In 820, the processing engine 112B (e.g., the obtaining module 403) may obtain a preliminary model. The preliminary model may have one or more preliminary parameters. The preliminary parameters of the preliminary model may be a default setting of the O2O service system 100 or be adjustable in different situations.

[00132] In some embodiments, the preliminary model may be an RNN model, for example, an LSTM model. In certain embodiments, the LSTM model may include a plurality of units, each of which includes a cell, an input gate, an output gate, and a forget gate. The cell may be configured to "remember" values over arbitrary time intervals. The input gate may be configured to control the input of new information that flows into the cell. The forget gate may be configured to determine the information that remains in the cell.

The output gate may be configured to determine the information that is outputted from the cell.

[00133]In 830, the processing engine 112B (e.g., the training module 404) may generate the service demand prediction model by training the preliminary model using the first sample sets of data related to the sample sub-areas.

[00134]For illustration purposes, the sample historical time slices may be denoted as a time sequence $\{t_1, t_2, \dots, t_n\}$, the first sample sets of data corresponding to the sample historical time slices related to a sample sub-area may be denoted as $\{d_1, d_2, \dots, d_n\}$, and the historical numbers of service orders initiated in the sample sub-area in the sample historical time slices may be denoted as $\{o_1, o_2, \dots, o_n\}$. In some embodiments, during the model training, for each sample sub-area, the training module 404 may determine a predicted o_i in the sample sub-area in t_i by inputting $\{d_1, d_2, \dots, d_{i-1}\}$ corresponding to $\{t_1, t_2, \dots, t_{i-1}\}$ into the preliminary model. For example, the training module 404 may determine a predicted o_n in the last sample historical time slice (e.g., last 30 minutes) by inputting $\{d_1, d_2, \dots, d_{n-1}\}$ corresponding to $\{t_1, t_2, \dots, t_{n-1}\}$ into the preliminary model. The training module 404 may further determine a difference between the predicted o_i and known o_i of the plurality of sample sub-areas, which may also be referred to as a loss function for brevity. Exemplary loss functions may include an absolute percent error (MAPE), a mean squared error (MSE), a root mean square error (RMSE), a cross entropy loss, or the like. According to the loss function, the training module 404 may further adjust the preliminary model (e.g., adjust the preliminary parameters) until the loss function reaches a desired value. After the loss function reaches the desired value, the adjusted preliminary binary model may be designated as the service demand prediction model.

[00135]FIG. 9 is a flowchart illustrating an exemplary process for generating a service supply prediction model according to some embodiments of the present disclosure. At least a portion of process 900 may be implemented on the computing device 200 as

illustrated in FIG. 2 or the mobile device 300 as illustrated in FIG. 3. In some embodiments, one or more operations of process 900 may be implemented in the O2O service system 100 as illustrated in FIG. 1. In some embodiments, one or more operations in the process 800 may be stored in a storage device (e.g., the storage device 160, the ROM 230, the RAM 240, the storage 390) as a form of instructions, and invoked and/or executed by the server 110 (e.g., the processing engine 112B in the server 110, or the processor 220 of the computing device 200) or the CPU 340 of the mobile device. In some embodiments, the service supply prediction model may be used in operation 740 to predict the number of available service providers in the sub-area in the predetermined time period.

[00136] In 910, the processing engine 112B (e.g., the obtaining module 403) may obtain a plurality of second sample sets of data related to a plurality of sample sub-areas. In some embodiments, the sample sub-areas may include one or more of the sub-areas determined in operation 510. Additionally or alternatively, the sample sub-areas may include one or more sub-areas segmented by the processing engine 112A from another area.

[00137] Each of the second sample sets of data may corresponding to a sample historical time slice and include information related to the sample sub-areas. In some embodiments, each second sample set of data may at least include a historical number of available service providers in the each sample sub-area in the corresponding sample historical time slice. The historical number of available service providers in the a sample sub-area may be substantially similar to the historical number of available service providers in a sub-area as described in connection with operation 730, and the descriptions thereof are not repeated

[00138] In some embodiments, each of the second sample sets of data related to the sample sub-areas may include other information related to each sample sub-area, such as

peak-hours in the sample sub-area, location information related to the sample sub-area, road information related to the sample sub-area, traffic information of the sample sub-area in the corresponding sample historical time slice, weather information related to the sample sub-area in the corresponding sample historical time slice, news information related to the sample sub-area in the corresponding sample historical time slice, policy information related to the sample sub-area in the corresponding sample historical time slice, date information related to the corresponding sample historical time slice, a first number of available service providers entering the sample sub-area in the corresponding sample historical time slice, or a second number of available service providers leaving the sample sub-area in the corresponding sample historical time slice. The information related to a sample sub-area included each second sample set of data may be substantially similar to the information related to a sub-area included in a second set of data as described in connection with operation 730, and the descriptions thereof are not repeated.

[00139] In 920, the processing engine 112B (e.g., the obtaining module 403) may obtain a preliminary model. Operation 920 may be performed in a similar manner with operation 820, and the descriptions thereof are not repeated here.

[00140] In 930, the processing engine 112 (e.g., the training module 404) may generate the service supply prediction model by training the preliminary model using the plurality of second sample sets of data related to the plurality of sample sub-areas. Operation 930 may be performed in a similar manner with operation 830, and the descriptions thereof are not repeated here.

[00141] It should be noted that the above description of the process 800 and/or the process 900 is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teachings of the present disclosure.

However, those variations and modifications do not depart from the scope of the present disclosure. In some embodiments, one or more operations in the processes 800 and 900 may be omitted and/or one or more additional operations may be added to the process 900 and/or the process 900. For example, an operation may be added after 830 to test the service demand prediction model or after 930 to test the service supply prediction model.

[00142] FIG. 10 is a flowchart illustrating an exemplary process for determining a dispatching cost of dispatching available service providers from a second sub-area to a first sub-area according to some embodiments of the present disclosure. At least a portion of process 1000 may be implemented on the computing device 200 as illustrated in FIG. 2 or the mobile device 300 as illustrated in FIG. 3. In some embodiments, one or more operations of process 1000 may be implemented in the O2O service system 100 as illustrated in FIG. 1. In some embodiments, one or more operations in the process 1000 may be stored in a storage device (e.g., the storage device 160, the ROM 230, the RAM 240, the storage 390, etc.) as a form of instructions, and invoked and/or executed by the server 110 (e.g., the processing engine 112A in the server 110, or the processor 220 of the computing device 200) or the CPU 340 of the mobile device 300. In some embodiments, part or all of the process 1000 may be performed to achieve operation 530 as described in connection with FIG. 5.

[00143] In 1010, the processing engine 112A (e.g., the determination module 402) may plan a route from the second sub-area to the first sub-area.

[00144] The route between the first and second sub-areas may be a route between a first location of the first sub-area and a second location of the sub-area. The first location may be any location within the first sub-area, and the second location may be any location within the second sub-area. For example, the first location and the second location may be a center point of the first sub-area and the second sub-area, respectively. The route

between the first and second sub-areas may be a linear route or a route that extends along one or more roads in the first and second sub-areas.

[00145] In some embodiments, the determination module 402 may plan a plurality of possible routes from the second sub-area to the first sub-area. The determination module 402 may then select one route among the possible routes as the route from the second sub-area to the first sub-area. The selection of the route may be performed randomly or according to a certain rule. Merely by way of example, the route with the shortest route distance or travelling time may be selected as the route from the second sub-area to the first sub-area.

[00146] In 1020, the processing engine 112A (e.g., the determination module 402) may obtain road condition information related to the road.

[00147] Exemplary road condition information related to the route may include a distance of the route, an estimated travelling time along the route, a traffic condition along the route, a number of intersections along the route, a number of the traffic lights along the route, the type of a road (e.g., high way, expressway, elevated road, main road, secondary road) along the route, traffic accident information related to the route, traffic restrictions along the route, or the like, or any combination thereof. As used herein, the traffic condition along the route may refer to the congestion of the route, which may be measured by, such as the travel speed of service providers along the route, the traffic volume along the route, travel time of along the route, or the like, or any combination thereof. In some embodiments, the road condition information may include at least one of the distance of the route, the estimated travelling time along the route, or the traffic condition along the route. In some embodiments, the road condition information related to the route may include real-time road condition information and/or historical road condition information within a certain time period (e.g., 30 minutes, 1 hour) before the present moment.

[00148] In some embodiments, at least part of the road condition information may be obtained from an external source by, for example, the processing engine 112A (e.g., determination module 402) via the network 120. For example, the processing engine 112A (e.g., determination module 402) may obtain traffic condition information, traffic accident information, road type information of one or more roads along the route from, for example, a map platform, a traffic website or database via the network 120. Additionally or alternatively, at least part of the road condition information related to the route may be determined by the processing engine 112A (e.g., the determination module 402). For example, the determination module 402 may determine the estimated travelling time along the route based on various factors, such as the distance of the route, the traffic condition along the route, the number of intersections along the route, the number of the traffic lights along the route, the type of a road along the route, traffic accident information, traffic restrictions and/or rules, weather information, or the like, or any combination thereof.

[00149] In some embodiments, the determination module 402 may determine the traffic condition along the route based on the travel speed of service providers along the route. To this end, the determination module 402 may segment the route into a plurality of road segments. Different road segments may have the same length or different lengths. For each of the road segments, the determination module 402 may determine a travel speed of one or more service providers passing through the road segment. In some embodiments, the service provider(s) passing through the road segment may include one or more service providers who is at the road segment at the present moment and/or who passed through the road segment within a time period (e.g., 30 minutes) before the present moment. The travel speed of service provider(s) passing through the road segment may be, for example, an average (or median) value of the instantaneous speed of the service provider(s) at the road segment at present moment, an average (or median) value of the average speed of the service provider(s) passing through the road segment, or

the like, or any combination thereof. In some embodiments, the travel speed of one or more service provider(s) passing through a road segment may be determined according to Equation (5) as below:

$$V_i = \frac{\sum_j^J \text{Length}(\text{Traj}_{j,i}) * W_j}{\sum_j^J \text{Time}(\text{Traj}_{j,i})} \quad (5)$$

where i refers to the i^{th} road segment of the plurality of road segments, V_i refers to the travel speed of the service provider(s) passing through the i^{th} road segment, J refers to a number of the service provider(s) passing through the i^{th} road segment, j refers to the j^{th} service provider among the service provider(s) passing through the i^{th} road segment, $\text{Traj}_{j,i}$ refers to a travelling trajectory of the j^{th} service provider on the i^{th} road segment, $\text{Length}(\text{Traj}_{j,i})$ refers to a length of the $\text{Traj}_{j,i}$, $\text{Time}(\text{Traj}_{j,i})$ refers to a travel time corresponding to the $\text{Traj}_{j,i}$, and W_j refers to a weight regarding the $\text{Traj}_{j,i}$.

[00150] In some embodiments, the weight W_j may be related to the confidence factor of the $\text{Traj}_{j,i}$. The confidence factor may be determined by the processing engine 112A by analyzing the $\text{Traj}_{j,i}$. For example, the processing engine 112A may determine whether the $\text{Traj}_{j,i}$ includes an abnormal stop (e.g., the corresponding service provider stops for a period longer than a threshold), an uneven travelling speed, and/or an obvious different travel speed with respect to other traveling trajectory in the same road segments, or the like, or any combination thereof. Upon a determination that the $\text{Traj}_{j,i}$ includes the abnormal stop, the uneven travelling speed, and/or the obvious different travel speed, the processing engine 112A may assign a low confidence factor (e.g., 0) and a low weight W_j (e.g., 0) to the $\text{Traj}_{j,i}$.

[00151] After the travel speed of the service provider(s) passing through each road segment is determined, the determination module 402 may determine the traffic condition of the route based on the travel speeds corresponding to the plurality of road segments.

[00152] In some embodiments, the determination module 402 may determine a congestion level of each road segment based on the corresponding travel speed. For example, the

congestion levels of a road segment may include “extremely congested”, “congested”, “slow”, and “smooth”, each of which may correspond to a travel speed range. Merely by way of example, the travel speed range [0 km/h, 5 km/h] may correspond to “extremely congested” and the travel speed range [25 km/h, +∞] may correspond to “smooth”. The determination module 402 may then determine the traffic condition of the route based on the congestion levels of the road segments of the route. For example, the determination module 402 may determine the most common congestion level of the road segments as the traffic condition of the road. Additionally or alternatively, determination module 402 may determine an average travel speed of the travel speeds of the road segments, and determine the traffic condition of the route based on one or more travel speed thresholds. For example, if the average travel speed is less than 5 km/h, the traffic condition of the route may be extremely congested. As another example, if the average travel speed is greater than 25 km/h, the traffic condition of the route may be smooth.

[00153] In 1030, the processing engine 112A (e.g., the determination module 402) may determine the dispatch cost based on the road condition information.

[00154] In some embodiments, the determination module 402 may determine the dispatch cost based on one or more types of the road condition information as described in connection with operation 1120. Merely by way of example, the dispatch cost may have a positive relationship with the distance along the route (or the estimated travelling time along the route). The determination module 402 may assign a higher dispatch cost for a route A than a route B if the distance (or the estimated travelling time) along the route A is larger than the route B. As another example, the determination module 402 may determine the dispatch cost based on the estimated travelling time and the traffic condition along the route according to Equation (6) as below:

$$Cost = 1 + e^{\frac{\eta(1,2)}{60}} + \theta \quad (6)$$

where $eta(1,2)$ refers to the estimated travelling time along the route from the second sub-area to the first sub-area, θ refers to a penalty coefficient related to the traffic condition along the route. In some embodiments, θ may be determined based on the congestion level of the route as described in connection with operation 1020. The value of θ may increase with the congestion level of the route. Merely by way of example, the value of θ corresponding to the congestion levels “extremely congested”, “congested”, “slow”, and “smooth” may be 3, 2, 1, and 0, respectively.

[00155] It should be noted that the above description regarding the process 1000 is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teachings of the present disclosure. In some embodiments, one or more parameters of Equation (5) and/or Equation (6) may be omitted. For example, the θ in Equation (6) may be omitted and the dispatch cost may be determined solely on the $eta(1,2)$. Additionally or alternatively, the Equation (5) and/or Equation (6) may further include one or more additional parameters may be added. For example, a parameter related to the number of traffic accidents occurred along the route may be added into the Equation (6).

[00156] In some embodiments, the dispatch cost as illustrated in Equation (6) may be used to determine the dispatch utility for dispatching available service providers from the second sub-area to the first sub-area as described in connection with operation 540. Merely by way of example, the dispatch utility for dispatching available service providers from the second sub-area to the first sub-area may be determined according to Equation (7) as below:

$$Utility(1,2) = \frac{C[1]-C[2]}{1+e^{\frac{eta(1,2)}{60}}+\theta} \quad (7)$$

[00157] Having thus described the basic concepts, it may be rather apparent to those skilled in the art after reading this detailed disclosure that the foregoing detailed disclosure

is intended to be presented by way of example only and is not limiting. Various alterations, improvements, and modifications may occur and are intended to those skilled in the art, though not expressly stated herein. These alterations, improvements, and modifications are intended to be suggested by this disclosure, and are within the spirit and scope of the exemplary embodiments of this disclosure.

[00158] Moreover, certain terminology has been used to describe embodiments of the present disclosure. For example, the terms “one embodiment,” “an embodiment,” and/or “some embodiments” mean that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present disclosure. Therefore, it is emphasized and should be appreciated that two or more references to “an embodiment,” “one embodiment,” or “an alternative embodiment” in various portions of this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures or characteristics may be combined as suitable in one or more embodiments of the present disclosure.

[00159] Further, it will be appreciated by one skilled in the art, aspects of the present disclosure may be illustrated and described herein in any of a number of patentable classes or context including any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof. Accordingly, aspects of the present disclosure may be implemented entirely hardware, entirely software (including firmware, resident software, micro-code, etc.) or combining software and hardware implementation that may all generally be referred to herein as a “block,” “module,” “engine,” “unit,” “component,” or “system.” Furthermore, aspects of the present disclosure may take the form of a computer program product embodied in one or more computer readable media having computer readable program code embodied thereon.

[00160] A computer readable signal medium may include a propagated data signal with computer readable program code embodied therein, for example, in baseband or as part

of a carrier wave. Such a propagated signal may take any of a variety of forms, including electro-magnetic, optical, or the like, or any suitable combination thereof. A computer readable signal medium may be any computer readable medium that is not a computer readable storage medium and that may communicate, propagate, or transport a program for use by or in connection with an instruction execution system, apparatus, or device. Program code embodied on a computer readable signal medium may be transmitted using any appropriate medium, including wireless, wireline, optical fiber cable, RF, or the like, or any suitable combination of the foregoing.

[00161] Computer program code for carrying out operations for aspects of the present disclosure may be written in any combination of one or more programming languages, including an object oriented programming language such as Java, Scala, Smalltalk, Eiffel, JADE, Emerald, C++, C#, VB. NET, Python or the like, conventional procedural programming languages, such as the “C” programming language, Visual Basic, Fortran 1703, Perl, COBOL 1702, PHP, ABAP, dynamic programming languages such as Python, Ruby and Groovy, or other programming languages. The program code may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider) or in a cloud computing environment or offered as a service such as a software as a service (SaaS).

[00162] Furthermore, the recited order of processing elements or sequences, or the use of numbers, letters, or other designations, therefore, is not intended to limit the claimed processes and methods to any order except as may be specified in the claims. Although

the above disclosure discusses through various examples what is currently considered to be a variety of useful embodiments of the disclosure, it is to be understood that such detail is solely for that purpose, and that the appended claims are not limited to the disclosed embodiments, but, on the contrary, are intended to cover modifications and equivalent arrangements that are within the spirit and scope of the disclosed embodiments. For example, although the implementation of various components described above may be embodied in a hardware device, it may also be implemented as a software-only solution—e.g., an installation on an existing server or mobile device.

[00163] Similarly, it should be appreciated that in the foregoing description of embodiments of the present disclosure, various features are sometimes grouped together in a single embodiment, figure, or description thereof for the purpose of streamlining the disclosure aiding in the understanding of one or more of the various embodiments. This method of disclosure, however, is not to be interpreted as reflecting an intention that the claimed subject matter requires more features than are expressly recited in each claim. Rather, claimed subject matter may lie in less than all features of a single foregoing disclosed embodiment.

WE CLAIM:

1. A system for dispatching available service providers in an Online to Offline (O2O) service platform, comprising:

at least one non-transitory computer-readable storage medium including a set of instructions;

at least one processor in communication with the at least one non-transitory computer-readable storage medium, wherein when executing the instructions, the at least one processor is configured to direct the system to:

segment an area into a plurality of sub-areas, the plurality of sub-areas forming at least one pair of sub-areas, the at least one pair of sub-areas including a first sub-area and a second sub-area;

for each of the first sub-area and the second sub-area, determine a demand-supply gap of the sub-area in a predetermined time period;

determine a dispatch cost of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period; and

determine a dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period based on the demand-supply gap of the first sub-area in the predetermined time period, the demand-supply gap of the second sub-area in the predetermined time period, and the dispatch cost.

2. The system of claim 1, wherein the at least one pair of sub-areas are a plurality of pairs of sub-areas, and at least one processor is further configured to direct the system to:

for each pair of the plurality of pairs of sub-areas, determine a dispatch utility of dispatching available service providers from the second sub-area of the pair to the first sub-area of the pair in the predetermined time period;

rank the dispatch utility of the plurality of pairs of sub-areas; and

determine, based on the ranking result of the dispatch utility, a strategy for dispatching available service providers between the plurality of sub-areas.

3. The system of claim 1 or claim 2, wherein to determine the demand-supply gap of the sub-area in the predetermine time period, the at least one processor is further configured to direct the system to:

obtain a plurality of first sets of data related to the sub-area, each of the plurality of first sets of data corresponding to a first historical time slice and at least including a historical number of service orders initiated in the sub-area in the corresponding first historical time slice;

predict a number of service orders initiated in the sub-area in the predetermined time period by inputting the plurality of first sets of data related to the sub-area into a service demand prediction model;

obtain a plurality of second sets of data related to the sub-area, each of the plurality of second sets of data corresponding to a second historical time slice and at least including a historical number of available service providers in the sub-area in the corresponding second historical time slice;

predict the number of available service providers in the sub-area in the predetermined time period by inputting the plurality of second sets of data related to the sub-area into a service supply prediction model; and

determine, based on the number of service orders initiated in the sub-area in the predetermined time period and the number of available service providers in the sub-area in the predetermined time period, the demand-supply gap of the sub-area in the predetermined time period.

4. The system of claim 3, wherein to obtain the plurality of second sets of data related to the sub-area, the at least one processor is further configured to direct the system to:

for each of the plurality of second sets of data related to the sub-area,

obtain a first number of available service providers entering the sub-area in the corresponding second historical time slice;

obtain a second number of available service providers leaving the sub-area in the corresponding second historical time slice; and

determine, based on the first number and the second number, the historical number of available service providers of the sub-area in the corresponding second historical time slice.

5. The system of claim 4, wherein to determine, based on the first number and the second number, the historical number of available service providers of the sub-area in the corresponding second historical time slice, the at least one processor is further configured to direct the system to:

obtain a size of the sub-area; and

normalize the historical number of available service providers of the sub-area in the corresponding second historical time slice based on the first number, the second number, and the size of the sub-area.

6. The system of any one of claims 3 to 5, wherein each of the plurality of first sets of data related to the sub-area further includes at least one of information related to peak-hours in the sub-area, location information related to the sub-area, road information related to the sub-area, traffic information of the sub-area in the corresponding first historical time slice, weather information related to the sub-area in the corresponding first historical time slice, news information related to the sub-area in the corresponding first historical time

slice, policy information related to the sub-area in the corresponding first historical time slice, or date information related to the corresponding first historical time slice.

7. The system of any one of claims 3 to 5, wherein each of the plurality of second sets of data related to the sub-area further includes at least one of information related to peak-hours in the sub-area, location information related to the sub-area, road information related to the sub-area, traffic information of the sub-area in the corresponding second historical time slice, weather information related to the sub-area in the corresponding second historical time slice, news information related to the sub-area in the corresponding second historical time slice, policy information related to the sub-area in the corresponding second historical time slice, date information related to the corresponding second historical time slice, a first number of available service providers entering the sub-area in the corresponding second historical time slice, or a second number of available service providers leaving the sub-area in the corresponding second historical time slice.

8. The system of any one of claims 3 to 7, wherein the service demand prediction model is generated according to a service demand prediction model training process, the service demand prediction model training process comprising:

- obtaining a plurality of first sample sets of data related to a plurality of sample sub-areas, each of the plurality of first sample sets of data corresponding to a sample historical time slice and at least including a historical number of service orders initiated in each of the plurality of sample sub-areas in the corresponding sample historical time slice;

- obtaining a preliminary model; and

- generating the service demand prediction model by training the preliminary model using the plurality of first sample sets of data related to the plurality of sample sub-areas.

9. The system of any one of claims 3 to 7, wherein the service supply prediction model is generated according to a service supply prediction model training process, the service supply prediction model training process comprising:

obtaining a plurality of second sample sets of data related to a plurality of sample sub-areas, each of the plurality of second sample sets of data corresponding to a sample historical time slice and at least including a historical number of available service providers in each of the plurality of sample sub-areas in the corresponding sample historical time slice;

obtaining a preliminary model; and

generating the service supply prediction model by training the preliminary model using the plurality of second sample sets of data related to the plurality of sample sub-areas.

10. The system of anyone of claims 3 to 9, wherein at least one of the service demand prediction model or the service supply prediction model is a recurrent neural network (RNN) model.

11. The system of anyone of claims 3 to 10, wherein the plurality of first historical time slices and the plurality of second historical time slices are within 1 hour to 72 hours before the predetermined time period.

12. The system of any one of claims 1 to 11, wherein to determine the dispatch cost of dispatching available service providers from the second sub-area to the first sub-area, the at least one processor is further configured to direct the system to:

determine a route from the second sub-area to the first sub-area;

obtain road condition information related to the route; and

determine, based on the road condition information related to the route, the dispatch cost of dispatching available service providers from the second sub-area to the first sub-area.

13. The system of claim 12, wherein the road condition information related to the route includes at least one of a distance of the route, an estimated travelling time along the route, or a traffic condition along the route.

14. The system of claim 13, wherein the determining the dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period is performed according to an Equation as below:

$$Utility(1,2) = \frac{G[1] - G[2]}{1 + e^{\frac{eta(1,2)}{60}} + \theta}$$

wherein $G[1]$ refers to the demand-supply gap of the first sub-area in the predetermined time period, $G[2]$ refers to the demand-supply gap of the second sub-area in the predetermined time period, $eta(1,2)$ refers to the estimated travelling time along the route, θ refers to a penalty coefficient related to the traffic condition along the route, and $(1 + e^{\frac{eta(1,2)}{60}} + \theta)$ represents the dispatch cost of dispatching available service providers from the second sub-area to the first sub-area.

15. The system of claim 12, wherein the road condition related to the route includes the traffic condition along the route, and

to obtain the traffic condition along the route, the at least one processor is further configured to direct the system to:

segment the route into a plurality of road segments;

for each of the plurality of road segments, determine a travel speed of one or more service providers passing through the road segment; and

determine the traffic condition of the route based on the travel speeds corresponding to the plurality of road segments.

16. The system of claim 15, wherein for each of the plurality of road segments, the determining a travel speed of one or more service providers passing through the road segment is performed according to an Equation as below:

$$V_i = \frac{\sum_j^J Length(Traj_{j,i}) * W_j}{\sum_j^J Time(Traj_{j,i})}$$

where i refers to the i^{th} road segment of the plurality of road segments, V_i refers to the travel speed of one or more service providers passing through the i^{th} road segment, J refers to a number of service providers passing through the i^{th} road segment, j refers to the j^{th} service provider among the service providers passing through the i^{th} road segment, $Traj_{j,i}$ refers to a travelling trajectory of the j^{th} service provider on the i^{th} road segment, $Length(Traj_{j,i})$ refers to a length of the $Traj_{j,i}$, $Time(Traj_{j,i})$ refers to a travel time corresponding to the $Traj_{j,i}$, and W_j refers to a weight regarding the $Traj_{j,i}$.

17. A method implemented on a computing device having at least one processor, at least one computer-readable storage medium, and a communication platform connected to a network, comprising:

segmenting an area into a plurality of sub-areas, the plurality of sub-areas forming at least one pair of sub-areas, the at least one pair of sub-areas including a first sub-area and a second sub-area;

for each of the first sub-area and the second sub-area, determining a demand-supply gap of the sub-area in a predetermined time period;

determining a dispatch cost of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period; and

determining a dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period based on the demand-supply gap of the first sub-area in the predetermined time period, the demand-supply gap of the second sub-area in the predetermined time period, and the dispatch cost.

18. The method of claim 17, wherein the at least one pair of sub-areas are a plurality of pairs of sub-areas, and the method further comprises:

for each pair of the plurality of pairs of sub-areas, determining a dispatch utility of dispatching available service providers from the second sub-area of the pair to the first sub-area of the pair in the predetermined time period;

ranking the dispatch utility of the plurality of pairs of sub-areas; and

determining, based on the ranking result of the dispatch utility, a strategy for dispatching available service providers between the plurality of sub-areas.

19. The method of claim 17 or claim 18, wherein the determining the demand-supply gap of the sub-area in the predetermine time period comprises:

obtaining a plurality of first sets of data related to the sub-area, each of the plurality of first sets of data corresponding to a first historical time slice and at least including a historical number of service orders initiated in the sub-area in the corresponding first historical time slice;

predicting a number of service orders initiated in the sub-area in the predetermined time period by inputting the plurality of first sets of data related to the sub-area into a service demand prediction model;

obtaining a plurality of second sets of data related to the sub-area, each of the plurality of second sets of data corresponding to a second historical time slice and at least including a historical number of available service providers in the sub-area in the corresponding second historical time slice;

predicting the number of available service providers in the sub-area in the predetermined time period by inputting the plurality of second sets of data related to the sub-area into a service supply prediction model; and

determining, based on the number of service orders initiated in the sub-area in the predetermined time period and the number of available service providers in the sub-area in the predetermined time period, the demand-supply gap of the sub-area in the predetermined time period.

20. The method of claim 19, wherein the obtaining the plurality of second sets of data related to the sub-area comprises:

for each of the plurality of second sets of data related to the sub-area,

obtaining a first number of available service providers entering the sub-area in the corresponding second historical time slice;

obtaining a second number of available service providers leaving the sub-area in the corresponding second historical time slice; and

determining, based on the first number and the second number, the historical number of available service providers of the sub-area in the corresponding second historical time slice.

21. The method of claim 20, wherein the determining, based on the first number and the second number, the historical number of available service providers of the sub-area in the corresponding second historical time slice comprises:

obtaining a size of the sub-area; and

normalizing the historical number of available service providers of the sub-area in the corresponding second historical time slice based on the first number, the second number, and the size of the sub-area.

22. The method of any one of claims 19 to 21, wherein each of the plurality of first sets of data related to the sub-area further includes at least one of information related to peak-hours in the sub-area, location information related to the sub-area, road information related to the sub-area, traffic information of the sub-area in the corresponding first historical time slice, weather information related to the sub-area in the corresponding first historical time slice, news information related to the sub-area in the corresponding first historical time slice, policy information related to the sub-area in the corresponding first historical time slice, or date information related to the corresponding first historical time slice.

23. The method of any one of claims 19 to 21, wherein each of the plurality of second sets of data related to the sub-area further includes at least one of information related to peak-hours in the sub-area, location information related to the sub-area, road information related to the sub-area, traffic information of the sub-area in the corresponding second historical time slice, weather information related to the sub-area in the corresponding second historical time slice, news information related to the sub-area in the corresponding second historical time slice, policy information related to the sub-area in the corresponding second historical time slice, date information related to the corresponding second historical time slice, a first number of available service providers entering the sub-area in the corresponding second historical time slice, or a second number of available service providers leaving the sub-area in the corresponding second historical time slice.

24. The method of any one of claims 19 to 23, wherein the service demand prediction model is generated according to a service demand prediction model training process, the service demand prediction model training process including:

- obtaining a plurality of first sample sets of data related to a plurality of sample sub-areas, each of the plurality of first sample sets of data corresponding to a sample historical time slice and at least including a historical number of service orders initiated in each of the plurality of sample sub-areas in the corresponding sample historical time slice;

- obtaining a preliminary model; and

- generating the service demand prediction model by training the preliminary model using the plurality of first sample sets of data related to the plurality of sample sub-areas.

25. The method of any one of claims 19 to 23, wherein the service supply prediction model is generated according to a service supply prediction model training process, the service supply prediction model training process including:

- obtaining a plurality of second sample sets of data related to a plurality of sample sub-areas, each of the plurality of second sample sets of data corresponding to a sample historical time slice and at least including a historical number of available service providers in each of the plurality of sample sub-areas in the corresponding sample historical time slice;

- obtaining a preliminary model; and

- generating the service supply prediction model by training the preliminary model using the plurality of second sample sets of data related to the plurality of sample sub-areas.

26. The method of anyone of claims 19 to 25, wherein at least one of the service demand prediction model or the service supply prediction model is a recurrent neural network (RNN) model.

27. The method of anyone of claims 19 to 26, wherein the plurality of first historical time slices and the plurality of second historical time slices are within 1 hour to 72 hours before the predetermined time period.

28. The method of any one of claims 17 to 27, wherein the determining the dispatch cost of dispatching available service providers from the second sub-area to the first sub-area comprises:

determining a route from the second sub-area to the first sub-area;

obtaining road condition information related to the route; and

determining, based on the road condition information related to the route, the dispatch cost of dispatching available service providers from the second sub-area to the first sub-area.

29. The method of claim 28, wherein the road condition information related to the route includes at least one of a distance of the route, an estimated travelling time along the route, or a traffic condition along the route.

30. The method of claim 29, wherein the determining the dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period is performed according to an Equation as below:

$$Utility(1,2) = \frac{G[1] - G[2]}{1 + e^{\frac{eta(1,2)}{60}} + \theta}$$

wherein $G[1]$ refers to the demand-supply gap of the first sub-area in the predetermined time period, $G[2]$ refers to the demand-supply gap of the second sub-area in the predetermined time period, $eta(1,2)$ refers to the estimated travelling time along the

route, θ refers to a penalty coefficient related to the traffic condition along the route, and $(1 + e^{\frac{\eta(1,2)}{60}} + \theta)$ represents the dispatch cost of dispatching available service providers from the second sub-area to the first sub-area.

31. The method of claim 28, wherein the road condition related to the route includes the traffic condition along the route, and

the obtaining the traffic condition along the route comprises:

segmenting the route into a plurality of road segments;

for each of the plurality of road segments, determining a travel speed of one or more service providers passing through the road segment; and

determining the traffic condition of the route based on the travel speeds corresponding to the plurality of road segments.

32. The method of claim 31, wherein for each of the plurality of road segments, the determining a travel speed of one or more service providers passing through the road segment is performed according to an Equation as below:

$$V_i = \frac{\sum_j^J \text{Length}(\text{Traj}_{j,i}) * W_j}{\sum_j^J \text{Time}(\text{Traj}_{j,i})}$$

where i refers to the i^{th} road segment of the plurality of road segments, V_i refers to the travel speed of one or more service providers passing through the i^{th} road segment, J refers to a number of service providers passing through the i^{th} road segment, j refers to the j^{th} service provider among the service providers passing through the i^{th} road segment, $\text{Traj}_{j,i}$ refers to a travelling trajectory of the j^{th} service provider on the i^{th} road segment, $\text{Length}(\text{Traj}_{j,i})$ refers to a length of the $\text{Traj}_{j,i}$, $\text{Time}(\text{Traj}_{j,i})$ refers to a travel time corresponding to the $\text{Traj}_{j,i}$, and W_j refers to a weight regarding the $\text{Traj}_{j,i}$.

33. A non-transitory computer-readable storage medium embodying a computer program product, the computer program product comprising instructions configured to cause a computing device to:

segment an area into a plurality of sub-areas, the plurality of sub-areas forming at least one pair of sub-areas, the at least one pair of sub-areas including a first sub-area and a second sub-area;

for each of the first sub-area and the second sub-area, determine a demand-supply gap of the sub-area in a predetermined time period;

determine a dispatch cost of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period; and

determine a dispatch utility of dispatching available service providers from the second sub-area to the first sub-area in the predetermined time period based on the demand-supply gap of the first sub-area in the predetermined time period, the demand-supply gap of the second sub-area in the predetermined time period, and the dispatch cost.

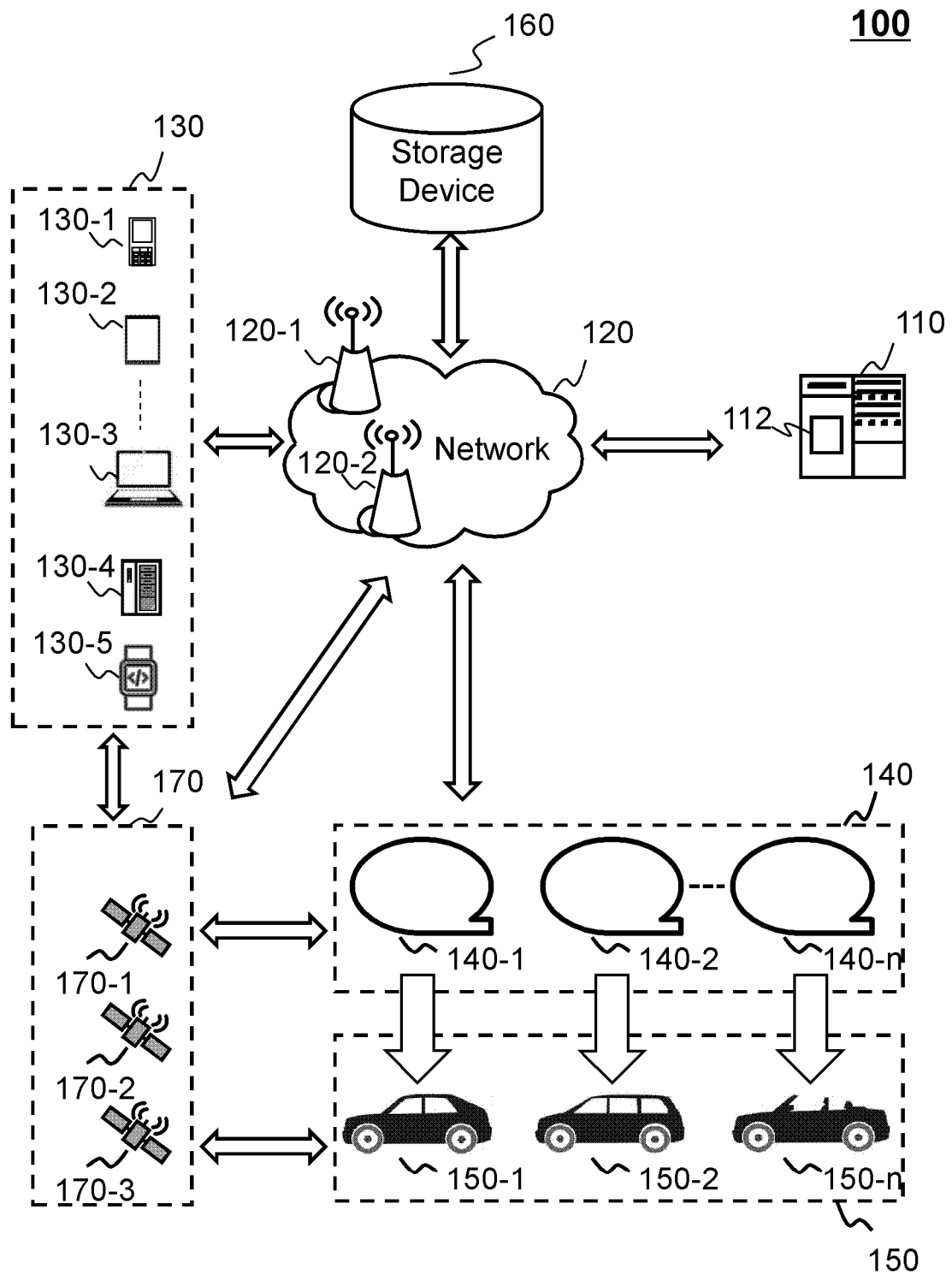
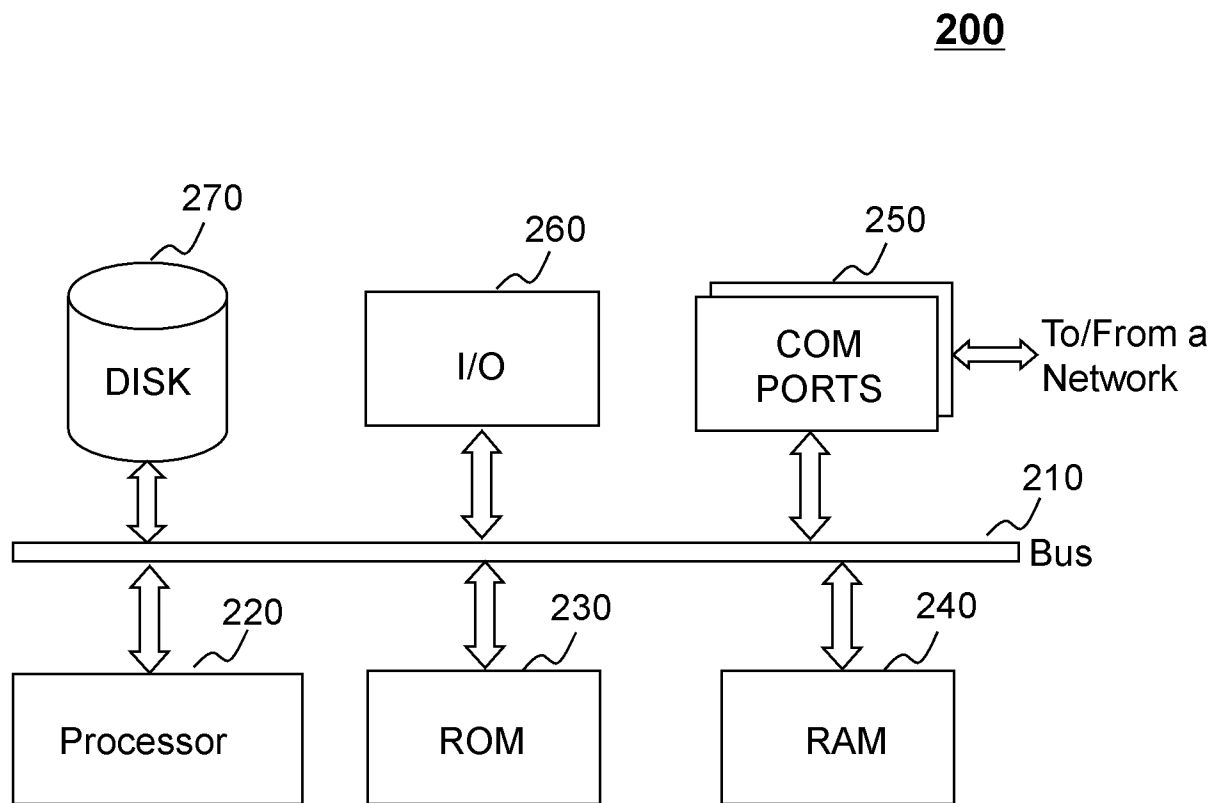
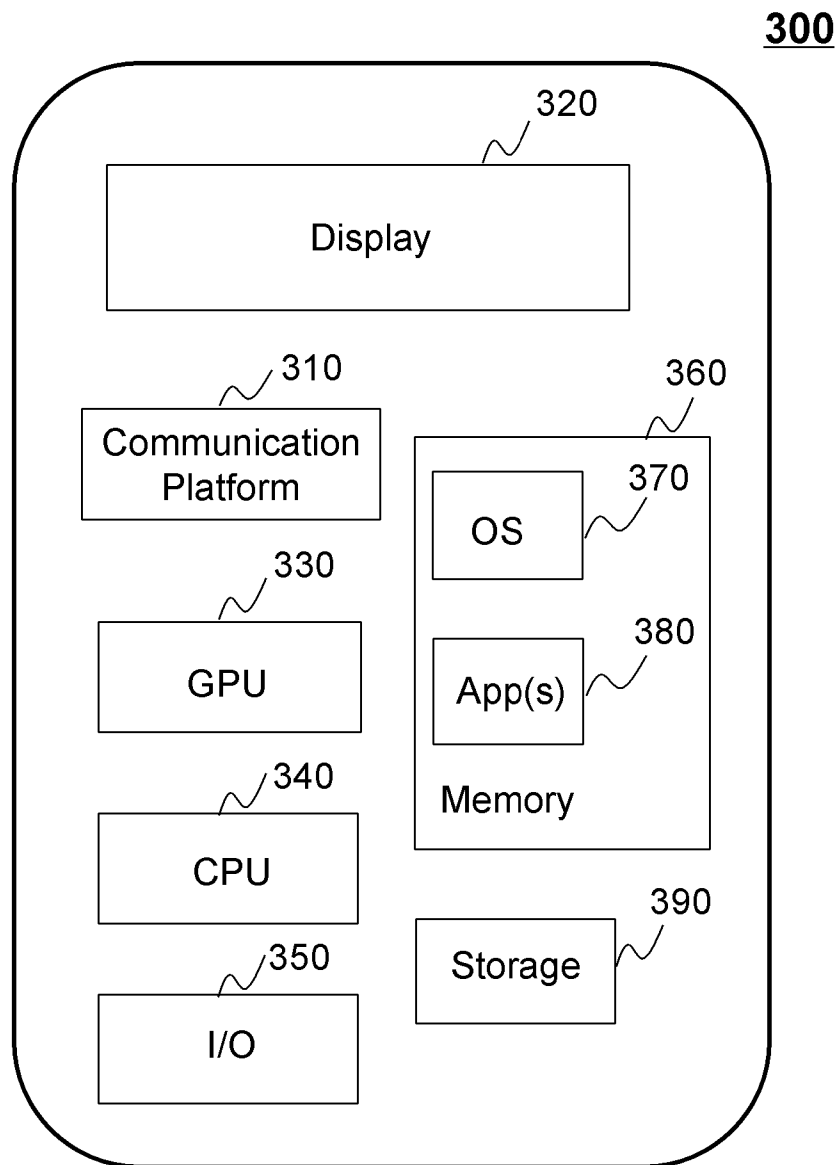
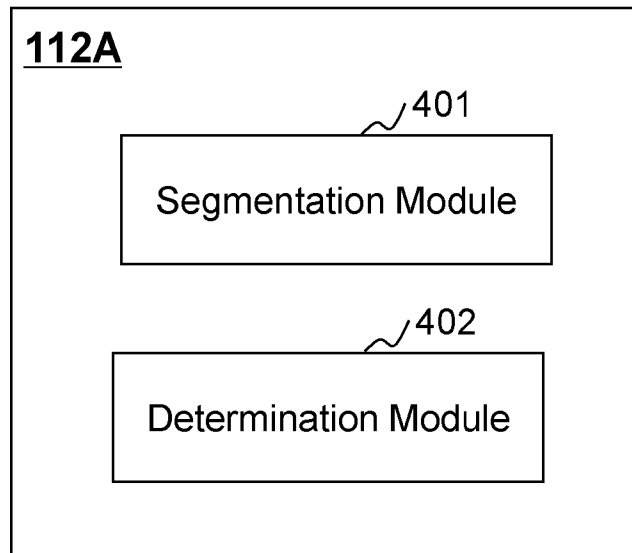
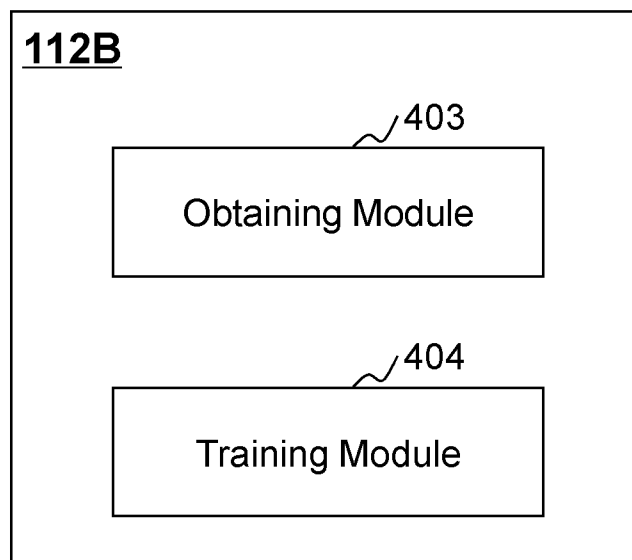
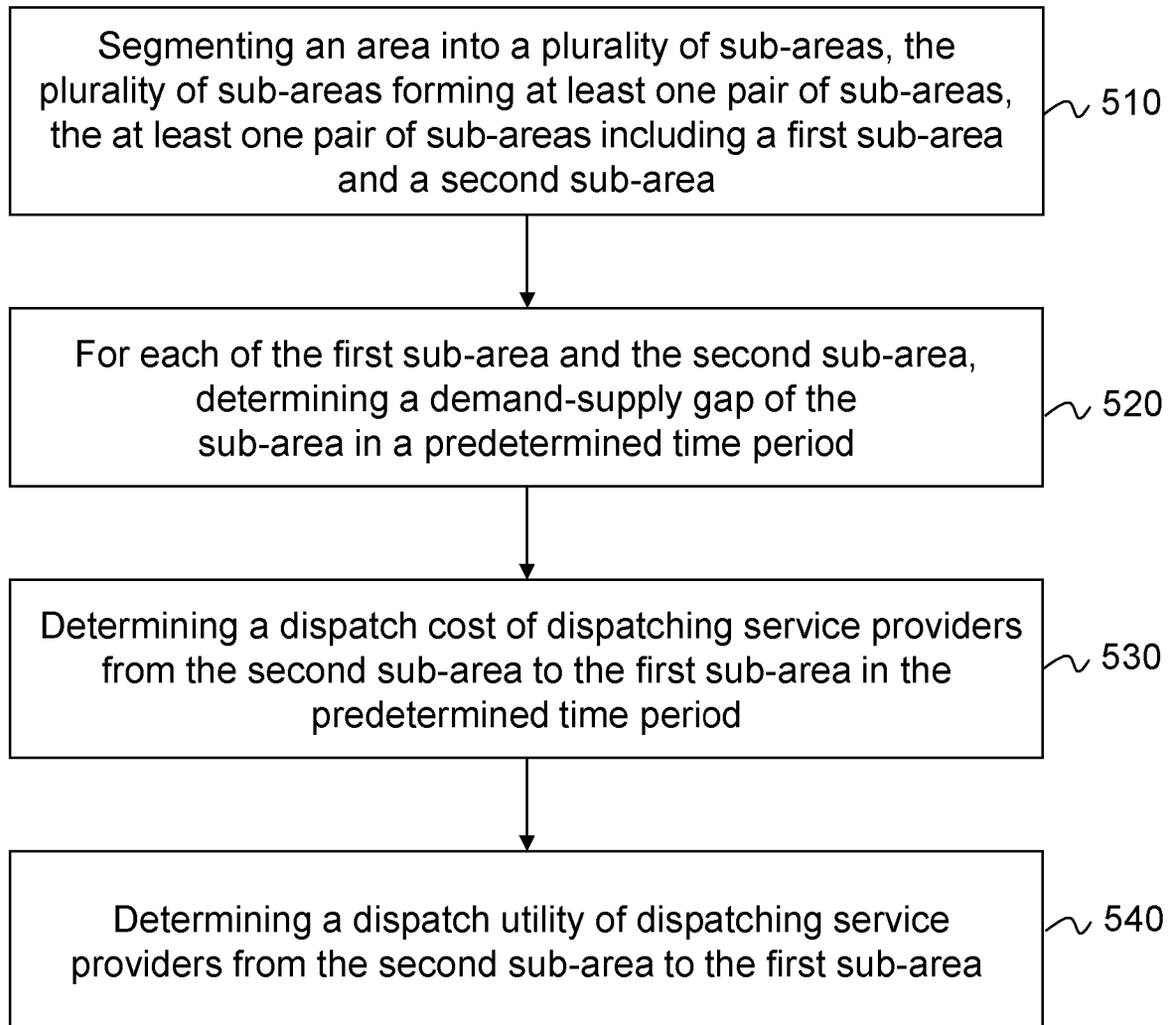


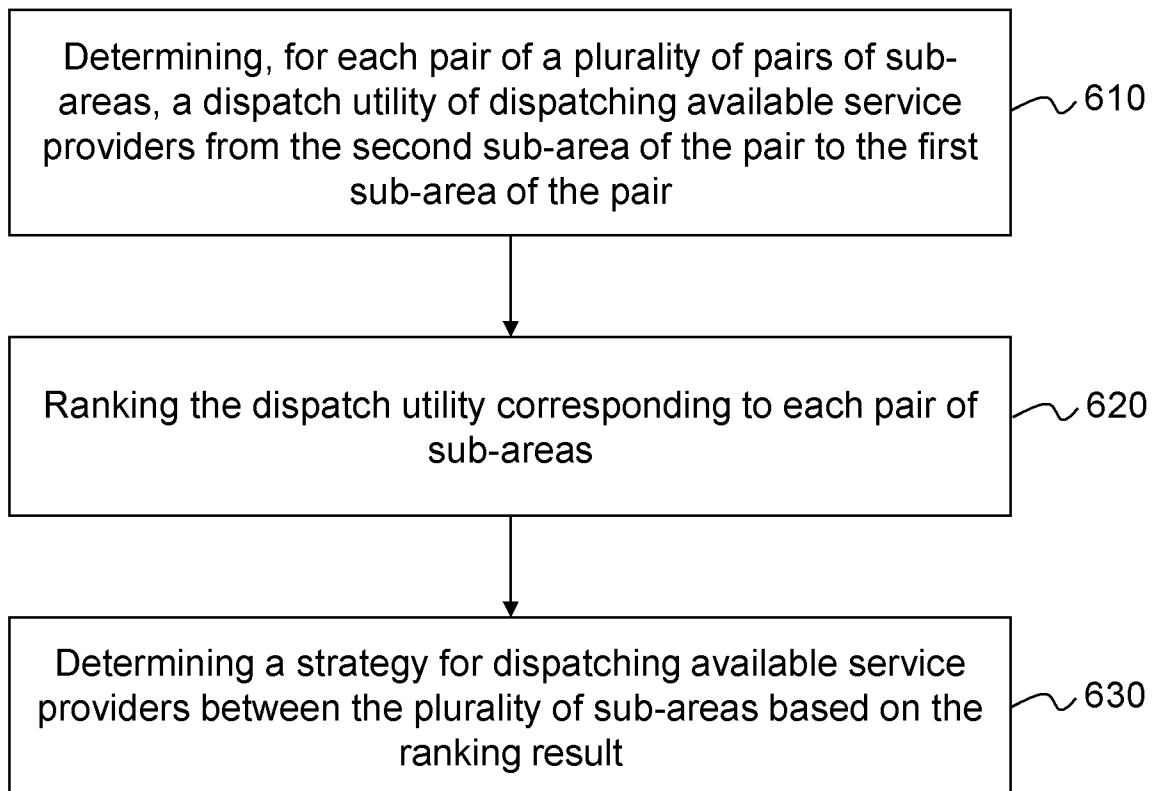
FIG. 1

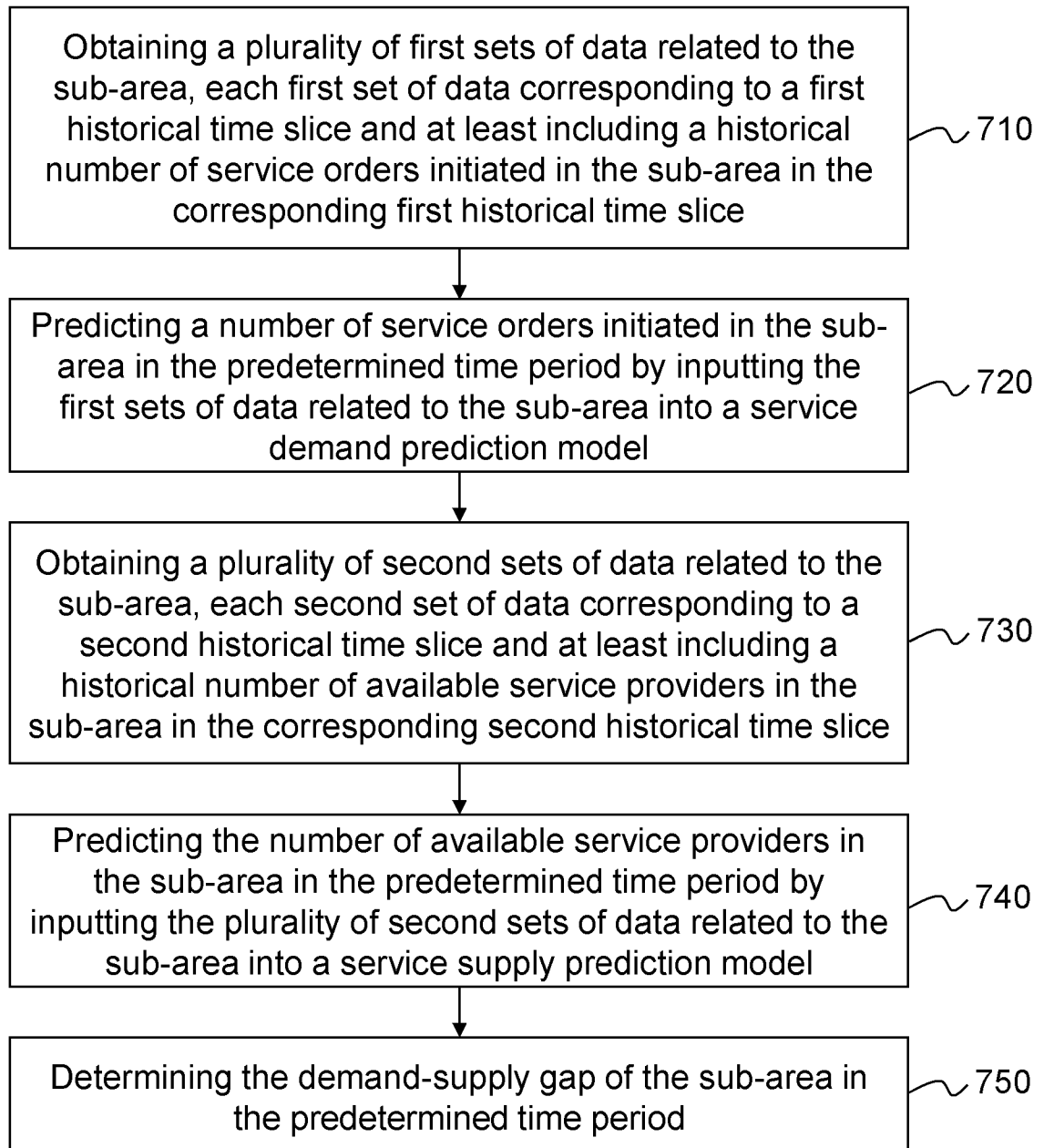
**FIG. 2**

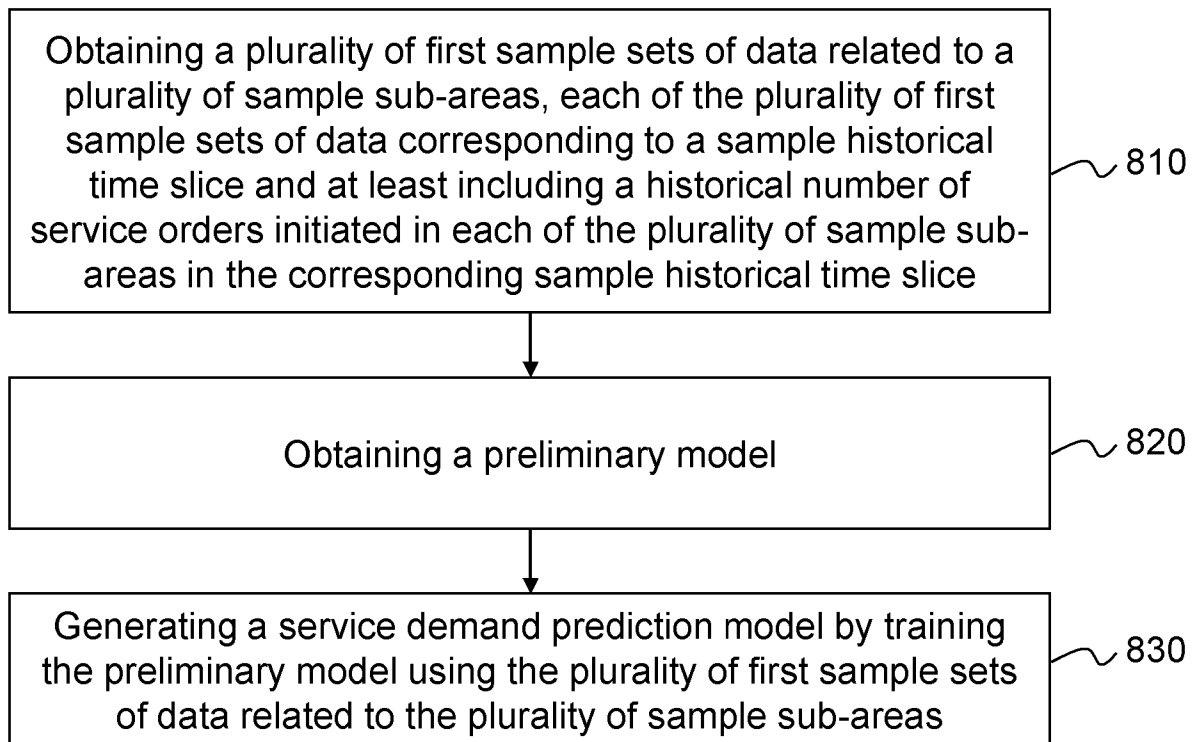
**FIG. 3**

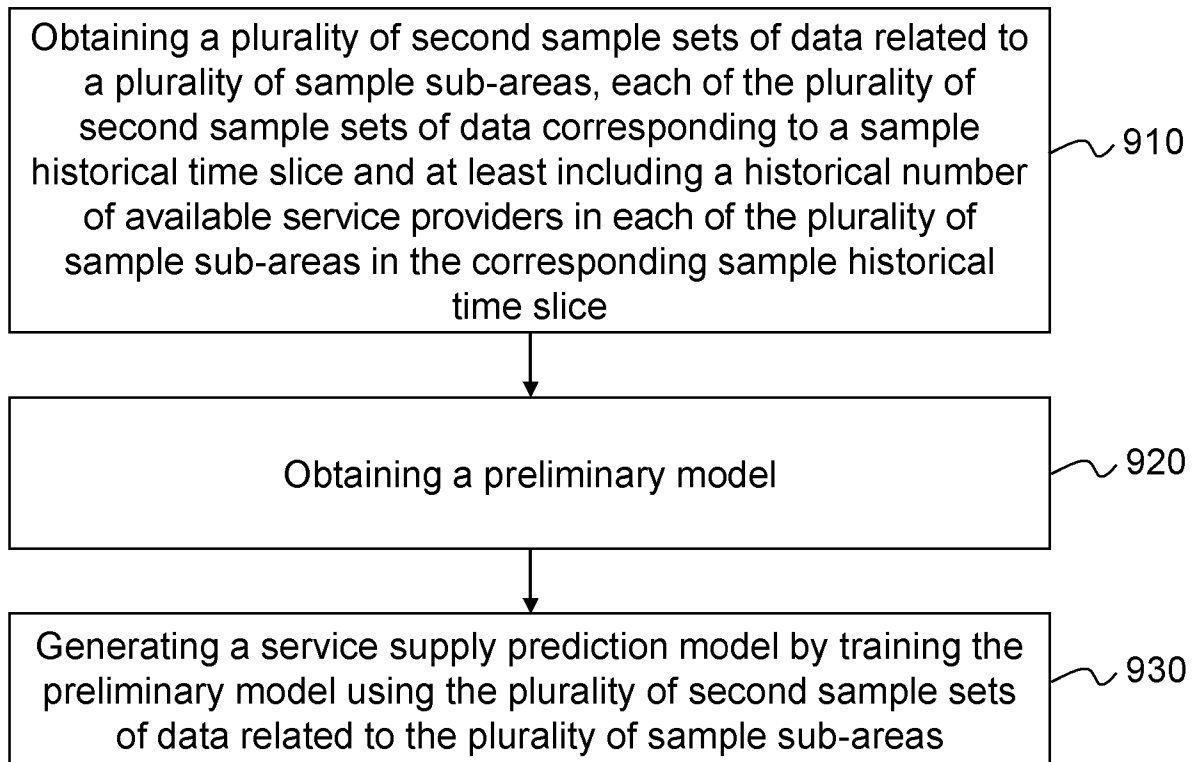
**FIG. 4A****FIG. 4B**

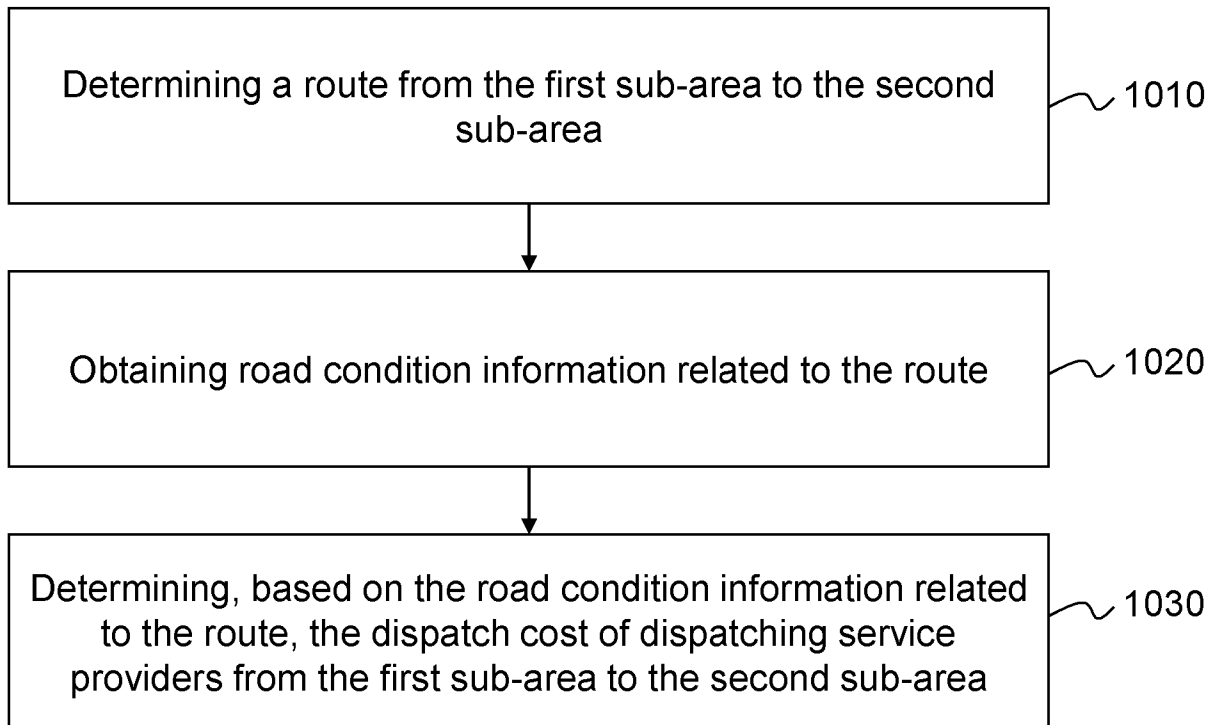
500**FIG. 5**

600**FIG. 6**

700**FIG. 7**

800**FIG. 8**

900**FIG. 9**

1000**FIG. 10**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/097205**A. CLASSIFICATION OF SUBJECT MATTER**

G06Q 30/06(2012.01)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)

G06Q; G08G; 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)

CNPAT;CNKI;WPI;EPODOC:vehicle, transport, dispatch+, area, region, supply, demand, cost, distance

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 108010306 A (BEIJING DIDI INFINITY TECHNOLOGY AND DEVELOPMENT CO., LTD.) 08 May 2018 (2018-05-08) paragraphs[0071]-[0126]	1-2, 12-13, 15, 17-18, 28-29, 31, 33
A	CN 107545486 A (DIDI CHINA TECHNOLOGY CO., LTD.) 05 January 2018 (2018-01-05) the whole document	1-33
A	US 2017091856 A1 (SKURT, INC.) 30 March 2017 (2017-03-30) the whole document	1-33

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

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

17 March 2019

Date of mailing of the international search report

28 March 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088
China

Authorized officer

CHEN,Xiyuan

Facsimile No. (86-10)62019451

Telephone No. 86-(10)-53961594

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/097205

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108010306	A	08 May 2018	US	20190057479	A1	21 February 2019
CN	107545486	A	05 January 2018	None			
US	2017091856	A1	30 March 2017	None			